UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Univariate parametric and nonparametric statistical quality control techniques with estimated process parameters

by

Schalk William Human

Submitted in partial fulfillment of
the requirements for the degree

Philosphiae Doctor (Mathematical Statistics)

in the
Faculty of Natural & Agricultural Sciences
University of Pretoria
Pretoria

July 2009

Declaration

I declare that the thesis that I hereby submit for the degree Philosphiae Doctor (Mathematical Statistics) at the University of Pretoria has not previously been submitted by me for degree purposes at any other university.

Signature _____ Date _____

Acknowledgements

Thanks go to many for making this research and degree possible.

I would like to sincerely thank my supervisor Prof. CF Smit and my co-supervisor Prof. S Chakraborti for their extraordinary efforts to my benefit throughout my tenure as a PhD student in the Department of Statistics at the University of Pretoria. Their professional knowledge, constructive guidance and years long experience were indispensable during the entire process of doing my research and preparing this dissertation.

I am truly grateful to my colleagues at the Department of Statistics at the University of Pretoria. I have thoroughly enjoyed collaborating and learning from you. Thank you for every bit of your support, encouragement and kindness.

I am also indebted to Prof. NAS Crowther for allowing me in the PhD program and his unconditional trust for granting me permission to take research leave, and STATOMET along with the NRF for supporting me financially. Thank you very much.

Thanks also go to the faculty members and the postgraduate students of the Department of Information Systems, Statistics and Management Science at the University of Alabama, USA. They made my research visits a real pleasure.

Special thanks to my family – especially my parents Will and Elsabé, and my brother Cornel. Your lifelong love, sacrifices and continuous support has meant the world to me!

My Creator and Saviour for the talents and grace bestowed upon me.

Summary

Chapter 1 gives a brief introduction to statistical quality control (SQC) and provides background information regarding the research conducted in this thesis.

We begin **Chapter 2** with the design of Shewhart-type Phase I S^2, S and R control charts for the situation when the mean and the variance are both unknown and are estimated on the basis of m independent rational subgroups each of size n available from a normally distributed process. The derivations recognize that in Phase I (with unknown parameters) the signaling events are dependent and that more than one comparison is made against the same estimated limits simultaneously; this leads to working with the joint distribution of a set of dependent random variables. Using intensive computer simulations, tables are provided with the charting constants for each chart for a given false alarm probability. Second an overview of the literature on Phase I parametric control charts for univariate variables data is given assuming that the form of the underlying continuous distribution is known. The overview presents the current state of the art and what challenges still remain. It is pointed out that, because the Phase I signaling events are dependent and multiple signaling events are to be dealt with simultaneously (in making an in-control or not-in-control decision), the joint distribution of the charting statistics needs to be used and the recommendation is to control the probability of at least one false alarm while setting up the charts.

In **Chapter 3** we derive and evaluate expressions for the run-length distributions of the Phase II Shewhart-type *p*-chart and the Phase II Shewhart-type *c*-chart when the parameters are estimated. We then examine the effect of estimating p and c on the performance of the *p*-chart and the *c*-chart via their run-length distributions and associated characteristics such as the average run-length, the false alarm rate and the probability of a "no-signal". An exact approach based on the binomial and the Poisson distributions is used to derive expressions for the Phase II run-length distributions and the related Phase II characteristics using expectation by conditioning (see e.g. Chakraborti, (2000)). We first obtain the characteristics of the run-length distributions conditioned on point estimates from Phase I and then find the unconditional characteristics by averaging over the distributions of the point estimators. The in-control and the out-of-

control properties of the charts are looked at. The results are used to discuss the appropriateness of the widely followed empirical rules for choosing the size of the Phase I sample used to estimate the unknown parameters; this includes the number of reference samples m and the sample size n.

Chapter 4 focuses on distribution-free control charts and considers a new class of nonparametric charts with runs-type signaling rules (i.e. runs of the charting statistics above and below the control limits) for both the scenarios where the percentile of interest of the distribution is known and unknown. In the former situation (or Case K) the charts are based on the sign test statistic and enhance the sign chart proposed by Amin et al. (1995); in the latter scenario (or Case U) the charts are based on the two-sample median test statistic and improve the precedence charts by Chakraborti et al. (2004). A Markov chain approach (see e.g. Fu and Lou, (2003)) is used to derive the run-length distributions, the average run-lengths, the standard deviation of the run-lengths etc. for our runs rule enhanced charts. In some cases, we also draw on the results of the geometric distribution of order k (see e.g. Chapter 2 of Balakrishnan and Koutras, (2002)) to obtain closed form and explicit expressions for the run-length distributions and/or their associated performance characteristics. Tables are provided for implementation of the charts and examples are given to illustrate the application and usefulness of the charts. The in-control and the out-of-control performance of the charts are studied and compared to the existing nonparametric charts using criteria such as the average run-length, the standard deviation of the run-length, the false alarm rate and some percentiles of the run-length, including the median run-length. It is shown that the proposed "runs rules enhanced" sign charts offer more practically desirable in-control average run-lengths and false alarm rates and perform better for some distributions.

Chapter 5 wraps up this thesis with a summary of the research carried out and offers concluding remarks concerning unanswered questions and/or future research opportunities.

Contents

Chapter 3 Attributes control charts: Case K and Case U 124

Chapter 4 Nonparametric Shewhart-type control charts with runs-type signaling rules: Case K and Case U 254

Chapter 1

Introduction and research objectives

1.0 Introduction

Statistical process control (SPC) refers to the collection of statistical procedures and problem-solving tools used to control and monitor the quality of the output of some production process, including the output of services (see e.g. Balakrishnan et al., (2006) p. 6678 and Montgomery, (2005) p. 148). The aim of SPC is to detect and eliminate or, at least reduce, unwanted variation in the output of a process. The benefits include saving time, increasing profits and an overall increase in the quality of products and services.

The quality of process output can be measured in various ways. Frequently the percentage or the fraction of items that does not conform to specifications is used. In many practical situations it is more convenient to measure the quality of the product or the service by the number of nonconformities per "inspection unit" or the "unit area of opportunity" such as the number of scratches on a plate of glass, the number of tears in a sheet of material or the number of errors made by a cash register attendant during a day. Sometimes the quality of a sample of items is measured by the mean (average) of the measurements or by some other measure of central tendency such as a percentile. Consider, for example, a beverage filling machine designed to fill each container (such as a bottle or a can) with 500ml of cool drink. Some containers will have slightly more than 500ml and some will have slightly less, in accordance with a fill volume distribution. If the filling machine begins to wear or, its inputs or its environment changes, the distribution of the net filling volume can change. If such a change is permanent and goes undetected more and more containers will be filled incorrectly, resulting in waste or containers filled below specifications. While in the former case the waste is in the form of "free" product for the consumer, typically waste consists of rework or scrap. We can measure the quality of the process, i.e. the ability of the beverage filling machine to fill the containers with 500ml of cool drink, in a number of ways. We can, for instance, take successive samples of containers and count the

number of containers with too much or too little cool drink, according to the required specifications. Alternatively, we could measure the amount of cool drink in each container and then calculate the average fill volume for each sample. Both these summary measures provide useful information regarding the functioning of the process; for example, if either the number of containers that are not filled according to the specifications or the average fill volume increases above or drop below certain critical points, action is required to find the root cause and rectify the problem.

SPC has long been applied in high-volume manufacturing processes such as the one described above. In recent times it has also been applied in government offices, by educators and administrators from the public and private sectors, by providers of healthcare services, and by those in the service industries (such as finance, hospitality and transportation) to name but a few. These are primarily service industries where the “volume” or the “speed” of production is less in comparison to the usual manufacturing process and the quality characteristics are less tangible and not easily measured on a numerical scale. The key idea, however, is that the principles and concepts of SPC can be applied to any repetitive process, i.e. a process wherein the same action is performed over-and-over with the intention to obtain the same “outcome” or “result” on each “trial”.

A wide range of statistical procedures are used in the various stages of SPC; these range from basic descriptive techniques and summary measures (such as histograms, stem-and-leaf diagrams, check sheets, scatter diagrams etc.) to more advanced procedures (such as process optimization, evolutionary operation and design of experiments). Many of the statistical procedures that are used in SPC have a long and rich history and/or fill a separate niche in the process control environment; these include, amongst many other procedures, acceptance sampling and sampling schemes, measurement systems analysis, calibration, process capability analysis and capability indices, reliability analysis, statistical and stochastic modeling, six sigma as well as statistical process control and statistical process monitoring using control charts. For an excellent reference source and a comprehensive overview on these and other related topics see, for example, the *Encyclopedia of Statistics in Quality and Reliability* edited by Ruggeri et al. and published in 2007 by John Wiley & Sons Ltd.

The collection of statistical tools is undoubtedly an important component of SPC but it should be kept in mind that they comprise merely its technical aspects. SPC, in general, builds an environment in which all the individuals of an organization seek continuous improvement in quality and productivity

and is best implemented and most successful when management becomes involved (Montgomery, (2005) p. 148).

Given the multifaceted structure of SPC, it is essential that a researcher accurately describes as far as it is possible the context and the exact nature of his research within the SPC domain. Therefore, it is appropriate to say that:

> This thesis focuses on improving existing control charting methodologies and developing new control charts; more specifically, it focuses on univariate parametric and nonparametric Shewhart-type Phase I and Phase II variables control charts and attributes control charts (for samples of size $n > 1$) when process parameters are estimated.

To have a better handle on the precise meaning of the above statement and the focus of this thesis, the rest of Chapter 1 is devoted to explaining what a control chart is and discusses the similarities and/or dissimilarities between the major types of control charts. This exposition includes a discussion on:

(i) Shewhart-type charts vs. EWMA-type and CUSUM-type charts,
(ii) Univariate charts vs. multivariate charts,
(iii) Variables charts vs. attributes charts,
(iv) Phase I charts vs. Phase II charts, and
(v) Parametric charts vs. nonparametric charts.

Following the discussion concerning the different types of control charts, we describe in more detail what is done in each of the remaining chapters of this thesis.

It is important to note that the author of this thesis does not intend to present a full-blown discussion and/or overview on all the aspects of SPC in Chapter 1. Instead, we cover only the key aspects to equip the reader with the necessary terminology (principles) in order to grasp what is to be covered in the rest of this thesis. We hope that a discussion regarding points (i) to (v) listed above will give the reader the necessary background of the underlying basic ideas about this vast area.

Also, note that, we focus on control charts for samples of size $n > 1$ and use the phrases “rational subgroup” and “random sample” interchangeably throughout the thesis but, strictly speaking, a rational subgroup is not necessarily a random sample (see e.g. the discussion in Montgomery, (2005) on p. 162).

Control chart

A control chart is a statistical procedure (or scheme) that can be depicted graphically for on-line process monitoring of a measurable characteristic (such as the mean measurement value or the percentage nonconforming items) with the objective to show whether the process is operating within the limits of expected variation (see e.g. Ruggeri, Kenett and Faltin (2007) p. 429) . The simplest and most widely used control chart is the Shewhart-type of chart; this chart is named after the father of quality control i.e. Dr. Walter A. Shewhart (1891-1967) of Bell Telephone Laboratories, who developed the chart in the 1930's and laid the foundation of modern statistical process control in his book *Economic Control of Quality of Manufactured Product* that was originally published in 1931.The the wider use and popularity of control charts outside manufacturing, which lead to Quality Management and Six Sigma, can be attributed to Deming (1986).

Shewhart-type control chart

A typical Shewhart-type control chart is shown in Figure 1.1. The chart is a basic graphical display of the successive values of a summary measure (statistic) calculated from a sample of measurements taken on a key quality characteristic and plotted on the vertical axis versus the sample number or time on the horizontal axis. The control chart usually has a centerline (*CL*) and two horizontal lines, one line on either side of the centerline. The line above the centerline is called the upper control limit (*UCL*) whereas the line below the centerline is called the lower control limit (*LCL*). These three lines are placed on the control chart to aid the user in making an informed and objective decision whether a process is in-control or not; this decision is primarily based on the pattern of the points plotted on the chart and/or their position relative to the control limits. Notice that it is customary to join the points on a control chart using straight-line segments for easier visualization over time.

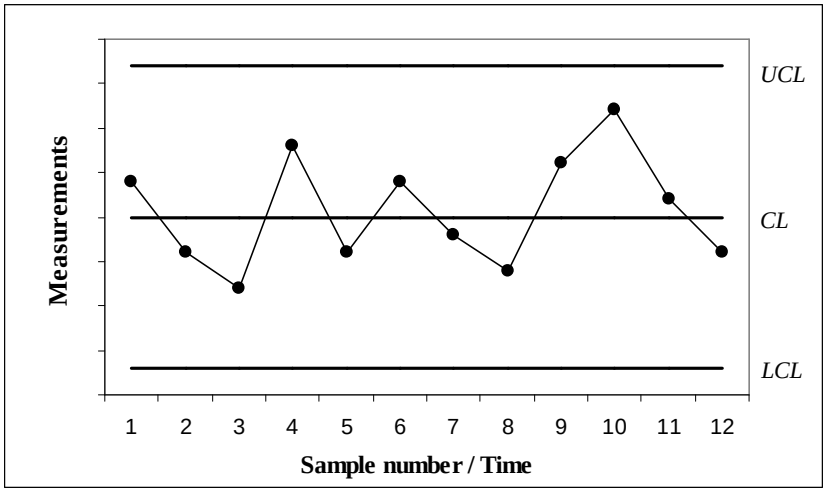


Figure 1.1: A Shewhart-type of control chart

The basic assumption underlying control chart analysis is that the variation in the quality of products or services is due in part to common causes (or chance causes) and in part to special causes (or assignable causes) – see Deming (1986). The term common cause refers to the inherent or the natural variability that is present in a process. This is also referred to as the uncontrollable or the ever present "background noise" that might be due to the cumulative effect of many small and undetectable (but unavoidable) causes. Special causes are those sources of variability that are not part of the common causes (or natural variability of a process) and therefore directly affect the quality of a process.

Combining these two sources of variation, i.e. common causes and assignable causes of variation, accounts for the total variation present in a process. Based on this point of view, a process is considered to be in-control if it is operating only in the presence of common causes and when special causes are part of the process variability, the process is said to be out-of-control. The fundamental idea of the Shewhart-type of control chart entails identifying and removing, to an extent that is economically viable, the assignable causes of variation.

Control charts play a crucial role in detecting whether a process is in-control or out-of-control. The standard Shewhart-type control charts are based on inspecting samples at equally spaced time intervals and issuing an alarm (a signal) if the "result of the sample" is considerably worse (i.e. larger or smaller) than what one can expect if the process was operating on target. For example, a single point (plotting statistic) that plots outside the control limits i.e. lies above the upper control limit or lies below the lower control limit, is usually interpreted as a signal (an alarm) of a possible special cause. The alarm signals that the process is deemed to be in an out-of-control state, which is indicative of deteriorated performance of the process. Investigation is thus required to find the origin of the source of the variation and, if necessary, action is needed for its elimination. On the other hand, if no point plots outside the control limits we continue drawing successive samples from the output of the process to monitor the process.

EWMA and CUSUM charts

More technically sophisticated control charts than the Shewhart-type of chart have been proposed and are widely used in practice; the most popular being the exponentially weighted moving average (EWMA) and the cumulative sum (CUSUM) control charts. The EWMA and CUSUM control charts

are different from the Shewhart-type of chart in that they are memory-based charts which sequentially combine the information from multiple (past) samples with the present (or current) sample information in the decision making process. The Shewhart-type of chart, however, uses only the information available from the most recent (last) sample. For the essential theoretical underpinning of the CUSUM control chart the reader may consult the original articles by Page (1954, 1961) or the book by Hawkins and Olwell (1998). The seminal article by Roberts (1959), who introduced the EWMA chart, as well the articles by Crowder (1987, 1989) and Lucas and Saccucci (1990) provide good discussions on the EWMA chart. For an application-orientated perspective on the CUSUM and EWMA charts, the books by Montgomery (2005) and Ryan (2000) are worth reading.

Multivariate control charts

Some practical situations require the simultaneous monitoring and control of two or more related (correlated) quality characteristics. The usual practice (see Ryan, (2000) p. 253) is to monitor each characteristic separately; this results in a univariate control chart for each variable but, may be inefficient or may lead to erroneous conclusions (see Ryan, (2000) p. 254 and Montgomery, (2005) p. 486). Control charts to deal with multiple measurements (variables) were therefore developed.

The control charts for the monitoring and control of multiple variables parallel the charts for a single variable. Hence, there are multivariate extensions to the univariate Shewhart, the univariate EWMA and the univariate CUSUM charts. The corresponding multivariate charts are labeled the Hotelling's T^2 chart, the multivariate EWMA (abbreviated MEWMA) control chart and the multivariate CUSUM chart.

In this thesis, we focus on univariate control charts. An overview of multivariate control charts, which includes a discussion on the Hotelling's T^2 chart, the MEWMA chart and the multivariate CUSUM chart, can be found in Ruggeri, Kenett and Faltin (2007). For an applied and self-contained text that provides a detailed coverage of the practical and theoretical aspects of Hotelling's T^2 chart, the book by Mason and Young (2002) gives a good exposition.

Variables and Attributes control charts

A quality characteristic that can be measured on a numerical scale is called a variable. Examples include width, length, temperature, volume, speed etc. When monitoring a variable we need to monitor

both its location (i.e. mean or average) and its spread (i.e. variance or standard deviation or range). Sample statistics most commonly used to monitor the location of a process are the sample mean and the sample median or some other percentile (order statistic), whereas the sample range, the sample standard deviation and the sample variance are regularly used to monitor the process variation.

In situations where it is not practical or the quality characteristics cannot conveniently be represented numerically, we typically classify each item as either conforming or nonconforming to the specifications on the particular quality characteristic(s) of interest; such types of quality characteristics are called attributes. Some examples of quality characteristics that are attributes, are the number of nonconforming parts manufactured during a given time period or the number of tears in a sheet of material.

The *p*-chart and the *np*-chart are attribute charts that are based on the binomial distribution and are used to monitor the proportion (fraction) of nonconforming items in a sample and the number of nonconforming items in a sample, respectively. Another type of attribute chart is the *c*-chart, which is based on the Poisson distribution, and is useful for monitoring the number of occurrences of nonconformities (defects) over some interval of time or area of opportunity, rather than the proportion of nonconforming items in a sample.

A thorough bibliography of articles related to attributes control charts can be found in Woodall (1997).

Phase I and Phase II control charts

The statistical process control regime is typically implemented in two stages: Phase I (the so-called retrospective phase) and Phase II (the prospective or the monitoring phase). In Phase I, the primary interest is to better understand the process and to assess process stability; the latter step often consists of trying to bring a process in-control by analysing historical or preliminary data, locating and eliminating any assignable causes of variation. A process operating at or around a desirable level or specified target with no assignable causes of variation is said to be stable or in statistical control, or simply in-control. Once control is established to the satisfaction of the user, any unknown quantities (parameters) are estimated from the in-control data (also called reference data), leading to the setting up of control charts so that effective on-line process monitoring can begin in Phase II.

In Phase I the goal is to make sure that a process is operating at or near acceptable target(s) under some natural (common) causes of variation and that no special causes or concerns are present. Phase I analysis is usually an iterative process in which control charts play an important role. The control limits obtained early in Phase I are viewed as trial limits and are often revised and refined to ensure that the process is in-control. If target values of the parameters of interest are known (often referred to as the standards known case or Case K), one needs to ensure that the process is operating at or close to these given targets subject only to common causes of variation. If the parameters are unknown, establishing control of the process involves estimation of the parameters as well as setting up or estimating the control limits. This situation is often referred to as the standards unknown case (or Case U). Both of these situations (Case K, U) can occur in practice but Case U occurs more often, particularly when not much historical knowledge or expert opinion is available.

The decision problem under a Phase I control charting scenario is similar, in principle, to that in a multi-sample test of homogeneity problem, where one tests whether the data from various groups come from the same distribution (in-control process). Champ and Jones (2004) have noted this fact, for example. Under this motivation, the false alarm probability (FAP), i.e. the probability of at least one false alarm, is used to construct and evaluate Phase I control charts. Thus a Phase I control chart is designed by specifying a nominal false alarm probability, say FAP_0.

In Phase II the control chart is used to monitor the process on-line in order to detect the occurrence of any assignable causes of variation (such as process shifts) so that any necessary corrective actions can be taken quickly. The operation of the Phase II chart involves: (i) taking successive samples from the output of the process, (ii) calculating the specified sample statistic from each sample, and (iii) comparing the value of each sample statistic (i.e. the plotting statistic), one after the other, with the Phase II control limits. If a point plots outside the control limits an alarm signals and a search for assignable causes typically follows. Because we want the Phase II chart to signal quickly when a change takes place and not signal too often when the process is actually in-control (which is when no shift or change has taken place) the design objective in Phase II focuses on the performance of the chart (i.e. how efficient the chart is in detecting changes) and therefore concentrates on the distribution of the run-length random variable associated with the chart.

The run-length is defined as the number of samples to be collected or the number of points to be plotted on the chart before the first or next out-of-control signal is observed. The discrete random variable defining the run-length is called the *run-length random variable* and the distribution of this

random variable is called the *run-length distribution*. The characteristics of this distribution give us more insight into the performance of a chart and can be used to design a Phase II chart. Hence, in Phase II when designing the chart, we typically specify some attribute of the in-control Phase II run-length distribution to be complied with, such as the average run-length, and determine the appropriate Phase II control limits that gives the desired performance.

Parametric and Nonparametric control charts

In the process control environment of variables data (i.e. data that can be measured on a continuous numerical scale) parametric control charts are typically used; these charts are based on the assumption that the process output follows a specific distribution, for example, a normal distribution. Often this assumption cannot be verified or is not met. It is well-known that if the underlying process distribution is not normal, the control limits are no longer valid so that the performance of the parametric charts can be degraded. Such considerations provide reasons for the development and application of easy to use and more flexible and robust control charts that are not specifically designed under the assumption of normality or any other parametric distribution. Distribution-free or nonparametric control charts can serve this broader purpose.

A thorough review of the literature on nonparametric control charts can be found in Chakraborti et al. (2001, 2007). The term nonparametric is not intended to imply that there are no parameters involved, quite to the contrary. While the term distribution-free seems to be a better description of what one expects these charts to accomplish, nonparametric is perhaps the term more often used; in this thesis, both terms (distribution-free and nonparametric) are used since for our purposes they mean the same.

The main advantage of nonparametric charts is their general flexibility i.e. their application does not require knowledge of the specific probability distribution for the underlying process. In addition, nonparametric control charts are likely to share the robustness properties of the well-known nonparametric tests and confidence intervals; these properties entail, among others, that outliers and/or deviations from assumptions like symmetry far less impact them.

A formal definition of a nonparametric or distribution-free control chart could be given in terms of its run-length distribution, namely that, if the in-control run-length distribution is the same for every continuous probability distribution, the chart is called distribution-free or nonparametric (see e.g. Chakraborti et al. 2001, 2007).

1.1 Research objectives

We now turn our attention to the specific research questions studied in the remaining chapters of this thesis, which consists of Chapters 2, 3, 4 and 5. Each of Chapters 2, 3 and 4 focuses on a particular aspect of Shewhart-type Phase I and Phase II variables and attributes control charts when process parameters are estimated; these three chapters form the heart of this thesis. Chapter 5 provides a summary of the research done in this thesis and offers concluding remarks on some unanswered questions and/or future research.

1.1.1 Chapter 2

Chapter 2 focuses on Phase I Shewhart-type variables control charts to monitor the spread (i.e. the variance, the standard deviation or the range) of a process.

Consider setting up a Shewhart-type Phase I control chart for the variance or the standard deviation or the range of a process that follows a normal distribution with an unknown mean, μ, and an unknown variance, σ^2, based on the availability of m independent rational subgroups (samples) each of size n taken when the process was thought to be in-control.

Constructing a Shewhart-type Phase I control chart for a spread parameter typically entails:

(i) Estimating the unknown parameters (if they are not known or unspecified),
(ii) Calculating or estimating the Phase I control limits,
(iii) Plotting the estimated Phase I control limits and the Phase I charting statistics on the control chart, and then
(iv) Simultaneously comparing all the Phase I charting statistics with the estimated Phase I control limits.

If any of the charting statistics plot on or outside the estimated control limits, the corresponding subgroups are suspected to be from an out-of-control process. These subgroups are then examined, possibly discarded and steps (i) to (iv) are repeated. This iterative, trial-and-error process usually continues until all the remaining charting statistics plot between the latest control limits and show no non-random pattern. Once this state is reached, the remaining data are considered to be from an in-control process and this final Phase I data set (often referred to as in-control or reference data) is used to estimate the process variance or the standard deviation or the range, which is subsequently used in

setting up the Shewhart-type Phase I control charts for the mean. Note that, if a Phase I charting statistic plots on or outside the estimated Phase I limits but no assignable cause can be found that warrants its removal, it is typically not discarded. To illustrate the above methodology and the way it is currently applied in practice, consider the data of Table 1.1.

Table 1.1 displays $m = 20$ rational subgroups each of size $n = 5$ simulated from a normal distribution; for our current purpose the mean and the variance of the normal distribution(s) from which the samples were simulated are not mentioned because we assume that both these parameters are unknown. Also shown in Table 1.1 are the sample variances, S_i^2, the sample standard deviations, S_i, and the sample ranges, R_i, for $i = 1,2,...,20$. We use these data to construct Shewhart-type Phase I control charts for the variance, the standard deviation and the range. The purpose of setting up the Phase I charts is to inquire whether all 20 samples are from a normal distribution(s) with equal variances or equal standard deviations.

Table 1.1: Data for constructing Shewhart-type Phase I control charts for the variance, the standard deviation and the range

Sample number / Time (i)	X_{i1}	X_{i2}	X_{i3}	X_{i4}	X_{i5}	S_i^2	S_i	R_i
1	23.0	27.8	21.5	24.3	18.9	10.93	3.31	8.90
2	14.2	25.9	27.3	17.9	19.1	30.77	5.55	13.10
3	24.7	16.6	22.8	26.9	21.5	15.03	3.88	10.30
4	23.6	20.8	28.4	18.6	24.5	13.95	3.74	9.80
5	14.1	20.9	18.2	19.0	28.7	28.85	5.37	14.60
6	23.0	13.4	29.4	28.4	11.6	68.83	8.30	17.80
7	19.5	14.9	23.3	12.1	11.2	26.20	5.12	12.10
8	16.8	25.5	19.2	19.7	23.6	12.39	3.52	8.70
9	15.1	18.1	22.3	18.4	23.0	10.64	3.26	7.90
10	17.5	16.0	19.1	26.8	23.1	19.42	4.41	10.80
11	26.2	24.3	22.0	21.4	25.9	4.82	2.20	4.80
12	15.9	23.2	17.8	16.6	13.8	12.41	3.52	9.40
13	14.8	17.0	19.1	13.1	15.0	5.32	2.31	6.00
14	13.8	18.3	25.0	18.2	18.5	16.03	4.00	11.20
15	28.2	23.2	16.6	18.8	18.7	21.53	4.64	11.60
16	12.9	20.0	32.2	16.4	26.1	59.47	7.71	19.30
17	22.0	11.9	21.5	21.1	17.9	17.80	4.22	10.10
18	21.1	19.4	16.3	21.8	14.3	10.23	3.20	7.50
19	16.2	21.4	25.5	14.2	28.0	34.67	5.89	13.80
20	12.5	17.2	17.9	14.4	16.5	4.92	2.22	5.40

Phase I S^2 chart

First, consider constructing a Shewhart-type Phase I S^2 control chart for the variance. In this case the unknown process variance, σ^2, is estimated using the unbiased pooled variance estimator

$$S_p^2 = \frac{1}{m}\sum_{i=1}^{m} S_i^2 \qquad (1\text{-}1)$$

where $S_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_{ij} - \overline{X}_i)^2$ for $i = 1,2,...,m$ denotes the i^{th} sample variance.

The charting statistics for the S^2 chart are the sample variances S_i^2, $i = 1,2,...,m$ and the estimated Phase I probability limits (see e.g. Montgomery, (2005) p. 231) are

$$U\hat{C}L = \frac{S_p^2 \chi^2_{\alpha,n-1}}{n-1} \qquad \hat{C}L = S_p^2 \qquad L\hat{C}L = \frac{S_p^2 \chi^2_{1-\alpha,n-1}}{n-1} \qquad (1\text{-}2)$$

where $\chi^2_{\alpha,n-1}$ is the $100(1-\alpha)^{th}$ percentile of the chi-square distribution with $n-1$ degrees of freedom. Note that, typically one takes $\alpha = 0.00135$ and finds the chi-square percentiles so that the probability that a single point plotting outside the control limits is 0.0027 for any sample.

For the data in Table 1.1 we find that

$$S_p^2 = \frac{1}{20}\sum_{i=1}^{20} S_i^2 = \frac{1}{20}(10.93 + 30.77 + ... + 4.92) = 21.21.$$

Taking $\alpha = 0.00135$ with $n-1 = 5-1 = 4$ we calculate (using MS Excel) that $\chi^2_{0.99865,4} = 0.1058$ and $\chi^2_{0.00135,4} = 17.8004$; substituting these values of the percentiles and $S_p^2 = 21.21$ in (1-2) yields the values of the estimated Phase I control limits i.e.

$$U\hat{C}L = \frac{21.21 \times 17.8004}{4} = 94.38 \qquad \hat{C}L = 21.21 \qquad L\hat{C}L = \frac{21.21 \times 0.1058}{4} = 0.561.$$

The corresponding Phase I S^2 chart is shown in Figure 1.2. Because all the sample variances, i.e. S_i^2 for $i = 1,2,...,20$, displayed in Table 1.1, plot between the estimated control limits the process variance is considered to be in-control. Essentially, this decision implies that the underlying population

variances (from which the samples were obtained) are not significantly different but, as will be pointed out later, this conclusion might be wrong because of the fact that multiple comparisons (between the charting statistics and the same set of estimated control limits) are to be dealt with is not taken into account in making the in-control or not-in-control decision.

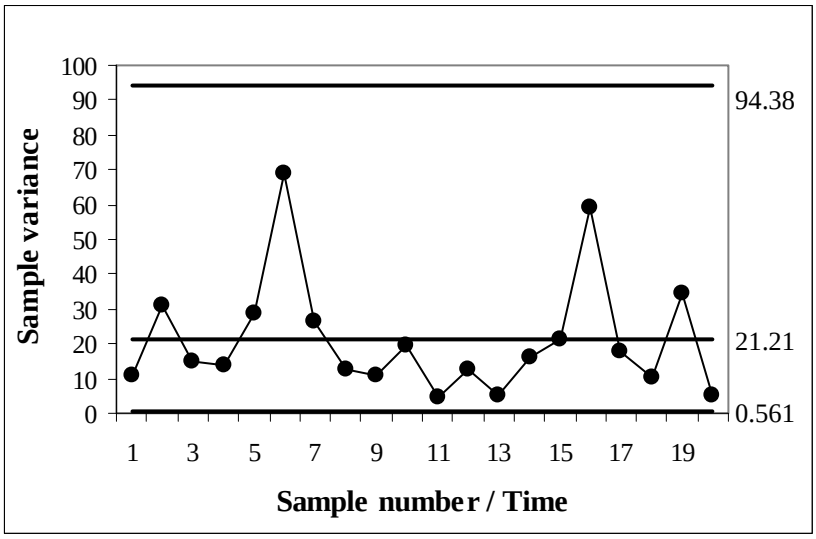


Figure 1.2: The Shewhart-type Phase I S^2 control chart for the data in Table 1.1

Phase I S chart

Next, consider setting up a Shewhart-type Phase I S control chart for the data in Table 1.1. In this case, the unknown process standard deviation, σ, is estimated using the unbiased point estimator

$$\hat{\sigma}_S = \frac{\bar{S}}{c_4} = \frac{1}{c_4}\left(\frac{1}{m}\sum_{i=1}^{m} S_i\right) \tag{1-3}$$

where $S_i = \sqrt{S_i^2}$ denotes the i^{th} sample standard deviation and c_4 denotes the unbiasing constant, which is tabulated, for example, in Appendix VI of Montgomery (2005).

The charting statistics for the S chart are the sample standard deviations, i.e. S_i, for $i = 1,2,...,m$. The estimated *k*-sigma control limits and the estimated centerline of the Phase I S chart are

$$U\hat{C}L = \bar{S} + k\frac{\bar{S}}{c_4}\sqrt{1-c_4^2} \qquad \hat{C}L = \bar{S} \qquad L\hat{C}L = \bar{S} - k\frac{\bar{S}}{c_4}\sqrt{1-c_4^2} \tag{1-4}$$

where the charting constant, k, is typically set equal to 3 so that we can write

$$U\hat{C}L = B_4\bar{S} \qquad \hat{C}L = \bar{S} \qquad L\hat{C}L = B_3\bar{S} \tag{1-5}$$

where $B_3 = 1 - \frac{3}{c_4}\sqrt{1-c_4^2}$ and $B_4 = 1 + \frac{3}{c_4}\sqrt{1-c_4^2}$ are constants and tabulated, for example, in Appendix VI of Montgomery, (2005).

For the data of Table 1.1 we get

$$\bar{S} = \frac{1}{20}\sum_{i=1}^{20} S_i = \frac{1}{20}(3.31 + 5.55 + ... + 2.22) = 4.317$$

and find that the charting constants, for $n = 5$, are $B_4 = 2.089$ and $B_3 = 0$.

We find the estimated 3-sigma control limits for the Phase I S chart by substituting $\bar{S} = 4.317$, $B_4 = 2.089$ and $B_3 = 0$ in (1-5), which gives

$$U\hat{C}L = 9.018 \qquad \hat{C}L = 4.317 \qquad L\hat{C}L = 0.$$

The corresponding Phase I S chart is shown in Figure 1.3. The points plotted on the chart are the twenty sample standard deviation i.e. S_i for $i = 1,2,...,20$, of Table 1.1. Because none of the points plots outside the control limits, the process standard deviation is deemed to be in-control.

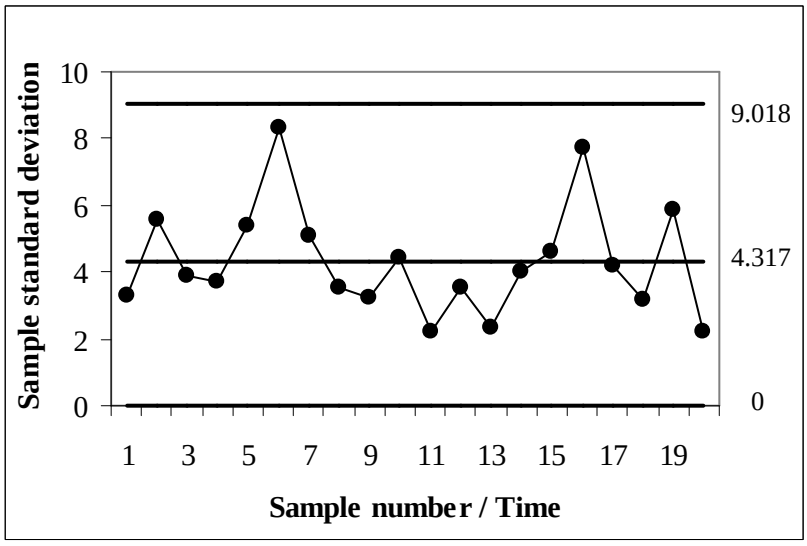


Figure 1.3: The Shewhart-type Phase I S control chart for the data in Table 1.1

Phase I R chart

Lastly, consider the R chart. This chart is popular in practice since the range is easy to calculate and it is known that for small samples, the range is an efficient estimator of the standard deviation of a normal distribution.

In case of the R control chart, the unknown process standard deviation, σ, is estimated using the unbiased point estimator

$$\hat{\sigma}_R = \frac{\overline{R}}{d_2} = \frac{1}{d_2}\left(\frac{1}{m}\sum_{i=1}^{m} R_i\right) \tag{1-6}$$

where $R_i = \max(X_{ij}) - \min(X_{ij})$, $j = 1,2,...,n$ is the i^{th} sample range and d_2 is an unbiasing constant which is tabulated, for example, in Appendix VI of Montgomery (2005).

For the Phase I R chart the charting statistics are the sample ranges i.e. R_i for $i = 1,2,...,m$, and the estimated *k*-sigma limits and the estimated centerline are (see e.g. Montgomery, (2005) p. 197 and p. 198)

$$U\hat{C}L = \left(1 + k\frac{d_3}{d_2}\right)\overline{R} \qquad \hat{C}L = \overline{R} \qquad L\hat{C}L = \left(1 - k\frac{d_3}{d_2}\right)\overline{R} \tag{1-7}$$

where d_3 is a known function of n (see e.g. Montgomery, (2005) p.198). In routine applications, the charting constant k is set equal to 3, which leads to a simpler representation of the estimated control limits of the R chart i.e.

$$U\hat{C}L = D_4\overline{R} \qquad \hat{C}L = \overline{R} \qquad L\hat{C}L = D_3\overline{R} \tag{1-8}$$

where $D_3 = 1 - 3\frac{d_3}{d_2}$ and $D_4 = 1 + 3\frac{d_3}{d_2}$ are constants and tabulated, for example, in Appendix VI of Montgomery, (2005).

For the data of Table 1.1 it is calculated that

$$\overline{R} = \frac{1}{20}\sum_{i=1}^{20} R_i = \frac{1}{20}(8.90 + 13.10 + ... + 5.40) = 10.66$$

and that $D_3 = 0$ and $D_4 = 2.114$; using these values the estimated control limits and the estimated centerline of the R chart are calculated using (1-8) and found to be

$$U\hat{C}L = 22.52 \qquad \hat{C}L = 10.66 \qquad L\hat{C}L = 0.$$

The Phase I R chart is shown in Figure 1.4. The points plotted on the chart are the sample ranges i.e. R_i for $i = 1,2,...,20$ listed in the last column of Table 1.1. Like the Phase I S^2 chart and the Phase I S chart, there is no indication that the process spread is out-of-control. One would thus typically proceed with setting up the Shewhart-type Phase I $\overline{X}$ for the mean as described by Champ and Jones (2004).

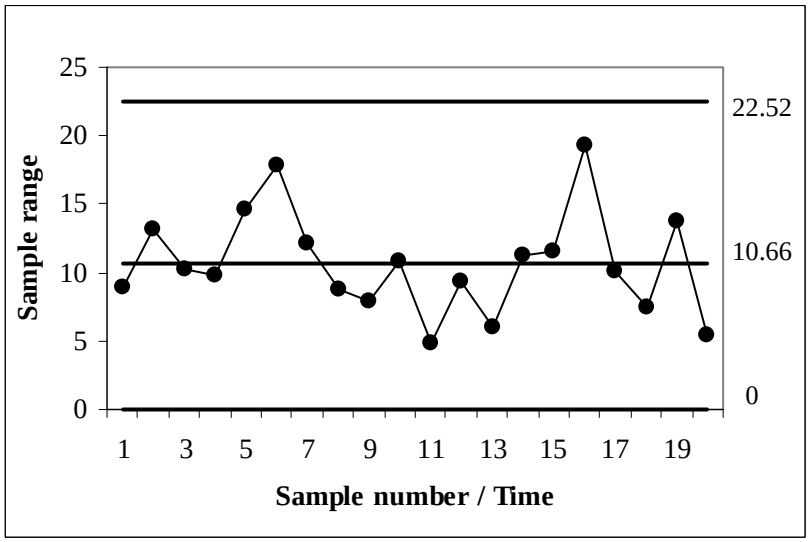


Figure 1.4: The Shewhart-type Phase I S control chart for the data in Table 1.1

There are a number of problems in setting up the Phase I control charts in the usual manner as described above. These problems are:

(i) The m charting statistics are simultaneously compared to the estimated control limits, which are functions of the estimated parameters and are therefore random variables themselves (this was indicated by the $\hat{}$- notation; read *hat-notation*). Since the charting statistics and the control limits are obtained using the same data, successive comparisons (over subgroups) of the charting statistics with the estimated control limits are dependent events. The signaling events (defined as the event when a charting statistic plots on or outside the control limits) for the i^{th} and the j^{th} subgroups (where $i \neq j = 1,2,...,m$) are therefore statistically dependent.

Thus, in order to correctly design a Phase I control chart in the unknown parameter case, both the dependence of the signaling events and the multiple nature of the comparisons inherent in the decision process must be taken into account. Both of these considerations require a certain

joint probability distribution and this joint distribution (and the associated manipulations thereof) are at the heart of the study of a Phase I control chart (which is done in Chapter 2).

(ii) The estimated control limits of the Phase I S^2, S and R charts ignores the dependency between the signaling events and are incorrectly calculated in such a way as to ensure that the false alarm rate (denoted FAR and defined as the probability for a single charting statistic to plot outside the control limits when the process is in-control) is approximately 0.0027. Given the inherently repetitive nature of a Phase I analysis and the fact that the charting statistics from all the subgroups are simultaneously compared with the same estimated control limits, using the *FAR* to design a Phase I chart is not a good idea since this naturally inflates the FAP i.e. the probability that at least one charting statistic plots outside the estimated control limits when the process is in-control.

The following example illustrates this problem in the context of the Shewhart-type $\overline{X}$ control chart in Case K: If there are 15 samples and one uses the traditional 3-sigma control limits for setting up a Phase I chart for the mean, $\overline{X}_i$, when standards are known (i.e. mean of μ_0 and variance equal to σ_0^2), the *FAR* is equal to

$$\begin{aligned} FAR &= 1 - \Pr(LCL < \overline{X}_i < UCL \mid IC) \\ &= 1 - \Pr(\mu_0 - 3\sigma_0/\sqrt{n} < \overline{X}_i < \mu_0 + 3\sigma_0/\sqrt{n} \mid IC) \\ &= 0.0027 \end{aligned}$$

for each sample, which is at a commonly desirable level, but the *FAP* is equal to

$$\begin{aligned} FAP &= \Pr(\text{ At least one false alarm}) \\ &= 1 - \Pr(\text{ No false alarm}) \\ &= 1 - (1 - 0.0027)^{15} \\ &= 0.0397 \end{aligned}$$

which may be deemed rather large. Thus the recommendation is to determine the Phase I control limits so that the FAP is controlled at some desirable (nominal) small value.

(iii) The estimated k-sigma control limits of the S chart and the R chart are based on the tacit assumption that the sampling distributions of the sample standard deviation and the sample range are symmetric. It is well-known that this is not the case; in fact, the sampling distributions are asymmetric.

By using the relevant joint distribution of the charting statistics to calculate the charting constants this common mistake can be corrected.

The above-mentioned problems with regard to the construction of the Phase I control charts for the variance, the standard deviation and the range lead to the question:

> How should one calculate the control limits of these three Phase I charts so that, when one simultaneously compares all m the charting statistics with the corresponding control limits, the probability that at least one point plots outside the limits, if the process is in-control, is equal to a nominal (desired) value?

This question is answered in Chapter 2 where we specifically study and design the Phase I S^2, S and R control charts assuming that the mean and the variance are both unknown and are estimated on the basis of m independent rational subgroups each of size n available from a normally distributed process. The derivations recognize that in Phase I (with unknown parameters) the signaling events are dependent and that more than one comparison is made against the same estimated limits simultaneously and leads to working with the joint distribution of a set of random variables. Using intensive computer simulations, tables are provided for the charting constants for each chart for a given false alarm probability of 0.01, 0.05 and 0.10, respectively.

In view of the problems currently associated with setting up Phase I control charts for the variance, the standard deviation and the range, an extensive overview of the literature on Shewhart-type Phase I parametric control charts for univariate variables data is presented assuming that the form of the underlying continuous distribution is known. The overview not only presents the current state of the art but also points out what challenges still remain.

1.1.2 Chapter 3

Chapter 3 focuses on the Phase II Shewhart-type *p*-chart and *c*-chart with unknown parameters. For completeness we also study the statistical properties of these charts assuming that the parameters are known.

Consider constructing a Phase II attributes *p*-chart or a Phase II attributes *c*-chart for the situation when the process parameters p and c are unknown and estimated from an in-control reference sample following a Phase I analysis.

The setting up of the Phase II *p*-chart and the Phase II *c*-chart, in general, entails:

(i) Obtaining a point estimate of the unknown process parameter based on the in-control Phase I data,

(ii) Estimating the Phase II control limits, and then

(iii) Comparing each Phase II charting statistic, one at a time and based on new incoming samples or inspection units, with the estimated Phase II control limits.

As long as no Phase II charting statistic plots on or outside the control limits, we continue to draw successive samples from the process output and monitor the process. However, as soon as a charting statistics plots on or outside the estimated Phase II control limits, we stop the charting procedure, declare the process out-of-control and start a search for assignable causes. To illustrate the steps outlined in (i) to (iii) listed above we look at an example based on the data of Table 1.2 next; this example demonstrates how the Phase II attributes *p*-chart is typically implemented in practice.

Column 1 of Table 1.2 lists the sample numbers; these range from 1 to 25. The X_i values in column 2 were obtained via simulation from a binomial distribution with parameters $n = 50$ and p. Note that, because we assume that the fraction nonconforming is unknown the value of p is not mentioned here. In this simulated scenario we can assume that these X_i's represent the number of nonconforming items in 25 consecutive Phase II samples, each of size 50, taken from the output of some process. The corresponding observed fractions nonconforming, i.e. $p_i = X_i / 50$, are displayed in column 3. To illustrate the approach outlined in steps (i) to (iii) listed above, assume that p was estimated from a Phase I study and found to be 0.175. Using this point estimate of the unknown

fraction nonconforming, typically denoted as $\bar{p}$, and the data from Table 1.2 we can construct a Shewhart-type Phase II p-chart.

Table 1.2: Data for constructing Shewhart-type Phase II p-chart for monitoring the fraction nonconforming items in samples of size $n = 50$

Sample number / Time (i)	Number of nonconforming items, X_i	Sample fraction nonconforming $p_i = X_i / 50$
1	14	0.28
2	8	0.16
3	12	0.24
4	9	0.18
5	12	0.24
6	13	0.26
7	11	0.22
8	10	0.20
9	16	0.32
10	10	0.20
11	7	0.14
12	10	0.20
13	11	0.22
14	14	0.28
15	9	0.18
16	4	0.08
17	10	0.20
18	8	0.16
19	12	0.24
20	7	0.14
21	11	0.22
22	10	0.20
23	10	0.20
24	13	0.26
25	9	0.18

In case of the Phase II p-chart, the estimated 3-sigma control limits and center line are

$$U\hat{C}L = \bar{p} + 3\sqrt{\bar{p}(1-\bar{p})/n} \qquad \hat{C}L = \bar{p} \qquad L\hat{C}L = \bar{p} - 3\sqrt{\bar{p}(1-\bar{p})/n}$$

where $\bar{p}$ denotes the point estimate of the unknown fraction nonconforming, p, obtained at the end of a successful Phase I analysis and n denotes the sample size (see e.g. Montgomery, (2005) p. 269). Note that, if the estimated lower control limit turns out to be negative, it is adjusted upward and set equal to zero.

The Phase II charting statistics are the fractions nonconforming in the samples, i.e. p_i for $i = 1,2,\ldots, 25$ calculated from successive Phase II samples taken from the output of the process.

Based on the point estimate $\bar{p} = 0.175$ the estimated 3-sigma control limits and centerline, for our example, are

$$U\hat{C}L = 0.175 + 3\sqrt{0.175(0.825)/50} = 0.3362$$

$$\hat{C}L = 0.175$$

$$L\hat{C}L = 0.175 - 3\sqrt{0.175(0.825)/50} = 0.0138.$$

The Phase II p - chart is shown in Figure 1.5. The points that are plotted on the control chart are the p_i's from column 3 of Table 1.2. Note that, unlike the Phase I charts discussed earlier, each charting statistic of a Phase II control chart is plotted one at a time as soon as it is calculated from the most recent (i.e. the latest or last) sample taken from the output of the process; this typically happen in real-time. Because none of the points plot outside the control limits, the process is deemed to be in-control and we can continue to draw successive samples from the process output and monitor the process over time.

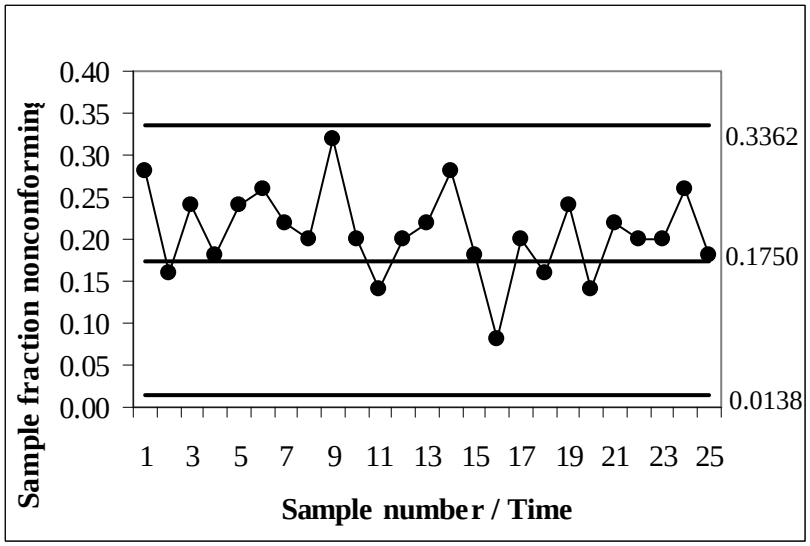


Figure 1.5: The Shewhart-type Phase II p -chart for the data in Table 1.2 with $\bar{p}$ = 0.1750

There is a major concern in setting up the Phase II *p*-chart in the usual manner as described above:

The point estimate $\bar{p}$ influences the performance of the Phase II *p*-chart and this influence is typically not taken into account when setting up a Phase II chart. To illustrate how significant the influence of $\bar{p}$ can be, suppose, for example, that a different Phase I sample was used to estimate p and that $\bar{p} = 0.16$, that is, $\bar{p} \neq 0.1750$. Under these circumstances the estimated Phase II control limits would be

$$U\hat{C}L = 0.16 + 3\sqrt{0.16(0.84)/50} = 0.3155$$

$$\hat{C}L = 0.16$$

$$L\hat{C}L = 0.16 - 3\sqrt{0.16(0.84)/50} = 0.0045$$

and the Phase II *p*-chart, based on the data in column 3 of Table 1.2, with these estimated control limits are shown in Figure 1.6. It is observed that, with the point estimate of $\bar{p} = 0.16$, the estimated control limits are narrower (than those in Figure 1.5) and that the chart signals on the 9th sample indicating that the process is out-of-control. Thus, with a different Phase I sample and/or a change in the value of the point estimate, $\bar{p}$, the Phase II *p*-chart can lead to a different decision regarding the state of the process. The same concern appears in application of the Phase II *c*-chart.

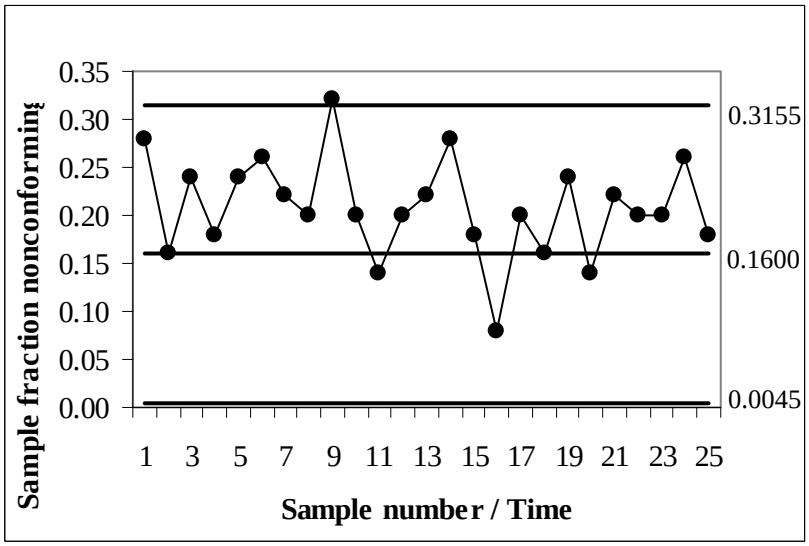


Figure 1.6: The Shewhart-type Phase II *p*-chart for the data in Table 1.2 with $\bar{p} = 0.1600$

The questions that emanate from the above-mentioned concern regarding the application of the Phase II *p*-chart and the Phase II *c*-chart is:

> How large should the Phase I reference samples (used to estimate the unknown parameters p and c) be so that the performance of the charts when the parameters are unknown and estimated, is comparable to their performance when the parameters are known? Are the widely-followed empirical guidelines (i.e. $m = 20$ or 25 with $n = 4$ or 5 to estimate the unknown parameters) reasonable?

To answer these questions, we investigate the effect of estimating the unknown parameters p and c on the performance of the charts in detail in Chapter 3. To do this, we derive and evaluate expressions for the run-length distributions of the Phase II Shewhart-type *p*-chart and the Phase II Shewhart-type *c*-chart when the parameters are estimated. We then examine the effect of estimating p and c on the performance of the *p*-chart and the *c*-chart via their run-length distributions and associated characteristics such as the average run-length, the false alarm rate and the probability of a "no-signal".

An exact approach based on the binomial and the Poisson distributions is used to derive expressions for the Phase II run-length distributions and the related Phase II characteristics using expectation by conditioning (see e.g. Chakraborti, (2000)). We first obtain the characteristics of the run-length distributions conditioned on point estimates from Phase I and then find the unconditional characteristics by averaging over the distributions of the point estimators.

Next, the in-control and the out-of-control properties of the charts are looked at. The results are used to discuss the appropriateness of the widely followed empirical rules for choosing the size of the Phase I sample, used to estimate the unknown parameters; this includes both the number of reference samples m and the sample size n.

1.1.3 Chapter 4

Chapter 4 focuses on improving some of the existing nonparametric control charts and designing new distribution-free charting procedures.

Consider the situation where monitoring the location parameter, θ, of a quality characteristic (that is measured on a continuous numerical scale) with an unknown continuous cumulative distribution function is of interest. In such a case the usual control chart procedures that are based on a particular parametric distribution are not appropriate and a distribution-free (or nonparametric) control chart procedure would be more useful.

If θ is the median and it is required to monitor whether or not θ changes (i.e. moves away) from its specified or known value, θ_0 (say), one can construct a control chart that uses the well-known sign test statistic as a charting statistic. A Shewhart-type control chart based on the sign test statistic was studied by Amin, Reynolds and Bakir (1995) and is known as the sign chart. Using the sign chart entails that one:

(i) Takes successive samples of size n from the process output,

(ii) Calculates the number of observations greater than or equal to the specified value, θ_0, within each sample, T_i (say), and then

(iii) Compares each charting statistic, T_i, one at a time, with appropriately chosen control limits.

The sign chart signals, like any other typical Shewhart-type control chart, if a single point plots on or outside the control limits, that is, on or below the lower control limit or, on or above the upper control limit.

The sign chart is easy to apply in practice and it requires the minimum number of assumptions (namely it only requires that the underlying process distribution be continuous and that a specified value, θ_0, for the location parameter, θ, is available) but, it has a serious shortcoming. For any sample size, n, the number of possible in-control average run-length values (ARL_0's) to choose from when designing the chart is limited and, furthermore, the maximum attainable ARL_0, for a two-sided chart, is 2^{n-1}. For $n=5$, which is often the recommended sample size used in practice, the maximum

ARL_0 is $2^{5-1}=16$; the corresponding false alarm rate is $1/ARL_0 = 1/16 = 0.0625$. Such a small ARL_0 or, alternatively, such a large FAR, implies that the chart would signal (erroneously) more often than a typical 3-sigma Shewhart-type chart and would lead to deteriorated performance of the charting procedure.

To allow the practitioner more flexibilily in designing the sign chart, that is, to have a wider range of ARL_0's and FAR's to choose from, in this thesis we enhance the sign chart by proposing new runs-type signaling rules (i.e. decision rules). These signaling rules are based on runs of the charting statistics outside the control limits. Similar signaling rules were successfully used, for example, by Chakraborti and Eryilmaz (2007) to solve a similar weakness of the signed-rank chart introduced by Bakir (2004). In addition to the signaling rules, we further improve the sign chart of Amin, Reynolds and Bakir (1995) and consider the situation where it is required to monitor percentiles other than the median.

If expert knowledge is not available and a value for the location parameter, θ, can not be specified one can not use the sign chart; this situation requires a different nonparametric control chart i.e. one requires a nonparametric control chart that can handle the scenario where monitoring a continuous random variable with an unknown cumulative distribution function and an unknown or an unspecified value for the location parameter is of interest. Chakraborti, Van der Laan and Van de Wiel (2004) considered a class of nonparametric control charts capable of solving this problem. Their charts, called precedence charts, are based on the two-sample median test statistic, which requires the availability of an in-control Phase I reference sample from which to estimate the control limits.

To construct a precedence chart entails:

(i) Arranging the observations $X_1, X_2, ..., X_m$ from the in-control Phase I reference sample of size m in ascending order i.e. $X_{1:m}, X_{2:m}, ..., X_{m:m}$, where $X_{i:m}$ is the i^{th} smallest observation in the group of m;

(ii) Estimating the lower control limit and the upper control limit by $\hat{LCL} = X_{a:m}$ and $\hat{UCL} = X_{b:m}$, where $X_{a:m}$ and $X_{b:m}$ (with $1 \le a < b \le m$) are suitably selected order statistics from the Phase I sample;

(iii) Obtaining new incoming Phase II samples each of size n (independently from one another and from the Phase I sample);

(iv) Calculating the charting statistic, $Y_{j:n}$, which is the j^{th} order statistic of the Phase II sample and depends on the quantile being monitored, and then

(v) Comparing each $Y_{j:n}$, one at a time, with the estimated control limits.

The precedence chart signals if a single point plots on or outside the control limits.

In this thesis we extend the precedence charts of Chakraborti et al. (2004) by incorporating the same signaling rules (or tests), involving runs of the charting statistic, that we used with the sign chart. These signaling rules, as mentioned earlier, are appealing because Chakraborti and Eryilmaz (2007) showed that when the in-control median is specified (and it is not necessary to estimate the median) the incorporation of similar rules provide more practical and powerful control charts. Similar extensions have been considered in the literature for Shewhart-type control charts in the case of the normal distribution (see e.g. Nelson, (1984), Klein, (2000) and Shmueli and Cohen, (2003)).

To summarize:

> In Chapter 4 a new class of nonparametric control charts with runs-type signaling rules for the situations where the location parameter of the distribution is known and unknown is considered. In the former situation the charts are based on the sign test statistic and enhance the sign chart proposed by Amin et al. (1995); in the latter situation the charts are based on the two-sample median test statistic and improve the precedence charts by Chakraborti et al. (2004).

To design the nonparametric control charts and study their performance, their run-length distributions are required. The run-length distributions and the associated performance characteristics for the "runs rule enhanced" charts are derived by using a Markov chain approach (see e.g. Fu and Lou, (2003)) and, in some cases, we also draw on the results of the geometric distribution of order k (see e.g. Balakrishnan and Koutras, (2002), Chapter 2). To implement the charts in practice we provide tables with the necessary charting constants and/or control limits and examples are given to illustrate the application and usefulness of the charts.

Lastly, the in-control and the out-of-control performance of the new distribution-free charts are studied and compared to the existing nonparametric charts, using the average run-length, the standard deviation of run-length, the false alarm rate and some percentiles of the run-length, including the median run-length. It is shown that the newly developed runs rules enhanced sign charts offer more practically desirable in-control average run-lengths and false alarm rates than the sign chart of Amin, Reynolds and Bakir (1995) and the precedence charts of Chakraborti, Van der Laan and Van de Wiel (2004) and, perform better than the Shewhart $\overline{X}$ chart and a number of existing nonparametric charts for some distributions.

Layout of the Thesis

The rest of this thesis is structured as follows. In Chapter 2 we look at Phase I variables control charts; this includes the design of the Phase I S^2, S and R charts and an in-depth overview of the literature on Phase I parametric control charts for univariate variables data. In Chapter 3 we study the Phase II Shewhart-type *p*-chart and the Phase II Shewhart-type *c*-chart. In Chapter 4 we design a new class of nonparametric Shewhart-type control charts with runs-type signaling rules (i.e. runs of the charting statistics above and below the control limits) for the scenarios where the percentile of interest of the distribution is either known or unknown. Lastly, in Chapter 5 we wrap up this thesis with a summary of the research carried out and offer concluding remarks concerning unanswered questions and/or future research opportunities. In Chapter 5 we also list the research outputs related to and based on this thesis; this list includes the details of the technical reports and the peer-reviewed articles that were published, the articles that were accepted for publication, the local and the international conferences where papers were presented and the draft articles that were submitted and are currently under review.

Chapter 2

Variables control charts: Phase I

2.0 Chapter overview

Introduction

In practice the statistical process control (SPC) regime is implemented in two phases: Phase I, the so-called retrospective phase, and Phase II, the prospective or the monitoring phase (see e.g. Woodall, (2000)).

In Phase I the primary interest is to better understand the process and to assess process stability. The latter step consists of trying to bring a process in-control by analyzing historical data in order to locate and eliminate assignable causes of variation. A process operating at or around a desirable level or specified target with no assignable causes of variation is said to be stable, or in statistical control, or simply in-control (IC).

Montgomery (2005) p. 199 describes the process of establishing control in Phase I as iterative and that the control limits are viewed as trial limits. Once statistical control is established to the satisfaction of the user, any unknown quantities or parameters are estimated from the in-control data which leads to the setting-up of control charts so that effective process monitoring can begin in Phase II.

In addition to the use of various exploratory (e.g. graphical) and confirmatory (e.g. testing of hypotheses) statistical tools, control charts play a crucial role in a Phase I analysis. They help in getting a better view of what is going on over time and assist in diagnosing the source(s) of assignable causes so that their effect can be minimized or removed.

Motivation

The success of process monitoring in Phase II depends critically on the success of the corresponding Phase I analysis. In this regard, the effects of parameter estimation based on Phase I reference data on the performance of Phase II control charts have been studied by several authors (see e.g. Jensen et al. (2006) for an overview). These studies emphasize the importance for a proper understanding of the issues while setting accurate Phase I control limits.

The most familiar control charts in practice include those for the mean and the spread i.e. the variance and/or the standard deviation, of an assumed (at least approximately) normally distributed process. While Champ and Jones (2004) studied the Shewhart-type Phase I $\bar{X}$ chart for the mean, we study and design Shewhart-type Phase I S^2, S and R charts for the process variance and/or standard deviation. The spread charts are particularly important since an estimate of the variance or the standard deviation is usually necessary in setting up the control chart for the mean. Thus, the spread of the process must be monitored and controlled before (or at least simultaneously) attempting to monitor the mean.

Despite the fact that Phase I analysis is such an important component of SPC, not all authors make a clear distinction between Phases I and II or discuss the various ramifications in the current teaching and practice of SPC. Moreover, although several authors studied some statistical aspects of Phase I control charting methods, a search of the standard textbooks on SPC methods (with some exceptions, such as Montgomery, (2005) p. 199) did not reveal much, if any, discussion of this important topic. It would therefore be helpful and beneficial for researchers, instructors (educators) and practitioners to know what the issues are, what the present state of the art is and what challenges still remain. To this end, an overview of the literature on univariate parametric Shewhart-type Phase I variables control charts for the mean and the variance is given, under the assumption that the form of the underlying continuous process distribution is known.

Methodology

A key to the Phase I analysis, as Champ and Jones (2004) stated, "requires a different paradigm than studying the prospective monitoring of a process". One implication of this statement is that the metric of a control chart's performance must be carefully chosen depending on which phase of the analysis (I or II) one is referring to.

Because Phase I control charting is about ensuring that a process is in-control, it is in principle similar to a multi-sample hypothesis testing problem for homogeneity that tests if the data from several independent samples come from the same (in-control) distribution or process. With this motivation, the false alarm probability, denoted FAP, is the criterion typically used to measure and evaluate control limits in Phase I. The FAP is defined as the probability of at least one false alarm (signal) when the process is actually in-control.

The Phase I S^2, S and R control charts that are developed in this chapter are designed and/or implemented by specifying a nominal false alarm probability, say FAP_0, and then determining the charting constants (Phase I control limits) so that the FAP is less than or equal to the FAP_0. The derivations take into account that the signaling events (when a charting statistic falls outside either of the control limits) are dependent and use the relevant joint probability distribution of the Phase I charting statistics while computing the FAP.

Layout of Chapter 2

We begin with a general discussion of Phase I, in which we describe the goals and discuss some of the standard methods for designing and implementing Shewhart-type Phase I charts; this is done Section 2.1. The design of the Shewhart-type Phase I S^2, S and R charts are then studied in Section 2.2. This is followed by an overview of the literature on univariate parametric Shewhart-type Phase I control charts for the mean and the spread of variables data in Section 2.3. Finally, we conclude Chapter 2 with a summary and recommendations for future research.

2.1 Phase I SPC

Introduction

Much of the preliminary statistical analysis is done in Phase I. This includes planning, administration, design of the study, data collection, data management, exploratory work (including graphical and numerical analysis, goodness-of-fit analysis, and so on) to ensure that the process is truly in a state of statistical control (see e.g. Woodall, (2000) and Montgomery, (2005) p. 168 and p. 199). The goal is to make sure that a process is operating at or near an acceptable target(s) under some natural or common causes of variation and that no special causes or concerns are present. In this regard Phase I control charts play an important role. While Champ and Jones (2004) studied the Shewhart-type Phase I $\bar{X}$ chart for the mean, we study and design Shewhart-type Phase I S^2, S and R charts for the process variance and/or standard deviation.

Case K and Case U

If target values of the parameters of interest are known, one needs to ensure that the process is operating at or around these given targets subject only to common causes of variation. This situation is referred to as the "Standards Known Case" and denoted Case K.

If the parameters are unknown, establishing control involves estimation of the parameters and the control limits; this causes that both the charting statistics and the control limits of a Phase I chart are random variables. This situation is referred to as the "Standards Unknown Case" and denoted Case U.

Although both of these situations can occur in practice, Case U occurs more often, particularly when not much historical information or expert opinion is available.

Phase I control charting

At the beginning of a Phase I analysis it is assumed that m independent random samples or rational subgroups are available, each of size $n > 1$, taken sequentially over time from a process with a continuous cumulative distribution function (c.d.f) $F(x;\underline{\boldsymbol{\theta}})$, where F is a known function and $\underline{\boldsymbol{\theta}} = (\theta_1, \theta_2, ..., \theta_k)$, $k \geq 1$, is a vector of unknown parameters. Symbolically, we write $X_{ij} \sim iidF(x;\underline{\boldsymbol{\theta}})$

where iid denotes "independently and identically distributed", and X_{ij} denotes the j^{th} observation in the i^{th} sample for $i = 1,2,...,m$ and $j = 1,2,...,n$.

Generally speaking, depending on the parameter of interest, we calculate a charting statistic C_i for $i = 1,2,...,m$ from each subgroup and calculate the point estimates $\hat{\underline{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_k)$ using the mn individual observations combined (pooled). Using statistical distribution theory and some given performance criterion, an estimated lower and upper control limit, denoted

$$\hat{LCL} = g_1(\hat{\underline{\theta}}) \qquad \text{and} \qquad \hat{UCL} = g_2(\hat{\underline{\theta}})$$

are then obtained, where g_1 and g_2 are two specified functions of $\hat{\underline{\theta}}$ such that $\hat{LCL} < \hat{UCL}$.

A plot of the charting statistics (from all m the subgroups) together with the estimated control limits constitutes the Phase I control chart.

If all m the charting statistics plot between the control limits and no systematic pattern is present, the process is considered to be in control (IC). On the contrary, if any one or more of the C_i's fall on or outside the estimated control limits, the process is declared to be out-of-control (OOC) and some action or intervention is required. This entails, for example, that the OOC samples are re-examined, possibly discarded and the remaining samples are then re-checked for control. Revised values are subsequently obtained for the estimators as well as the control limits from the remaining samples and the corresponding charting statistics are then plotted against the revised limits.

This iterative trial-and-error process continues until all the charting statistics plot inside the latest control limits for the samples at hand. Once this state is reached, the remaining data are thought to be from an in-control process and this final Phase I data set is used to find appropriate control limits for Phase II monitoring of the process. This final Phase I data is referred to as in-control data or a set of reference data.

Note that, if at any stage during Phase I control charting it happens that some of the Phase I charting statistics plot outside the estimated control limits but no assignable cause(s) can be found that justify their removal, the process may be considered in-control and the observations from these samples are then included in the reference data.

Remark 1

(i) Intuitively, the charting statistic C_i for the i^{th} sample is taken to be an efficient estimator of the parameter of interest.

For example, if it is assumed that the underlying process distribution is normal with an unknown mean μ and an unknown variance σ^2, the data would be represented as $X_{ij} \sim iidN(\mu, \sigma^2)$ for $i = 1,2,...,m$ and $j = 1,2,...,n$; thus, if the unknown process mean μ is the parameter of interest, the i^{th} sample mean $\bar{X}_i$ (the best estimator) is a natural charting statistic.

(ii) Estimation of the parameters is an important step in setting-up control charts in Case U. Unbiased estimators are preferred and if more than one such estimator is available, the minimum variance unbiased (MVU) estimators should be used.

For example, if it is assumed that $X_{ij} \sim iidN(\mu, \sigma^2)$ for $i = 1,2,...,m$ and $j = 1,2,...,n$ where both μ and σ^2 are unknown, it is common practice to use the overall mean of the pooled sample

$$\bar{\bar{X}} = \frac{1}{m}\sum_{i=1}^{m}\bar{X}_i = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}X_{ij}$$

to estimate μ and use the pooled variance

$$\bar{V} = \frac{1}{m}\sum_{i=1}^{m}S_i^2 = \frac{1}{m(n-1)}\sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij} - \bar{X}_i)^2$$

to estimate σ^2. In this case, we would say that $\hat{\underline{\theta}} = (\bar{\bar{X}}, \bar{V})$ estimates $\underline{\theta} = (\mu, \sigma^2)$.

(iii) Since the Phase I charting statistics and the estimated control limits are obtained using the same data, successive comparisons (over subgroups) of the charting statistics with the estimated control limits are dependent events. This implies that the signaling events (i.e. the event that a charting statistic falls on or outside the control limits) or the non-signaling events (i.e. the event that a charting statistic plots between the control limits) for the i^{th} and the j^{th} subgroups, where $i \neq j = 1,2,...,m$, are statistically dependent.

To illustrate the dependency of the non-signaling events, assume, for example, that $X_{ij} \sim iidN(\mu,\sigma^2)$ for $i=1,2,...,m$ and $j=1,2,...,n$ where both μ and σ^2 are unknown and that we are interested in monitoring the mean. In this case, the sample means $\bar{X}_i$ for $i=1,2,...,m$ would be the Phase I charting statistics and the MVU's are $\hat{\underline{\theta}}=(\bar{\bar{X}},\bar{V})$, which would be used to estimate the unknown parameters $\underline{\theta}=(\mu,\sigma^2)$. Thus, writing the estimated control limits as functions of $\hat{\underline{\theta}}$ i.e.

$$\hat{LCL}=g_1(\bar{\bar{X}},\bar{V}) \qquad \text{and} \qquad \hat{UCL}=g_2(\bar{\bar{X}},\bar{V})$$

it is clear that the events

$$\{g_1(\bar{\bar{X}},\bar{V})<\bar{X}_{t_1}<g_2(\bar{\bar{X}},\bar{V})\,|\,IC\} \qquad \text{and} \qquad \{g_1(\bar{\bar{X}},\bar{V})<\bar{X}_{t_2}<g_2(\bar{\bar{X}},\bar{V})\,|\,IC\}$$

where $t_1 \neq t_2 = 1,2,...,m$ are dependent, because the overall mean $\bar{\bar{X}}$ and the pooled variance $\bar{V}$ are functions of all the X_{ij}'s.

It is important to note that, because the m charting statistics are compared to the control limits simultaneously, the false alarm probability (which is the probability for one or more of the charting statistics to plot outside the control limits when the process is in-control) is expected to be inflated. Thus, in order to correctly design a Phase I control chart in Case U, the dependence of the signaling events and the multiple nature of the comparisons must be taken into account.

(iv) It is believed that at the end of a successful Phase I analysis the practitioner will have a set of in-control data or reference data which can be used to estimate any unknown parameters and to obtain a set of control limits to be used in Phase II process monitoring.
Without any loss of generality, it is assumed that m denotes the final number of reference samples at the end of a successful Phase I analysis. Thus the reference data set is assumed to have $N=mn$ individual observations.

(v) For greater generality, only two-sided charts are considered in the discussions that follow. In applications where a one-sided chart is more meaningful or preferred, these discussions can be suitably adapted.

2.1.1 Design and implementation of two-sided Shewhart-type Phase I charts

Introduction

The decision problem under a Phase I control charting scenario is similar, in principle, to a multi-sample test of homogeneity problem where one tests whether the data from various samples come from the same in-control distribution or in-control process (see e.g. Champ and Jones, (2004)).

Under this motivation, the false alarm probability (FAP), which is the probability of at least one false alarm when the process is in-control, is used to construct and evaluate Phase I control charts and not the false alarm rate (FAR), which is the probability for a single charting statistic to plot outside the control limits when the process is in-control.

False alarm probability

An out-of-control situation is indicated when a charting statistic falls either on or above the estimated upper control limit or plots on or below the estimated lower control limit. This important event is called a signal or a signaling event.

To study the false alarm probability it is convenient to consider the complementary event, that is, when a subgroup does not signal, called the non-signaling event. Thus, if

$$E_i = \{L\hat{C}L < C_i < U\hat{C}L\} \quad \text{for} \quad i = 1,2,...,m$$

denote the non-signaling event for the i^{th} subgroup, the false alarm probability can be expressed as

$$\begin{aligned} FAP &= \Pr(\text{At least one false alarm from the } m \text{ samples}) \\ &= 1 - \Pr(\text{No false alarm from the } m \text{ samples}) \\ &= 1 - \Pr(\bigcap_{i=1}^{m} \{E_i\} \mid IC) \qquad (2\text{-}1) \\ &= 1 - \Pr(\bigcap_{i=1}^{m} \{L\hat{C}L < C_i < U\hat{C}L\} \mid IC) \\ &= 1 - \int_{l}^{u}\int_{l}^{u}\cdots\int_{l}^{u} f_{C_1,C_2,...,C_m}(c_1,c_2,...,c_m)dc_1dc_2...dc_m \end{aligned}$$

where $f_{C_1,C_2,...,C_m}(c_1,c_2,...,c_m)$ denotes the joint probability density function (p.d.f) of the charting statistics $C_1,C_2,...,C_m$ when the process is in-control and for notational convenience the estimated control limits are written as $L\hat{C}L = l$ and $U\hat{C}L = u$.

False alarm rate

The false alarm rate, which is the probability of a single charting statistic plotting outside the control limits when the process is in-control, can be expressed as

$$FAR = 1 - \Pr(L\hat{C}L < C_i < U\hat{C}L \mid IC) = 1 - \int_l^u g_{C_i}(c_i)dc_i \qquad (2\text{-}2)$$

where $g_{C_i}(c_i)$ denotes the marginal p.d.f of any of the charting statistics C_i for $i = 1,2,...,m$ when the process is in-control.

Remark 2

(i) It is clear from (2-1) that the *FAP* involves *m* non-signaling events simultaneously. Also, because the control limits are estimated and the charting statistics are all compared with the same pair of control limits, the non-signaling events are dependent. Hence, calculation of the *FAP* requires knowledge of the joint (multivariate) distribution of the charting statistics, when the process is in-control; this is highlighted in the last step of (2-1). The derivation of this joint distribution and the subsequent determination of the control limits (associated charting constants) form the main stumbling blocks in the study and the design of Phase I control charts.

(ii) Expression (2-2) shows that the *FAR* involves only a single sample and a single signaling event. Calculation of the *FAR* therefore requires only the marginal (univariate) distribution of the i^{th} charting statistic C_i when the process is in-control. This in-control marginal distribution is typically the same for all $i = 1,2,....,m,$ so that it can be called a common *FAR*.

(iii) Given the inherent repetitive nature of a Phase I analysis and the fact that the charting statistics from all *m* the subgroups are judged simultaneously, the *FAP* is a more useful and logical metric. The *FAP* is as a result the recommended chart design criterion adopted

in Phase I so that a Phase I control chart is designed by specifying a nominal false alarm probability, say FAP_0, as apposed to specifying a nominal FAR, say FAR_0.

(iv) The objective and the design criterion in Phase I is different from designing Phase II control charts (see e.g. Chapter 3), based on in-control Phase I data, where one would specify some attribute of the in-control Phase II run-length distribution, such as the average run-length (ARL_0), to determine the control limits.

Also, even though the FAR is a commonly used performance measure in practice, it is most often used in the design of Phase II control charts. When the FAR is used in designing a Phase I control chart it should be done with caution and the user should be aware of the effects on the performance of the Phase I chart; this is discussed in more detail in the next section.

Implementation of two-sided Shewhart-type Phase I charts

Implementation of Phase I charts requires the determination of the control limits and/or the charting constants. Different approaches exist and may be used. Each approach is based on an assumption regarding the dependence or independence of the charting statistics coupled with a particular performance criterion. Four approaches are considered and they are:

(i) *FAP*-based control limits,
(ii) *FAR*-based control limits,
(iii) Approximate *FAR*-based control limits, and
(iv) Bonferroni control limits.

Methods (ii), (iii) and (iv) assume that the charting statistics are independent and use the FAR, in some way or another, in their design. However, while method (ii) incorrectly ignores the fact that several charting statistics are compared to the control limits at the same time, methods (iii) and (iv) do not. Furthermore, while method (ii) focuses exclusively on controlling the FAR at a nominal value of FAR_0, methods (iii) and (iv) indirectly controls the FAP by adjusting the FAR.

Method (i), on the other hand, is distinctly different from methods (ii), (iii) and (iv) as it correctly controls the FAP by explicitly taking into account the dependency between the charting statistics and the signaling events.

Method (i): *FAP*-based control limits

The *FAP*-based control limits are the optimal pair of limits because they are derived from the relevant joint p.d.f of the charting statistics which correctly accounts for: (i) the fact that the Phase I control limits are estimated, (ii) the Phase I signaling events are dependent, and (iii) that multiple charting statistics are compared with the estimated control limits in a single step.

In this approach the design of a Phase I control chart requires the user to specify a desirable nominal *FAP* value and then find the corresponding control limits. This means that one needs to solve for that combination(s) of values of $l = l_{FAP_0}$ and $u = u_{FAP_0}$ such that

$$FAP_0 = 1 - \int_{l_{FAP_0}}^{u_{FAP_0}} \int_{l_{FAP_0}}^{u_{FAP_0}} \cdots \int_{l_{FAP_0}}^{u_{FAP_0}} f_{C_1,C_2,\ldots,C_m}(c_1, c_2, \ldots, c_m) dc_1 dc_2 \ldots dc_m \qquad (2\text{-}3)$$

where FAP_0 is the nominal value of *FAP* (typically set equal to 0.01, 0.05 or 0.10 in practice) and l_{FAP_0} and u_{FAP_0} denote the lower and the upper *FAP*-based control limits, respectively.

Finding the two unknowns l_{FAP_0} and u_{FAP_0} from expression (2-3), uniquely, poses a problem without additional restrictions. For example, in some cases the charting statistics are symmetrically distributed around zero (without any loss of generality) and then it makes sense to use symmetric control limits, that is, setting $L\hat{C}L = -U\hat{C}L = -d$, say, where d is then obtained by solving

$$FAP_0 = 1 - \int_{-d}^{d} \int_{-d}^{d} \cdots \int_{-d}^{d} f_{C_1,C_2,\ldots,C_m}(c_1, c_2, \ldots, c_m) dc_1 dc_2 \ldots dc_m . \qquad (2\text{-}4)$$

If the plotting statistics are not symmetrically distributed and two-sided control charts are desired, one possibility is to use the equal-tailed conservative approach in which half the nominal *FAP* is assigned in each tail i.e. half above the $U\hat{C}L$ and half below the $L\hat{C}L$, respectively. This approach can be more clearly explained by expressing the *FAP* of (2-1) as

$$\begin{aligned} FAP &= 1 - \int_{l}^{u} \int_{l}^{u} \cdots \int_{l}^{u} f_{C_1,C_2,\ldots,C_m}(c_1, c_2, \ldots, c_m) dc_1 dc_2 \ldots dc_m \\ &= \Pr\left(\min(C_1, C_2, \ldots, C_m) \le l \text{ or } \max(C_1, C_2, \ldots, C_m) \ge u \mid IC\right) \qquad (2\text{-}5a) \\ &= \begin{aligned} &\Pr(\min(C_1, C_2, \ldots, C_m) \le l \mid IC) + \Pr(\max(C_1, C_2, \ldots, C_m) \ge u \mid IC) \\ &- \Pr(\min(C_1, C_2, \ldots, C_m) \le l \text{ and } \max(C_1, C_2, \ldots, C_m) \ge u \mid IC). \end{aligned} \end{aligned}$$

Expression (2-5a) follows since the event that at least one of the charting statistics plot either above or below the control limits can be equivalently expressed as the union of two events: (i) that the minimum of the charting statistics lies below the $L\hat{C}L$, and (ii) that the maximum of the charting statistics lies above the $U\hat{C}L$. From (2-5a) it follows that an upper bound for the FAP is given by

$$FAP \leq \Pr(\min(C_1, C_2, ..., C_m) \leq l \mid IC) + \Pr(\max(C_1, C_2, ..., C_m) \geq u \mid IC). \qquad (2\text{-}5b)$$

The equal-tails conservative method entails that one finds those values of the FAP-based control limits $l = l_{FAP_0}$ and $u = u_{FAP_0}$ such that

$$\Pr\left(\min(C_1, C_2, ..., C_m) \leq l_{FAP_0} \mid IC\right) = FAP_0 / 2 \qquad (2\text{-}6a)$$

and

$$\Pr\left(\max(C_1, C_2, ..., C_m) \geq u_{FAP_0} \mid IC\right) = FAP_0 / 2\,. \qquad (2\text{-}6b)$$

and ensures that the false alarm probability is not greater than FAP_0.

Using expressions (2-6a) and (2-6b) rather than (2-3) or (2-5a) to solve for l_{FAP_0} and u_{FAP_0} may be advantageous in some cases in the sense that it involves calculating the percentiles of some univariate distributions i.e. the distributions of the minimum and the maximum, which might be computationally easier than using the multivariate p.d.f in (2-3) or using the joint distribution of the minimum and the maximum in (2-5b) . However, finding closed form expressions for the p.d.f.'s of

$$C_{\min} = \min(C_1, C_2, ..., C_m) \qquad \text{and} \qquad C_{\max} = \max(C_1, C_2, ..., C_m)$$

to evaluate analytically is complex as the C_i's are statistically dependent random variables.

Attained False Alarm Rate

Given the FAP-based control limits l_{FAP_0} and u_{FAP_0} that satisfy (2-3) or (2-6a) and (2-6b), one may be interested in calculating the attained false alarm rate ($AFAR$). The $AFAR$ is the resultant probability for a single charting statistic to plot outside the FAP-based control limits when the process is in-control, and defined as

$$AFAR = 1 - \Pr(l_{FAP_0} < C_i < u_{FAP_0} \mid IC) = 1 - \int_{l_{FAP_0}}^{u_{FAP_0}} g_{C_i}(c_i)\,dc_i\,. \qquad (2\text{-}7)$$

Using (2-7) one can compute the $AFAR$ given the marginal distribution $g_{C_i}(c_i)$ of C_i for $i = 1,2,...,m$ is known.

Method (ii): *FAR*-based control limits

Classically (see e.g. Hillier, (1969) and Yang and Hillier, (1970)) a Phase I chart is designed by controlling the false alarm rate at a nominal FAR value, FAR_0 say. This entails finding that/those combination(s) of values for the control limits such that

$$FAR_0 = 1 - \Pr(l_{FAR_0} < C_i < u_{FAR_0} \mid IC) = 1 - \int_{l_{FAR_0}}^{u_{FAR_0}} g_{C_i}(c_i)dc_i \qquad (2\text{-}8)$$

where l_{FAR_0} and u_{FAR_0} denote the lower and the upper FAR-based control limits, respectively.

It is evident from (2-8) that the *FAR*-based control limits use only the marginal distribution $g_{C_i}(c_i)$ for $i = 1,2,...,m$ of a single charting statistic to find the Phase I control limits and overlooks the fact that multiple charting statistics are simultaneously compared with the estimated control limits. Hence, this approach is flawed can be improved upon.

Attained False Alarm Probability

Given the values l_{FAR_0} and u_{FAR_0} that satisfy (2-8), the attained false alarm probability ($AFAP$) may be calculated from (2-3) as

$$AFAP = 1 - \int_{l_{FAR_0}}^{u_{FAR_0}} \int_{l_{FAR_0}}^{u_{FAR_0}} \cdots \int_{l_{FAR_0}}^{u_{FAR_0}} f_{C_1,C_2,...,C_m}(c_1, c_2, ..., c_m)dc_1dc_2...dc_m . \qquad (2\text{-}9)$$

The $AFAP$ is the resultant probability that at least one charting statistic will plot outside the *FAR*-based control limits when the process is in-control.

Remark 3

> Solving for the control limits from (2-8) and then using the resultant *FAR*-based control limits to construct a Phase I chart not only ignores the fact that multiple charting statistics are compared to the control limits at the same time, but also ignores the dependency between the signaling events. The attained false alarm probably might thus be far-off the desired FAP_0 and, as a result, this approach is not recommended. Typically the $AFAP$ is inflated and larger than the chosen FAR_0.

If only the marginal distribution of C_i for $i = 1,2,...,m$ is available and one wishes to design a Phase I chart so that the FAP is close to FAP_0 one may use the approximate *FAR*-based control limits or the Bonferroni control limits to find approximate Phase I control limits. Both these approaches assume that the C_i's are independent and works with the (exact) marginal distribution of the charting statistic, but they do take into account that more than one signaling event need to be dealt with.

Method (iii): Approximate *FAR*-based control limits

A simple and popular alternative to the exact *FAP*-based control limits is the approximate *FAR*-based control limits. The latter is often used and yields an approximate solution (see e.g. Champ and Jones, (2004)).

In this approach one approximates the Phase I control limits by ignoring the dependence among the signaling events, but account for the fact that multiple comparisons are made at the same time.

In particular, when the number of subgroups m is large, the correlation among the charting statistics approaches zero and the charting statistics are approximately independent. Then, from (2-1) and (2-2) it can be seen that

$$FAP \approx 1 - \prod_{i=1}^{m} \Pr(E_i \mid IC) = 1 - [\Pr(E_i \mid IC)]^m = 1 - (1 - FAR)^m \qquad (2\text{-}10)$$

so that

$$FAR \approx 1 - (1 - FAP)^{\frac{1}{m}}. \qquad (2\text{-}11)$$

Expressions (2-10) and (2-11) show the relationship between the FAP and the FAR for large m and can be used to ensure that the $FAP \approx FAP_0$ by controlling the FAR.

For example, it follows from (2-11) that for symmetrically distributed charting statistics the approximate *FAR*-based control limits are given by the

$$\frac{1}{2}[1 - (1 - FAP_0)^{\frac{1}{m}}]100^{\text{th}} \qquad \text{and} \qquad [1 - \frac{1}{2}\{1 - (1 - FAP_0)^{\frac{1}{m}}\}]100^{\text{th}}$$

percentiles of the marginal in-control distribution of a single charting statistic and would yield a false alarm probability of approximately FAP_0.

For asymmetric approximate *FAR*-based control limits one may use the

$$w[1-(1-FAP_0)^{\frac{1}{m}}]100^{\text{th}} \qquad \text{and} \qquad [1-(1-w)\{1-(1-FAP_0)^{\frac{1}{m}}\}]100^{\text{th}}$$

percentiles of the marginal in-control distribution of C_i for $i=1,2,...,m$, with $0 \le w \le 1$.

Method (iv): Bonferroni control limits

A fourth approach is to use a Bonferroni-type adjustment when calculating the Phase I limits. This method also yields an approximate solution, but is applicable whether or not the charting statistics are symmetrically distributed and ensures that the false alarm probability is at most as specified (see e.g. Ryan, (1989) p. 74 - 76).

It follows from Bonferroni's inequality (see e.g. Casella and Berger, (2002) p. 13) that one can find an upper bound for the false alarm probability as a function of the false alarm rate; this upper bound is given by

$$FAP = 1-\Pr(\bigcap_{i=1}^{m}\{E_i\} \mid IC) \le m-\sum_{i=1}^{m}\Pr(\{E_i\} \mid IC) = m-\sum_{i=1}^{m}(1-FAR) = mFAR\,. \qquad (2\text{-}12)$$

If it is desired that the $FAP \le FAP_0$, it is seen from (2-12) that setting $mFAR = FAP_0$ i.e. setting the false alarm rate equal to $FAR = FAP_0/m$, would meet the requirement. In this case, the symmetrically placed control limits are given by the

$$[FAP_0/2m]100^{\text{th}} \qquad \text{and} \qquad [1-(FAP_0/2m)]100^{\text{th}}$$

percentiles of the marginal in-control distribution of a single charting statistic.

Remark 4

(i) In some situations it may be reasonable to assume that the marginal distribution of the charting statistics C_i for $i = 1,2,...,m$ is normal (or, at least approximately so) and then use the percentiles of a normal distribution to find the control limits in the *FAR*-based approach, the approximate *FAR*-based approach and/or the Bonferroni approach, instead of using the exact marginal distribution $g_{C_i}(c_i)$ for $i = 1,2,...,m$. This, however, might result in a Phase I control chart with incorrectly placed limits, especially when m and n are not large.

Although the assumption of normality might be acceptable in some cases, there are scenarios (e.g. the S^2, S and R charts) where the marginal distribution of the charting statistic (e.g. the sample variance or the sample standard deviation or the sample range) is markedly non-normal; this is particularly true when small sample sizes are used.

(ii) The approximate *FAR*-based limits and the Bonferroni limits are easier to calculate than the exact *FAP*-based limits because one does not need to work with the joint distribution of the entire set of charting statistics. However, these two sets of limits are generally not suitable if the number of subgroups m is small.

Comparison of methods (i), (ii), (iii) and (iv) to design a Shewhart-type Phase I control chart

The four methods to design a Shewhart-type Phase I chart are illustrated in Figure 2.1 for the $\overline{X}$ chart in Case U. For this illustration a set of simulated data from the standard normal distribution was used and it was assumed that $m = 15$ random samples each of size $n = 5$ are available. The charting statistics are the sample means $\overline{X}_i$ for $i = 1,2,...,15$ of the simulated data in this research. The details on how to calculate the four pairs of control limits are given in Champ and Jones, (2004).

It is seen that there can be more false alarms if one uses the *FAR*-based control limits, denoted LCL(FAR) and UCL(FAR), than when one uses the *FAP*-based control limits. This is simply because the *FAR*-based control limits are tighter than the *FAP*-based control limits, denoted LCL(FAP) and UCL(FAP).

In contrast, it is noticed that the approximate *FAR*-based control limits (i.e. LCL(Approx FAR) and UCL(Approx FAR)) and the Bonferroni control limits (i.e. LCL(Bon) and UCL(Bon)) almost coincide and are both slightly wider than the *FAP*-based control limits. It is thus likely that one can observe less false alarms if one uses the approximate *FAR*-based control limits or the Bonferroni control limits instead of the *FAP*-based control limit. Although less false alarms might be appealing from a practical point of view, if the control limits are too wide, unwanted variation might go undetected.

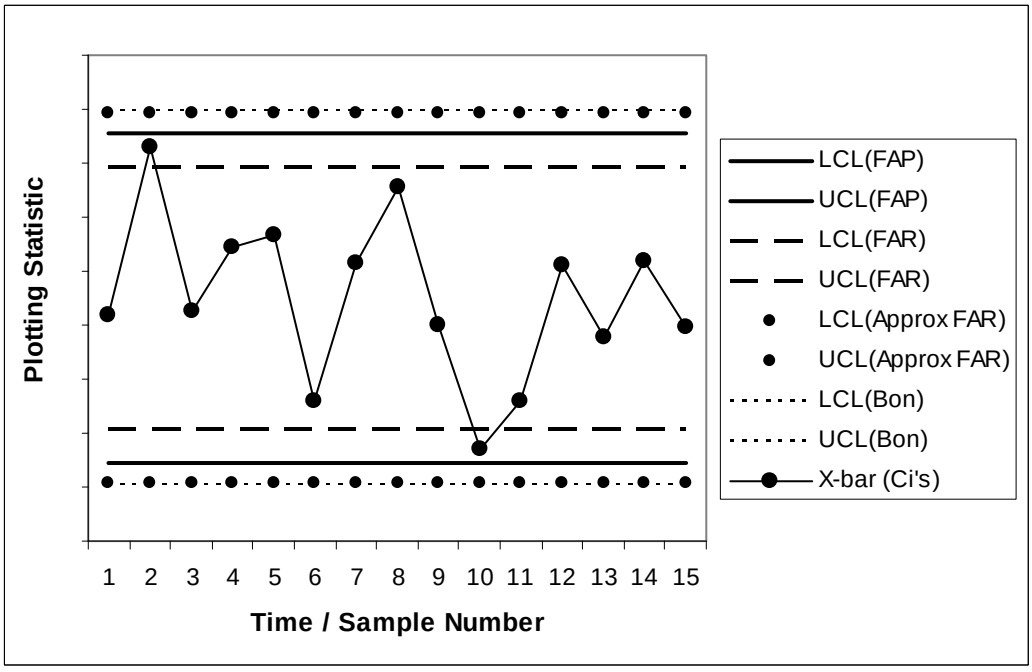


Figure 2.1: Shewhart-type Phase I $\overline{X}$ control charts in Case U

2.2 Shewhart-type S^2, S and R charts: Phase I

Introduction

The most familiar control charts in practice include those for the mean and the variance of an assumed (at least approximately) normally distributed process. While Champ and Jones (2004) studied the Shewhart-type Phase I $\bar{X}$ chart for the mean, we study and design Shewhart-type Phase I S^2, S and R charts for the spread.

Control charts for the spread are particularly important since an estimate of the variance or the standard deviation is necessary in creating a control chart for the mean. The spread must therefore be monitored and controlled before (or at least simultaneously) attempting to monitor the mean.

Assumptions

Suppose that m independent rational subgroups each of size $n>1$ are available from a normal distribution with an unknown mean μ and an unknown variance σ^2. The data are represented as $X_{ij} \sim iidN(\mu,\sigma^2)$ for $i=1,2,...,m$ and $j=1,2,...,n$ where X_{ij} is the j^{th} observation from the i^{th} subgroup.

Point estimators for the unknown standard deviation and the unknown variance and their probability distributions

Estimation of the mean and the variance affects the performance of the S^2, S and R charts because their control limits are defined in terms of the unknown variance and the charting statistics (in case of the S^2 chart and the S chart) also depend on the unknown mean. Furthermore, the sampling distributions of the charting statistics are affected since the degrees of freedom of the chi-square distribution and the chi distribution changes from n to $n-1$ when the mean is estimated.

Point estimators

Two unbiased point estimators for the process standard deviation and one for the process variance are

$$\hat{\sigma}_R = \frac{\overline{R}}{d_2} = \frac{1}{d_2}\left(\frac{1}{m}\sum_{i=1}^{m} R_i\right), \tag{2-13}$$

$$\hat{\sigma}_S = \frac{\overline{S}}{c_4} = \frac{1}{c_4}\left(\frac{1}{m}\sum_{i=1}^{m} S_i\right) \tag{2-14}$$

and

$$\hat{\sigma}^2_{S_p} = \overline{V} = \frac{1}{m}\sum_{i=1}^{m} S_i^2, \tag{2-15}$$

respectively, where

$$R_i = \max(X_{i1}, X_{i2}, ..., X_{in}) - \min(X_{i1}, X_{i2}, ..., X_{in}),$$

denotes the i^{th} sample range,

$$S_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_{ij} - \overline{X}_i)^2 \qquad \text{and} \qquad \overline{X}_i = \frac{1}{n}\sum_{j=1}^{n} X_{ij}$$

denote the i^{th} sample variance and the sample mean, respectively and the i^{th} sample standard deviation is defined as

$$S_i = \sqrt{S_i^2}$$

for $i = 1,2,...,m$.

The unbiasing constants d_2 and c_4 are tabulated, for example, in Appendix VI of Montgomery (2005).

The first estimator in (2-13) is typically used when the R chart is used to monitor spread. The second estimator in (2-14) is used in the application of the S chart, while the third estimator in (2-15) is a pooled variance estimator and is used in the S^2 chart.

Distribution of the point estimators

Under the assumption that the process follows a normal distribution, it is well-known that

$$\frac{m(n-1)\overline{V}}{\sigma^2} \sim \chi^2_{m(n-1)}$$

so that

$$\overline{V} \sim \frac{\sigma^2 \chi^2_{m(n-1)}}{m(n-1)};$$

this is an exact result.

The exact distribution of $\hat{\sigma}_R^2$ is complicated. Patnaik (1950) presented a method for approximating the distribution of $\frac{w\hat{\sigma}_R^2}{c^2\sigma^2}$ by that of a χ^2_w distribution where $c = c(w)$ is a constant and a function of w. This is done using a technique called "moment matching" (see e.g. Casella and Berger, (2002) p. 314) and involves setting the first two moments of the distribution of $\frac{\hat{\sigma}_R}{\sigma}$ equal to those of $\frac{c\chi_w}{\sqrt{w}}$ where χ_w is a random variable which follows a chi distribution i.e. the square root of a chi-square random variable with w degrees of freedom (see e.g. Johnson, Kotz and Balakrishnan, (1994, 1995)). Then, approximately,

$$\hat{\sigma}_R \sim \frac{\sigma c \chi_w}{\sqrt{w}}.$$

Using a similar approach, one can show that approximately

$$\hat{\sigma}_S \sim \frac{\sigma d \chi_t}{\sqrt{t}},$$

where the constant $d = d(t)$ is a function of t (see e.g. Champ and Jones, (2004)).

Values of the constants c and d for $\hat{\sigma}_R$ and $\hat{\sigma}_S$ were numerically approximated and tabulated by Champ and Jones (2004) for $m = 3(1)10(5)40$ and $n = 3(1)10$.

2.2.1 Phase I S^2 chart

Introduction

The application of the S^2 chart in Case K and its operation in prospective process monitoring in Case U, are discussed in various SPC textbooks (see e.g. Ryan, (1989) and Montgomery, (2005) p. 231). Here we study the retrospective use of the Phase I S^2 chart in Case U.

Charting statistics and control limits

For the Phase I S^2 chart the charting statistics are the sample variances S_i^2 for $i = 1,2,...,m$ and one uses the probability limits

$$L\hat{C}L = \frac{\bar{V}\chi^2_{1-\alpha_L,n-1}}{n-1} \qquad \hat{C}L = \bar{V} \qquad U\hat{C}L = \frac{\bar{V}\chi^2_{\alpha_U,n-1}}{n-1} \tag{2-16}$$

where $\chi^2_{\xi,n-1}$ is the $100(1-\xi)^{\text{th}}$ percentile of the chi-square distribution with $n-1$ degrees of freedom (see e.g. Montgomery, (2005) p. 231).

Typically one would take $\alpha_L = \alpha_U = 0.00135$ and find the chi-square percentiles with the idea that the false alarm rate is approximately $FAR_0 = 0.0027$. However, in Phase I charting, as noted earlier, it is better to control the false alarm probability at some nominal value i.e. $FAP = FAP_0$, which results in some false alarm rate

$$AFAR = AFAR_L + AFAR_U \tag{2-17}$$

where

$$AFAR_L = \Pr(S_i^2 \le L\hat{C}L_{FAP_0} \mid IC) \quad \text{and} \quad AFAR_U = \Pr(S_i^2 \ge U\hat{C}L_{FAP_0} \mid IC) \tag{2-18}$$

are the probabilities that a charting statistic plots on or outside the estimated lower and upper limits, respectively, which result from controlling the false alarm probability. The resultant FAR is called the attained false alarm rate and denoted $AFAR$. Depending on the values of FAP_0, m and n, the $AFAR$ can be substantially different from 0.0027, as will be seen later.

Design of the Phase I S^2 chart for a nominal *FAP*

The objective when designing a Phase I S^2 chart is to know where to place the control limits of the S^2 chart so that if the m charting statistics are simultaneously plotted on the chart, the probability of at least one charting statistic plotting on or outside the estimated control limits, when the process is in-control, is at most equal to FAP_0. This goal is identical to ensure that the FAP of the Phase I S^2 chart is less than or equal to a pre-specified FAP i.e. $FAP \leq FAP_0$; for this we need an expression to calculate the FAP.

First we derive *FAP*-based control limits for the S^2 chart, which is an exact solution and takes account of the dependence between the signaling events. We then also find approximate *FAR*-based control limits, which is an approximate solution and is suitable when the number of Phase I subgroups m is large.

False alarm probability of the Phase I S^2 chart

The exact false alarm probability of the Phase I S^2 chart, under the assumption of i.i.d. observations from a normal distribution with the mean and the variance both unknown, is

$$\begin{aligned}
FAP &= \Pr(\text{At least one false alarm from the } m \text{ samples}) \\
&= 1-\Pr(\text{No false alarm from the } m \text{ samples}) \\
&= 1-\Pr\left(\bigcap_{i=1}^{m}\{L\hat{C}L < S_i^2 < U\hat{C}L\} \mid IC\right) \\
&= 1-\Pr\left(\bigcap_{i=1}^{m}\left\{\frac{\overline{V}\,\chi^2_{1-\alpha_L,n-1}}{n-1} < S_i^2 < \frac{\overline{V}\,\chi^2_{\alpha_U,n-1}}{n-1}\right\} \mid IC\right) \\
&= 1-\Pr\left(\bigcap_{i=1}^{m}\left\{\frac{\chi^2_{1-\alpha_L,n-1}}{n-1} < \frac{S_i^2}{\overline{V}} < \frac{\chi^2_{\alpha_U,n-1}}{n-1}\right\} \mid IC\right) \\
&= 1-\Pr\left(\bigcap_{i=1}^{m}\left\{\frac{\chi^2_{1-\alpha_L,n-1}}{m(n-1)} < \frac{\frac{(n-1)S_i^2}{\sigma^2}}{\frac{m(n-1)\overline{V}}{\sigma^2}} < \frac{\chi^2_{\alpha_U,n-1}}{m(n-1)}\right\} \mid IC\right) \\
&= 1-\Pr\left(\bigcap_{i=1}^{m}\{a < Y_i < b\} \mid IC\right) \\
&= 1-\int_a^b\int_a^b\cdots\int_a^b f(y_1,y_2,\ldots,y_m)\,dy_1dy_2\ldots dy_m
\end{aligned} \tag{2-19}$$

where

$$a = \frac{\chi^2_{1-\alpha_L,n-1}}{m(n-1)} \qquad \text{and} \qquad b = \frac{\chi^2_{\alpha_U,n-1}}{m(n-1)} \tag{2-20}$$

are called the charting constants, and

$$Y_i = \frac{\frac{(n-1)S_i^2}{\sigma^2}}{\frac{m(n-1)\overline{V}}{\sigma^2}} = \frac{\frac{(n-1)S_i^2}{\sigma^2}}{\sum_{j=1}^{m}\frac{(n-1)S_j^2}{\sigma^2}} = \frac{X_i}{\sum_{j=1}^{m}X_j} \tag{2-21}$$

where the X_i for $i=1,2,\ldots,m$ denotes independent chi-square random variables each with $n-1$ degrees of freedom and $f(y_1,y_2,\ldots,y_m)$ denotes the joint p.d.f of $(Y_1,Y_2,\ldots,Y_m)$ when the process is in-control.

Remark 5

(i) From the definitions of the estimated control limits $L\hat{C}L$ and $U\hat{C}L$ and the charting constants a and b in (2-16) and (2-20), respectively, it is seen that one may also write the control limits as

$$L\hat{C}L = ma\overline{V} \qquad \text{and} \qquad U\hat{C}L = mb\overline{V} \tag{2-22}$$

where m is the number of Phase I subgroups and $\overline{V}$ is defined in (2-15).

The control limits in (2-16) are defined in terms of the marginal distribution of the charting statistics (i.e. the percentiles of the χ^2_{n-1} distribution) and allows one to easily calculate the estimated control limits of a Phase II S^2 chart. In contrast, the alternative form of the estimated control limits in (2-22) simplifies the calculation of the limits for a Phase I S^2 chart, which is the focus here. Example 1 (given later) explains how the limits in (2-22) may be used.

(ii) From the derivation of the *FAP* in (2-19) it is clear that any two non-signaling or signaling events are dependent since the corresponding Y_i random variables are statistically dependent. This is so because each Y_i is a function of and depends on all the S_i^2's through $\overline{V} = \frac{1}{m}\sum_{i=1}^{m} S_i^2$ in their denominators (see e.g. expression (2-21)). The joint probability distribution of the charting statistics when the process is in-control i.e. $f(y_1, y_2, ..., y_m)$, is therefore needed to calculate and study the false alarm probability; this is highlighted by the last step in (2-19).

Joint probability distribution of $(Y_1, Y_2, ..., Y_m)$

Deriving an exact closed form expression for $f(y_1, y_2, ..., y_m)$ in order to calculate the *FAP* of the S^2 chart is difficult. A particular obstacle is the fact that the Y_i random variables are linearly dependent i.e. $\sum_{i=1}^{m} Y_i = 1$; this causes the joint distribution of $(Y_1, Y_2, ..., Y_m)$ to be singular and of dimension $m-1$. To calculate the *FAP* one can use the joint distribution of $(Y_1, Y_2, ..., Y_{m-1})$; this is looked at next.

Calculating the exact *FAP* of the S^2 chart

To analytically calculate the *FAP* one may begin with (2-19) and proceed as follows:

$$\begin{aligned} FAP &= 1 - \int_a^b \int_a^b \cdots \int_a^b f(y_1, y_2, ..., y_m) dy_1 dy_2 ... dy_m \\ &= 1 - \Pr\left(\bigcap_{i=1}^{m} \{a < Y_i < b\} \mid IC \right) \\ &= 1 - \Pr\left(\left\{ \bigcap_{i=1}^{m-1} \{a < Y_i < b\} \right\} \bigcap \left\{ a < 1 - \sum_{i=1}^{m-1} Y_i < b \right\} \mid IC \right) \\ &= 1 - \iint \cdots \int_{R(\mathbf{y})} f(y_1, y_2, ..., y_{m-1}) dy_1 dy_2 ... dy_{m-1} \end{aligned} \tag{2-23}$$

where the joint p.d.f of $(Y_1, Y_2, ..., Y_{m-1})$, which is denoted $f(y_1, y_2, ..., y_{m-1})$, is integrated over the region

$$R(\mathbf{y}) = \left\{ (y_1, y_2, ..., y_{m-1}) \mid a < y_i < b \text{ for } i = 1,2,...,m-1 \text{ and } a < 1 - \sum_{i=1}^{m-1} y_i < b \right\}.$$

Remark 6

(i) The first two steps in (2-23) are identical to the last two steps of (2-19) but in the reverse order.

(ii) The third step in (2-23) follows by using the fact that $\sum_{i=1}^{m} Y_i = 1$ and then writing $Y_m = 1 - \sum_{i=1}^{m-1} Y_i$.

(iii) From the third and the fourth steps in (2-23) it is evident that calculating the *FAP* does not necessarily require the joint distribution of $(Y_1, Y_2, ..., Y_m)$. Instead one may use the joint distribution of $(Y_1, Y_2, ..., Y_{m-1})$, which is known.

(iv) Given a closed form expression for the joint p.d.f $f(y_1, y_2, ..., y_{m-1})$, the last integral expression in (2-23) can be evaluated using a computer software package(s) capable of numerical integration (for e.g. Mathcad® or Mathematica®).

Joint probability distribution of $(Y_1, Y_2, ..., Y_{m-1})$

The joint distribution of Y_i for $i = 1,2,...,m-1$ is the type I or standard Dirichlet distribution and is regarded as a multivariate generalization of the beta distribution (see e.g. Chapter 49 of Kotz, Balakrishnan and Johnson, (2000)).

The standard Dirichlet distribution, in general, is denoted $(Y_1, Y_2, ..., Y_{m-1}) \sim D^I(\theta_1, \theta_2, ..., \theta_{m-1}; \theta_m)$ where θ_i for $i = 1,2,...,m$ are the parameters of the distribution.

The joint p.d.f of $(Y_1, Y_2, ..., Y_{m-1})$ is given by

$$f(y_1, y_2, ..., y_{m-1}) = \frac{\Gamma\left(\sum_{i=1}^{m} \theta_i\right)}{\prod_{i=1}^{m} \Gamma(\theta_i)} \prod_{i=1}^{m-1} y_i^{\theta_i - 1} \left(1 - \sum_{i=1}^{m-1} y_i\right)^{\theta_m - 1} \qquad (2\text{-}24)$$

where $\{0 \le y_i$ for $i = 1,2,...,m-1$ and $\sum_{i=1}^{m-1} y_i \le 1\}$ and the correlation between Y_i and Y_j for all $i \ne j = 1,2,...,m-1$, is given by

$$\text{corr}(Y_i, Y_j) = -\sqrt{\theta_i \theta_j / (\sum_{k=1}^{m} \theta_k - \theta_i)(\sum_{k=1}^{m} \theta_k - \theta_j)} \qquad (2\text{-}25)$$

(see e.g. Chapter 49 of Kotz, Balakrishnan and Johnson, (2000)).

Substituting (2-24) in (2-23), with each $\theta_i = \frac{n-1}{2}$ for $i = 1,2,...,m$, one can analytically calculate the FAP of the S^2 chart.

FAP-based control limits for the S^2 chart

Calculating the FAP of the Phase I S^2 chart and designing the chart is not the same. Calculating the FAP requires one to evaluate expression (2-23). The design, as noted earlier, requires one to find the proper position of the control limits so that the FAP is less than or equal to FAP_0. This implies that one has to find combinations of values for a and b, denoted $\hat{a}$ and $\hat{b}$, so that

$$FAP(a=\hat{a},b=\hat{b}) = 1 - \int_{\hat{a}}^{\hat{b}}\int_{\hat{a}}^{\hat{b}}\ldots\int_{\hat{a}}^{\hat{b}} f(y_1,y_2,\ldots,y_m)dy_1dy_2\ldots dy_m \le FAP_0 \qquad (2\text{-}26)$$

or, equivalently, such that

$$FAP(a=\hat{a},b=\hat{b}) = 1 - \underset{\hat{R}(\mathbf{y})}{\int\int\ldots\int} f(y_1,y_2,\ldots,y_{m-1})dy_1dy_2\ldots dy_{m-1} \le FAP_0 \qquad (2\text{-}27)$$

where

$$\hat{R}(\mathbf{y}) = \left\{(y_1,y_2,\ldots,y_{m-1}) \mid \hat{a} < y_i < \hat{b} \text{ for } i = 1,2,\ldots,m-1 \text{ and } \hat{a} < 1 - \sum_{i=1}^{m-1} y_i < \hat{b}\right\}. \qquad (2\text{-}28)$$

Remark 7

(i) The equivalence of expressions (2-26) and (2-27) follows from the first and the last steps in (2-23).

(ii) Solving for $\hat{a}$ and $\hat{b}$ from (2-26) is not possible because, as mentioned earlier, a closed form expression for $f(y_1,y_2,\ldots,y_m)$ is not traceable. Also, solving for $\hat{a}$ and $\hat{b}$ from (2-27) involves multiple integrals (as little as $m=2$ but even as many as $m=300$ or 500) to be evaluated using numerical integration procedures; because this approach is computationally very demanding (i.e. computer intensive and time consuming) it is undesirable. As an alternative approach, we use a re-written form of (2-26) coupled with computer simulation experiments to solve for $\hat{a}$ and $\hat{b}$.

From expressions (2-5a) and (2-5b) it follows that expression (2-26) can also be written as

$$FAP(a=\hat{a},b=\hat{b}) = \Pr(\min(Y_1,Y_2,...,Y_m) \le \hat{a} \mid IC) + \Pr(\max(Y_1,Y_2,...,Y_m) \ge \hat{b} \mid IC) - \Pr(\min(Y_1,Y_2,...,Y_m) \le \hat{a} \text{ and } \max(Y_1,Y_2,...,Y_m) \ge \hat{b} \mid IC) \quad (2\text{-}29a)$$

which implies that

$$FAP \le \Pr(\min(Y_1,Y_2,...,Y_m) \le \hat{a} \mid IC) + \Pr(\max(Y_1,Y_2,...,Y_m) \ge \hat{b} \mid IC). \quad (2\text{-}29b)$$

Expressions (2-29a) and (2-29b) follow because we need the probability that at least one of the Y_i's plots either on or below the estimated lower control limit or, on or above the estimated upper control limit. The first probability can be expressed in terms of the smallest of the Y_i's whereas the second in terms of the largest of the Y_i's. This is consistent with our earlier discussions in section 2.1.1.

Because, in general, the Y_i's are not symmetrically distributed the two probabilities on the right in (2-29b) will not be equal in general. This creates a problem since two unknowns cannot ordinarily be determined uniquely from a single condition.

To simplify matters we follow an equal-tailed conservative approach in that $\hat{a}$ and $\hat{b}$ are found such that each term on the right in (2-29) is at most $\frac{FAP_0}{2}$. Thus, we find $\hat{a}$ and $\hat{b}$ so that

$$\Pr(\min(Y_1,Y_2,...,Y_m) \le \hat{a} \mid IC) \le \frac{FAP_0}{2} \quad \text{and} \quad \Pr(\max(Y_1,Y_2,...,Y_m) \ge \hat{b} \mid IC) \le \frac{FAP_0}{2}. \quad (2\text{-}30)$$

The distribution theory of the largest and the smallest of order statistics of the Y_i's is fairly involved and is not attempted here (see e.g. Eisenhart, Hastay and Wallis, (1947)). Instead we use intensive computer simulations (accounting for the dependence among the charting statistics) to solve the equations in (2-30) and find the charting constants $\hat{a}$ and $\hat{b}$.

Simulation algorithm for determining the *FAP*-based control limits of the S^2 chart

The steps of the computer simulation algorithm to find $\hat{a}$ and $\hat{b}$ are:

Step 1: Generate 100,000 vector valued observations from the joint distribution of $(Y_1, Y_2, ..., Y_m)$. To obtain one such observation we generate m independent χ^2_{n-1} random variables for a given *m* and *n* (denoted by X_i for $i = 1,2,...,m$), calculate the sum $SUM_1 = \sum_{j=1}^{m} X_j$, and then calculate $Y_i = \frac{X_i}{SUM_1}$ for $i = 1,2,...,m$. The vector $(Y_1, Y_2, ..., Y_m)$ is one such observation.

Step 2: Find $Y_{\max} = \max(Y_1, Y_2, ..., Y_m)$ and $Y_{\min} = \min(Y_1, Y_2, ..., Y_m)$.

Step 3:

Let

$$\hat{a} = \max\left\{u : \hat{\Pr}_1(\min(Y_1, Y_2, ..., Y_m) \le u \mid IC) \le \frac{FAP_0}{2}\right\}$$

and

$$\hat{b} = \min\left\{u : \hat{\Pr}_2(\max(Y_1, Y_2, ..., Y_m) \ge u \mid IC) \le \frac{FAP_0}{2}\right\}$$

where $0 < u < 1$; this means that we choose $\hat{a}$ to be that value of u such that the proportion of $Y_{\min}$'s less than or equal to $\hat{a}$ is at most $\frac{FAP_0}{2}$ and we choose $\hat{b}$ to be that value of u such that the proportion of $Y_{\max}$'s greater than or equal to $\hat{b}$ is less than or equal to $\frac{FAP_0}{2}$.

Note that, in the above expressions for $\hat{a}$ and $\hat{b}$, the "$\hat{\Pr}_1$" and the "$\hat{\Pr}_2$" denote relative frequencies and are therefore, strictly speaking, empirical probabilities i.e.

$$\hat{\Pr}_1(u) = \frac{\text{number of simulated } Y_{\min} \text{ values} \le u}{100{,}000} \quad \text{and} \quad \hat{\Pr}_2(u) = \frac{\text{number of simulated } Y_{\max} \text{ values} \ge u}{100{,}000}.$$

Remark 8

Step 3 of the simulation algorithm is a conservative equal-tailed approach that ensures that the false alarm probability is not greater than the nominal FAP.

***FAP*-based charting constants for S^2 chart**

The values of $\hat{a}$ and $\hat{b}$ for $m = 3(1)10,15,20,25$ and $n = 3(1)10$ such that the FAP does not exceed 0.01, 0.05 and 0.10 are given in Tables 2.1, 2.2 and 2.3, respectively. Note that, for some combinations of m and n it is seen that $\hat{a} = 0$ so that the estimated lower control limit equals zero; this implies that the Phase I S^2 chart would have only an upper control limit.

To find the position of the Phase I control limits one replaces a with $\hat{a}$, substitute $\hat{b}$ for b and replace $\overline{V}$ with its observed value $\bar{v}$ into (2-22).

Example 1

Suppose that $m = 7$ Phase I samples are available each of size $n = 6$ and that it is desired that $FAP \leq FAP_0 = 0.05$.

From Table 2.1 we obtain $\hat{a} = 0.0115$ and $\hat{b} = 0.4271$; thus $m\hat{a} = (7)(0.0115) = 0.0805$ and $m\hat{b} = (7)(0.4271) = 2.9897$ so that the estimated lower and upper control limits of the Phase I S^2 chart in (2-22) are found to be

$$L\hat{C}L = 0.0805\bar{v} \qquad \text{and} \qquad U\hat{C}L = 2.9897\bar{v}$$

respectively. These limits ensure a $FAP \leq 0.05$.

Table 2.1: Values of $\hat{a}$ and $\hat{b}$ so that the false alarm probability of the Phase I S^2 chart is less than or equal to 0.01 when m = 3(1)10,15,20,25 and n = 3(1)10

	Sample size (n)							
m	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
3	0.0008 0.9595	0.0053 0.9080	0.0135 0.8604	0.0219 0.8199	0.0320 0.7881	0.0418 0.7621	0.0501 0.7384	0.0600 0.7176
4	0.0004 0.8910	0.0029 0.8107	0.0078 0.7520	0.0141 0.7074	0.0208 0.6669	0.0271 0.6418	0.0336 0.6144	0.0389 0.5954
5	0.0002 0.8213	0.0018 0.7284	0.0056 0.6683	0.0099 0.6200	0.0146 0.5800	0.0194 0.5496	0.0240 0.5286	0.0288 0.5101
6	0.0001 0.7615	0.0013 0.6624	0.0040 0.5952	0.0074 0.5494	0.0109 0.5105	0.0149 0.4812	0.0184 0.4630	0.0223 0.4445
7	0.0001 0.7011	0.0010 0.5997	0.0031 0.5384	0.0059 0.4915	0.0087 0.4588	0.0122 0.4322	0.0155 0.4124	0.0179 0.3956
8	0.0000 0.6533	0.0008 0.5521	0.0024 0.4869	0.0050 0.4472	0.0074 0.4161	0.0099 0.3909	0.0125 0.3722	0.0156 0.3543
9	0.0000 0.6042	0.0006 0.5132	0.0019 0.4533	0.0041 0.4108	0.0062 0.3832	0.0084 0.3567	0.0107 0.3367	0.0133 0.3222
10	0.0000 0.5715	0.0005 0.4749	0.0017 0.4182	0.0033 0.3793	0.0051 0.3495	0.0075 0.3296	0.0095 0.3100	0.0115 0.2950
15	0.0000 0.4366	0.0002 0.3533	0.0009 0.3092	0.0019 0.2766	0.0030 0.2542	0.0043 0.2363	0.0055 0.2221	0.0067 0.2114
20	0.0000 0.3564	0.0001 0.2822	0.0005 0.2440	0.0012 0.2172	0.0020 0.1990	0.0029 0.1846	0.0038 0.1737	0.0047 0.1648
25	0.0000 0.2985	0.0001 0.2370	0.0004 0.2022	0.0008 0.1806	0.0015 0.1650	0.0021 0.1521	0.0028 0.1432	0.0035 0.1361

Table 2.2: Values of $\hat{a}$ and $\hat{b}$ so that the false alarm probability of the Phase I S^2 chart is less than or equal to 0.05 when m = 3(1)10,15,20,25 and n = 3(1)10

	Sample size (n)							
m	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
3	0.0041	0.0156	0.0301	0.0433	0.0568	0.0694	0.0790	0.0889
	0.9088	0.8413	0.7877	0.7483	0.7173	0.6896	0.6682	0.6498
4	0.0020	0.0089	0.0179	0.0273	0.0361	0.0444	0.0524	0.0580
	0.8151	0.7303	0.6722	0.6303	0.5964	0.5712	0.5514	0.5335
5	0.0012	0.0056	0.0123	0.0191	0.0257	0.0319	0.0380	0.0431
	0.7329	0.6433	0.5854	0.5440	0.5116	0.4867	0.4673	0.4507
6	0.0008	0.0041	0.0089	0.0144	0.0197	0.0245	0.0292	0.0337
	0.6673	0.5765	0.5183	0.4789	0.4492	0.4253	0.4063	0.3929
7	0.0005	0.0031	0.0069	0.0115	0.0156	0.0198	0.0239	0.0272
	0.6067	0.5203	0.4662	0.4271	0.4008	0.3789	0.3617	0.3482
8	0.0004	0.0024	0.0055	0.0094	0.0129	0.0162	0.0197	0.0230
	0.5625	0.4740	0.4216	0.3876	0.3612	0.3415	0.3249	0.3116
9	0.0003	0.0019	0.0046	0.0078	0.0109	0.0139	0.0167	0.0196
	0.5210	0.4389	0.3882	0.3546	0.3301	0.3119	0.2960	0.2839
10	0.0002	0.0016	0.0039	0.0066	0.0093	0.0121	0.0146	0.0171
	0.4866	0.4072	0.3599	0.3282	0.3034	0.2863	0.2724	0.2597
15	0.0001	0.0008	0.0020	0.0036	0.0053	0.0070	0.0086	0.0100
	0.3676	0.3026	0.2629	0.2374	0.2190	0.2054	0.1943	0.1857
20	0.0000	0.0005	0.0013	0.0023	0.0035	0.0047	0.0059	0.0070
	0.2972	0.2405	0.2085	0.1872	0.1720	0.1609	0.1521	0.1446
25	0.0000	0.0003	0.0009	0.0017	0.0026	0.0035	0.0044	0.0052
	0.2492	0.2016	0.1734	0.1554	0.1430	0.1328	0.1256	0.1195

Table 2.3: Values of $\hat{a}$ and $\hat{b}$ so that the false alarm probability of the Phase I S^2 chart is less than or equal to 0.10 when m = 3(1)10,15,20,25 and n = 3(1)10

	Sample size (n)							
m	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
3	0.0083	0.0248	0.0431	0.0586	0.0740	0.0863	0.0967	0.1066
	0.8709	0.7979	0.7441	0.7078	0.6778	0.6516	0.6325	0.6161
4	0.0042	0.0141	0.0256	0.0366	0.0467	0.0555	0.0639	0.0700
	0.7689	0.6823	0.6283	0.5904	0.5586	0.5363	0.5176	0.5023
5	0.0025	0.0092	0.0175	0.0254	0.0329	0.0397	0.0463	0.0516
	0.6817	0.5969	0.5445	0.5063	0.4780	0.4549	0.4383	0.4237
6	0.0016	0.0067	0.0128	0.0193	0.0256	0.0307	0.0356	0.0401
	0.6159	0.5323	0.4803	0.4451	0.4185	0.3975	0.3813	0.3685
7	0.0012	0.0049	0.0100	0.0153	0.0201	0.0247	0.0290	0.0326
	0.5598	0.4795	0.4309	0.3971	0.3729	0.3529	0.3376	0.3257
8	0.0009	0.0039	0.0080	0.0125	0.0167	0.0204	0.0241	0.0274
	0.5164	0.4374	0.3910	0.3595	0.3353	0.3183	0.3036	0.2922
9	0.0007	0.0031	0.0066	0.0105	0.0140	0.0173	0.0204	0.0235
	0.4779	0.4029	0.3582	0.3285	0.3069	0.2904	0.2768	0.2663
10	0.0005	0.0026	0.0056	0.0089	0.0120	0.0150	0.0178	0.0204
	0.4456	0.3731	0.3314	0.3035	0.2819	0.2669	0.2543	0.2439
15	0.0002	0.0012	0.0029	0.0049	0.0067	0.0087	0.0104	0.0120
	0.3352	0.2768	0.2420	0.2198	0.2035	0.1914	0.1811	0.1737
20	0.0001	0.0007	0.0019	0.0031	0.0045	0.0058	0.0071	0.0083
	0.2699	0.2204	0.1922	0.1734	0.1597	0.1500	0.1420	0.1357
25	0.0000	0.0005	0.0013	0.0023	0.0033	0.0043	0.0053	0.0062
	0.2276	0.1846	0.1603	0.1442	0.1328	0.1237	0.1175	0.1119

Attained false alarm rate

To calculate the attained FAR of the Phase I S^2 chart designed such that its $FAP \leq FAP_0$, one needs the marginal distribution of Y_i for $i = 1,2,...,m$.

It can be verified (see e.g. Chapter 49 of Kotz, Balakrishnan and Johnson, (2000)) that each Y_i for $i = 1,2,...,m$ follows a standard or type I beta distribution with parameters $\frac{n-1}{2}$ and $\frac{(m-1)(n-1)}{2}$.

The beta distribution, in general, is denoted $Y_i \sim Beta(u,v)$, where $u,v > 0$ are the parameters of the distribution (see e.g. Gupta and Nadarajah, (2004)).

Given the *FAP*-based control limits $a = \hat{a}$ and $b = \hat{b}$, the attained false alarm rate can be calculated as

$$\begin{aligned} AFAR &= 1 - \Pr(\hat{a} < Y_i < \hat{b} \mid IC) \\ &= 1 - [\Pr(Y_i \leq \hat{b} \mid IC) - \Pr(Y_i \leq \hat{a} \mid IC)] \\ &= [1 - I_{\hat{b}}(\frac{n-1}{2}, \frac{(m-1)(n-1)}{2})] + I_{\hat{a}}(\frac{n-1}{2}, \frac{(m-1)(n-1)}{2}) \\ &= AFAR_U + AFAR_L \end{aligned} \tag{2-31}$$

where $$I_x(u,v) = [B(u,v)]^{-1} \int_0^x t^{u-1}(1-t)^{v-1} dt, \quad 0 < x < 1$$

is the c.d.f of the beta distribution, also known as the incomplete beta function (see e.g. Gupta and Nadarajah, (2004)).

Some *AFAR* values for the Phase I S^2 chart are shown in Table 2.4 for some selected values of m and n.

Table 2.4: *AFAR* values for the S^2 chart for selected values of *m* and *n* when $FAP_0 = 0.05$

	Sample size (*n*)			
m	**4**	**6**	**8**	**10**
15	0.00322	0.00328	0.00335	0.00327
20	0.00252	0.00240	0.00249	0.00256
25	0.00182	0.00196	0.00201	0.00195

Example 2

Consider again Example 1. It was found that $\hat{a} = 0.0115$ and $\hat{b} = 0.4271$. Since $m = 7$ and $n = 6$ it follows that $Y_i \sim Beta(2.5, 15)$ for $i = 1,2,...,m$ so that the attained false alarm rate of this Phase I S^2 chart corresponding to a $FAP_0 = 0.05$ can be calculated using (2-31) and is equal to

$$\begin{aligned} AFAR &= [1 - I_{0.4271}(2.5, 15)] + I_{0.0115}(2.5, 15) \\ &= 0.003654 + 0.003737 \\ &= 0.007391. \end{aligned}$$

The $AFAR$ is different from the typical and anticipated 0.0027 and is a result of the parameter estimation and the simultaneous comparisons. Note that the tail false alarm probabilities $AFAR_U$ and $AFAR_L$ are unequal in this case; this is so since the $Beta(2.5, 15)$ distribution is asymmetric.

Remark 9

Because marginally each $Y_i \sim Beta(\frac{n-1}{2}, \frac{(m-1)(n-1)}{2})$ for $i = 1,2,...,m$ it follows that each $\theta_i = \frac{n-1}{2}$ for $i = 1,2,...,m$ in (2-24) and (2-25). Thus, in our situation (with equal sample sizes) the correlation between any Y_i and Y_j is equal to

$$\text{corr}(Y_i, Y_j) = -1/(m-1) \quad \text{for all} \quad i \neq j = 1,2,...,m-1 \tag{2-32}$$

and follows by substituting $\frac{n-1}{2}$ for θ_i where $i = 1,2,3,...,m-1$ in (2-25).

The result of (2-32) means that any two of the Y_i's are equally and negatively correlated; this corresponds to the result of Champ and Jones (2004) in case of the Shewhart-type Phase I $\overline{X}$ chart. Most importantly, as m increases $\text{corr}(Y_i, Y_j) = -1/(m-1)$ tends to zero; the significance thereof is described below.

Approximate *FAR*-based control limits for the Phase I S^2 chart

Because the correlation between Y_i and Y_j approaches zero as the number of subgroups m increases, we can approximate a and b assuming that the Y_i's are independently distributed when m is large. In particular, for $m \geq 25$ the correlation between any pair of Y_i's is less than 0.05 in absolute value.

If the Y_i's are independent and identically distributed the false alarm probability equals

$$\begin{aligned} FAP &= 1 - \Pr\left(\bigcap_{i=1}^{m} \{a < Y_i < b\} \mid IC\right) \\ &\approx 1 - \prod_{i=1}^{m} \Pr(a < Y_i < b \mid IC) \\ &= 1 - \Pr(a < Y_i < b \mid IC)^m \\ &= 1 - (1 - FAR)^m . \end{aligned} \tag{2-33}$$

It follows from (2-33) that

$$FAR = 1 - \Pr(a < Y_i < b \mid IC) \approx 1 - (1 - FAP)^{\frac{1}{m}} \tag{2-34}$$

so that, under the equal-tailed approach, we may approximate a and b for a specified FAP_0 such that

$$P(Y_i \leq a \mid IC) = P(Y_i \geq b \mid IC) = \frac{1}{2}[1 - (1 - FAP_0)^{\frac{1}{m}}] . \tag{2-35}$$

Thus, a and b can be approximated by the

$$\frac{1}{2}[1 - (1 - FAP_0)^{\frac{1}{m}}]100^{\text{th}} \quad \text{and} \quad [1 - \frac{1}{2}\{1 - (1 - FAP_0)^{\frac{1}{m}}\}]100^{\text{th}}$$

percentiles, of a $Beta(\frac{(n-1)}{2}, \frac{(m-1)(n-1)}{2})$ distribution; these approximate FAR-based control limits are denoted by $\tilde{a}$ and $\tilde{b}$, respectively.

Moreover, it follows from (2-31) and (2-33) that the (approximate) attained false alarm rate is

$$AFAR = 1 - P(\tilde{a} < Y_i < \tilde{b} \mid IC)^{\frac{1}{m}} \approx 1 - (1 - FAP)^{\frac{1}{m}} .$$

Approximate *FAR*-based charting constants

Tables 2.5, 2.6 and 2.7 give the approximate values of a and b, denoted $\tilde{a}$ and $\tilde{b}$, for m = 25,30,50,100,300, n = 3(1)10 and a false alarm probability of 0.01, 0.05 and 0.10, respectively.

Table 2.5: Values of $\tilde{a}$ and $\tilde{b}$ so that the false alarm probability of the Phase I S^2 chart approximately equals 0.01 when m = 25,30,50,100,300 and n = 3(1)10

	Sample size (n)							
m	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
25	0.0000	0.0001	0.0004	0.0009	0.0015	0.0022	0.0029	0.0035
	0.2986	0.2374	0.2029	0.1805	0.1646	0.1526	0.1432	0.1356
30	0.0000	0.0001	0.0003	0.0007	0.0012	0.0017	0.0023	0.0028
	0.2590	0.2046	0.1742	0.1545	0.1406	0.1302	0.1220	0.1154
50	0.0000	0.0000	0.0001	0.0003	0.0006	0.0009	0.0012	0.0015
	0.1713	0.1333	0.1125	0.0991	0.0898	0.0828	0.0774	0.0730
100	0.0000	0.0000	0.0001	0.0001	0.0002	0.0004	0.0005	0.0006
	0.0952	0.0730	0.0610	0.0535	0.0482	0.0443	0.0413	0.0388
300	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0002
	0.0361	0.0273	0.0226	0.0197	0.0176	0.0161	0.0150	0.0140

Table 2.6: Values of $\tilde{a}$ and $\tilde{b}$ so that the false alarm probability of the Phase I S^2 chart approximately equals 0.05 when m = 25,30,50,100,300 and n = 3(1)10

	Sample size (n)							
m	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
25	0.0000	0.0003	0.0009	0.0017	0.0026	0.0035	0.0044	0.0053
	0.2493	0.2004	0.1729	0.1550	0.1423	0.1327	0.1251	0.1190
30	0.0000	0.0002	0.0007	0.0013	0.0020	0.0028	0.0035	0.0042
	0.2162	0.1728	0.1486	0.1328	0.1217	0.1133	0.1067	0.1014
50	0.0000	0.0001	0.0003	0.0006	0.0010	0.0014	0.0018	0.0022
	0.1432	0.1129	0.0963	0.0856	0.0780	0.0724	0.0680	0.0644
100	0.0000	0.0000	0.0001	0.0002	0.0004	0.0006	0.0007	0.0009
	0.0801	0.0623	0.0526	0.0465	0.0422	0.0389	0.0365	0.0345
300	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0002	0.0002
	0.0309	0.0236	0.0197	0.0173	0.0156	0.0143	0.0134	0.0126

Table 2.7: Values of $\tilde{a}$ and $\tilde{b}$ so that the false alarm probability of the Phase I S^2 chart approximately equals 0.10 when *m* = 25,30,50,100,300 and *n* = 3(1)10

	Sample size (*n*)							
m	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
25	0.0001	0.0006	0.0014	0.0024	0.0034	0.0044	0.0054	0.0063
	0.2265	0.1834	0.1592	0.1433	0.1320	0.1235	0.1168	0.1114
30	0.0001	0.0004	0.0010	0.0018	0.0026	0.0035	0.0042	0.0050
	0.1965	0.1583	0.1369	0.1229	0.1130	0.1056	0.0998	0.0950
50	0.0000	0.0002	0.0005	0.0009	0.0013	0.0018	0.0022	0.0026
	0.1306	0.1038	0.0890	0.0794	0.0727	0.0676	0.0637	0.0605
100	0.0000	0.0001	0.0002	0.0003	0.0005	0.0007	0.0009	0.0011
	0.0734	0.0575	0.0489	0.0433	0.0394	0.0365	0.0343	0.0325
300	0.0000	0.0000	0.0000	0.0001	0.0001	0.0002	0.0002	0.0003
	0.0285	0.0220	0.0184	0.0162	0.0147	0.0135	0.0126	0.0119

Comparing the approximate charting constants $\tilde{a}$ and $\tilde{b}$ from Tables 2.5, 2.6 and 2.7 with the exact (simulated) charting constants $\hat{a}$ and $\hat{b}$ in Tables 2.1, 2.2 and 2.3, for $m = 25$, we see that the values are almost identical and the approximation is thus reasonable.

2.2.2 Phase I *S* chart

Introduction

In situations where it is desirable to estimate and monitor the process spread using the sample standard deviation the S chart is used.

Charting statistics and control limits

The charting statistics for the Phase I S chart are the sample standard deviations S_i for $i = 1,2,...,m$. The estimated *k*-sigma control limits and the centerline of the Phase I S chart are

$$L\hat{C}L = \overline{S} - k_L \frac{\overline{S}}{c_4}\sqrt{1-c_4^2} \qquad \hat{C}L = \overline{S} \qquad U\hat{C}L = \overline{S} + k_U \frac{\overline{S}}{c_4}\sqrt{1-c_4^2}\,. \qquad (2\text{-}36)$$

Typically the charting constants $k_L, k_U \geq 0$ are taken to be $k_L = k_U = k = 3$; in these scenarios the constants B_3 and B_4 are defined as

$$B_3 = 1 - \frac{3}{c_4}\sqrt{1-c_4^2} \qquad \text{and} \qquad B_4 = 1 + \frac{3}{c_4}\sqrt{1-c_4^2} \qquad (2\text{-}37)$$

and the control limits are written as

$$L\hat{C}L = B_3\overline{S} \qquad \hat{C}L = \overline{S} \qquad U\hat{C}L = B_4\overline{S}$$

(see e.g. Montgomery, (2005) p. 224).

For more flexibility and to account for the fact that the sampling distribution of S_i is not symmetric we assume that k_L is not necessarily equal to k_U.

Design of the Phase I *S* chart for a nominal *FAP*

The Phase I S chart is applied in a manner similar to the S^2 chart. The aim is also the same i.e. we want to find values for the charting constants k_L and k_U so that if the m charting statistics are simultaneously plotted on the S chart the probability that at least one of the S_i's for $i = 1,2,...,m$ plot outside $L\hat{C}L$ and/or $U\hat{C}L$ is at most equal to FAP_0. To design the Phase I S chart we need an expression for the false alarm probability.

False alarm probability of the Phase I *S* chart

The false alarm probability of the Phase I S chart is derived as follows:

$$
\begin{aligned}
FAP &= 1 - \Pr\left(\bigcap_{i=1}^{m}\{L\hat{C}L < S_i < U\hat{C}L\} \mid IC\right) \\
&= 1 - \Pr\left(\bigcap_{i=1}^{m}\left\{\overline{S}\left(1 - \frac{k_L}{c_4}\sqrt{1-c_4^2}\right) < S_i < \overline{S}\left(1 + \frac{k_U}{c_4}\sqrt{1-c_4^2}\right)\right\} \mid IC\right) \\
&= 1 - \Pr\left(\bigcap_{i=1}^{m}\left\{1 - \frac{k_L}{c_4}\sqrt{1-c_4^2} < \frac{S_i}{\overline{S}} < 1 + \frac{k_U}{c_4}\sqrt{1-c_4^2}\right\} \mid IC\right) \\
&= 1 - \Pr\left(\bigcap_{i=1}^{m}\left\{\frac{1 - \frac{k_L}{c_4}\sqrt{1-c_4^2}}{m} < \frac{\frac{\sqrt{n-1}S_i}{\sigma}}{\sum_{j=1}^{m}\frac{\sqrt{n-1}S_j}{\sigma}} < \frac{1 + \frac{k_U}{c_4}\sqrt{1-c_4^2}}{m}\right\} \mid IC\right) \\
&= 1 - \Pr\left(\bigcap_{i=1}^{m}\left\{\frac{B_3^{**}}{m} < V_i < \frac{B_4^{**}}{m}\right\} \mid IC\right) \\
&= 1 - \Pr\left(\bigcap_{i=1}^{m}\{c < V_i < d\} \mid IC\right)
\end{aligned}
\qquad (2\text{-}38)
$$

where

$$
c = \frac{B_3^{**}}{m} \qquad \text{and} \qquad d = \frac{B_4^{**}}{m} \qquad (2\text{-}39)
$$

with

$$
B_3^{**} = 1 - \frac{k_L}{c_4}\sqrt{1-c_4^2} \qquad \text{and} \qquad B_4^{**} = 1 + \frac{k_U}{c_4}\sqrt{1-c_4^2} \qquad (2\text{-}40)
$$

and

$$V_i = \frac{\frac{\sqrt{n-1}S_i}{\sigma}}{\sum_{j=1}^{m}\frac{\sqrt{n-1}S_j}{\sigma}} = \frac{W_i}{\sum_{j=1}^{m}W_j}$$

where

$$W_i = \frac{\sqrt{n-1}S_i}{\sigma} \qquad \text{for} \qquad i = 1,2,...,m$$

are independent and identically distributed chi random variables each with $n-1$ degrees of freedom.

Remark 10

If $k_L = k_U = 3$ it follows from (2-37) and (2-40) that $B_3^{**} = B_3$ and $B_4^{**} = B_4$.

FAP-based control limits for the _S_ chart

The design of a Phase I S chart requires one to find that combination(s) of values for c and d, denoted $\hat{c}$ and $\hat{d}$, so that

$$FAP = 1 - \Pr\left(\bigcap_{i=1}^{m}\{\hat{c} < V_i < \hat{d}\} \mid IC\right) \leq FAP_0 \qquad (2\text{-}41)$$

and then obtain k_L and k_U needed for calculating the estimated control limits.

A major problem in the analytical determination of $\hat{c}$ and $\hat{d}$ from (2-41) is finding a closed form expression for the joint distribution of $(V_1, V_2, ..., V_m)$. In this regard, note that, the V_i's are correlated and linearly dependent. In particular, it is seen from the definition of V_i for $i = 1,2,...,m$ that $\sum_{i=1}^{m} V_i = 1$. As a result, even for an in-control process, the joint distribution of $(V_1, V_2, ..., V_m)$ is complicated.

To overcome these obstacles in designing the Phase I S chart we make use of the equal-tails approach (described in section 2.1.1) coupled with computer simulation (as was the case for the S^2 chart) and obtain the charting constants k_L and k_U of the S chart by first solving for $\hat{c}$ and $\hat{d}$ from

$$\Pr\left(\min(V_1, V_2, ..., V_m) \leq \hat{c} \mid IC\right) \leq \frac{FAP_0}{2} \quad \text{and} \quad \Pr\left(\max(V_1, V_2, ..., V_m) \geq \hat{d} \mid IC\right) \leq \frac{FAP_0}{2} \qquad (2\text{-}42)$$

and then calculating the values of k_L and k_U as

$$k_L = \frac{c_4(1 - m\hat{c})}{\sqrt{1 - c_4^2}} \quad \text{and} \quad k_U = \frac{c_4(m\hat{d} - 1)}{\sqrt{1 - c_4^2}}, \qquad (2\text{-}43)$$

where equation (2-43) follows from the definitions of c, d, B_3^{**} and B_4^{**} given in (2-39) and (2-40), respectively.

The steps of the simulation algorithm for determining the charting constants of the S chart are similar to those of the S^2 chart and found by modifying steps 1, 2 and 3, described earlier, in a natural way.

Simulation algorithm for determining the *FAP*-based control limits of the *S* chart

Step 1: Generate 100,000 vector valued observations from the joint distribution of $(V_1, V_2, ..., V_m)$.

To obtain one such observation we generate m independent χ^2_{n-1} random variables for a given *m* and *n* (denoted by X_i for $i = 1,2,...,m$), calculate $\sqrt{X_i}$ for $i = 1,2,...,m$, calculate their sum $SUM_2 = \sum_{j=1}^{m} \sqrt{X_j}$, and then calculate $V_i = \frac{\sqrt{X_i}}{SUM_2}$ for $i = 1,2,...,m$. The vector $(V_1, V_2, ..., V_m)$ is one such observation.

Step 2: Find $V_{\max} = \max(V_1, V_2, ..., V_m)$ and $V_{\min} = \min(V_1, V_2, ..., V_m)$.

Step 3:

Let

$$\hat{c} = \max\left\{u : \hat{\Pr}_1(\min(V_1, V_2, ..., V_m) \le u \mid IC) \le \frac{FAP_0}{2}\right\}$$

and

$$\hat{d} = \min\left\{u : \hat{\Pr}_2(\max(V_1, V_2, ..., V_m) \ge u \mid IC) \le \frac{FAP_0}{2}\right\}$$

where $0 < u < 1$ and the empirical probabilities are defined as

$$\hat{\Pr}_1(u) = \frac{\text{number of simulated } Y_{\min} \text{ values } \le u}{100{,}000} \quad \text{and} \quad \hat{\Pr}_2(u) = \frac{\text{number of simulated } Y_{\max} \text{ values } \ge u}{100{,}000},$$

respectively.

FAP-based charting constants for the _S_ chart

Tables 2.8, 2.9 and 2.10 display the values of k_L and k_U for $m = 3(1)10,15,20,25,30,50,100,300$ and $n = 3(1)10$ so that the false alarm probability of the S chart do not exceed 0.01, 0.05 and 0.10, respectively.

For $n = 3$ and $m \geq 100$, the tabulated values of k_L is 1.9128, which results in a lower control limit of zero, when $FAP = 0.01$. Similar observations can also be made when $FAP = 0.05$ or 0.10 when $n = 3$ for $m = 300$. This is interesting to note because for the usual S chart, the lower control limit is negative for $n \leq 5$ and is therefore adjusted upwards to be equal to zero - see e.g. the values of the constant B_3 in Appendix VI of Montgomery, (2005) p. 725.

Table 2.8: Values of k_L and k_U so that the false alarm probability of the Phase I S chart is less than or equal to 0.01 when m = 3(1)10,15,20,25,30,50,100,300 and n = 3(1)10

	Sample size (n)							
m	3	4	5	6	7	8	9	10
3	1.7974	2.0132	2.1105	2.1789	2.2311	2.2435	2.2640	2.2613
	2.6131	2.6140	2.5992	2.5624	2.5600	2.5331	2.5293	2.5132
4	1.8156	2.0613	2.1898	2.2775	2.3335	2.3624	2.3937	2.4059
	3.0413	2.9607	2.9106	2.8788	2.8437	2.8098	2.7910	2.8183
5	1.8325	2.1004	2.2413	2.3320	2.3981	2.4360	2.4658	2.4855
	3.2957	3.1594	3.1079	3.0557	3.0376	3.0082	2.9980	2.9801
6	1.8393	2.1219	2.2841	2.3877	2.4403	2.4742	2.5277	2.5442
	3.4962	3.3369	3.2539	3.2110	3.1641	3.1270	3.1230	3.1168
7	1.8472	2.1421	2.3097	2.4108	2.4872	2.5438	2.5762	2.6188
	3.6037	3.4471	3.3952	3.3155	3.2988	3.2363	3.2270	3.2228
8	1.8516	2.1513	2.3364	2.4347	2.5253	2.5728	2.6113	2.6624
	3.7261	3.5653	3.4495	3.3996	3.3634	3.3353	3.3083	3.2828
9	1.8560	2.1667	2.3411	2.4582	2.5443	2.6001	2.6613	2.6637
	3.7940	3.6226	3.5531	3.4457	3.4407	3.4107	3.3772	3.3607
10	1.8631	2.1679	2.3612	2.4743	2.5614	2.6199	2.6728	2.7119
	3.8428	3.6961	3.5955	3.5475	3.4696	3.4773	3.4533	3.4412
15	1.8697	2.2022	2.4039	2.5407	2.6311	2.7083	2.7654	2.8146
	4.1240	3.9034	3.8242	3.7640	3.6958	3.6999	3.6504	3.6235
20	1.8822	2.2176	2.4356	2.5856	2.6736	2.7671	2.8226	2.8670
	4.2196	4.0562	3.9179	3.8660	3.8437	3.7754	3.7529	3.7639
25	1.8841	2.2271	2.4590	2.6134	2.7127	2.7874	2.8482	2.9026
	4.3659	4.1107	4.0364	3.9743	3.9288	3.9097	3.8633	3.8352
30	1.8841	2.2413	2.4659	2.6196	2.7382	2.8186	2.8896	2.9215
	4.4166	4.2410	4.0887	4.0423	3.9968	3.9373	3.9461	3.9065
50	1.8937	2.2627	2.5072	2.6753	2.8063	2.8886	2.9566	3.0179
	4.6576	4.4069	4.3119	4.2526	4.2179	4.1397	4.1195	4.0867
100	1.9128	2.2982	2.5623	2.7526	2.8913	2.9806	3.0355	3.1436
	4.8776	4.6675	4.5461	4.4846	4.4560	4.4156	4.3364	4.3172
300	1.9128	2.3693	2.6725	2.8145	2.9934	3.1277	3.2326	3.3113
	5.2028	4.9518	4.8491	4.7939	4.7622	4.7100	4.5729	4.6106

Table 2.9: Values of k_L and k_U so that the false alarm probability of the Phase I S chart is less than or equal to 0.05 when m = 3(1)10,15,20,25,30,50,100,300 and n = 3(1)10

	Sample size (n)							
m	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
3	1.6500	1.7722	1.8187	1.8579	1.8698	1.8748	1.8867	1.8878
	2.1442	2.1207	2.0851	2.0669	2.0743	2.0529	2.0420	2.0278
4	1.7001	1.8594	1.9319	1.9658	2.0001	2.0121	2.0326	2.0354
	2.4644	2.3968	2.3628	2.3406	2.3253	2.2976	2.2817	2.2852
5	1.7330	1.9132	2.0003	2.0614	2.0834	2.1103	2.1268	2.1418
	2.6693	2.5849	2.5362	2.5021	2.4933	2.4709	2.4658	2.4541
6	1.7521	1.9428	2.0526	2.1149	2.1382	2.1806	2.2060	2.2123
	2.8455	2.7228	2.6786	2.6524	2.6212	2.5905	2.5837	2.5811
7	1.7682	1.9762	2.0860	2.1446	2.2015	2.2321	2.2478	2.2873
	2.9597	2.8384	2.7858	2.7483	2.7345	2.6954	2.6751	2.6888
8	1.7797	1.9959	2.1226	2.1872	2.2505	2.2696	2.2928	2.3338
	3.0390	2.9360	2.8588	2.8404	2.8056	2.7730	2.7627	2.7530
9	1.7871	2.0217	2.1402	2.2188	2.2658	2.3053	2.3456	2.3506
	3.1243	3.0149	2.9530	2.8918	2.8682	2.8477	2.8344	2.8175
10	1.7980	2.0281	2.1656	2.2392	2.2926	2.3292	2.3692	2.3975
	3.1886	3.0730	3.0004	2.9567	2.9219	2.9106	2.8896	2.8795
15	1.8210	2.0814	2.2303	2.3273	2.3913	2.4433	2.4757	2.5002
	3.4354	3.3028	3.2167	3.1794	3.1345	3.1149	3.1005	3.0891
20	1.8363	2.1134	2.2813	2.3815	2.4491	2.5022	2.5387	2.5736
	3.5846	3.4355	3.3558	3.2969	3.2791	3.2529	3.2168	3.2107
25	1.8458	2.1324	2.3075	2.4124	2.4916	2.5390	2.5920	2.6197
	3.7060	3.5421	3.4646	3.4176	3.3930	3.3485	3.3213	3.3008
30	1.8497	2.1489	2.3254	2.4433	2.5239	2.5758	2.6294	2.6574
	3.7739	3.6297	3.5432	3.4856	3.4560	3.4295	3.4021	3.3658
50	1.8650	2.1916	2.3832	2.5052	2.6022	2.6678	2.7398	2.7664
	4.0073	3.8382	3.7471	3.6650	3.6567	3.6245	3.6071	3.5837
100	1.8937	2.2271	2.4521	2.5980	2.7212	2.7966	2.8383	2.8921
	4.2846	4.0989	3.9950	3.9588	3.9118	3.8637	3.8239	3.8142
300	1.9128	2.2982	2.5899	2.7217	2.8913	3.0174	2.9960	3.0598
	4.6863	4.4542	4.3532	4.2372	4.2519	4.1581	4.0998	4.1076

Table 2.10: Values of k_L and k_U so that the false alarm probability of the Phase I S is less than or equal to 0.10 when m = 3(1)10,15,20,25,30,50,100,300 and n = 3(1)10

	Sample size (n)							
m	3	4	5	6	7	8	9	10
3	1.5461	1.6251	1.6559	1.6704	1.6861	1.6838	1.6904	1.6904
	1.9032	1.8691	1.8446	1.8285	1.8212	1.8100	1.7996	1.7990
4	1.6136	1.7277	1.7810	1.8062	1.8259	1.8369	1.8497	1.8526
	2.1813	2.1276	2.0984	2.0759	2.0668	2.0489	2.0310	2.0354
5	1.6574	1.7971	1.8570	1.9021	1.9151	1.9392	1.9533	1.9637
	2.3757	2.3065	2.2717	2.2407	2.2297	2.2189	2.2076	2.2005
6	1.6844	1.8348	1.9138	1.9646	1.9811	2.0150	2.0334	2.0362
	2.5276	2.4427	2.4042	2.3778	2.3559	2.3322	2.3211	2.3196
7	1.7053	1.8750	1.9567	2.0060	2.0467	2.0647	2.0795	2.1054
	2.6477	2.5581	2.5138	2.4863	2.4631	2.4352	2.4185	2.4277
8	1.7246	1.8973	1.9970	2.0512	2.0926	2.1136	2.1256	2.1628
	2.7330	2.6498	2.5855	2.5633	2.5389	2.5199	2.5072	2.5015
9	1.7355	1.9257	2.0212	2.0796	2.1219	2.1464	2.1788	2.1846
	2.8162	2.7142	2.6629	2.6273	2.6018	2.5894	2.5790	2.5610
10	1.7464	1.9381	2.0499	2.1062	2.1498	2.1821	2.2076	2.2299
	2.8826	2.7768	2.7194	2.6877	2.6566	2.6494	2.6334	2.6239
15	1.7808	2.0068	2.1311	2.2067	2.2586	2.2943	2.3278	2.3493
	3.1312	3.0185	2.9439	2.9103	2.8794	2.8610	2.8403	2.8313
20	1.8018	2.0471	2.1821	2.2639	2.3267	2.3697	2.3968	2.4311
	3.2900	3.1559	3.0858	3.0433	3.0206	2.9953	2.9645	2.9592
25	1.8124	2.0672	2.2179	2.3119	2.3641	2.4102	2.4540	2.4835
	3.3904	3.2637	3.1960	3.1547	3.1209	3.0909	3.0749	3.0493
30	1.8267	2.0921	2.2427	2.3413	2.4117	2.4544	2.4993	2.5317
	3.4870	3.3525	3.2787	3.2258	3.2008	3.1645	3.1537	3.1394
50	1.8458	2.1442	2.3144	2.4279	2.5001	2.5574	2.6018	2.6406
	3.7204	3.5776	3.4991	3.4330	3.4015	3.3669	3.3705	3.3322
100	1.8745	2.2034	2.3970	2.5361	2.6192	2.6862	2.7595	2.8083
	4.0169	3.8619	3.7471	3.7114	3.6737	3.6429	3.5874	3.5628
300	1.9128	2.2982	2.5072	2.6289	2.7893	2.9070	2.9960	3.0598
	4.3994	4.2410	4.1052	4.0516	4.0478	3.9373	3.9816	3.8562

Attained false alarm rate

To analytically calculate the attained false alarm rate of the S chart for given m, n and FAP_0 the marginal distribution of V_i i.e. the ratio of a chi random variable, W_i, to the sum of m independent chi random variables, $\sum_{j=1}^{m} X_j$, is needed. Currently, the distribution of $\sum_{j=1}^{m} X_j$ for $m \geq 3$ is unknown; only the distributions of the sum and the ratio of $m = 2$ correlated chi variates are known and available in the literature (see e.g. Krishnan, (1967)). Thus, we used computer simulation to determine the $AFAR$ for selected values of m and n when $FAP_0 = 0.05$. These values are shown in Table 2.11.

Table 2.11: The *AFAR* values for the S chart for selected *m*, *n* values when $FAP_0 = 0.05$

	Sample size (*n*)			
m	**4**	**6**	**8**	**10**
15	0.00324	0.00343	0.00329	0.00320
20	0.00252	0.00261	0.00247	0.00258
25	0.00209	0.00177	0.00212	0.00192
50	0.00096	0.00112	0.00098	0.00069
100	0.00062	0.00051	0.00046	0.00061

From Table 2.11 it is seen that, for a fixed FAP_0 of 0.05, the $AFAR$ (i) decreases as the number of samples, m, increases, for a fixed sample size *n*, and (ii) stays fairly constant for a fixed *m* but with increasing n. Also, note that for $m = 20$ and $FAP_0 = 0.05$, the $AFAR$ is close to 0.0027 for all values of *n* considered.

2.2.3 Phase I R chart

Introduction

Finally we consider the R chart. This chart is popular in the industry since the range is easy to calculate and for small n it is known that the range is a fairly efficient estimator of the standard deviation of a normal distribution.

Charting statistics and control limits

For the Phase I R chart the charting statistics are the sample ranges R_i for $i = 1,2,...,m$ and we define the estimated k-sigma control limits and the centerline of the R chart as

$$L\hat{C}L = D_3^{**}\overline{R} \qquad \hat{C}L = \overline{R} \qquad U\hat{C}L = D_4^{**}\overline{R} \tag{2-44}$$

where

$$D_3^{**} = 1 - k_L\frac{d_3}{d_2} \quad \text{and} \quad D_4^{**} = 1 + k_U\frac{d_3}{d_2}. \tag{2-45}$$

and $k_L, k_U \geq 0$ are the charting constants.

If $k_L = k_U = 3$ then $D_3^{**} = D_3$ and $D_4^{**} = D_4$ where

$$D_3 = 1 - 3\frac{d_3}{d_2} \quad \text{and} \quad D_4 = 1 + 3\frac{d_3}{d_2} \tag{2-46}$$

(see e.g. Montgomery, (2005) p. 197 and p. 198).

As in the case of the S chart, the definitions of D_3^{**} and D_4^{**} in (2-45) extend the usual R chart by accounting for the fact that the sampling distribution of R_i is asymmetric and thus allows for a charting constant(s) other than 3.

False alarm probability of the *R* chart

An expression for the false alarm probability of the R chart is needed to design the chart. Such an expression is obtained in a similar manner as that of the S^2 chart and the S chart and is derived as follows:

$$\begin{aligned}
FAP &= 1-\Pr\left(\bigcap_{i=1}^{m}\{L\hat{C}L < R_i < U\hat{C}L\} \mid IC\right)\\
&= 1-\Pr\left(\bigcap_{i=1}^{m}\{D_3^{**}\bar{R} < R_i < D_4^{**}\bar{R}\} \mid IC\right)\\
&= 1-\Pr\left(\bigcap_{i=1}^{m}\left\{D_3^{**} < \frac{R_i}{\bar{R}} < D_4^{**}\right\} \mid IC\right)\\
&= 1-\Pr\left(\bigcap_{i=1}^{m}\left\{\frac{D_3^{**}}{m} < \frac{R_i}{\sum_{j=1}^{m} R_j} < \frac{D_4^{**}}{m}\right\} \mid IC\right)\\
&= 1-\Pr\left(\bigcap_{i=1}^{m}\left\{\frac{D_3^{**}}{m} < U_i < \frac{D_4^{**}}{m}\right\} \mid IC\right)\\
&= 1-\Pr\left(\bigcap_{i=1}^{m}\{e < U_i < f\} \mid IC\right)
\end{aligned} \tag{2-47}$$

where

$$e = \frac{D_3^{**}}{m} \qquad \text{and} \qquad f = \frac{D_4^{**}}{m} \tag{2-48}$$

and

$$U_i = \frac{R_i}{\sum_{j=1}^{m} R_j}. \tag{2-49}$$

***FAP*-based control limits for the *R* chart**

As in the case of the *S* chart, the charting constants k_L and k_U of the R chart are obtained by first finding that combination of values for the charting constants $e = \hat{e}$ and $f = \hat{f}$ such that

$$\Pr\left(\min(U_1, U_2, ..., U_m) \le \hat{e} \mid IC\right) \le \frac{FAP_0}{2} \quad \text{and} \quad \Pr(\max(U_1, U_2, ..., U_m) \ge \hat{f} \mid IC) \le \frac{FAP_0}{2} \qquad (2\text{-}50)$$

and then, once e and f are found, we calculate the charting constants k_L and k_U from

$$k_L = (1 - \hat{e}m)\frac{d_2}{d_3} \qquad \text{and} \qquad k_U = (\hat{f}m - 1)\frac{d_2}{d_3} \qquad (2\text{-}51)$$

which follow directly from the definitions of e, f, D_3^{**} and D_4^{**} given in (2-48) and (2-45), respectively

Here, similar to the S chart, there are problems in the analytical determination of the constants $\hat{e}$ and $\hat{f}$; these obstacles arise due to the fact that even for an in-control process, the marginal distribution of U_i as well as the joint distribution of $(U_1, U_2, ..., U_m)$ are complicated. Thus, we again find the values of $\hat{e}$ and $\hat{f}$ via computer simulation.

Simulation algorithm for determining the *FAP*-based control limits of the *R* chart

The steps of the simulation algorithm to find $\hat{e}$ and $\hat{f}$ were as follows:

Step 1: Generate 100,000 vector valued observations from the joint distribution of $(U_1, U_2, ..., U_m)$. To obtain one such observation we: (i) generate m independent random samples each of size n from the standard normal distribution, (ii) calculate the range R_i, $i = 1,2,...,m$ for each sample, (iii) obtain the sum $SUM_3 = \sum_{j=1}^{m} R_j$, and (iv) calculate $U_i = \frac{R_i}{SUM_3}$ for $i = 1,2,...,m$. The vector $(U_1, U_2, ..., U_m)$ is one such observation.

Step 2: Find $U_{\max} = \max(U_1, U_2, ..., U_m)$ and $U_{\min} = \min(U_1, U_2, ..., U_m)$.

Step 3:

Let

$$\hat{e} = \max\left\{u : \hat{\Pr}_1(\min(U_1, U_2, ..., U_m) \le u \mid IC) \le \frac{FAP_0}{2}\right\}$$

and

$$\hat{f} = \min\left\{u : \hat{\Pr}_2(\max(U_1, U_2, ..., U_m) \ge u \mid IC) \le \frac{FAP_0}{2}\right\}$$

where $0 < u < 1$ and the empirical probabilities "$\Pr_1$" and "$\Pr_2$"are as defined earlier.

FAP-based charting constants for the _R_ chart

Tables 2.12, 2.13 and 2.14 display the values of the charting constants k_L and k_U for $m = 3(1)10,15,20,25,30,50,100,300$ and $n = 3(1)10$ so that the false alarm probability does not exceed 0.01, 0.05 and 0.10, respectively.

Note that, similar to the S chart, when $FAP = 0.01$, $n = 3$ and $m \geq 100$, the tabulated value of k_L is 1.9065, which results in a lower control limit equal to zero, for the Phase I R chart. Similar observations can also be made when $FAP = 0.05$ or 0.10 for $m \geq 300$ and when $n = 3$. For the usual R chart with symmetrically placed limits, on the other hand, the lower control limit is negative for $n \leq 6$ and is therefore adjusted upwards to be equal to zero - see e.g. the values of the constant D_3 in Appendix VI of Montgomery (2005) p. 725.

Table 2.12: Values of k_L and k_U so that the false alarm probability of the Phase I R chart is less than or equal to 0.01 when m = 3(1)10,15,20,25,30,50,100,300 and n = 3(1)10

	Sample size (n)							
m	3	4	5	6	7	8	9	10
3	1.7881	1.9853	2.0686	2.1222	2.1340	2.1564	2.1782	2.1890
	2.5799	2.6011	2.5898	2.5591	2.5560	2.5255	2.5127	2.5265
4	1.8112	2.0487	2.1569	2.2172	2.2463	2.2845	2.2760	2.2863
	3.0321	2.9575	2.8946	2.8711	2.8618	2.8664	2.8450	2.8501
5	1.8246	2.0719	2.2143	2.2695	2.3112	2.3366	2.3488	2.3771
	3.2993	3.1739	3.1538	3.1107	3.0805	3.0484	3.0619	3.0799
6	1.8368	2.0997	2.2318	2.3194	2.3599	2.3991	2.4120	2.4323
	3.5111	3.3599	3.3006	3.2727	3.2454	3.2296	3.2383	3.2356
7	1.8411	2.1138	2.2681	2.3398	2.3985	2.4318	2.4407	2.4697
	3.6426	3.4942	3.4457	3.4167	3.3980	3.3525	3.3640	3.3344
8	1.8455	2.1226	2.2808	2.3738	2.4229	2.4665	2.4789	2.5026
	3.7124	3.6070	3.5041	3.4735	3.4798	3.4636	3.4258	3.4356
9	1.8516	2.1418	2.3045	2.3965	2.4398	2.4814	2.5146	2.5203
	3.8091	3.6639	3.5978	3.5685	3.5464	3.5150	3.5394	3.5449
10	1.8570	2.1432	2.3098	2.4055	2.4735	2.5102	2.5326	2.5528
	3.8817	3.7647	3.6748	3.6187	3.6356	3.6247	3.6169	3.6264
15	1.8665	2.1783	2.3610	2.4638	2.5352	2.5762	2.6171	2.6397
	4.1505	3.9741	3.9184	3.9056	3.8580	3.8504	3.8834	3.8311
20	1.8722	2.1947	2.3852	2.5101	2.5839	2.6248	2.6612	2.6879
	4.2783	4.1461	4.0705	4.0580	4.0381	4.0344	4.0139	4.0165
25	1.8779	2.2052	2.4095	2.5250	2.6131	2.6647	2.6925	2.7420
	4.3803	4.2291	4.1593	4.1088	4.1225	4.0882	4.1260	4.0937
30	1.8837	2.2134	2.4095	2.5489	2.6326	2.6908	2.7384	2.7613
	4.4708	4.3216	4.2455	4.2014	4.2427	4.1837	4.2087	4.2250
50	1.8875	2.2462	2.4633	2.5848	2.6943	2.7602	2.7936	2.8386
	4.6996	4.5158	4.4824	4.4973	4.4472	4.4441	4.4293	4.4606
100	1.9065	2.2696	2.5037	2.6595	2.7592	2.8470	2.9038	2.9351
	4.9189	4.7731	4.7112	4.7214	4.7068	4.7219	4.7050	4.7116
300	1.9065	2.3398	2.6114	2.7193	2.8566	2.9512	3.0141	3.0510
	5.3002	5.1709	5.0612	5.0800	5.0315	5.0690	5.0358	5.0592

Table 2.13: Values of k_L and k_U so that the false alarm probability of the Phase I R chart is less than or equal to 0.05 when m = 3(1)10,15,20,25,30,50,100,300 and n = 3(1)10

	Sample size (n)							
m	3	4	5	6	7	8	9	10
3	1.6451	1.7523	1.7900	1.8210	1.8175	1.8273	1.8353	1.8286
	2.1401	2.1070	2.0891	2.0607	2.0564	2.0516	2.0485	2.0411
4	1.6945	1.8428	1.8974	1.9292	1.9490	1.9610	1.9555	1.9634
	2.4594	2.4025	2.3594	2.3475	2.3320	2.3318	2.3113	2.3187
5	1.7273	1.8917	1.9653	1.9991	2.0239	2.0380	2.0400	2.0584
	2.6863	2.5901	2.5683	2.5385	2.5190	2.5033	2.5032	2.5084
6	1.7487	1.9298	2.0089	2.0577	2.0833	2.1012	2.1099	2.1218
	2.8316	2.7478	2.7061	2.7043	2.6514	2.6505	2.6473	2.6516
7	1.7611	1.9565	2.0571	2.0992	2.1327	2.1498	2.1551	2.1697
	2.9593	2.8718	2.8351	2.8143	2.7959	2.7668	2.7697	2.7694
8	1.7693	1.9748	2.0783	2.1348	2.1606	2.1859	2.2054	2.2091
	3.0505	2.9575	2.9140	2.8854	2.8696	2.8664	2.8494	2.8548
9	1.7813	1.9965	2.1082	2.1653	2.1885	2.2095	2.2301	2.2353
	3.1416	3.0405	2.9920	2.9715	2.9533	2.9307	2.9307	2.9401
10	1.7902	2.0075	2.1187	2.1784	2.2268	2.2464	2.2606	2.2708
	3.2049	3.1096	3.0502	3.0360	3.0156	3.0137	3.0141	3.0123
15	1.8150	2.0590	2.1874	2.2621	2.3112	2.3314	2.3525	2.3674
	3.4670	3.3424	3.2804	3.2870	3.2591	3.2567	3.2549	3.2402
20	1.8303	2.0918	2.2291	2.3129	2.3632	2.3956	2.4186	2.4330
	3.6072	3.5003	3.4459	3.4364	3.4214	3.4095	3.3964	3.3985
25	1.8350	2.1058	2.2614	2.3457	2.4021	2.4390	2.4627	2.4910
	3.7320	3.6033	3.5671	3.5261	3.5220	3.5154	3.5103	3.5047
30	1.8436	2.1222	2.2802	2.3696	2.4281	2.4720	2.4958	2.5180
	3.8131	3.7039	3.6478	3.6187	3.6194	3.5900	3.5912	3.5994
50	1.8589	2.1643	2.3287	2.4354	2.5157	2.5519	2.5914	2.6068
	4.0514	3.9074	3.8767	3.8697	3.8466	3.8365	3.8412	3.8427
100	1.8875	2.1994	2.3960	2.5101	2.5969	2.6734	2.6833	2.7420
	4.3278	4.1882	4.1459	4.1237	4.1225	4.1316	4.1168	4.1323
300	1.9065	2.2696	2.5306	2.6296	2.7592	2.8470	2.9038	2.9351
	4.7282	4.6094	4.5766	4.5421	4.5445	4.5483	4.4844	4.4799

Table 2.14: Values of k_L and k_U so that the false alarm probability of the Phase I R chart is less than or equal to 0.10 when m = 3(1)10,15,20,25,30,50,100,300 and n = 3(1)10

	Sample size (n)							
m	3	4	5	6	7	8	9	10
3	1.5353 1.8970	1.6105 1.8620	1.6325 1.8460	1.6426 1.8267	1.6461 1.8149	1.6513 1.8110	1.6522 1.8092	1.6456 1.7989
4	1.6091 2.1826	1.7136 2.1339	1.7499 2.0966	1.7690 2.0882	1.7828 2.0788	1.7887 2.0762	1.7805 2.0599	1.7873 2.0654
5	1.6549 2.3851	1.7747 2.3129	1.8266 2.2856	1.8542 2.2710	1.8649 2.2528	1.8766 2.2394	1.8765 2.2477	1.8846 2.2399
6	1.6812 2.5319	1.8232 2.4586	1.8780 2.4267	1.9160 2.4246	1.9314 2.3865	1.9492 2.3859	1.9489 2.3826	1.9596 2.3828
7	1.6997 2.6470	1.8582 2.5885	1.9327 2.5411	1.9633 2.5277	1.9873 2.5050	2.0016 2.4970	2.0033 2.4892	2.0102 2.4991
8	1.7159 2.7378	1.8812 2.6673	1.9620 2.6189	1.9985 2.5985	2.0256 2.5917	2.0415 2.5915	2.0525 2.5760	2.0546 2.5767
9	1.7298 2.8327	1.9018 2.7520	1.9919 2.7061	2.0308 2.6837	2.0541 2.6728	2.0721 2.6494	2.0779 2.6561	2.0893 2.6551
10	1.7407 2.8922	1.9210 2.8124	2.0083 2.7648	2.0589 2.7492	2.0872 2.7365	2.1005 2.7290	2.1136 2.7274	2.1241 2.7188
15	1.7778 3.1582	1.9818 3.0476	2.0864 2.9936	2.1500 2.9957	2.1846 2.9718	2.2012 2.9650	2.2201 2.9682	2.2342 2.9563
20	1.7960 3.3097	2.0216 3.2008	2.1376 3.1606	2.2053 3.1496	2.2463 3.1357	2.2707 3.1178	2.2937 3.1097	2.3017 3.1282
25	1.8064 3.4270	2.0473 3.3166	2.1739 3.2777	2.2486 3.2422	2.2885 3.2380	2.3175 3.2376	2.3341 3.2438	2.3558 3.2248
30	1.8207 3.5156	2.0660 3.4090	2.1995 3.3571	2.2710 3.3408	2.3307 3.3370	2.3575 3.3192	2.3745 3.3155	2.3906 3.3213
50	1.8398 3.7559	2.1175 3.6383	2.2614 3.5940	2.3607 3.5858	2.4183 3.5707	2.4477 3.5588	2.4811 3.5471	2.4910 3.5723
100	1.8684 4.0418	2.1760 3.9308	2.3422 3.8767	2.4503 3.8548	2.5320 3.8629	2.5692 3.8539	2.6098 3.8595	2.6261 3.8620
300	1.9065 4.4422	2.2696 4.3286	2.4498 4.2536	2.6296 4.2731	2.6618 4.2524	2.7428 4.2358	2.7936 4.2639	2.8192 4.2482

Attained false alarm rate

To calculate the attained false alarm rate of the R chart given m, n and a specified FAP the (marginal) distribution of U_i i.e. the ratio of a range to the sum of m independent ranges, one of which is U_i, is required. Again as noted earlier, this distribution is complex and not available. Instead, we used simulation to determine the $AFAR$ for selected values of m and n when $FAP_0 = 0.05$. These values are shown in Table 2.15.

Table 2.15: *AFAR* values for the *R* chart for selected *m*, *n* values when $FAP_0 = 0.05$

	Sample size (n)			
m	**4**	**6**	**8**	**10**
15	0.00311	0.00322	0.00357	0.00352
20	0.00233	0.00238	0.00239	0.00245
25	0.00205	0.00216	0.00207	0.00196
50	0.00099	0.00096	0.00105	0.00110
100	0.00044	0.00057	0.00033	0.00039

The findings in case of the R chart is similar to that of the S chart i.e. from Table 2.15 we see that for a fixed FAP, the attained false alarm rate (i) decreases as the number of samples m increases, for a fixed sample size *n*, and (ii) stays fairly constant for a fixed *m* but with increasing n. Also, note that for $m = 20$ and $FAP_0 = 0.05$, the $AFAR$ is close to 0.0027 for all n considered.

Example 3

To illustrate the calculations of the control limits for the Phase I S^2, S and R charts we use a dataset from Montgomery (2005), page 223, on the inside diameter measurements for automobile engine piston rings. The data consists of $m = 25$ samples each of size $n = 5$ and are shown in Table 2.16. Also shown in Table 2.16 are the sample mean $\overline{X}_i$, the sample range R_i, the sample standard deviation S_i and the sample variance S_i^2 for each sample. The unit of measurement is millimetre (*mm*) and we omit mentioning this below to avoid repetition.

Table 2.16: Inside diameter measurements (in *mm*) for automobile engine piston rings*

	Observations					Sample statistics			
Sample number (*i*)	X_1	X_2	X_3	X_4	X_5	$\overline{X}_i$	R_i	S_i	S_i^2
1	74.030	74.002	74.019	73.992	74.008	74.010	0.038	0.0148	0.0002182
2	73.995	73.992	74.001	74.011	74.004	74.001	0.019	0.0075	0.0000563
3	73.988	74.024	74.021	74.005	74.002	74.008	0.036	0.0147	0.0002175
4	74.002	73.996	73.993	74.015	74.009	74.003	0.022	0.0091	0.0000825
5	73.992	74.007	74.015	73.989	74.014	74.003	0.026	0.0122	0.0001493
6	74.009	73.994	73.997	73.985	73.993	73.996	0.024	0.0087	0.0000758
7	73.995	74.006	73.994	74.000	74.005	74.000	0.012	0.0055	0.0000305
8	73.985	74.003	73.993	74.015	73.988	73.997	0.030	0.0123	0.0001502
9	74.008	73.995	74.009	74.005	74.004	74.004	0.014	0.0055	0.0000307
10	73.998	74.000	73.990	74.007	73.995	73.998	0.017	0.0063	0.0000395
11	73.994	73.998	73.994	73.995	73.990	73.994	0.008	0.0029	0.0000082
12	74.004	74.000	74.007	74.000	73.996	74.001	0.011	0.0042	0.0000178
13	73.983	74.002	73.998	73.997	74.012	73.998	0.029	0.0105	0.0001093
14	74.006	73.967	73.994	74.000	73.984	73.990	0.039	0.0153	0.0002342
15	74.012	74.014	73.998	73.999	74.007	74.006	0.016	0.0073	0.0000535
16	74.000	73.984	74.005	73.998	73.996	73.997	0.021	0.0078	0.0000608
17	73.994	74.012	73.986	74.005	74.007	74.001	0.026	0.0106	0.0001117
18	74.006	74.010	74.018	74.003	74.000	74.007	0.018	0.0070	0.0000488
19	73.984	74.002	74.003	74.005	73.997	73.998	0.021	0.0085	0.0000717
20	74.000	74.010	74.013	74.020	74.003	74.009	0.020	0.0080	0.0000637
21	73.982	74.001	74.015	74.005	73.996	74.000	0.033	0.0122	0.0001477
22	74.004	73.999	73.990	74.006	74.009	74.002	0.019	0.0074	0.0000553
23	74.010	73.989	73.990	74.009	74.014	74.002	0.025	0.0119	0.0001423
24	74.015	74.008	73.993	74.000	74.010	74.005	0.022	0.0087	0.0000757
25	73.982	73.984	73.995	74.017	74.013	73.998	0.035	0.0162	0.0002617

**Note*: Table 2.16 is a modified version of Table 5.3 in Montgomery (2005).

To illustrate how to construct the Phase I charts for a small number of samples (which is often the case in practice) we use only the first 10 samples. Afterwards all 25 samples are used to illustrate the construction of the charts for a larger number of samples.

Using only the first $m = 10$ samples the unbiased point estimates for the process standard deviation, calculated using (2-13) and (2-14), are found to be

$$\hat{\sigma}_R = \frac{1}{2.326}\left(\frac{1}{10}\sum_{i=1}^{10} R_i\right) = 0.010232 \qquad \text{and} \qquad \hat{\sigma}_S = \frac{1}{0.94}\left(\frac{1}{10}\sum_{i=1}^{10} S_i\right) = 0.010280$$

respectively.

The values of $\hat{\sigma}_R$ and $\hat{\sigma}_S$ are displayed in the first panel (labeled $m = 10$) of Table 2.17 along with the charting constants k_L and k_U for the Phase I S chart and the Phase I R chart which ensures that the false alarm probability of these charts is at most 0.05; these charting constants were obtained from Tables 2.9 and 2.13, respectively. The estimated lower control limit, the estimated centerline and the estimated upper control limit (which are also shown in Table 2.17) are calculated from (2-36) and (2-44), respectively.

The unbiased point estimate of the process variance (based on the first 10 samples only) is calculated using (2-15) i.e.

$$\overline{V} = \frac{1}{10}\sum_{i=1}^{10} S_i^2 = 0.000105$$

and is listed in the first panel (labeled $m = 10$) of Table 2.18. Also shown in Table 2.18 are the values for the charting constants $\hat{a}$ and $\hat{b}$, obtained from Table 2.2, so that the false alarm probability of the Phase I S^2 chart is less than or equal to 0.05. The estimated control limits and estimated centerline were computed according to (2-22).

For all $m = 25$ samples similar calculations were carried out. The unbiased point estimates, the charting constants, the estimated control limits and the estimated centerlines are given in the second panel (labeled $m = 25$) of each of Tables 2.17 and 2.18, respectively.

For large m the 0.001025^{th} and the 0.998975^{th} percentiles of the univariate type I or standard beta distribution with parameters 2 and 48 was used to approximate the charting constants a and b in case of the S^2 chart. These percentiles and the ensuing estimated control limits are also shown in the third panel of Table 2.18. A Phase I S^2 chart designed with these limits has a false alarm probability

approximately equal to 0.05. It is seen that for $m = 25$, the univariate beta approximation is reasonably good compared to the simulation results.

The resultant Shewhart-type Phase I R, S and S^2 charts for $m = 10$ and $m = 25$ are shown in panels (a), (b) and (c) of Figures 2.2 and 2.3, respectively. It appears that the process standard deviation is in control and it would be safe to use 0.00999, which is the centerline of the Phase I R chart, as an unbiased estimate of the process standard deviation to calculate the Phase I mean control chart proposed by Champ and Jones (2004) and to check to see if the process mean is in-control.

The Shewhart-type Phase I S^2, assuming independence of the of the charting statistics, is shown in Figure 2.4; it is seen to be almost identical to the Shewhart-type Phase I S^2 for $m = 25$.

Table 2.17: Parameter estimates and chart constants for the *R* chart and the *S* chart

	***m* = 10**		***m* = 25**	
	***S* chart**	***R* chart**	***S* chart**	***R* chart**
Unbiased Point Estimate	0.010280	0.010232	0.010000	0.009991
k_L	2.1656	2.1187	2.3075	2.2614
k_U	3.0004	3.0502	3.4646	3.5671
$L\hat{C}L$	0.002068	0.005069	0.001527	0.003718
$\hat{C}L$	0.009663	0.023800	0.009400	0.023240
$U\hat{C}L$	0.020187	0.050766	0.021219	0.054033

Table 2.18: Parameter estimates and chart constants for the S^2 chart

	***m* = 10**	***m* = 25**	**Univariate beta distribution assuming independence ($m \geq 25$)**
Unbiased Point Estimate	0.000105	0.000101	
$\hat{a}$	0.0039	0.0009	$a = 0.001025^{th}$ *percentile* = 0.0009
$\hat{b}$	0.3599	0.1734	$b = 0.998975^{th}$ *percentile* = 0.1729
$L\hat{C}L$	0.000004	0.000002	0.000002
$\hat{C}L$	0.000105	0.000101	0.000101
$U\hat{C}L$	0.000378	0.000436	0.000434

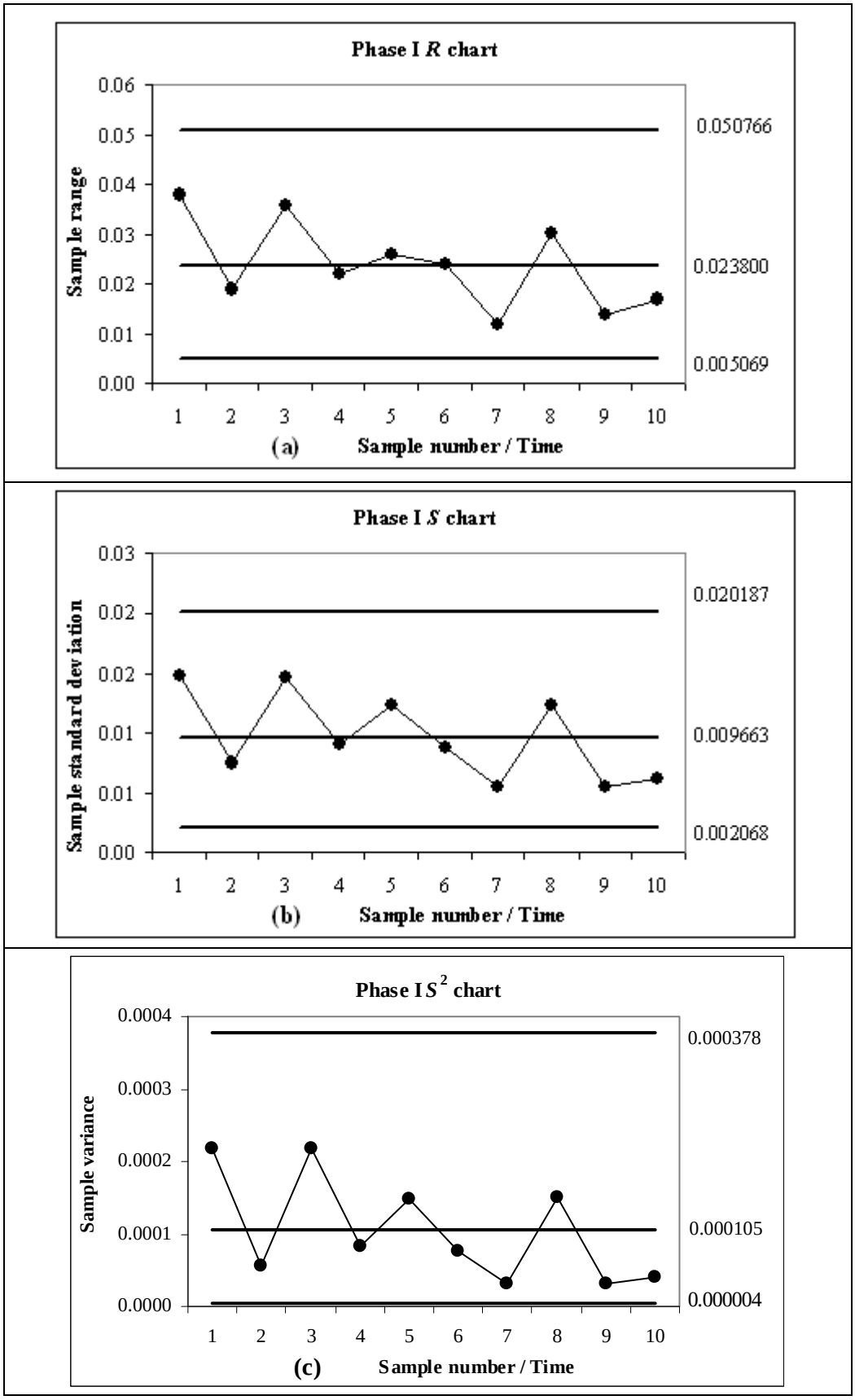


Figure 2.2: Shewhart-type Phase I R, S and S^2 charts for $m = 10$

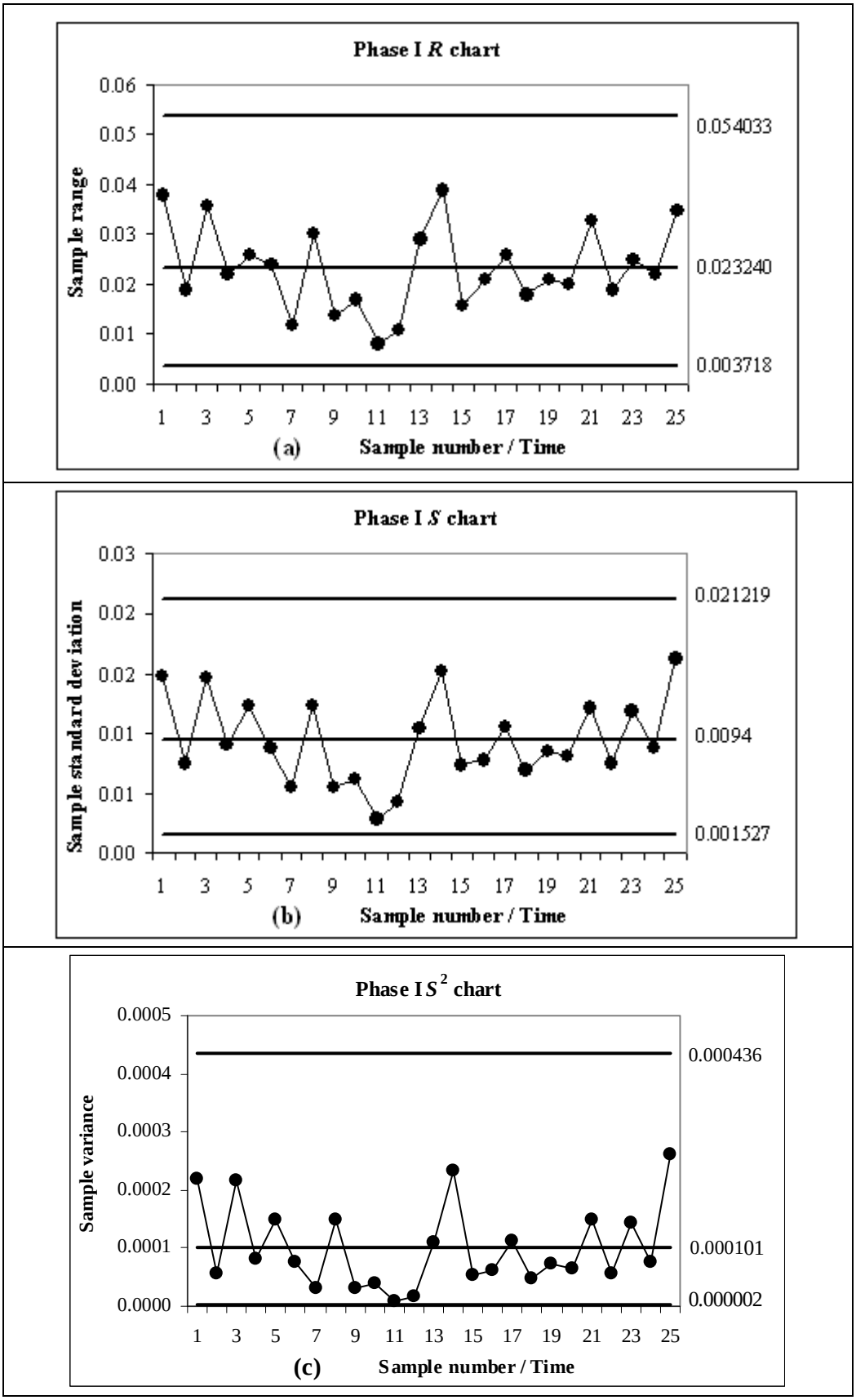


Figure 2.3: Shewhart-type Phase I R, S and S^2 charts for $m = 25$

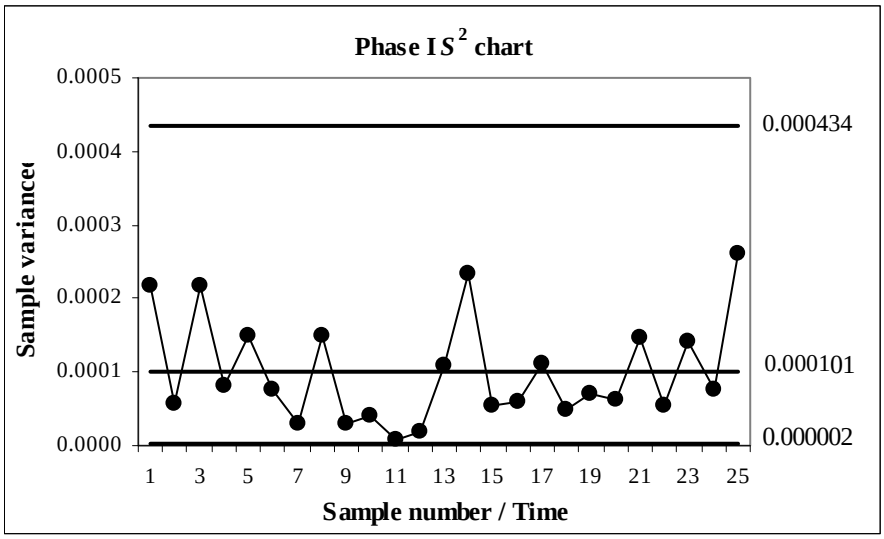


Figure 2.4: Shewhart-type Phase I S^2 assuming independence of the charting statistics

2.3 Literature review: Univariate parametric Shewhart-type Phase I variables charts for location and spread

Phase I control charts form an integral part of SPC, but only a few authors make a clear distinction between Phases I and II, and no more than one of the popular SPC textbooks (e.g. Montgomery, (2005) p. 199 and p. 204) briefly talks about this important topic.

By giving an overview of the literature on univariate parametric Shewhart-type Phase I variables control charts for the mean, standard deviation and variance this gap would be filled. The overview would be particularly helpful to researchers, instructors and practitioners as they would get to know what the issues related to Phase I are, what the present state of the art is and what challenges and future research still remain.

2.3.1 Phase I charts for the normal distribution

Introduction

First we review Shewhart-type Phase I control charts for the mean, standard deviation and variance when the underlying distribution is normally distributed.

Assumptions

Let $X_{ij} \sim iidN(\mu, \sigma^2)$ represent the Phase I data where X_{ij} for $i = 1,2,...,m$ and $j = 1,2,...,n$ denotes the j^{th} observation from the i^{th} subgroup, μ denotes the unknown mean and σ^2 denotes the unknown variance.

(a) King (1954): "Probability Limits for the Average Chart When Process Standards Are Unspecified"

One of the first authors to consider the Phase I problem was King (1954) who studied the Phase I $\overline{X}$ chart for Case U.

Using the overall average

$$\overline{\overline{X}} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij} = \frac{1}{m}\sum_{i=1}^{m}\overline{X}_i$$

and the average of the subgroup ranges

$$\overline{R} = \frac{1}{m}\sum_{i=1}^{m} R_i ,$$

(where $\overline{X}_i$ is the i^{th} sample mean and R_i is the i^{th} sample range) to estimate the unknown mean μ and the unknown standard deviation σ, respectively, he suggests replacing the traditional estimated 3-sigma control limits

$$U\hat{C}L / L\hat{C}L = \overline{\overline{X}} \pm A_2\overline{R}$$

where $A_2 = \dfrac{3}{d_2\sqrt{n}}$ is a function of the sample size n only, with the limits

$$U\hat{C}L / L\hat{C}L = \overline{\overline{X}} \pm C\overline{R}$$

where the charting constant $C = \dfrac{k_m}{d_2\sqrt{n}}$ is a function of the number of rational subgroups m and the sample size n.

Essentially, King proposes to replace the number "3" in the expression for A_2 by a factor k_m which depends on and is a function of m. So to calculate the limits proposed by King one has to find k_m and then substitute it in C, which is then used to calculate the limits.

King graphically provides approximate values of C for $m = 3(1)25$ and $n = 2,3,4,5,10$ so that the false alarm probability of the Phase I $\overline{X}$ chart is approximately 0.05.

The values of k_3 and k_4 are (apparently) obtained from theoretical considerations, whereas the values of k_m for $m \geq 5$ are obtained using the fact that the false alarm probability for the Phase I $\overline{X}$ chart can be written as

$$FAP_{\bar{X}} = 1 - \Pr(|\overline{X}_1 - \overline{\overline{X}}| \le k_m \sigma/\sqrt{n}, \ldots, |\overline{X}_m - \overline{\overline{X}}| \le k_m \sigma/\sqrt{n} \mid IC)$$

$$= 1 - \Pr(\bigcap_{i=1}^{m} \{|\overline{X}_i - \overline{\overline{X}}| \le k_m \sigma/\sqrt{n}\} \mid IC)$$

$$= 1 - \Pr\left(\frac{\max |\overline{X}_i - \overline{\overline{X}}|}{\sigma/\sqrt{n}} \le k_m \text{ for } i = 1,2,\ldots m \mid IC \right)$$

where σ denotes the unknown (but constant) process standard deviation.

King constructs the 95th percentile of the sampling distribution of $\frac{\max |\overline{X}_i - \overline{\overline{X}}|}{\sigma/\sqrt{n}}$ for $i = 1,2,\ldots,m$ (for values of $m \ge 5$) to find k_m and then approximates the multiplier C by taking $C = \frac{k_m}{d_2\sqrt{n}}$.

King observes that the proposed charting constant C approaches A_2 rather rapidly as m increases, but he also notes that his approximation is based on ignoring the sampling fluctuations of $\overline{R}$ and that some of his C values are obtained from simulations.

Remark 11

Despite the shortcomings in the approach used by King, his idea to design a Phase I control chart using a nominal (specified) false alarm probability turned out to be the correct approach.

(b) Hillier (1969): " $\overline{X}$ - and *R*-Chart Control Limits Based on A Small Number of Subgroups"

Hillier (1969) proposes a method for finding the control limits for the Phase I $\overline{X}$ chart and the Phase I R chart that can be reliably used regardless of how few Phase I subgroups are available.

Hillier acknowledges the fact that the Phase I signaling events are dependent and suggests that the conventional factors A_2, D_3 and D_4 usually given in SPC textbooks and used in setting-up the control limits at

$$U\hat{C}L_{\overline{X}} / L\hat{C}L_{\overline{X}} = \overline{\overline{X}} \pm A_2\overline{R}$$

for the $\overline{X}$ chart and

$$L\hat{C}L_R = D_3\overline{R} \qquad \text{and} \qquad U\hat{C}L_R = D_4\overline{R}$$

for the R chart, be replaced by more appropriate charting constants A_2^{**}, D_3^{**} and D_4^{**} so that the false alarm rates of the $\overline{X}$ and the R charts are controlled at

$$FAR_{\overline{X}} = \alpha_2 \qquad \text{and} \qquad FAR_R = \alpha_3 + \alpha_4,$$

respectively; where α_2 is the $FAR_{\overline{X}}$ for the $\overline{X}$ chart and α_3 and α_4 are the probabilities that a Phase I sample range R_i for each $i = 1,2,...,m$ plots below or above the estimated lower and upper control limits of the R chart, respectively.

The factor A_2^{**} is derived by studying the probability expression of the false alarm rate for the Phase I $\overline{X}$ chart, which is given by

$$\begin{aligned} FAR_{\overline{X}} &= 1 - \Pr(L\hat{C}L_{\overline{X}} < \overline{X}_i < U\hat{C}L_{\overline{X}} \mid IC) \\ &= 1 - \Pr(\overline{\overline{X}} - A_2^{**}\overline{R} < \overline{X}_i < \overline{\overline{X}} + A_2^{**}\overline{R} \mid IC) \\ &= 1 - \Pr(-A_2^{**} < \frac{\overline{X}_i - \overline{\overline{X}}}{\overline{R}} < A_2^{**} \mid IC) \quad \text{for each } i = 1,2,...,m. \end{aligned}$$

If the false alarm rate of the $\overline{X}$ chart should be controlled at α_2, it implies that

$$\alpha_2 = 1 - \Pr(-A_2^{**} < \frac{\overline{X}_i - \overline{\overline{X}}}{\overline{R}} < A_2^{**} \mid IC) \quad \text{for} \quad i = 1,2,...,m$$

and that one has to solve for A_2^{**}; for this one requires the distribution of $\frac{\overline{X}_i - \overline{\overline{X}}}{\overline{R}}$.

To this end, Hillier notes that, under the assumption of normality,

$$(X_i - \overline{\overline{X}}) \sim N(0, \frac{m-1}{mn}\sigma^2) \quad \text{for} \quad i = 1,2,...,m$$

(which is an exact result) and approximates the distribution of $\frac{v\overline{R}^2}{c^2\sigma^2}$ by that of a χ_v^2 distribution where c (which is a constant) and v are functions of m and n (see Patnaik, (1950)). Hillier then uses the fact that the numerator $(X_i - \overline{\overline{X}})$ for $i = 1,2,...,m$ and the denominator $\overline{R}$ are independent, to write the probability expression of the false alarm rate of the Phase I $\overline{X}$ chart as

$$\alpha_2 = 1 - \Pr(-c\sqrt{\frac{mn}{m-1}}A_2^{**} < \frac{(\overline{X}_i - \overline{\overline{X}})/\sqrt{\frac{m-1}{mn}\sigma^2}}{\sqrt{\frac{v\overline{R}^2/c^2\sigma^2}{v}}} < c\sqrt{\frac{mn}{m-1}}A_2^{**} \mid IC)$$

$$\approx 1 - \Pr(-c\sqrt{\frac{mn}{m-1}}A_2^{**} < T_v < c\sqrt{\frac{mn}{m-1}}A_2^{**} \mid IC)$$

where

$$\frac{(\overline{X}_i - \overline{\overline{X}})/\sqrt{\frac{m-1}{mn}\sigma^2}}{\sqrt{\frac{v\overline{R}^2/c^2\sigma^2}{v}}} = T_v$$

has approximately a Student's *t*-distribution with v degrees of freedom when the process is in-control.

Approximate values for A_2^{**} can therefore be obtained by setting

$$c\sqrt{\frac{mn}{m-1}}A_2^{**} = t_{\frac{\alpha_2}{2},v}$$

and solving for A_2^{**} i.e.

$$A_2^{**} = \frac{1}{c}\sqrt{\frac{m-1}{mn}}t_{\frac{\alpha_2}{2},v}$$

where $t_{\frac{\alpha_2}{2},v}$ denotes the value such that, if the random variable T_v has a *t*-distribution with v degrees of freedom, then $\Pr(-t_{\frac{\alpha_2}{2},v} < T_v < t_{\frac{\alpha_2}{2},v} \mid IC) = 1-\alpha_2$ where α_2 is the desired false alarm rate.

The constants D_3^{**} and D_4^{**} for the Phase I R chart are obtained in a similar manner by studying an expression for the false alarm rate of the *R* chart. The details can be found in Hillier, (1969).

Tables with values of A_2^{**}, D_3^{**} and D_4^{**} are provided by Hillier (1969) for subgroups of size $n = 5$ when $m = 2(1)10,15,20,25,50,100,\infty$ and α_2 and/or α_3 and/or α_4 are equal to 0.001, 0.0027, 0.01, 0.025 and 0.05, respectively.

The implementation of Hillier's procedure is straightforward. First one chooses the desired values of α_2, α_3 and α_4 and calculate the recommended control limits using the appropriate values of A_2^{**}, D_3^{**} and D_4^{**}. Then, for each Phase I subgroup, one checks if both its average $\bar{X}_i$ and its range R_i fall inside the control limits for the $\bar{X}$ chart and between those of the *R* chart. If they do not, the particular subgroup(s) are discarded (only if an assignable cause was found) and the overall mean $\bar{\bar{X}}$, the mean range $\bar{R}$ and the control limits are re-calculated using the remaining subgroups where the factors A_2^{**}, D_3^{**} and D_4^{**} are based on the updated value of m i.e. the number of Phase I subgroups still being used to calculate $\bar{\bar{X}}$ and $\bar{R}$. This iterative procedure is continued until all the remaining subgroup means and subgroup ranges fall between the control limits of both the charts. Once this state is reached one may calculate the appropriate control limits for prospective monitoring of the process in Phase II.

Note that, if at any stage during Phase I control charting it happens that some of the Phase I charting statistics plot outside the estimated control limits but no assignable cause(s) can be found that justify their removal, the process may be considered in-control and the observations from these samples are then included in the reference data.

(c) Yang and Hillier (1970): "Mean and Variance Control Chart Limits Based on a Small Number of Subgroups"

Yang and Hillier (1970) extend and improve the method proposed by Hillier (1969) to find probability limits for the Phase I $\bar{X}$ chart using the average (pooled) sample variance $\bar{V} = \frac{1}{m}\sum_{i=1}^{m} S_i^2$ (instead of the mean range $\bar{R}$) where S_i^2 for $i = 1,2,...,m$ is the i^{th} subgroup variance; they also develop Phase I limits for a variance chart and a standard deviation chart based on $\bar{V}$.

Phase I $\bar{X}$ chart

In particular, Yang and Hillier (1970) recommend that instead of calculating the estimated control limits of the Phase I $\bar{X}$ chart in the usual way i.e.

$$U\hat{C}L / L\hat{C}L = \bar{\bar{X}} \pm A_2 \bar{R}$$

one should replace $\bar{R}$ with $\sqrt{\bar{V}}$ and substitute A_4^{**} for A_2 and calculate the control limits as

$$U\hat{C}L / L\hat{C}L = \bar{\bar{X}} \pm A_4^{**} \sqrt{\bar{V}} \ .$$

The charting constant A_4^{**} comes from studying the false alarm rate of the Phase I $\bar{X}$ chart, which is given by

$$\begin{aligned} FAR_{\bar{X}} &= 1 - \Pr(L\hat{C}L_{\bar{X}} < \bar{X}_i < U\hat{C}L_{\bar{X}} \mid IC) \\ &= 1 - \Pr(\bar{\bar{X}} - A_4^{**}\sqrt{\bar{V}} < \bar{X}_i < \bar{\bar{X}} + A_4^{**}\sqrt{\bar{V}} \mid IC) \\ &= 1 - \Pr(-A_4^{**} < \frac{\bar{X}_i - \bar{\bar{X}}}{\sqrt{\bar{V}}} < A_4^{**} \mid IC) \quad \text{for} \quad i = 1,2,...,m. \end{aligned}$$

To control the $FAR_{\bar{X}}$ at a level of α implies that one has to find that value of A_4^{**} such that

$$\alpha = 1 - \Pr(-A_4^{**} < \frac{\bar{X}_i - \bar{\bar{X}}}{\sqrt{\bar{\bar{V}}}} < A_4^{**} \mid IC) \quad \text{for} \quad i = 1,2,...,m;$$

this requires one to find the distribution of $\frac{\bar{X}_i - \bar{\bar{X}}}{\sqrt{\bar{\bar{V}}}}$.

In this regard, the authors note that, under the assumption of normality, the numerator

$$(X_i - \bar{\bar{X}}) \sim N(0, \frac{m-1}{mn}\sigma^2) \quad \text{for} \quad i = 1,2,...,m$$

and that the random variable

$$\frac{m(n-1)\bar{V}}{\sigma^2} \sim \chi^2_{m(n-1)};$$

they then write the false alarm rate of the Phase I $\bar{X}$ chart as

$$FAR_{\bar{X}} = 1 - \Pr(-\sqrt{\frac{mn}{m-1}}A_4^{**} < \frac{(\bar{X}_i - \bar{\bar{X}})/\sqrt{\frac{m-1}{mn}\sigma^2}}{\sqrt{\frac{m(n-1)\bar{V}/\sigma^2}{m(n-1)}}} < \sqrt{\frac{mn}{m-1}}A_4^{**} \mid IC)$$

$$= 1 - \Pr(-\sqrt{\frac{mn}{m-1}}A_4^{**} < T_{m(n-1)} < \sqrt{\frac{mn}{m-1}}A_4^{**} \mid IC) \quad \text{for} \quad i = 1,2,...,m$$

where $T_{m(n-1)}$ is a random variable which has a Student's *t*-distribution with $m(n-1)$ degrees of freedom; this is an exact result.

The charting constant can thus be calculated by solving for A_4^{**} from

$$\alpha = 1 - \Pr(-\sqrt{\frac{mn}{m-1}}A_4^{**} < T_{m(n-1)} < \sqrt{\frac{mn}{m-1}}A_4^{**} \mid IC) \quad \text{for} \quad i = 1,2,...,m.$$

This is done by setting

$$\sqrt{\frac{mn}{m-1}}A_4^{**} = t_{\frac{\alpha}{2},m(n-1)}$$

and solving for A_4^{**} i.e.

$$A_4^{**} = \sqrt{\frac{m-1}{mn}}t_{\frac{\alpha}{2},m(n-1)}$$

where $t_{\frac{\alpha}{2},m(n-1)}$ denotes the value such that

$$\Pr(-t_{\frac{\alpha}{2},m(n-1)} < T_v < t_{\frac{\alpha}{2},m(n-1)} \mid IC) = 1-\alpha$$

and $FAR_{\overline{X}} = \alpha$ is the desired false alarm rate.

The authors provide a table with values of A_4^{**} for subgroups of size $n=5$ when $m = 2(1)10,15,20,25,50,100,\infty$ and α equal to 0.001, 0.002, 0.01 and 0.05, respectively.

Phase I variance chart

For the variance chart based on $\overline{V} = \frac{1}{m}\sum_{i=1}^{m} S_i^2$ Yang and Hillier (1970) propose that one uses S_i^2 for $i = 1,2,...,m$ as charting statistics and that the estimated control limits be calculated as

$$\hat{LCL}_{S^2} = B_7^{**}\overline{V} \qquad \text{and} \qquad \hat{UCL}_{S^2} = B_8^{**}\overline{V}$$

where B_7^{**} and B_8^{**} are the charting constants.

The charting constants B_7^{**} and B_8^{**} are found using the fact that the random variable

$$\frac{S_i^2/\sigma^2}{[m(n-1)\overline{V} - (n-1)S_i^2]/(m-1)(n-1)\sigma^2} = \frac{(m-1)S_i^2}{m\overline{V} - S_i^2}$$

has an F-distribution with degrees of freedom equal to $(n-1)$ and $(m-1)(n-1)$ i.e.

$$\frac{(m-1)S_i^2}{m\bar{V}-S_i^2} \overset{d}{=} F_{n-1,(m-1)(n-1)} \quad \text{for } i=1,2,...,m\,.$$

Solving algebraically for S_i^2 in terms of the random variables $\bar{V}$ and $F_{n-1,(m-1)(n-1)}$ one finds that

$$S_i^2 \overset{d}{=} \frac{mF_{n-1,(m-1)(n-1)}}{m-1+F_{n-1,(m-1)(n-1)}}\bar{V} \quad \text{for each } i=1,2,...,m$$

which is a strictly increasing and monotone function of $F_{n-1,(m-1)(n-1)}$ for $m>1$.

The proposed control limits of Yang and Hillier for their Phase I variance chart to retrospectively test the initial subgroups (using $\bar{V}$) are thus obtained by setting

$$B_7^{**} = \frac{mF_{1-\alpha_L,n-1,(m-1)(n-1)}}{m-1+F_{1-\alpha_L,n-1,(m-1)(n-1)}} \qquad \text{and} \qquad B_8^{**} = \frac{mF_{1-\alpha_U,n-1,(m-1)(n-1)}}{m-1+F_{1-\alpha_U,n-1,(m-1)(n-1)}}$$

where $F_{\beta,n-1,(m-1)(n-1)}$ is the fractile such that, if the random variable F_{n_1,n_2} has an F-distribution with n_1 and n_2 degrees of freedom, then $\Pr(F_{n_1,n_2} > F_{\beta,n-1,(m-1)(n-1)}) = \beta$ and α_L (α_U) is the desired probability that a Phase I sample variance S_i^2 for $i=1,2,...,m$ plots below (above) the estimated control limit.

Phase I standard deviation chart

Because S_i^2 is expressed in terms of a strictly increasing and monotone function of $F_{n-1,(m-1)(n-1)}$ for $m>1$, the authors proposed that the adjusted control limits of their Phase I standard deviation chart be calculated as

$$\hat{LCL} = \sqrt{B_7^{**}\bar{V}} \qquad \text{and} \qquad \hat{UCL} = \sqrt{B_8^{**}\bar{V}}\,.$$

One would then compare each sample standard deviation S_i for $i=1,2,...,m$ with the estimated limits $\sqrt{B_7^{**}\bar{V}}$ and $\sqrt{B_8^{**}\bar{V}}$, respectively.

Tables with values of B_7^{**} and B_8^{**} are provided for subgroups of size $n=5$ when $m=1(1)10,15,20,25,50,100,\infty$ and α_L and/or α_U are equal to 0.001, 0.005, 0.025, respectively.

Remark 12

It is important to note that, unlike King (1954), neither Hillier (1969) nor Yang and Hillier (1970) consider the correlation (i.e. dependency) between the signaling events that result from the use of estimated process parameters, and they control the false alarm rate of each subgroup and not the false alarm probability (like King, (1954)).

The control limits by Hillier (1969) and Yang and Hillier (1970) are referred to as the "standard limits". Yang and Hillier (1970) also suggest a second method for constructing Phase I charts referred to as "individual limits". In the latter approach each subgroup is tested one at a time while treating the other $m-1$ subgroups as in-control. The control limits for the plotted charting statistic at time i are therefore functions of the other $m-1$ samples and require recalculating m different sets of control limits.

(d) Chou and Champ (1995): "A comparison of two Phase I control charts", and Champ and Chou (2003): "Comparison of Standard and Individual Limits Phase I Shewhart $\bar{X}$, *R*, and *S* Charts"

The standard limits and the individual limits are studied in detail by Chou and Champ (1995) and Champ and Chou (2003). These authors discuss, evaluate and compare the standard limits and the individual limits Shewhart-type Phase I $\bar{X}$ charts assuming normality when the process parameters are unknown.

Champ and Chou (2003) also show that the individual limits and the standard limits Shewhart-type Phase I R charts can be designed to be equivalent; a result that they show also holds for the individual limits and the standard limits Shewhart-type Phase I S charts.

Standard limits Phase I $\bar{X}$ chart

In particular, Champ and Chou (2003) define the estimated control limits of the standard limits Shewhart-type Phase I $\bar{X}$ chart as

$$U\hat{C}L_{\bar{X}} / L\hat{C}L_{\bar{X}} = \bar{\bar{X}} \pm k_{\bar{X}} \frac{\sqrt{\bar{V}}}{c_{4,m}\sqrt{n}}$$

where

$$\bar{\bar{X}} = \frac{1}{m}\sum_{i=1}^{m} \bar{X}_i \qquad \text{and} \qquad \bar{V} = \frac{1}{m}\sum_{i=1}^{m} S_i^2$$

are the overall mean and the pooled variance (which includes all m the Phase I samples),

$$c_{4,m} = \frac{\sqrt{2}\Gamma((m(n-1)+1)/2)}{\sqrt{m(n-1)}\Gamma(m(n-1)/2)}$$

is the unbiasing constant and the charting constant

$$k_{\bar{X}} = \sqrt{(m-1)/mc}_{4,m} t_{m(n-1),0,\alpha/(2m)}$$

was chosen using Boole's inequality such that, if the process is in-control, the probability that at least one sample mean $\bar{X}_i$ for $i = 1,2,...,m$ is outside the control limits is at most α, $0 < \alpha < 1$ and where $t_{m(n-1),0,\alpha/(2m)}$ is the $[1-\alpha/(2m)]100^{\text{th}}$ percentage point of the univariate central Student's t-distribution with $m(n-1)$ degrees of freedom.

Individual limits Phase I $\overline{X}$ chart

The estimated control limits of the standard limits Shewhart-type Phase I $\overline{X}$ chart are re-expressed as

$$U\hat{C}L_{\overline{X}} / L\hat{C}L_{\overline{X}} = \overline{\overline{X}} \pm A_4^{***}\sqrt{\overline{V}}$$

where

$$A_4^{***} = \sqrt{(m-1)/(mn)}\,t_{m(n-1),0,\alpha/(2m)},$$

and then, Champ and Chou (2003) define the estimated control limits of the individual limits Shewhart-type Phase I $\overline{X}$ chart as

$$U\hat{C}L_{\overline{X},[i]} / L\hat{C}L_{\overline{X},[i]} = \overline{\overline{X}}_{[i]} \pm A_{4,[i]}^{***}\sqrt{\overline{V}_{[i]}}$$

where

$$\overline{\overline{X}}_{[i]} = \frac{1}{m-1}\sum_{k=1,k\neq i}^{m}\overline{X}_k = \frac{1}{m-1}\left(\sum_{k=1}^{m}\overline{X}_k - \overline{X}_i\right)$$

and

$$\overline{V}_{[i]} = \frac{1}{m-1}\sum_{k=1,k\neq i}^{m}S_k^2 = \frac{1}{m-1}\left(\sum_{k=1}^{m}S_k^2 - S_i^2\right)$$

are the overall mean and the pooled variance when only the i^{th} sample is removed, respectively and

$$A_{4,[i]}^{***} = \sqrt{m/((m-1)n)}\,t_{(m-1)(n-1),0,\alpha/(2m)}.$$

Tables with values of A_4^{***} and $A_{4,[i]}^{***}$ when $m = 5(10)25$ and $n = 2(1)10$ for $\alpha = 0.05$ are provided.

Performance comparison

In the performance comparison of the standard limits versus the individual limits of the Shewhart-type Phase I $\overline{X}$ chart, the authors evaluate the effectiveness of the two charts in identifying an out-of-control process. A simple out-of-control scenario is chosen where one of the samples is assumed to be out-of-control and the other $m-1$ samples are in-control. Without any loss of generality they take the first sample to be out-of-control and assume that

$$\overline{X}_1 \sim N(\mu_1, \frac{\sigma^2}{n}) \qquad \text{and} \qquad \overline{X}_i \sim N(\mu, \frac{\sigma^2}{n}) \text{ for } i = 2,3,...,m$$

where $\mu_1 = \mu + \delta\frac{\sigma}{\sqrt{n}}$ so that the first sample is out-of-control and reflected only in its mean.

The probabilities that the first and the second (without loss of generality) sample means fall outside the standard limits are shown to be

$$p^{***}_{\bar{X},1} = 1 - \tilde{T}_{m(n-1),\theta_1}\left(\sqrt{\frac{mn}{m-1}}A^{***}_4\right) + \tilde{T}_{m(n-1),\theta_1}\left(-\sqrt{\frac{mn}{m-1}}A^{***}_4\right)$$

and

$$p^{***}_{\bar{X},2} = 1 - \tilde{T}_{m(n-1),\theta_2}\left(\sqrt{\frac{mn}{m-1}}A^{***}_4\right) + \tilde{T}_{m(n-1),\theta_2}\left(-\sqrt{\frac{mn}{m-1}}A^{***}_4\right)$$

respectively, whereas the probabilities that the first and the second sample means fall outside the individual limits are shown to be

$$p^{***}_{\bar{X},[1]} = 1 - \tilde{T}_{(m-1)(n-1),\theta_{[1]}}\left(\sqrt{\frac{mn}{m-1}}A^{***}_{4,[1]}\right) + \tilde{T}_{(m-1)(n-1),\theta_{[1]}}\left(-\sqrt{\frac{mn}{m-1}}A^{***}_{4,[1]}\right)$$

and

$$p^{***}_{\bar{X},[2]} = 1 - \tilde{T}_{(m-1)(n-1),\theta_{[2]}}\left(\sqrt{\frac{mn}{m-1}}A^{***}_{4,[1]}\right) + \tilde{T}_{(m-1)(n-1),\theta_{[2]}}\left(-\sqrt{\frac{mn}{m-1}}A^{***}_{4,[1]}\right),$$

respectively, where

$$\theta_{[1]} = \theta_1 = \sqrt{(m-1)/m}\delta,$$

$$\theta_{[2]} = -\theta_2 = -\sqrt{1/[m(m-1)]}\delta$$

and $\tilde{T}_{v,\theta}$ denotes the c.d.f of a univariate non-central Student's t-distribution with v degrees of freedom and non-centrality parameter θ.

Tables with values of $p^{***}_{\bar{X},1}$, $p^{***}_{\bar{X},2}$, $p^{***}_{\bar{X},[1]}$ and $p^{***}_{\bar{X},[2]}$ are provided for $\delta = 0.0(0.1)2.0$, $m = 5(5)25$ and $\alpha = 0.05$ and used in the performance comparison.

Champ and Chou (2003) notes that, in general, $p^{***}_{\bar{X},1} > p^{***}_{\bar{X},[1]}$ for all size shifts δ in the mean, number of samples m, and the sample size n. They also point out that although $p^{***}_{\bar{X},2} \approx p^{***}_{\bar{X},[2]}$, in general, $p^{***}_{\bar{X},2}$ is slightly larger than $p^{***}_{\bar{X},[2]}$.

Based on these results, they conclude that the standard limits $\bar{X}$ chart slightly outperforms the corresponding individual limits $\bar{X}$ chart and highlight the fact that there is more work involved in setting up the individual limits chart.

Standard limits and Individual limits Phase I *R* and *S* charts

Lastly, Champ and Chou (2003) consider the Shewhart-type Phase I R and S charts. Specifically, they show that *if* the standard limits and the individual limits Phase I R charts are defined as

$$U\hat{C}L_{R,i} = k_{R,U,i}\frac{\overline{R}}{d_2} \qquad L\hat{C}L_{R,i} = k_{R,L,i}\frac{\overline{R}}{d_2}$$

and

$$U\hat{C}L_{R,[i]} = k_{R,U,[i]}\frac{\overline{R}_{[i]}}{d_2} \qquad L\hat{C}L_{R,[i]} = k_{R,L,[i]}\frac{\overline{R}_{[i]}}{d_2}$$

respectively, and one lets

$$k_{R,L,[i]} = \frac{m-1}{mk_{R,L,i}^{-1}-1} \quad \text{and} \quad k_{R,U,[i]} = \frac{m-1}{mk_{R,U,i}^{-1}-1},$$

then the standard limits and the individual limits Phase I R charts are equivalent; where $\overline{R} = \frac{1}{m}\sum_{j=1}^{m} R_j$ is the mean range of all m the Phase I samples and $\overline{R}_{[i]} = \frac{1}{m-1}\sum_{j=1, j\neq i}^{m} R_j = \frac{1}{m-1}(\sum_{j=1}^{m} R_j - R_i)$ is the mean range excluding only the i^{th} sample.

Similarly, the authors show that if the standard limits and the individual limits Phase I S charts are defined as

$$U\hat{C}L_{S,i} = k_{S,U,i}\frac{\overline{S}}{c_4} \qquad L\hat{C}L_{S,i} = k_{S,L,i}\frac{\overline{S}}{c_4}$$

and

$$U\hat{C}L_{S,[i]} = k_{S,U,[i]}\frac{\overline{S}_{[i]}}{c_4} \qquad L\hat{C}L_{S,[i]} = k_{S,L,[i]}\frac{\overline{S}_{[i]}}{c_4},$$

respectively, or, if the standard limits and the individual limits Phase I S charts are defined as

$$U\hat{C}L_{S,i} = k_{S,U,i}\frac{\sqrt{\overline{V}}}{c_{4,m}} \qquad L\hat{C}L_{S,i} = k_{S,L,i}\frac{\sqrt{\overline{V}}}{c_{4,m}}$$

and

$$U\hat{C}L_{S,[i]} = k_{S,U,[i]}\frac{\sqrt{\overline{V}_{[i]}}}{c_{4,m}} \qquad L\hat{C}L_{S,[i]} = k_{S,L,[i]}\frac{\sqrt{\overline{V}_{[i]}}}{c_{4,m}},$$

respectively, the standard limits and the individual limits Phase I S charts are equivalent if one takes

$$k_{S,L,[i]} = \frac{m-1}{mk_{S,L,i}^{-1} - 1} \qquad \text{and} \qquad k_{S,U,[i]} = \frac{m-1}{mk_{S,U,i}^{-1} - 1};$$

where $\bar{S} = \frac{1}{m}\sum_{i=1}^{m} S_i$ is the average standard deviation of all the Phase I samples and

$$\bar{S}_{[i]} = \frac{1}{m-1}\sum_{k=1,k\neq i}^{m} S_j = \frac{1}{m-1}\left(\sum_{k=1}^{m} S_k - S_i\right)$$

is the average standard deviation excluding only the i^{th} sample.

(e) Champ and Jones (2004): "Designing Phase I $\bar{X}$ Charts with Small Sample Sizes"

One of the problems with the approach by Chou and Champ (1995) and Champ and Chou (2003) in developing their Phase I $\bar{X}$ chart is that they use Boole's inequality and do not explicitly take account of the large number of simultaneous comparisons inherent in Phase I. Champ and Jones (2004) recognized this by setting up *FAP*-based probability limits for the Shewhart-type Phase I $\bar{X}$ chart in Case U, when the mean and the standard deviation are both unknown. They use three unbiased estimators $\hat{\sigma}$ of σ in the calculations of the control limits, which are defined as

$$U\hat{C}L / L\hat{C}L = \bar{\bar{X}} \pm k\frac{\hat{\sigma}}{\sqrt{n}}.$$

The estimators $\hat{\sigma}$ are (i) the average sample range $\frac{\bar{R}}{d_2}$, (ii) the average sample standard deviation $\frac{\bar{S}}{c_4}$, and (iii) the square root of the pooled sample variance $\frac{\sqrt{\bar{V}}}{c_{4,m}}$, where $\bar{R} = \frac{1}{m}\sum_{i=1}^{m} R_i$, $\bar{S} = \frac{1}{m}\sum_{i=1}^{m} S_i$ and $\bar{V} = \frac{1}{m}\sum_{i=1}^{m} S_i^2$.

The authors show that the joint distribution of the m standardized subgroups means

$$T_{v,i} = c\sqrt{\frac{m}{m-1}}\left(\frac{\bar{X}_i - \bar{\bar{X}}}{\hat{\sigma}/\sqrt{n}}\right) \quad \text{for} \quad i = 1,2,...,m,$$

follows either an exact or an approximate (depending on the estimator used for σ) equi-correlated central multivariate t-distribution with correlation $-1/(m-1)$, where the degrees of freedom v and the unbiasing constant c (which varies depending on the particular estimator (i), (ii) or (iii) used for σ, see Champ and Jones (2004) for details) are both functions of m and n.

The exact false alarm probability of the Phase I $\bar{X}$ chart is shown to be

$$\begin{aligned} FAP_{\bar{X}} &= 1 - \Pr\left(\bigcap_{i=1}^{m}\left\{\bar{\bar{X}} - k\frac{\hat{\sigma}}{\sqrt{n}} < \bar{X}_i < \bar{\bar{X}} + k\frac{\hat{\sigma}}{\sqrt{n}}\right\} \mid IC\right) \\ &= 1 - \Pr\left(\bigcap_{i=1}^{m}\{-d < T_{v,i} < d\} \mid IC\right) \\ &= 1 - \int_{-d}^{d}\int_{-d}^{d}\cdots\int_{-d}^{d} f_{T_{v,1},T_{v,2},...,T_{v,m}}(t_{v,1}, t_{v,2},..., t_{v,m})\, dt_{v,1}\, dt_{v,2}...dt_{v,m} \end{aligned}$$

where $f_{T_{v,1},T_{v,2},...,T_{v,m}}(t_{v,1},t_{v,2},...,t_{v,m})$ is the joint density of $T_{v,1},T_{v,2},...,T_{v,m}$, $d = kc\sqrt{\frac{m}{m-1}}$ and k equals either k_R or k_S or k_V depending on which unbiased estimator of σ was used.

Using a modified version of a program by Nelson (1982) for the equi-correlated multivariate *t*-distribution, Champ and Jones (2004) provide tables for the charting constants k_R, k_S and k_V for $m = 4(1)10,15$ and $n = 3(1)10$ for a nominal false alarm probability of 0.1, 0.05 and 0.01, respectively. Note that, although Champ and Jones (2004) followed an exact approach, the accuracy of the values of the charting constants k_R, k_S and k_V that they obtained, depend on the accuracy of the program by Nelson (1982).

Champ and Jones (2004) use simulations to compare the performance of their control limits of the $\bar{X}$ chart when $m \geq 20$ with: (i) approximate limits using univariate Student's t critical values, and (ii) approximate limits assuming that each $T_{v,i}$ approximately follows a standard normal distribution. Although both of these approximate procedures are easy to use, the latter is not recommended unless the number of subgroups m is at least 30.

(f) Neduraman and Pignatiello (2005): "On Constructing Retrospective $\overline{X}$ Control Chart Limits"

Neduraman and Pignatiello (2005) adopted the analysis of means (ANOM) approach (see e.g. the book by Nelson, Wludyka and Copeland, (2005)) to construct a Shewhart-type Phase I $\overline{X}$ chart for the mean while maintaining the false alarm probability at a desired level. They also compare the performance of their ANOM based control limits with that of Bonferroni-adjusted control limits through computer simulation experiments and make recommendations as to when each of the approaches may be used.

Their chart is based on the result that if

$$T_i = \frac{\overline{X}_i - \overline{\overline{X}}}{\sqrt{(m-1)\overline{V}/mn}} \quad \text{for} \quad i = 1,2,...,m,$$

then the standardized charting statistics $(T_1, T_2, ..., T_m)$ has an equi-correlated multivariate t-distribution with common correlation $-1/(m-1)$. Using this result they find critical values, denoted by $h_{FAP_0,m,v}$, such that

$$\begin{aligned} FAP_0 &= 1 - \Pr(\bigcap_{i=1}^{m} \{-h_{FAP_0,m,v} < T_i < h_{FAP_0,m,v}\} \mid IC) \\ &= 1 - \Pr(\max_{1 \le i \le m} |T_i| < h_{FAP_0,m,v} \mid IC) \end{aligned}$$

where $v = m(n-1)$ represents the degrees of freedom of the variance estimator $\overline{V}$ and FAP_0 is the nominal false alarm probability. The Phase I ANOM based control limits are given by

$$U\hat{C}L / L\hat{C}L = \overline{\overline{X}} \pm h_{FAP_0,m,v}\sqrt{(m-1)\overline{V}/mn}$$

where the plotting statistics are the usual sample means $\overline{X}_i$ for $i = 1,2,...,m$.

The authors provide tables for the critical values $h_{FAP_0,m,v}$ for $m = 5(5)30(10)50,75,100$, $n = 5,7,10$ and $FAP_0 = 0.0027, 0.01, 0.05$.

Finally, Neduraman and Pignatiello (2005) compare the performance of their ANOM based control limits with those obtained by a Bonferroni-type adjustment via computer simulation.

The Bonferroni-adjusted control limits are obtained by setting the false alarm rate for each subgroup equal to FAP_0 / m so that the estimated Bonferroni-adjusted control limits are given by

$$U\hat{C}L / L\hat{C}L = \overline{\overline{X}} \pm z_{FAP_0/(2m)}(\overline{S} / c_4) / \sqrt{n}$$

where $\overline{S}$ denotes the average of the m sample standard deviations, c_4 is the unbiasing constant and $z_{FAP_0/2m}$ is the $(1-FAP_0/2m)100^{\text{th}}$ percentage point of the standard normal distribution.

Neduraman and Pignatiello (2005) found that: (i) for small subgroup sizes the ANOM based control limits perform better than the Bonferroni-adjusted limits in that it maintains the false alarm probability at the desired level for all subgroup sizes considered, (ii) that the estimated (or empirical) false alarm probability of the ANOM approach is relatively close to the desired level, whereas it is higher than the desired level when the Bonferroni-adjusted limits are used for small sample sizes, but (iii) for large n, the two sets of limits perform relatively similarly.

The authors recommend that the exact ANOM control limits be used for small subgroup sizes and that either approach may be used for larger subgroup sizes to control the overall probability of a false alarm (i.e. the FAP). Note, however, that these authors incorrectly base the two-sided Bonferroni-adjusted control limits on $z_{FAP_0/m}$ and not on $z_{FAP_0/2m}$.

Remark 13

(i) The ANOM based control limits of Neduraman and Pignatiello (2005) are derived using $\overline{V}$ whereas the Bonferroni-adjusted limits, to which they compare their ANOM based limits, are based on $\overline{S}$.

(ii) It is apparent that the ANOM based approach of Neduraman and Pignatiello (2005) is similar in spirit to that of Champ and Jones (2004), particularly when using $\overline{V}$ as an estimator of σ^2. It should be noted that while Champ and Jones (2004) used the unbiased estimator $\sqrt{\overline{V}}/c_4$ of σ, Nedumaran and Pignatiello (2005) did not; they simply used $\sqrt{\overline{V}}$.

(iii) In the approach by Neduraman and Pignatiello (2005) and that of Champ and Jones (2004) we are working with a singular multivariate *t*-distribution (since $\sum_{i=1}^{m} T_{v,i} = 1$ and $\sum_{i=1}^{m} T_i = 1$) with a negative and common correlation of $-1/(m-1)$. So, the computer programs used to find the critical values must take account of the singularity of the joint distribution.

2.3.2 Phase I charts for other settings

Control charts for rational subgroups of size $n > 1$ from a normal distribution is important, but there are situations where (a) the assumption of normality is not valid, for example, when the time between some events (such as failures) is monitored and it is well-known that the exponential distribution is a better model, and (b) in some cases it is more natural to analyze the individual observations as they are collected so that the sample size $n = 1$ (see e.g. Montgomery, (2005)) . Two methods that are useful in these situations are considered next.

(a) Jones and Champ (2002): "Phase I control chart for times between events"

Phase I charts have been considered for distributions other than the normal that is useful in SPC applications. Jones and Champ (2002) proposed Phase I charts to monitor the time between events for the standards known and unknown cases. These charts are referred to as Phase I exponential charts.

Assuming that the occurrence of defects in a continuous process variable can be well modelled by a Poisson process and denoting the time of occurrence of the i^{th} defect by T_i, with the time between successive defects denoted by $X_i = T_i - T_{i-1}$, it is well-known that $X_i \sim iidEXP(\mu_i)$ for $i = 1,2,...,m$.

Standard known: Case K

In the standards known case $\mu_i = \mu_0$ for all $i = 1,2,...,m$, the charting statistics are the X_i's and the control limits for the Phase I exponential chart are given by

$$L\hat{C}L = k_L\mu_0 \qquad \text{and} \qquad U\hat{C}L = k_U\mu_0$$

where μ_0 is the known (specified) value of μ and the charting constants k_L and k_U are selected such that $0 < k_L < k_U$.

Jones and Champ (2002) show that the Phase I exponential chart in Case K can be designed by choosing values for k_L and k_U such that the probability of an alarm in case of an out-of-control process is greater than the desired false alarm probability FAP_0 i.e. choosing k_L and k_U such that

$$1 - \Pr(\bigcap_{i=1}^{m}\{k_L\mu_0 < X_i < k_U\mu_0\} \mid \text{at least one } \mu_i \neq \mu_0, \forall i) \geq FAP_0 .$$

Because the above equation is satisfied when $k_L\mu_0$ and $k_U\mu_0$ are taken to be the τ^{th} and the $(1-\alpha+\tau)^{\text{th}}$ percentage points of the exponential distribution with mean μ_0, it follows that

$$k_L = -\ln(1-\tau) \qquad \text{and} \qquad k_U = -\ln(\alpha-\tau)$$

with $\alpha = 1-(1-FAP_0)^{1/m}$, $0<\tau<\alpha$, and where τ is determined such that $(1-\tau)\ln(1-\tau)-(\alpha-\tau)\ln(\alpha-\tau)=0$.

Tables with values of τ, k_L and k_U for various values of FAP_0 and m are provided that can be used to easily calculate the control limits.

Standard unknown: Case U

For the standards unknown case the authors design exact lower one-sided charts (details omitted) as well as approximate two-sided Phase I exponential charts so that the false alarm probability is at most α. This is done using the fact that $X_i/\hat{\mu}$ is related to the univariate F-distribution, when the process is in-control, through

$$\frac{X_i}{\hat{\mu}} = \frac{m}{1+(m-1)F_{2(m-1),2}}$$

where the random variable $F_{2(m-1),2}$ follows an F - distribution with $2(m-1)$ and 2 degrees of freedom.

Using this result together with Boole's inequality it is shown that

$$\Pr\left(\bigcap_{i=1}^{m}\left\{\frac{m\hat{\mu}}{1+(m-1)F_{2(m-1),2,1-\alpha-\tau}} < X_i < \frac{m\hat{\mu}}{1+(m-1)F_{2(m-1),2,\tau}}\right\} \mid IC\right) \geq 1-\alpha$$

where $\alpha = FAP_0/m$, $0<\tau<\alpha$, and $F_{2(m-1),2,1-\alpha+\tau}$ and $F_{2(m-1),2,\tau}$ are the $(1-\alpha+\tau)^{\text{th}}$ and τ^{th} percentage points of the F-distribution with $2(m-1)$ and 2 degrees of freedom, respectively.

Consequently, the estimated control limits for the approximate two-sided Phase I exponential chart are given by

$$\hat{LCL} = \frac{m\hat{\mu}}{1+(m-1)F_{2(m-1),2,1-\alpha/m-\tau}} \qquad \text{and} \qquad \hat{UCL} = \frac{m\hat{\mu}}{1+(m-1)F_{2(m-1),2,\tau}}$$

with $\hat{\mu} = \frac{1}{m}\sum_{i=1}^{m} X_i$ and $0<\tau<\alpha/m$, respectively.

The performance of the Phase I exponential charts are evaluated by Jones and Champ (2002) using computer simulation experiments and assuming that n of the m X_i's ~ $EXP(\mu + c\mu)$ are out-of-control while the remaining $m-n$ X_i's ~ $EXP(\mu)$ are in-control.

For the standards known case a table containing values of the probability of at least one signal for various values of FAP_0, n and c, and samples of size $m = 30$, is provided. For the standards unknown case similar tables are provided which contain values of the proportion of charts with at least one signal for various values of FAP_0, n and c, and samples of size $m = 30$.

The authors point out that the sensitivity of the Phase I exponential charts is inversely related to the FAP_0 value and it should therefore not be set too low or the charts may not achieve the desired level of sensitivity.

(b) Change-point modeling and other control charting methods

In some applications it is natural to collect and record the data as they are observed individually. In this setting, some authors have suggested formulating the question of whether or not a process is in-control as a change-point problem. This formulation typically assumes that the observations up to and including a point in time (called the change-point) are i.i.d. (with the same mean and variance) with some known distribution (such as the normal) while the observations after the change-point are also i.i.d. with the same distribution but with a different mean and/or variance.

For example, when the common distribution is normal, one writes

$$X_i \sim \begin{cases} iidN(\mu_1, \sigma_1^2) & \text{for } i = 1,...,\tau \\ iidN(\mu_2, \sigma_2^2) & \text{for } i = \tau+1,...,n \end{cases}$$

where X_i for $i = 1,2,...,m$ denotes an individual observation and $0 \le \tau < n$ is the unknown change-point (in time). The goal is to be able to detect and/or locate the change-point as well as measure the magnitude of the change as quickly as possible.

The change-point problem has a rich history in the statistics literature. In the SPC context, there are several papers, including Hawkins (1977), Sullivan and Woodall (1996), Hawkins, Qiu and Kang (2003) as well as Hawkins and Zamba (2005). Because the majority of these methods are based on the likelihood ratio testing procedure and because only the typical Phase I setting (i.e. checking whether one or more Phase I plotting statistics plot outside the control limits) is the focus here, a detailed discussion is not given.

Other control charting methods for Case U include, for example

(i) *Q*-charts (Quesenberry, (1991)),

(ii) control charts using sequential sampling schemes (see e.g. Zhang, Xie and Goh, (2006)), and

(iii) the model-based control charts (Koning, (2006)).

The *Q*-charts and charts based on sequential sampling schemes can be used in situations where self-starting techniques are needed, for example, in low-volume, job-shop (short-run) processes and/or in start-up situations. While these charts are useful in these situations, they are not applied in a typical Phase I setting.

2.4 Concluding remarks: Summary and recommendations

The focus in this chapter was primarily univariate variables Shewhart-type Phase I control charts.

In particular, we

(i) looked at what a Shewhart-type Phase I control chart is and how it is typically designed,

(ii) studied the design of Phase I control charts for process spread, and

(iii) gave an overview on the literature of univariate parametric Shewhart-type Phase I control chart for location and spread.

Section 2.1 gave a general discussion on Shewhart-type Phase I control charts in which the goals of Phase I control charting and the methods for designing and implementing Shewhart-type Phase I charts were described.

It turned out that the *FAP*-based control limits are the best to use when designing a Phase I chart because they correctly account for the fact that the Phase I signaling events are dependent and that multiple signaling events have to be dealt with simultaneously to make an in-control or not in-control decision; as a result it is recommended that the exact joint probability distribution of the charting statistics should be used (where possible) to control the false alarm probability when designing a Phase I chart.

The approximate *FAR*-based limits and the Bonferroni control limits were both shown to be close competitors of the *FAP*-based control limits; however, these two sets of control limits are both slightly wider than the *FAP*-based control limits and might lead to fewer alarms. In situations where the exact joint probability distribution is not available either of these two simpler (approximate) sets of control limits may be used; in such scenarios the marginal in-control distribution of each charting statistic is required.

Lastly, it was shown that the *FAR*-based control limits ignore the dependency of the Phase I charting statistics and overlooks the fact that multiple charting statistics are to be dealt with simultaneously; as a result, it is likely that one my observe more false alarms that what is typically expected and this approach should therefore not be used in designing a Shewhart-type Phase I control chart.

The techniques used in designing the Phase I S^2, S and R charts of section 2.2 recognized that multiple signaling events are involved and that the comparisons of the charting statistics with the estimated control limits are not independent. The design of the S^2 chart for $m < 25$ needs to be based on a multivariate singular beta distribution, also known as the type I or standard Dirichlet distribution, with common correlation $-1/(m-1)$; whereas for $m \geq 25$, percentiles of the univariate type I or standard beta distribution may be used as an approximation. For the R and the S charts, the design of the charts depends on some joint probability distribution(s) that are currently unknown.

Using computer simulations, the necessary charting constants for each chart were calculated so that the false alarm probabilities of the charts do not exceed 0.01, 0.05 and 0.10, respectively. For other desirable nominal false alarm probabilities the methods given in section 2.2 can be used to find the appropriate charting constants.

It is recommended that practitioners use the charting constants provided in Tables 2.1, 2.2, 2.3, 2.5, 2.6, 2.7, 2.8, 2.9, 2.10, 2.12, 2.13, 2.14 when computing the control limits of the Phase I S^2, S and R charts. The connection between the false alarm rate and the false alarm probability in a number of selective cases was also examined in order to provide some guidance to the user.

Finally, In section 2.3 we gave an overview on univariate parametric Shewhart-type Phase I control charts for location and spread. It is believed that this would be to the benefit of all users of control charts in that it informs them what the present state of the art is and what future research still remains.

Although the Phase I control charts included in the overview are all based on the assumption that the observations are i.i.d., one can argue that autocorrelation can be present in a number of potential applications. Thus further research on Phase I control charts for autocorrelated data (see e.g. Maragah and Woodall, (1992) and Boyles, (2000)) will be of great benefit to the SPC practitioner. Also, even though the overview focused on variables data, attributes data are common in some applications and as a result Phase I charts for attributes data (see e.g. Borror and Champ (2001)) are also useful and more work needs to be done in this area. Moreover, since not much is typically known or can be assumed about the underlying process distribution in a Phase I setting, nonparametric Phase I control charts would be of practical benefit and should be investigated.

It should be noted that, a clear consensus does not appear to exist as to how Phase I charts should be compared and contrasted. In Phase II, control chart performance is typically measured in terms of some attribute of the run-length distribution. In Phase I, the preferred performance metric is the probability of at least one signal. So for the in-control case, one can compare two or more charts by comparing their *FAP*'s. In the out-of-control case, if there are two control charts with the same or

roughly the same *FAP*, one can examine the probability of at least one signal when there is a shift in the process parameter and the chart with a higher probability of a signal should be preferred. This would be in line with comparing the power of two tests that are of the same size. Champ and Jones (2004) undertook the in-control *FAP* comparison in a simulation study whereas Jones and Champ (2002) looked at the out-of-control comparison of Phase I control charts.

2.5 Appendix 2A: SAS® programs

2.5.1 SAS® program to find the charting constants for the Phase I S^2 chart

```
proc iml;

sim=100000;
m=5;
dof=4;
x=j(sim,m,.);
y=j(sim,2,.);
call randgen(x,'CHISQ',dof);

do i=1 to sim;
sum=x[i,+];
y[i,1]=max(x[i,])/sum;
y[i,2]=min(x[i,])/sum;
end;

out=j(2000,2,.);

do alpha=0.0001 to 0.2 by 0.0001;
a=cinv(alpha/2,dof)/(m*dof);
b=cinv(1-alpha/2,dof)/(m*dof);
r=10000*alpha;
t=j(sim,3,.);
t[,1] = y[,1] > j(sim,1,b);
t[,2] = y[,2] < j(sim,1,a);
t[,3] = t[,1]|t[,2];
FAP   = t[+,3]/sim;
out[r,1] = alpha;
out[r,2] = FAP;
end;

create FAP_Ssq from out[colname={alpha  FAP}];
append from out;
quit;

proc export data=FAP_Ssq
outfile="c:\FAP_Ssq.xls" replace;
run;
```

2.5.2 SAS® program to find the charting constants for the Phase I *S* chart

```
proc iml;

sim=10;
m=5;
n=5;
x=j(sim,m,0);
y=j(sim,2,.);

do i=1 to sim;
          do j=1 to m;
                    z=j(1,n,0);
                    call randgen(z,'NORMAL');
                    x[i,j]=sqrt((ssq(z)-sum(z)*sum(z)/n)/(n-1));
          end;
end;

do i=1 to sim;
sum=x[i,+];
y[i,1]=max(x[i,])/sum;
y[i,2]=min(x[i,])/sum;
end;

out=j(350,2,.);
do k=0.01 to 3.5 by 0.01;
lcl=(1-k*sqrt(1-0.94*0.94)/0.94)/m;
ucl=(1+k*sqrt(1-0.94*0.94)/0.94)/m;
r=100*k;
t=j(sim,3,.);
t[,1] = y[,1] > j(sim,1,ucl);
t[,2] = y[,2] < j(sim,1,lcl);
t[,3] = t[,1]|t[,2];
FAP = t[+,3]/sim;
out[r,1] = k;
out[r,2] = FAP;
end;

create FAP_S from out[colname={k FAP}];
append from out;

proc export data=FAP_S
outfile="c:\FAP_S.xls" replace;
quit;
```

2.5.3 SAS® program to find the charting constants for the Phase I *R* chart

```
proc iml;

sim=10;
m=5;
n=5;
x=j(sim,m,0);
y=j(sim,2,.);

do i=1 to sim;
          do j=1 to m;
    z=j(1,n,0);
    call randgen(z,'NORMAL');
          x[i,j]=max(z)-min(z);
          end;
end;

do i=1 to sim;
sum=x[i,+];
y[i,1]=max(x[i,])/sum;
y[i,2]=min(x[i,])/sum;
end;

out=j(350,2,.);

do k=0.01 to 3.5 by 0.01;
ucl=(1+k*0.864/2.326)/m;
lcl=(1-k*0.864/2.326)/m;
r=100*k;
t=j(sim,3,.);
t[,1] = y[,1] > j(sim,1,ucl);
t[,2] = y[,2] < j(sim,1,lcl);
t[,3] = t[,1]|t[,2];
FAP   = t[+,3]/sim;
out[r,1] = k;
out[r,2] = FAP;
end;

create FAP_R from out[colname={k FAP}];
append from out;

proc export data=FAP_R
outfile="c:\FAP_R.xls" replace;
quit;
```

Chapter 3

Attributes control charts: Case K and Case U

3.0 Chapter overview

Introduction

When studying categorical quality characteristics the items or the units of product are inspected and classified simply as conforming (they meet certain specifications) or nonconforming (they do not meet the specifications). The classification is typically carried out with respect to one or more of the specifications on some desired characteristics. We label such characteristics "attributes" and call the data collected "attributes data" (see e.g. Chapter 6, p.265 of Montgomery, (2005)).

The *p*-chart and the *c*-chart are well known and commonly used attributes control charts. The *p*-chart is based on the binomial distribution and works with the fraction of nonconforming items in a sample. The *c*-chart is based on the Poisson distribution and deals with the number of nonconformities in an inspection unit. Several statistical process control (SPC) textbooks including the ones by Farnum (1994), Ryan (2000) and Montgomery (2005) describe these charts.

Motivation

The *p*-chart and *c*-chart are particularly useful in the service industries and in non-manufacturing quality improvements efforts since many of the quality characteristics found in these environments are in actual fact attributes. SPC with attributes data therefore constitutes an important area of research and applications (see e.g. Woodall (1997) for a review).

The classical application of the *p*-chart and the *c*-chart requires that the parameters of the distributions are known. In many situations the true fraction nonconforming, p, and the true average number of nonconformities in an inspection unit, c, are unknown or unspecified and need to be

estimated from a reference sample or historical (past) data. While there are empirical rules and guidelines for setting up the charts, little is known about their run-length distributions when the fact that the parameters are estimated is taken into account. Understanding the effect of estimating the parameters on the in-control (IC) and the out-of-control (OOC) performance of the charts are therefore of interest from a practical and a theoretical point of view.

In this chapter we derive and evaluate expressions for the run-length distributions of the Shewhart-type *p*-chart and the Shewhart-type *c*-chart when the parameters are estimated. An exact approach based on the binomial and the Poisson distributions is used since in many applications the values of p and c are such that the normal approximation to the binomial and the Poisson distributions is quite poor, especially in the tails. The results are used to discuss the appropriateness of the widely followed empirical rules for choosing the size of the Phase I sample used to estimate the unknown parameters; this includes both the number of reference samples (or inspection units) m and the sample size n. Note that, in our developments, we assume that the size of each subgroup or the size of each inspection unit stays constant over time.

Methodology

We examine the effect of estimating p and c on the performance of the *p*-chart and the *c*-chart via their run-length distributions and associated characteristics such as the average run-length (ARL), the false alarm rate (FAR) and the probability of a "no-signal". Exact expressions are derived for the Phase II run-length distributions and the related Phase II characteristics using expectation by conditioning (see e.g. Chakraborti, (2000)). We first obtain the characteristics of the run-length distributions conditioned on point estimates from Phase I and then find the unconditional characteristics by averaging over the distributions of the point estimators. This two-step analysis provides valuable insight into the specific as well as the overall effects of parameter estimation on the performance of the charts in Phase II.

The conditional characteristics let us focus on specific values of the estimators and look at the performance of the charts in more detail for the particular value(s) at hand. The unconditional characteristics characterize the overall performance of the charts i.e. averaged over all possible values of the estimators.

In practice we will obviously have only a single realization for each of the point estimators and the characteristics of the conditional run-length distribution therefore provide important information

specific only to our particular situation; but, since each user will have his own values for each of the point estimators the conditional run-length performance will be different from user to user. The unconditional run-length, on the other hand, lets us look at the bigger picture, averaged over all possible values of the point estimators, and is therefore the same for all users.

Layout of Chapter 3

This chapter consists of two main sections and an appendix. The first section is labeled "The *p*-chart and the *c*-chart for standards known (Case K)" and the second section is called "The *p*-chart and the *c*-chart for standards unknown (Case U)". In the first section we study the charts when the parameters are known. The second section focuses on the situation when the parameters are unknown and forms the heart of Chapter 3. In both sections we study the *p*-chart and the *c*-chart in unison; this points out the similarity and the differences between the charts and helps one to understand the theory and/or methodology better.

Appendix 3A gives an example of each chart and contains a discussion on the characteristics of the *p*-chart and the *c*-chart in Case K. To the author's knowledge none of the standard textbooks and/or articles currently available in the literature give a detailed discussion of the Case K *p*-chart's and the Case K *c*-chart's characteristics.

3.1 The *p*-chart and the *c*-chart for standards known (Case K)

Introduction

Case K is the scenario where known values for the parameters are available. This will happen in high volume manufacturing processes where ample reliable information is available so that it is possible to specify values for the parameters.

Studying Case K not only sets the stage for the situation when the parameters are unknown (Case U), but the characteristics and the performance of the charts in Case K are also important. In particular, it helps us understand the operation and the performance of the charts in the simplest of cases (when the parameters are known) and provides us with benchmark values which we can use to determine the effect of estimating the parameters on the operation and the performance of the charts in Case U (when the parameters are unknown).

The *p*-chart is used when we monitor the fraction of nonconforming items in a sample of size $n \geq 1$ and is based on the binomial distribution. The *c*-chart is based on the Poisson distribution and used when we focus on monitoring the number of nonconformities in an inspection unit, where the inspection unit may consist of one or more than one physical unit.

Assumptions

We derive and study the characteristics of the charts in Case K assuming that: (i) the sample size and the size of an inspection unit (whichever is applicable) stay constant over time, (ii) the nonconforming items occur independently i.e. the occurrence of a nonconforming item at a particular point in time does not affect the probability of a nonconforming item in the time periods that immediately follow, and (iii) the probability of observing a nonconformity in an inspection unit is small, yet the number of possible nonconformities in an inspection unit is infinite.

To this end, let $X_i \sim iidBin(n, p)$ for $i = 1,2,...$ denote the number of nonconforming items in a sample of size $n \geq 1$ with true fraction nonconforming $0 < p < 1$; the sample fraction nonconforming is then defined as $p_i = X_i / n$. Similarly, let $Y_i \sim iidPoi(c)$, $c > 0$ for $i = 1,2,...$ denote the number of nonconformities in an inspection unit where c denotes the true average number of nonconformities in an inspection unit.

Charting statistics

The charting statistics of the *p*-chart is the sample fraction nonconforming $p_i = X_i / n$ for $i = 1,2,\ldots$; the charting statistics of the *c*-chart is the number of nonconformities Y_i for $i = 1,2,\ldots$, in an inspection unit.

Control limits

For known values of the true fraction nonconforming and the true average number of nonconformities in an inspection unit, denoted by p_0 and c_0 respectively, the upper control limits (*UCL*'s), the centerlines (*CL*'s), and the lower control limits (*LCL*'s) of the traditional *p*-chart and the traditional *c*-chart are

$$UCL_p = p_0 + 3\sqrt{p_0(1-p_0)/n} \qquad CL_p = p_0 \qquad LCL_p = p_0 - 3\sqrt{p_0(1-p_0)/n} \tag{3-1}$$

and

$$UCL_c = c_0 + 3\sqrt{c_0} \qquad CL_c = c_0 \qquad LCL_c = c_0 - 3\sqrt{c_0} \tag{3-2}$$

respectively (see e.g. Montgomery, (2005) p. 268 and p. 289).

The control limits in (3-1) and (3-2) are k-sigma limits (where $k = 3$) and based on the tacit assumption that both the binomial distribution and the Poisson distribution are well approximated by the normal distribution.

The subscripts "*p*" and "*c*" in (3-1) and (3-2) are used to distinguish the control limits of the two charts; where no confusion is possible the subscripts are dropped.

Implementation

The actual operation of the charts consist of: (i) taking independent samples and independent inspection units at equally spaced successive time intervals, (ii) computing the charting statistics, and then (iii) plotting the charting statistics (one at a time) reflected on the vertical axis of the control charts versus the sample number and the inspection unit number $i = 1,2,...$ reflected on the horizontal axis.

The control limits are also displayed on the charts so that every time a new charting statistic is plotted it is in actual fact compared to the control limits. The aim is to detect when (or if) the true process parameters p and c change (moves away) from their known or specified or target values p_0 and c_0, respectively.

Signaling and non-signaling events

The event when a charting statistic (point) plots outside the control limits, which is called a signaling event and denoted by A_i for $i = 1,2,...$, is interpreted as evidence that the parameter is no longer equal to its specified value. The charting procedure therefore stops, a signal (alarm) is given, and we declare the process out-of-control (OOC) i.e. we say that $p \neq p_0$ or state that $c \neq c_0$. Investigation and corrective action is typically required to find and eliminate the possible assignable cause(s) and/or source(s) of variability responsible for the behavior.

The complimentary event is when a plotted point lies between (within) the control limits and labeled a non-signaling event or a “no-signal”. In case of a no-signal the charting procedure continues, no user intervention is necessary, and we consider the process to be in-control (IC) i.e. we say that $p = p_0$ or declare that $c = c_0$. We denote the non-signaling event by

$$A_i^C : \{LCL < Q_i < UCL\}$$

where $Q_i = p_i$ or Y_i for $i = 1,2,...$ and LCL and UCL denote the control limits in either (3-1) or (3-2).

Note that, in a hypothesis-testing framework, concluding that the process is out-of-control when the process is actually in-control is called a type I error; similarly, concluding that the process is in-control when it is really out-of-control is a called a type II error.

3.1.1 Probability of a no-signal

Introduction

The probability of a no-signal refers to the probability of a non-signaling event and is denoted by

$$\beta = \Pr(A_i^C) \quad \text{for} \quad i = 1,2,\ldots.$$

The probability of a no-signal is important because: (i) it is the key for the derivation of the run-length distribution, and (ii) plays a central role when we assess the performance of a control chart. Once we have the probability of a no-signal, the run-length distribution is completely known.

Probability of a no-signal: *p*-chart

The probability of a no-signal on the *p*-chart is the probability of the event

$$\{LCL_p < p_i < UCL_p\} \quad \text{for} \quad i = 1,2,\ldots. \tag{3-3}$$

Since p is known and equal to p_0 the control limits LCL_p and UCL_p are known values (constants) which makes $p_i = X_i / n$ the only random quantity in (3-3).

The cumulative distribution function of the sample fraction nonconforming p_i is known and given by

$$\Pr(p_i \le a) = \Pr(X_i / n \le a) = \Pr(X_i \le na) = \sum_{j=0}^{[na]} \Pr(X_i = j) = \sum_{j=0}^{[na]} \binom{n}{j} p^j (1-p)^{n-j}$$

for $0 \le a \le 1$, $0 < p < 1$ and where $[na]$ denotes the largest integer not exceeding na. Because the distribution of p_i is defined in terms of that of $X_i \sim Bin(n, p)$ we re-express the non-signaling event in (3-3) as

$$\{nLCL_p < X_i < nUCL_p\}$$

and use the properties of the distribution of X_i to derive the probability of a no-signal.

Thus, at the i^{th} observation the non-signaling probability for the *p*-chart is a function of and depends on p, p_0 and n, and is derived as follows

$$\begin{aligned}\beta(p, p_0, n) &= \Pr(LCL_p < p_i < UCL_p) \\ &= \Pr(nLCL_p < X_i < nUCL_p) \\ &= \Pr(X_i < nUCL_p) - \Pr(X_i \le nLCL_p)\end{aligned}$$

$$= \begin{cases} H(b; p, n) & \text{if } nLCL_p < 0 \\ H(b; p, n) - H(a; p, n) & \text{if } nLCL_p \ge 0 \end{cases} \qquad (3\text{-}4)$$

$$= \begin{cases} 1 - I_p(b+1, n-b) & \text{if } nLCL_p < 0 \\ I_p(a+1, n-a) - I_p(b+1, n-b) & \text{if } nLCL_p \ge 0 \end{cases}$$

$$= 1 - I_p(b+1, n-b) - 1_{\{nLCL_p : nLCL_p \ge 0\}}(nLCL_p)(1 - I_p(a+1, n-a))$$

for $0 < p, p_0 < 1$, where UCL_p and LCL_p are defined in (3-1) and both are functions of n and p_0,

$$H(b; p, n) = \Pr(X_i \le b) = \sum_{j=0}^{b} \binom{n}{j} p^j (1-p)^{n-j}$$

denotes the cumulative distribution function (c.d.f) of the $Bin(n, p)$ distribution,

$$I_t(u, v) = (\beta(u, v))^{-1} B(t; u, v) \quad \text{for} \quad 0 < t < 1 \qquad \text{and} \qquad B(t; u, v) = \int_0^t s^{u-1}(1-s)^{v-1} ds \quad \text{for} \quad u, v > 0$$

denotes the c.d.f of the $Beta(u, v)$ distribution (also known as the incomplete beta function) with $\beta(u, v) = B(1; u, v)$,

$$1_{\{x : x \ge 0\}}(x) = \begin{cases} 1 & \text{if } x \ge 0 \\ 0 & \text{if } x < 0 \end{cases},$$

and where

$$a = [nLCL_p] \quad \& \quad b = \begin{cases} \min\{nUCL_p - 1, n\} & \text{if } nUCL_p \text{ is an integer} \\ \min\{[nUCL_p], n\} & \text{if } nUCL_p \text{ is not an integer} \end{cases} \qquad (3\text{-}5)$$

and $[x]$ denotes the largest integer not exceeding x.

Remark 1

(i) Making use of the c.d.f of the beta distribution and the indicator function $1_{\{x:x\geq 0\}}(x)$ helps us write the probability of a no-signal in a more compact way (see e.g. the last line of (3-4)).

(ii) The relationship between the c.d.f of the binomial distribution and the c.d.f of the type I or standard beta distribution is evident from (3-4) and given by

$$H(b;n,p) = 1 - I_p(b+1, n-b) = I_{1-p}(n-b, b+1).$$

(iii) The charting constants a and b in (3-5) are suitably modified to take account of the fact that the $Bin(n,p)$ distribution assigns nonzero probabilities only to integers from 0 to n.

(iv) To cover both the in-control and the out-of-control scenarios we do not assume that the specified value for the fraction nonconforming p_0 in (3-4) is necessarily equal to the true fraction nonconforming p.

Probability of a no-signal: *c*-chart

The probability of a no-signal on the *c*-chart is the probability that the event

$$\{LCL_c < Y_i < UCL_c\} \quad \text{for} \quad i = 1,2,... \tag{3-6}$$

occurs. Since c is specified and equal to c_0 the control limits LCL_c and UCL_c are constants. As a result Y_i is the only random variable in (3-6). Because the distribution of Y_i is known (assumed) to be Poisson with parameter (in general) c, we derive the probability of a no-signal on the *c*-chart (directly) in terms of the distribution of Y_i .

The probability of a no-signal on the *c*-chart is a function of and depends on c and c_0, and is derived as follows

$$\begin{aligned}\beta(c,c_0) &= \Pr(LCL_c < Y_i < UCL_c) \\ &= \Pr(Y_i < UCL_c) - \Pr(Y_i \le LCL_c) \\ &= G(f;c) - G(d;c) \\ &= \Gamma_{f+1}(c) - \Gamma_{d+1}(c)\end{aligned} \tag{3-7}$$

for $c, c_0 > 0$, where UCL_c and LCL_c are defined in (3-2) and both are functions of c_0,

$$G(f;c) = \Pr(Y_i \le f) = \sum_{j=0}^{f} \frac{e^{-c}c^j}{j!}$$

denotes the c.d.f of the $Poi(c)$ distribution,

$$\Gamma_t(u) = (\Gamma(t))^{-1}\Gamma(t;u) \qquad \text{where} \qquad \Gamma(t;u) = \int_t^{\infty} s^{u-1}e^{-s}ds \quad \text{for} \quad t,u > 0$$

denotes the upper incomplete gamma function,

$$\Gamma(t) = (t-1)!$$

for positive integer values of t, and where

$$d = \max\{0,[LCL_c]\} \quad \& \quad f = \begin{cases} UCL_c - 1 & \text{if } UCL_c \text{ is an integer} \\ [UCL_c] & \text{if } UCL_c \text{ is not an integer.} \end{cases} \tag{3-8}$$

Remark 2

(i) The relationship between the c.d.f of the Poisson distribution and the lower incomplete gamma function is evident from (3-7) and given by $G(f;c) = \Gamma_{f+1}(c)$.

(ii) The constants d and f in (3-8) incorporate the fact that the $Poi(c)$ distribution only assigns nonzero probabilities to non-negative integers.

(iii) We do not assume that c in (3-7) is necessarily equal to c_0; this enables us to study both the in-control and the out-of-control properties of the *c*-chart.

3.1.2 Operating characteristic and the OC-curve

The Operating Characteristic (OC) or the β-risk is the probability that a chart does not signal on the first sample or the first inspection unit following a sustained (permanent) step shift in the parameter and thus failing to detect the shift. For the *p*-chart the OC is the probability of a no-signal $\beta(p, p_0, n)$ with $p \neq p_0$ and for the *c*-chart the OC is the probability $\beta(c, c_0)$ with $c \neq c_0$.

A graphical display (plot) of the OC as a function of $0 < p < 1$ (in case of the *p*-chart), or as a function of $c > 0$ (in case of the *c*-chart), is called the operating characteristic curve or simply the OC-curve. The OC-curve lets us see a chart's ability to detect a shift in the process parameter and therefore describes the performance of the chart.

3.1.3 False alarm rate

As an alternative to the OC-curve we can graph the probability of a signal as a function of p for values of $0 < p < 1$ or as a function of c for values of $c > 0$. The probability of a signal is $1 - \beta$ i.e. one minus the probability of a no-signal, and is in some situations intuitively easier understood than the OC.

For the *p*-chart the probability of a signal is $1 - \beta(p, p_0, n)$ where $\beta(p, p_0, n)$ is defined in (3-4) and for the *c*-chart the probability of a signal is $1 - \beta(c, c_0)$ where $\beta(c, c_0)$ is given in (3-7).

When we substitute p with p_0 in $1-\beta(p, p_0, n)$ and replace c with c_0 in $1-\beta(c, c_0)$ we obtain the false alarm rate (FAR) of the charts, that is,

$$FAR(p_0, p_0, n) = 1-\beta(p_0, p_0, n) \qquad \text{and} \qquad FAR(c_0, c_0) = 1-\beta(c_0, c_0).$$

The false alarm rate is the probability of a signal when the process is in-control (i.e. no shift occurred) and often used a measure of a control chart's in-control performance.

The OC-curve and the probability of a signal as functions of p or c i.e. given a shift in the process, focus on the probability of a single event and involves only one charting statistic. A more popular and perhaps more useful method to evaluate and examine the performance of a control chart is its run-length distribution.

3.1.4 Run-length distribution

The number of rational subgroups to be collected or the number of charting statistics to be plotted on a control chart before the first or next signal, is called the run-length of a chart. The discrete random variable defining the run-length is called the run-length random variable and denoted by N. The distribution of N is called the run-length distribution.

Characteristics of the run-length distribution give us more insight into the performance of a chart. The characteristics of the run-length distribution most often looked at are, for example, its moments (such as the expected value and the standard deviation) as well as the percentiles or the quartiles (see e.g. Shmueli and Cohen, (2003)).

If no shift occurred (i.e. $p = p_0$ or $c = c_0$) the distribution of N is called the in-control run-length distribution. In contrast, if the process did encounter a shift (i.e. $p \neq p_0$ or $c \neq c_0$) the distribution of N is labeled the out-of-control run-length distribution. To distinguish between the in-control and the out-of-control situations the notations N_0 and N_1 are used; this notation is also used for the characteristics of the run-length distribution.

Assuming that the rational subgroups are independent and that the probability of a signal is the same for all samples (inspection units) the run-length distribution is given by

$$\Pr(N = j) = \beta^{j-1}(1-\beta) \qquad j = 1,2,... \tag{3-9}$$

where β denotes the probability of a no-signal defined in (3-4) or (3-7).

The distribution in (3-9) is recognized as the geometric distribution (of order 1) with probability of "success" $1-\beta$ so that we write, symbolically, $N \sim Geo(1-\beta)$. The success probability is the probability of a signal and, as mentioned before, completely characterizes the geometric (run-length) distribution.

Various statistical characteristics of the run-length distribution provide insight into how a control chart functions and performs. Typically we want the chart to signal quickly once a change takes place and not signal too often when the process is actually in-control, which is when no shift or no change has occurred. We are interested in the typical value as well as the spread or the variation in the run-length distribution.

3.1.5 Average run-length

A popular measure of the central tendency of a distribution is the expected value (mean) or the average. Accordingly, the average has been the most popular index or measure of a control chart's performance and is called the average run-length (*ARL*). The *ARL* is defined as the expected number of rational subgroups that must be collected before the chart signals.

When the process is in-control the expected number of charting statistics that must be plotted before the control chart signals erroneously is called the in-control average run-length and denoted by ARL_0. The out-of-control average run-length is denoted by ARL_1 and is the expected number of charting statistics to be plotted before a chart signals after the process has gone out-of-control. Obviously, for an efficient control chart the in-control average run-length should be large and the out-of-control average run-length should be small.

From the properties of the geometric distribution the *ARL* is the expected value of N so that

$$ARL = E(N) = 1/(1-\beta). \qquad (3\text{-}10)$$

Therefore, when the signaling events are independent and have the same probability the *ARL* of the chart is simply the reciprocal of the probability of a signal $1-\beta$. If the process is in-control, the in-control *ARL* is equal to the reciprocal of the *FAR*, that is, $ARL_0 = 1/FAR$. It is this simple relationship between the average run-length and the probability of a signal, or the in-control average run-length and the false alarm rate, that accounts for the popularity of the (in-control) average run-length and the probability of a signal (false alarm rate) as measures of a control chart's performance.

3.1.6 Standard deviation and percentiles of the run-length

Other characteristics of the run-length distribution are also of interest. For example, in addition to the mean we should also look at the standard deviation of the run-length distribution to get an idea about the variation or spread.

Using results for the geometric distribution, the standard deviation of the run-length, denoted by *SDRL*, is given by

$$SDRL = \text{stdev}(N) = \sqrt{\beta}/(1-\beta). \qquad (3\text{-}11)$$

Since the geometric distribution is skewed to the right the mean and the standard deviation become questionable measures of central tendency and spread so that additional descriptive measures are useful. For example, the percentiles, such as the median and the quartiles (which are more robust or outlier resistant), can provide valuable information about the location as well as the variation in the run-length distribution.

Because the run-length distribution is discrete, the $100q^{\text{th}}$ percentile $(0 < q < 1)$ is defined as the smallest integer j such that the cumulative probability is at least q, that is, $\Pr(N \leq j) \geq q$. The median run-length (denoted by *MDRL*) is the 50^{th} percentile so that $q = 0.5$, whereas the first quartile (Q_1) is the 25^{th} percentile so that $q = 0.25$.

3.1.7 In-control and out-of-control run-length distributions

The characteristics of the in-control run-length distributions are essential in the design and implementation of a control chart. Furthermore, for out-of-control performance comparisons we need the out-of-control run-length distributions and/or characteristics. For example, the in-control average run-lengths of the charts are typically fixed at an acceptably high level so that the number of false alarms or the false alarm rate is reasonably small. The chart with the smallest or the lowest out-of-control average run-length for a certain change (or shift of a specified size) in the process parameter is then selected to be the winner (i.e. the best performing chart). Alternatively, we can fix the false alarm rate of the charts at an acceptably small value and then select that chart with the highest probability of a signal (given a specified shift in the parameter) as the winner.

Note that, the average run-length and the probability of a signal are two equivalent performance measures in that they both lead to the same decision and follows from the relationship between the average run-length and the probability of a signal given in (3-10).

The run-length distributions and some related characteristics of the run-length distributions of the *p*-chart and the *c*-chart, which all conveniently follow from the properties of the geometric distribution of order 1, are summarized in Table 3.1 and Table 3.2, respectively.

The characteristics of the *p*-chart and the *c*-chart are seen to be all functions of and depend entirely on the probability of a no-signal, that is, $\beta(p, p_0, n)$ or $\beta(c, c_0)$; once we have expressions and/or numerical values for the two probabilities $\beta(p, p_0, n)$ and $\beta(c, c_0)$ the run-length distributions are completely known.

The in-control run-length distributions and the in-control characteristics of the run-length distributions are obtained when $p = p_0$ and $c = c_0$. The out-of-control run-length distributions and the out-of-control characteristics are found by setting $p \neq p_0$ and $c \neq c_0$, respectively.

An in-depth analysis and discussion of the in-control run-length distributions of the *p*-chart and the *c*-chart in Case K (and their related in-control properties) are given in Appendix 3A. From time to time we will refer to the results therein; especially when we study and look at the effects of parameter estimation on the performance of the charts in Case U.

Table 3.1: The probability mass function (p.m.f), the cumulative distribution function (c.d.f), the false alarm rate (*FAR*), the average run-length (*ARL*), the standard deviation of the run-length (*SDRL*) and the quantile function (qf) of the run-length distribution of the *p*-chart in Case K

p.m.f	$\Pr(N_p = j; p, p_0, n) = \beta(p, p_0, n)^{j-1}(1-\beta(p, p_0, n)) \qquad j = 1,2,\ldots$	(3-12)
c.d.f	$\Pr(N_p \le j; p, p_0, n) = 1-(\beta(p, p_0, n))^{j} \qquad j = 1,2,\ldots$	(3-13)
FAR	$FAR(p_0, n) = 1-\beta(p_0, p_0, n)$	(3-14)
ARL	$ARL(p, p_0, n) = E(N_p) = 1/(1-\beta(p, p_0, n))$	(3-15)
SDRL	$SDRL(p, p_0, n) = \text{stdev}(N_p) = \sqrt{\beta(p, p_0, n)}\,/(1-\beta(p, p_0, n))$	(3-16)
qf	$Q_{N_p}(q; p, p_0, n) = \inf\{\text{int}\, x : \Pr(N_p \le x; p, p_0, n) \ge q\} \quad 0 < q < 1$	(3-17)

Table 3.2: The probability mass function (p.m.f), the cumulative distribution function (c.d.f), the false alarm rate (*FAR*), the average run-length (*ARL*), the standard deviation of the run-length (*SDRL*) and the quantile function (qf) of the run-length distribution of the *c*-chart in Case K

p.m.f	$\Pr(N_c = j; c, c_0) = \beta(c, c_0)^{j-1}(1-\beta(c, c_0)) \qquad j = 1,2,\ldots$	(3-18)
c.d.f	$\Pr(N_c \le j; c, c_0) = 1-(\beta(c, c_0))^{j} \qquad j = 1,2,\ldots$	(3-19)
FAR	$FAR(c_0) = 1-\beta(c_0, c_0)$	(3-20)
ARL	$ARL(c, c_0) = E(N_c) = 1/(1-\beta(c, c_0))$	(3-21)
SDRL	$SDRL(c, c_0) = \text{stdev}(N_c) = \sqrt{\beta(c, c_0)}\,/(1-\beta(c, c_0))$	(3-22)
qf	$Q_{N_c}(q; c, c_0) = \inf\{\text{int}\, x : \Pr(N_c \le x; c, c_0) \ge q\} \quad 0 < q < 1$	(3-23)

3.2 The *p*-chart and the *c*-chart for standards unknown (Case U)

Introduction

Case U is the scenario when the parameters p and c are unknown. Case U occurs more often in practice than Case K particularly when not much historical knowledge or expert opinion is available. In the service industries, non-manufacturing environments and job-shop environments, which all involve low-volume of "production", it often happens that there is a scarcity of historical data.

Setting up a control chart in Case U consists of two phases: Phase I and Phase II. The former is the so-called retrospective phase whereas the latter is labeled the prospective or the monitoring phase (see e.g. Woodall, (2000)). In Phase I the parameters and the control limits are estimated from an in-control reference sample or calibration sample. In Phase II, new incoming subgroups are collected independently from the Phase I reference sample. The charting statistic for each Phase II subgroup is then calculated and individually compared to the estimated Phase II control limits until the first point plots outside the limits. The goal is to detect when (or if) the process parameters change.

We study and analyze the performance of the *p*-chart and *c*-chart following a Phase I analysis. In other words, we focus on the run-length distributions and the associated characteristics of the run-length distributions of the *p*-chart and the *c*-chart in Phase II.

3.2.1 Phase I of the Phase II *p*-chart and *c*-chart

The charting procedures to ensure that the Phase I data is representative of the in-control state of the process were discussed in Chapter 2. Here we consider the matter only in very general terms and assume that such in-control Phase I data is available; this implies that each sample and each inspection unit in the reference sample has identical (unknown) parameters.

Phase I data and assumptions

The Phase I data is the in-control reference sample or the historical (past) data that is used to estimate the unknown parameters. In case of the *p*-chart the Phase I data consists of m mutually independent samples each of size $n \geq 1$. The Phase I data for the *c*-chart consists of m mutually independent inspection units.

To this end, let $X_i \sim iidBin(n, p)$ for $i = 1,2,...,m$ denote the number of nonconforming items in the i^{th} reference sample of size $n \geq 1$ with unknown true fraction nonconforming $0 < p < 1$. The sample fraction nonconforming of each preliminary sample is $p_i = X_i / n$ for $i = 1,2,...,m$. Similarly, let $Y_i \sim iidPoi(c)$, $c > 0$ for $i = 1,2,...,m$ denote the number of nonconformities in the i^{th} reference inspection unit where c denotes the unknown true average number of nonconformities in an inspection unit.

Phase I point estimators for *p* and *c*

The average of the m Phase I sample fractions nonconforming $p_1, p_2, ..., p_m$ and the average of the numbers of nonconformities in each Phase I inspection unit $Y_1, Y_2, ..., Y_m$, are used to estimate p and c, respectively. In other words, we estimate p by

$$\bar{p} = \frac{1}{m}\sum_{i=1}^{m} p_i = \frac{1}{mn}\sum_{i=1}^{m} X_i = \frac{U}{mn} \tag{3-24}$$

and c by

$$\bar{c} = \frac{1}{m}\sum_{i=1}^{m} Y_i = \frac{V}{m} \tag{3-25}$$

where the random variable

$$U = \sum_{i=1}^{m} X_i \sim Bin(mn, p)$$

denotes the total number of nonconforming items in the entire set of mn reference observations and the random variable

$$V = \sum_{i=1}^{m} Y_i \sim Poi(mc)$$

denotes the total number of nonconformities in the entire set of m reference inspection units.

Remark 3

(i) It can be verified that the point estimators $\bar{p}$ and $\bar{c}$ in (3-24) and (3-25) are: (a) the maximum likelihood estimators (MLE's), and (b) the minimum variance unbiased estimators (MVUE's), of p and c, respectively (see e.g. Johnson, Kemp and Kotz, (2005) p. 126 and p. 174).

In particular, note that, the expected value and the variance of $\bar{p}$ are

$$E(\bar{p}) = \frac{E(U)}{mn} = \frac{mnp}{mn} = p,$$

and

$$\text{var}(\bar{p}) = \frac{\text{var}(U)}{(mn)^2} = \frac{mnp(1-p)}{(mn)^2} = \frac{p(1-p)}{mn},$$

respectively, whereas the expected value and the variance of $\bar{c}$ are

$$E(\bar{c}) = \frac{E(V)}{m} = \frac{mc}{m} = c,$$

and

$$\text{var}(\bar{c}) = \frac{\text{var}(V)}{m^2} = \frac{mc}{m^2} = \frac{c}{m},$$

respectively.

(ii) It is essential to note that the distribution of U depends on the unknown parameter p and the distribution of V depends on the unknown parameter c so that it is technically correct to write

$$U \mid p \sim Bin(mn, p) \qquad \text{and} \qquad V \mid c \sim Poi(mc)\,.$$

This observation will become vital when we study the unconditional run-length distributions and the characteristics of the unconditional run-length distribution in later sections.

3.2.2 Phase II *p*-chart and *c*-chart

A Phase II chart refers to the operation and implementation of a chart following a Phase I analysis in which any unknown parameters were estimated from the Phase I reference sample.

Phase II estimated control limits

It is standard practice to replace p_0 with $\bar{p}$ in (3-1) and substitute $\bar{c}$ for c_0 in (3-2) when the parameters p and/or c are unknown (see e.g. Ryan, (2000) p. 155 and p. 169 and, Montgomery, (2005) p. 269 and p. 290). The estimated upper control limits ($U\hat{C}L$'s), the estimated centerlines ($\hat{C}L$'s), and the estimated lower control limits ($L\hat{C}L$'s) of the *p*-chart and the *c*-chart are therefore given by

$$U\hat{C}L_p = \bar{p} + 3\sqrt{\bar{p}(1-\bar{p})/n} \qquad \hat{C}L_p = \bar{p} \qquad L\hat{C}L_p = \bar{p} - 3\sqrt{\bar{p}(1-\bar{p})/n} \tag{3-26}$$

and

$$U\hat{C}L_c = \bar{c} + 3\sqrt{\bar{c}} \qquad \hat{C}L_c = \bar{c} \qquad L\hat{C}L_c = \bar{c} - 3\sqrt{\bar{c}} \tag{3-27}$$

respectively.

By the invariance property of MLE's the estimated control limits in (3-26) and (3-27) are the MLE's of the control limits of (3-1) and (3-2) in Case K (see e.g. Theorem 7.2.10 in Casella and Berger, (2002) p. 320). However, unlike in Case K, the Phase II estimated control limits are functions of and depend on the point estimators (variables) $\bar{p}$ or $\bar{c}$ and are random variables. We therefore need to account for the variability in the estimated control limits while determining and understanding the chart's properties.

Phase II charting statistics

Let $p_i = X_i/n$ for $i = m+1, m+2, \ldots$ denote the Phase II charting statistics for the *p*-chart where $X_i \sim iidBin(n, p_1)$ denote the number of nonconforming items in the i^{th} Phase II sample of size $n \geq 1$ with fraction nonconforming $0 < p_1 < 1$. Similarly, let $Y_i \sim iidPoi(c_1)$, $c_1 > 0$ for $i = m+1, m+2, \ldots$ denote the number of nonconformities in the i^{th} Phase II inspection unit where c_1 denotes the average number of nonconformities in an inspection unit in Phase II. These Y_i's are the Phase II charting statistics of the *c*-chart.

Remark 4

(i) The *p*-chart

It is important to note that the application of the *p*-chart in Case U depends on three parameters: the unknown true fraction nonconforming p, the point estimate $\bar{p}$ and p_1.

In Phase II we denote p with p_1 so that p_1 denotes the probability of an item being nonconforming in the prospective monitoring phase and p denotes the probability of an item being nonconforming in the retrospective phase. To maintain greater generality and to cover both the in-control (IC) and the out-of-control (OOC) cases, we do not assume that p_1 is necessarily equal to p. We therefore write $p_1 = p$ for the IC scenario and $p_1 \neq p$ for the OOC case.

Also, in Phase I we estimate p by $\bar{p}$, which (due to sampling variability) is not necessarily equal to p; we write this as $p = \bar{p}$ and $p \neq \bar{p}$. When $p = \bar{p}$ we say that p is estimated without error.

This is a key observation. Because we use $\bar{p}$ to calculate the estimated control limits, in Phase II we are actually comparing p_1 against $\bar{p}$ and not against p; this leads to the following four unique scenarios:

(i) $p_1 = p = \bar{p}$: the process is IC in Phase II and p is estimated without error,
(ii) $p_1 \neq p = \bar{p}$: the process is OOC in Phase II and p is estimated without error,
(iii) $p_1 = p \neq \bar{p}$: the process is IC in Phase II and p is *not* estimated without error, and
(iv) $p_1 \neq p \neq \bar{p}$: the process is OOC in Phase II and p is *not* estimated without error.

To simplify matters we assume, without loss of generality, that the process operates IC in Phase II and $\bar{p}$ is not necessarily equal to p; this is scenario (iii) listed above.

(ii) The *c*-chart

For the *c*-chart in Case U we have a similar situation as that for the *p*-chart i.e. the application of the *c*-chart in Case U depends on three parameters: the true (but unknown) average number of nonconformities in an inspection unit c, the point estimate $\bar{c}$ and c_1.

In Phase II we denote c with c_1 so that c_1 denotes the average number of nonconformities in an inspection unit in the prospective monitoring phase and c denotes the average number of nonconformities in an inspection unit in the retrospective phase. To maintain greater generality and to cover both the in-control (IC) and the out-of-control (OOC) cases, we do not assume that c_1 is necessarily equal to c, which we write as $c_1 = c$ for the IC scenario and $c_1 \neq c$ for the OOC case.

In Phase I however we estimate c by $\bar{c}$, which (due to sampling variability) is not necessarily equal to c and we write this as $c = \bar{c}$ and $c \neq \bar{c}$. When $c = \bar{c}$ we say that c is estimated without error.

Now, because we use $\bar{c}$ to calculate the estimated control limits, in Phase II we are actually comparing c_1 against $\bar{c}$ and not c; this leads to the following four unique scenarios for the Phase II *c*-chart:

(i) $c_1 = c = \bar{c}$: the process is IC in Phase II and c is estimated without error,

(ii) $c_1 \neq c = \bar{c}$: the process is OOC in Phase II and c is estimated without error,

(iii) $c_1 = c \neq \bar{c}$: the process is IC in Phase II and c is *not* estimated without error, and

(iv) $c_1 \neq c \neq \bar{c}$: the process is OOC in Phase II and c is *not* estimated without error.

To simplify matters we assume, without loss of generality, that the process operates IC in Phase II and we assume that $\bar{c}$ is not necessarily equal to c; this is scenario (iii) listed above.

Phase II implementation and operation

The actual operation of the *p*-chart and the *c*-chart in Phase II consists of: (i) taking independent samples and independent inspection units (independent from the Phase I data), (ii) calculating the Phase II sample fractions nonconforming $p_i = X_i / n$ and the numbers of nonconformities in each Phase II inspection unit Y_i for $i = m+1, m+2, \ldots$, and then (iii) comparing these charting statistics (one at a time) to the estimated control limits in (3-26) and (3-27), respectively.

The moment that the first charting statistic plots on or outside the estimated limits a signal is given and the charting procedure stops. The process is then declared out-of-control and we say (in practice) that $p_1 \neq \bar{p}$ (in case of the *p*-chart) or state that $c_1 \neq \bar{c}$ (in case of the *c*-chart).

By comparing the Phase II charting statistics with the estimated control limits, the Phase II characteristics of the charts are (unlike in case K) affected by the variation in the point estimates $\bar{p} = U / mn$ and $\bar{c} = V / m$ where $U \mid p \sim Bin(mn, p)$ and $V \mid c \sim Poi(mc)$ are random variables but the values of m and n can be controlled or decided upon by the user.

The variation in the estimated control limits has significant implications on the properties of the charts. Most importantly the Phase II run-length distributions are no longer geometric since the Phase II signaling events are no longer independent. Intuitively, since estimating the limits introduces extra uncertainty it is expected that the run-length distributions in Case U will be more skewed to the right than the geometric. The additional variation must therefore be accounted for while determining and understanding the chart's properties. We give a systematic examination and detailed derivations of the Phase II run-length distributions of the *p*-chart and *c*-chart in what follows.

Phase II signaling event and Phase II non-signaling event

The event that occurs when a Phase II charting statistic plots outside the estimated control limits is called a Phase II signaling event and denoted by B_i for $i = m+1, m+2, \ldots$. In case of a Phase II signaling event, an alarm or signal is given and we declare the process out-of-control, that is, we say that $p_1 \neq \bar{p}$ or state that $c_1 \neq \bar{c}$. This means, for instance, that in practice we conclude that the probability p_1 of an item being nonconforming in Phase II is not equal to the estimated value $\bar{p}$.

The Phase II non-signaling event is the complementary event of the Phase II signaling event and occurs when a Phase II charting statistic plots within or between the estimated control limits. We denote the Phase II non-signaling event by

$$B_i^C : \{L\hat{C}L < Q_i < U\hat{C}L\}$$

where $Q_i = p_i$ or Y_i for $i = m+1, m+2, \ldots$ and $L\hat{C}L$ and $U\hat{C}L$ are the control limits in either (3-26) or (3-27), respectively.

In case of a Phase II non-signaling event no signal is given and we consider the process in-control, that is, we say that $p_1 = \bar{p}$ or state that $c_1 = \bar{c}$.

Dependency of the Phase II non-signaling events

If the Phase II signaling events were independent, the sequence of trials comparing each Phase II charting statistic Q_i with the estimated limits $U\hat{C}L$ and $L\hat{C}L$, would be a sequence of independent Bernoulli trials. The run-length between occurrences of the signaling event would therefore be a geometric random variable with probability of success equal to $\Pr(B_i)$. Moreover, the average run-length would be $ARL = 1/\Pr(B_i)$.

However, the signaling events B_i and B_j (or, equivalently, the non-signaling events B_i^C and B_j^C) are *not* mutually independent for $i \neq j = m+1, m+2, \ldots$ and the distribution of the run-length between the occurrences of the event B_i is as a result *not* geometric. In particular, because each Phase II p_i (or Y_i) for $i = m+1, m+2, \ldots$ is compared to the same set of estimated control limits, which are random variables, the signaling events are dependent.

To derive exact closed form expressions for the Phase II run-length distributions we use a two-step approach called the "method of conditioning" (see e.g. Chakraborti, (2000)). First we condition on the observed values of the random variables U and V to obtain the *conditional* Phase II run-length distribution and then use the *conditional* Phase II run-length distributions to obtain the marginal or *unconditional* Phase II run-length distributions.

To this end, note that, *given* (or *conditional on or having observed*) particular estimates of $\bar{p}$ and $\bar{c}$ (say $\bar{p}_{\text{obs}}$ and $\bar{c}_{\text{obs}}$), the Phase II non-signaling events *are mutually independent* each with the *same* probability so that the *conditional* Phase II run-length distributions *are* geometric. For instance, for a given or observed value of $\bar{p}$ (say $\bar{p}_{\text{obs}}$), the estimated Phase II control limits of the *p*-chart are constant i.e. they are *not* random variables, so that the conditional Phase II non-signaling events of the *p*-chart

$$\{\bar{p}-3\sqrt{\bar{p}(1-\bar{p})/n} < p_i < \bar{p}+3\sqrt{\bar{p}(1-\bar{p})/n} \mid \bar{p}=\bar{p}_{\text{obs}}\} \quad \text{for} \quad i=m+1,m+2,...$$

are mutually independent each with the *same* probability given by

$$1-\hat{\beta}_p = 1-\Pr(\bar{p}-3\sqrt{\bar{p}(1-\bar{p})/n} < p_i < \bar{p}+3\sqrt{\bar{p}(1-\bar{p})/n} \mid \bar{p}=\bar{p}_{\text{obs}}). \tag{3-28}$$

The same is true for the *c*-chart. That is, for an observed value of $\bar{c}$ (say $\bar{c}_{\text{obs}}$) the events

$$\{\bar{c}-3\sqrt{\bar{c}} < Y_i < \bar{c}+3\sqrt{\bar{c}} \mid \bar{c}=\bar{c}_{\text{obs}}\} \quad \text{for} \quad i=m+1,m+2,...$$

are mutually independent each with the *same* probability given by

$$1-\hat{\beta}_c = 1-\Pr(\bar{c}-3\sqrt{\bar{c}} < Y_i < \bar{c}+3\sqrt{\bar{c}} \mid \bar{c}=\bar{c}_{\text{obs}}). \tag{3-29}$$

The parameters of the *conditional* Phase II (geometric) run-length distributions are the conditional probabilities $1-\hat{\beta}_p$ and $1-\hat{\beta}_c$ so that, symbolically, we write

$$(N \mid \bar{p} = \bar{p}_{\text{obs}}) \sim Geo(1-\hat{\beta}_p) \qquad \text{and} \qquad (N \mid \bar{c} = \bar{c}_{\text{obs}}) \sim Geo(1-\hat{\beta}_c).$$

Thus, once the Phase I reference samples are gathered and the control limits are estimated, the Phase II run-length of a particular chart will follow some *conditional* distribution which will depend on the realization of the random variable $U = u$ or $V = v$, or, alternatively, on the observed values $\bar{p} = \bar{p}_{\text{obs}}$ or $\bar{c} = \bar{c}_{\text{obs}}$.

Note that the distributions of $U \mid p \sim Bin(mn, p)$ and $V \mid c \sim Poi(mc)$, or, equivalently, the distributions of $\bar{p}$ and $\bar{c}$, depend on the values of the unknown parameters p or c (see e.g. Remark 3(ii) as well as expressions (3-24) and (3-25), respectively). It is therefore better to write the conditional run-length distributions as

$$(N \mid \bar{p} = \bar{p}_{\text{obs}}, p) \sim Geo(1-\hat{\beta}_p) \qquad \text{and} \qquad (N \mid \bar{c} = \bar{c}_{\text{obs}}, c) \sim Geo(1-\hat{\beta}_c).$$

Moreover the conditional Phase II run-length distribution therefore provides only hypothetical information about the performance of a control chart with an estimated parameter. We can, for example, only assume some hypothetical value for p or c and then suppose that this estimate of p or c is the 25^{th} or the 75^{th} percentile of the sampling distributions of $\bar{p}$ or $\bar{c}$ so that the run-length distribution, conditioned on such a value, gives some insight into how a chart with this estimate performs in practice. This gives the user an idea of just how poorly or how well a chart will perform in a hypothetical case with an estimated parameter.

To overcome this abovementioned dilemma, the *marginal* or the *unconditional* run-length distribution can give a practitioner insight into a chart's general performance. The marginal distribution incorporates the additional variability which is introduced to the run-length through estimation of p or c by averaging over all possible values of the random variable U or V (while, of course, assuming a particular value for p or c). With the unconditional run-length distribution the practitioner therefore sees the overall effect of estimation on the run-length distribution *before* any data is collected.

3.2.3 Conditional Phase II run-length distributions and characteristics

The conditional run-length distributions and the associated conditional characteristics focus on the performance of the charts given $\bar{p} = p_{\text{obs}}$ and $\bar{c} = c_{\text{obs}}$.

Conditional probability of a no-signal

The probability of a no-signal in Phase II conditional on the point estimate $\bar{p} = p_{\text{obs}}$ or $\bar{c} = c_{\text{obs}}$ is called the conditional probability of a no-signal. This probability, which we previously denoted by $\hat{\beta}_p$ or $\hat{\beta}_c$, is in general denoted by

$$\hat{\beta} = \Pr(B_i^C \mid \hat{\theta}) \quad \text{for} \quad i = m+1, m+2, \ldots$$

where $\hat{\theta} = (\bar{p}, p)$ in case of the *p*-chart and $\hat{\theta} = (\bar{c}, c)$ in case of the *c*-chart.

The conditional probability of a no-signal, like in Case K (see e.g. Tables 3.1 and 3.2), completely characterizes the conditional Phase II run-length distribution and is thus the key to derive and examine the conditional Phase II run-length distributions of Case U. We derive exact expressions for $\hat{\beta}$ for both charts in what follows.

Conditional probability of a no-signal: *p*-chart

This probability is derived by conditioning on an observed value u of the random variable U or, equivalently, conditioning on an observed value $\bar{p}_{obs}$ of the point estimator $\bar{p} = U/mn$ (see e.g. (3-28)).

In doing so, the Phase II charting statistic $p_i = X_i/n$ for $i = m+1, m+2, \ldots$ is the only random variable in (3-28). The cumulative distribution function of p_i for $i = m+1, m+2, \ldots$, as mentioned earlier, is completely known and given by

$$\Pr(p_i \le a) = \Pr(X_i/n \le a) = \Pr(X_i \le na) = \sum_{j=0}^{[na]} \Pr(X_i = j) = \sum_{j=0}^{[na]} \binom{n}{j} p_1^j (1-p_1)^{n-j} \text{ for } 0 \le a \le 1 \text{ and}$$

p_1 denotes the true fraction nonconforming in Phase II (see Remark 4).

We therefore derive the conditional probability of a no-signal by first re-expressing the Phase II conditional non-signaling event in terms of X_i. This is done by making use of the relationship $X_i = np_i$. We then use the properties of X_i to derive an explicit and exact expression for the conditional probability of a no-signal.

For the *p*-chart the conditional probability of a no-signal in Phase II is

$$
\begin{aligned}
&\hat{\beta}(p_1, m, n \mid \bar{p} = \bar{p}_{\text{obs}}, p) \\
&= \Pr(L\hat{C}L_p < p_i < U\hat{C}L_p \mid \bar{p} = \bar{p}_{\text{obs}}, p) \\
&= \Pr(X_i < nU\hat{C}L_p | \bar{p} = \bar{p}_{\text{obs}}, p) - \Pr(X_i \le nL\hat{C}L_p | \bar{p} = \bar{p}_{\text{obs}}, p) \\
&= \Pr(X_i < n\{\bar{p} + 3\sqrt{\bar{p}(1-\bar{p})/n}\} | \bar{p} = \bar{p}_{\text{obs}}, p) - \Pr(X_i \le n\{\bar{p} - 3\sqrt{\bar{p}(1-\bar{p})/n}\} | \bar{p} = \bar{p}_{\text{obs}}, p) \\
&= \Pr(X_i < n\{\frac{U}{mn} + 3\sqrt{\frac{U}{mn}(1-\frac{U}{mn})/n}\} \mid U = u, p) - \Pr(X_i \le n\{\frac{U}{mn} - 3\sqrt{\frac{U}{mn}(1-\frac{U}{mn})/n}\} | U = u, p) \\
&= \Pr(X_i < m^{-1}(U + 3\sqrt{mU - n^{-1}U^2}) \mid U = u, p) - \Pr(X_i \le m^{-1}(U - 3\sqrt{mU - n^{-1}U^2}) \mid U = u, p) \qquad (3\text{-}30) \\
&= \begin{cases} 0 & \text{if } U = 0 \text{ or } U = mn \\ \overline{H}(\hat{b}, p_1, n) & \text{if } nL\hat{C}L_p < 0 \\ \overline{H}(\hat{b}, p_1, n) - \overline{H}(\hat{a}, p_1, n) & \text{if } nL\hat{C}L_p \ge 0 \end{cases} \\
&= \begin{cases} 0 & \text{if } U = 0 \text{ or } U = mn \\ 1 - I_{p_1}(\hat{b}+1, n-\hat{b}) & \text{if } nL\hat{C}L_p < 0 \\ I_{p_1}(\hat{a}+1, n-\hat{a}) - I_{p_1}(\hat{b}+1, n-\hat{b}) & \text{if } nL\hat{C}L_p \ge 0 \end{cases} \\
&= \begin{cases} 0 & \text{if } U = 0 \text{ or } U = mn \\ 1 - I_{p_1}(\hat{b}+1, n-\hat{b}) - 1_{\{nL\hat{C}L_p : nL\hat{C}L_p \ge 0\}}(nL\hat{C}L_p)(1 - I_{p_1}(\hat{a}+1, n-\hat{a})) & \text{if } U = 1,2,\ldots,mn-1 \end{cases}
\end{aligned}
$$

for $0 < p, \bar{p}, p_1 < 1$, where

$$\overline{H}(\hat{b}, p_1, n) = \sum_{j=0}^{\hat{b}} \binom{n}{j} p_1^j (1-p_1)^{n-j}$$

denotes the c.d.f of the $Bin(n, p_1)$ distribution and

$$\hat{a} = \hat{a}(m, n \mid U, p) = [nL\hat{C}L_p \,] \qquad (3\text{-}31a)$$

and

$$\hat{b} = \hat{b}(m, n \mid U, p) = \begin{cases} \min\{nU\hat{C}L_p - 1, n\} & \text{if } nU\hat{C}L \text{ is an integer} \\ \min\{[nU\hat{C}L_p], n\} & \text{if } nU\hat{C}L \text{ is not an integer.} \end{cases} \qquad (3\text{-}31b)$$

Remark 5

(i) The conditional probability of a no-signal for the *p*-chart is a function of and depends on

a. the fraction nonconforming in Phase II p_1,
b. the number of reference samples m,
c. the sample size n,
d. the point estimator $\bar{p}$ or, equivalently, the random variable U, and
e. the unknown true fraction nonconforming p; indirectly via the random variable $U \mid p \sim Bin(mn, p)$.

As noted earlier in Remark 4(i), p_1 is not necessarily equal to p and because of sampling variability $\bar{p}$ is typically different from p.

(ii) When none of the Phase I reference sample observations are nonconforming, that is, when $U = 0$ or $\bar{p} = 0$, it makes sense not to continue to Phase II but examine the situation in more detail. Similar logic applies to the other extreme, that is when all the observations are nonconforming so that $U = mn$ or $\bar{p} = 1$.

Based on this intuitive reasoning the conditional probability of a no-signal $\hat{\beta}(p_1, m, n \mid \bar{p}, p)$ is defined to be zero in both of these boundary situations. It then follows that the conditional probability of a signal $1 - \hat{\beta}(p_1, m, n \mid \bar{p}, p)$ is one. Effectively the control chart signals, in these cases, when p_i for $i = m+1, m+2, \ldots$ plots on or beyond either of the two estimated control limits or is equal to either 0 or n; this, in actual fact, implies that the *p*-chart signals on the first Phase II sample.

Conditional probability of a no-signal: *c*-chart

By conditioning on an observed value v of the random variable V or, equivalently, conditioning on an observed value $\bar{c}_{\text{obs}}$ of the point estimator $\bar{c} = V/m$, the Phase II charting statistic Y_i for $i = m+1, m+2, \ldots$ is the only random quantity (variable) in (3-29).

Because the distribution of Y_i is known (assumed) to be Poisson with parameter c_1, we use the properties of this distribution to derive an explicit and exact expression for the conditional probability of a no-signal for the *c*-chart.

The conditional probability of a no-signal in Phase II is

$$\begin{aligned}
&\hat{\beta}(c_1, m \mid \bar{c} = \bar{c}_{\text{obs}}, c) \\
&= \Pr(L\hat{C}L_c < Y_i < U\hat{C}L_c \mid \bar{c} = \bar{c}_{\text{obs}}, c) \\
&= \Pr(Y_i < U\hat{C}L_c \mid \bar{c} = \bar{c}_{\text{obs}}, c) - \Pr(Y_i \le L\hat{C}L_c \mid \bar{c} = \bar{c}_{\text{obs}}, c) \\
&= \Pr(Y_i < \bar{c} + 3\sqrt{\bar{c}} \mid \bar{c} = \bar{c}_{\text{obs}}, c) - \Pr(Y_i \le \bar{c} - 3\sqrt{\bar{c}} \mid \bar{c} = \bar{c}_{\text{obs}}, c) \\
&= \Pr\left(Y_i < \frac{V}{m} + 3\sqrt{\frac{V}{m}} \mid V = v, c\right) - \Pr\left(Y_i \le \frac{V}{m} - 3\sqrt{\frac{V}{m}} \mid V = v, c\right) \qquad \text{(3-32)} \\
&= \begin{cases} 0 & \text{if} \quad V = 0 \\ \bar{G}(\hat{f}; c_1) - \bar{G}(\hat{d}; c_1) & \text{if} \quad V = 1,2,3,\ldots \end{cases} \\
&= \begin{cases} 0 & \text{if} \quad V = 0 \\ \Gamma_{\hat{f}+1}(c_1) - \Gamma_{\hat{d}+1}(c_1) & \text{if} \quad V = 1,2,3,\ldots \end{cases}
\end{aligned}$$

for $c, \bar{c}, c_1 > 0$, where

$$\bar{G}(\hat{f}; c_1) = \sum_{j=0}^{\hat{f}} \frac{e^{-c_1} c_1^j}{j!}$$

denotes the c.d.f of the $Poi(c_1)$ distribution and

$$\hat{d} = \hat{d}(m \mid V, c) = \max\{0, [L\hat{C}L_c]\} \qquad \text{(3-33a)}$$

and

$$\hat{f} = \hat{f}(m \mid V, c) = \begin{cases} U\hat{C}L_c - 1 & \text{if } U\hat{C}L_c \text{ is an integer} \\ [U\hat{C}L_c] & \text{if } U\hat{C}L_c \text{ is not an integer.} \end{cases} \qquad \text{(3-33b)}$$

Remark 6

(i) The probability of a no-signal for the *c*-chart is a function of and depends on

a. the average number of nonconformities in an inspection unit in Phase II c_1,
b. the number of reference inspection units m from Phase I,
c. the point estimator $\bar{c}$ or, equivalently, the random variable V, and
d. the unknown true average number of nonconformities in an inspection unit c; indirectly via the random variable $V \mid c \sim Poi(mc)$.

Again, note that, c_1 is not necessarily equal to c, and since $\bar{c}$ is subject to sampling variation it is typically different from c.

(ii) When we observe no nonconformities in the Phase I reference sample i.e. when $V = 0$ or $\bar{c} = 0$, it is essential to pause and examine the situation in more detail. Thus, for $V = 0$ the conditional probability of a no-signal in Phase II is defined to be zero so that the conditional probability of a signal in Phase II is one.

Summary of the conditional run-length distributions and the related conditional characteristics

Given observed values u and v of the random variables U and V, the conditional run-length distributions of the charts are geometric with the probability of success equal to the conditional probability of a signal i.e.

$$1-\hat{\beta}(p_1, m, n \mid U = u, p) \qquad \text{and} \qquad 1-\hat{\beta}(c_1, m \mid V = v, c)$$

respectively.

This is so, because for given or fixed values of $U = u$ and $V = v$ the control limits can be calculated exactly and the analyses continue as if the parameters p and c are known. This is similar to the standards known case (Case K) where the run-length distribution was seen to be geometric. All the characteristics of the conditional run-length distributions therefore follow from the well-known properties of the geometric distribution. In particular, the conditional run-length distributions and the associated conditional characteristics for the *p*-chart and the *c*-chart are summarized in Table 3.3 and Table 3.4, respectively.

The conditional run-length distribution and the conditional characteristics of the run-length distributions all depend on either the observed value of the random variable U or that of V; these observed values cannot be controlled by the user and is a direct result of estimating p and c. Thus, as the values of U and V change (randomly), the conditional run-length distributions and the conditional characteristics of the run-length distributions will also change randomly. This implies, for example, that the conditional characteristics are random variables which all have their own probability distributions so that one can present a quantity such as the expected conditional *SDRL* i.e. $E_U(CSDRL(p_1, m, n \mid U, p)$ or $E_V(CSDRL(c_1, m \mid V, c)$. Although this is technically correct it is not the best approach; a better approach would be to calculate the unconditional standard deviation i.e.

$$USDRL = \sqrt{E_U(\text{var}(p_1, m, n \mid U, p)) + \text{var}_U(E(p_1, m, n \mid U, p))}$$

or

$$USDRL = \sqrt{E_V(\text{var}(c_1, m \mid V, c)) + \text{var}_V(E(c_1, m \mid V, c))}$$

which is computed from the marginal run-length distribution and incorporates both the expected conditional *SDRL* and the variation in the expected conditional *ARL*. We discuss this in more detail later when we examine the conditional and unconditional properties of the charts.

Table 3.3: The conditional probability mass function (c.p.m.f), the conditional cumulative distribution function (c.c.d.f), the conditional false alarm rate (*CFAR*), the conditional average run-length (*CARL*) and the conditional standard deviation of the run-length (*CSDRL*) of the *p*-chart in Phase II of Case U

c.p.m.f	$\Pr(N_p = j; p_1, m, n \mid U, p) = [\hat{\beta}(p_1, m, n \mid U, p)]^{j-1}[1 - \hat{\beta}(p_1, m, n \mid U, p)] \quad j = 1,2,....$	(3-34)
c.c.d.f	$\Pr(N_p \le j; p_1, m, n \mid U, p) = 1 - [\hat{\beta}(p_1, m, n \mid U, p)]^j \quad j = 1,2,....$	(3-35)
CFAR	$CFAR(p_1, m, n \mid U, p = p_1) = 1 - \hat{\beta}(p_1, m, n \mid U, p = p_1)$	(3-36)
CARL	$CARL(p_1, m, n \mid U, p) = 1/[1 - \hat{\beta}(p_1, m, n \mid U, p)]$	(3-37)
CSDRL	$CSDRL(p_1, m, n \mid U, p) = \sqrt{\hat{\beta}(p_1, m, n \mid U, p)}\,/[1 - \hat{\beta}(p_1, m, n \mid U, p)]$	(3-38)
cqf	$Q_{N_p}(q; p_1, m, n \mid U, p) = \inf\{\text{int}\, x : \Pr(N_p \le j; p_1, m, n \mid U, p) \ge q\} \quad 0 < q < 1$	(3-39)

Table 3.4: The conditional probability mass function (c.p.m.f), the conditional cumulative distribution function (c.c.d.f), the conditional false alarm rate (*CFAR*), the conditional average run-length (*CARL*) and the conditional standard deviation of the run-length (*CSDRL*) of the *c*-chart in Phase II of Case U

c.p.m.f	$\Pr(N_c = j; c_1, m \mid V, c) = [\hat{\beta}(c_1, m \mid V, c)]^{j-1}[1 - \hat{\beta}(c_1, m \mid V, c)] \quad j = 1,2,....$	(3-40)
c.c.d.f	$\Pr(N_c \le j; c_1, m \mid V, c) = 1 - [\hat{\beta}(c_1, m \mid V, c)]^j \quad j = 1,2,....$	(3-41)
CFAR	$CFAR(c_1, m \mid V, c = c_1) = 1 - \hat{\beta}(c_1, m \mid V, c = c_1)$	(3-42)
CARL	$CARL(c_1, m \mid V, c) = 1/[1 - \hat{\beta}(c_1, m \mid V, c)]$	(3-43)
CSDRL	$CSDRL(c_1, m \mid V, c) = \sqrt{\hat{\beta}(c_1, m \mid V, c)}\,/[1 - \hat{\beta}(c_1, m \mid V, c)]$	(3-44)
cqf	$Q_{N_c}(q; c_1, m \mid V) = \inf\{\text{int}\, x : \Pr(N_c \le j; c_1, m \mid V, c) \ge q\} \quad 0 < q < 1$	(3-45)

It is important to note that the conditional run-length distributions and the associated characteristics of the conditional run-length distributions do not only depend on the random variables U and V; they also indirectly depend on the unknown parameters p and c.

The dependency on U and V follows from the fact that we estimate p using $\bar{p} = U/mn$ and we estimate c using $\bar{c} = V/m$. The indirect dependency on p and c follows from the fact that the distribution of U (which is binomial with parameters mn and p) and the distribution of V (which is Poisson with parameter mc) depend on the unknown parameters p and c. To evaluate any of the conditional characteristics we need the observed values of U and V but we also need to assume values for p and c.

The aforementioned point is demonstrated in the following two examples which illustrate the operation and the implementation of the Phase II *p*-chart and the Phase II *c*-chart when we are given a particular Phase I sample.

Example 1: A Phase II *p*-chart

Consider Example 6.1 on p. 289 of Montgomery (2001) concerning a frozen orange juice concentrate that is packed in 6-oz cardboard cans. A machine is used to make the cans and the goal is to set up a control chart to improve i.e. decrease, the fraction of nonconforming cans produced by the machine. Since no specific value of the fraction nonconforming p is given the scenario is an example of Case U, that is, when the standard is unknown. The chart is therefore implemented in two stages.

Phase I

To establish the control chart $m = 30$ reference samples were taken each with $n = 50$ cans, selected in half hour intervals over a three-shift period in which the machine was in continuous operation. Once the Phase I control chart was established samples 15 and 23 were found to be out-of-control and eliminated after further investigation. Revised control limits were calculated using the remaining $m = 28$ samples. Based on the revised control limits sample 21 was found out-of-control, but since further investigations regarding sample 21 did not produce any reasonable or logical assignable cause it was not discarded. This is the retrospective phase (or Phase I) of the analysis. The final 28 samples were used to estimate the control limits and then monitor the process in Phase II.

Phase II (conditional)

Although the random variable U could theoretically take on any integer value from 0 to $mn = 28 \times 50 = 1400$, for the given set of reference data it was found that $U = 301$; this was the total number of nonconforming cans after discarding samples 15 and 23. It follows from (3-24) that the point estimate of p is $\bar{p} = 301/1400 = 0.215$.

The estimated control limits and centerline corresponding to $U = 301$ are found from (3-26) to be

$$U\hat{C}L_p = 0.215 + 3\sqrt{0.215(0.785)/50} = 0.3893 \text{ and } L\hat{C}L_p = 0.215 - 3\sqrt{0.215(0.785)/50} = 0.0407.$$

We find the constants $\hat{a}$ and $\hat{b}$ using (3-31) to be

$$\hat{b}(m = 28, n = 50 \,|\, U = 301, p) = 19 \quad \text{and} \quad \hat{a}(m = 28, n = 50 \,|\, U = 301, p) = 2.$$

Because U is unequal to 0 or mn it follows from (3-30) that the conditional probability of a no-signal in Phase II is

$$\hat{\beta}(p_1, m=28, n=50 \mid \bar{p}=0.215, p) = 1 - I_{p_1}(19, 50-19-1) - (1 - I_{p_1}(2, 50-2-1))$$
$$= I_{p_1}(2,47) - I_{p_1}(19,30)$$

for $0 < p, p_1 < 1$.

Assuming, without loss of generality, that the process is in-control at a fraction nonconforming of 0.2, that is, $p_1 = p = 0.2$, the conditional false alarm rate (*CFAR*) is equal to

$$1 - \hat{\beta}(p_1 = 0.2, 28, 50 \mid \bar{p}=0.215, p=0.2) = 1 - I_{0.2}(2,47) + I_{0.2}(19,30) = 0.002218.$$

The in-control conditional average run-length therefore equals

$$CARL_0 = 1/0.002218 = 450.89$$

and is found using (3-37).

Compared to the Case K *FAR* and *ARL* of 0.0027 and 369.84 (see e.g. Tables A3.4 and A3.5 of Appendix 3A) we see that our *p*-chart (here, in Case U, with $\bar{p} = 0.215$ and assuming that $p_1 = p = 0.2$) would signal less often, if the process is in-control, than what it would if p had in fact been known to be equal to 0.2.

However, note that, since each user has his/her own unique reference sample, the point estimate $\bar{p}$ will differ from one user to the next so that the performance of each user's chart will also vary. To this end, the unconditional characteristics are useful as they do not depend on any specific observed value of the point estimate. This, however, is looked at later when we continue Example 1 after having derived expressions for the unconditional characteristics of the *p*-chart's Phase II run-length distribution. ■

Example 2: A Phase II *c*-chart

Consider Example 6.3 on p. 310 in Montgomery (2001) about the quality control of manufactured printed circuit boards. Since c is not specified it had to be estimated. The chart was therefore implemented in two phases.

Phase I

A total of 26 successive inspection units each consisting of 100 individual items of product were obtained to estimate the unknown true average number of nonconformities in an inspection unit *c*. It was found that units number 6 and 20 were out-of-control and therefore eliminated. The revised control limits were calculated using the remaining $m = 24$ inspection units with the number of nonconformities in an inspection unit shown in Table 6.7 on p. 311 of Montgomery (2001). The revised control limits were used for monitoring the process in Phase II.

Phase II (conditional)

Theoretically the variable V, the total number of nonconformities in the 24 inspection units, could take on any positive integer value including zero i.e. $V \in \{0,1,2,...\}$. For the given Phase I data it is found that $V = 472$. Using (3-25) the average number of nonconformities in an inspection unit c is estimated as $\bar{c} = 472/24 = 19.67$ so that the estimated 3-sigma control limits are found from (3-27) to be

$$U\hat{C}L_c = 32.97 \quad \text{and} \quad L\hat{C}L_c = 6.36.$$

These estimated limits yield

$$\hat{d}(m = 24 \mid V = 472, c) = 6 \quad \text{and} \quad \hat{f}(m = 24 \mid V = 472, c) = 32.$$

Because V is unequal to zero it follows from (3-32) that the probability of a no-signal is

$$\hat{\beta}(c_1, 24 \mid \bar{c} = 19.67, c) = \Gamma_{33}(c_1) - \Gamma_7(c_1) \quad \text{for} \quad c, c_1 > 0.$$

For the given (observed) value of $V = 472$ one can investigate the chart's performance using the conditional properties. Assuming, without loss of generality, that the process operates in-control at an average of twenty nonconformities in an inspection unit, that is, $c_1 = c = 20$ is the true in-control average number of nonconformities in an inspection unit, the conditional false alarm rate i.e. the false alarm rate given $V = 472$, is found to be equal to

$$CFAR = 1 - \Gamma_{33}(20) - \Gamma_{7}(20) = 0.004983 .$$

The *CFAR* is approximately 72% larger than the value of 0.0029 one would have obtained in Case K for $c_0 = 20$ and is 85% higher than the nominal value 0.0027 (see e.g. Table A3.12 in Appendix 3A); this is true even though the estimated average number of nonconformities in an inspection unit ($\bar{c} = 19.67$) is within $|(19.67 - 20)/\sqrt{20}| = 0.07$ standard deviation units of the true average number of nonconformities in an inspection unit ($c = 20$). However, note that, like the *p*-chart of Example 1, each user typically has his/her own distinct Phase I data so that the performance of the *c*-chart in Case U will be different for each user. ■

To get an overall picture of a *p*-chart's or a *c*-chart's performance one needs to look at the unconditional properties of the chart; this is looked at later. First we look at the conditional run-length distribution and the related conditional characteristics of the *p*-chart and *c*-chart.

The characteristics of the conditional run-length distribution depend on and are functions of the random variables U or V; as a result, these characteristics are random variables themselves and vary as U or V changes.

To understand the effect of U or V on the characteristics of the conditional run-length distribution, it is instructive to study the conditional characteristics of the charts as functions of U and V as they show precisely how the conditional characteristics of each chart vary as the point estimates $\bar{p}$ and $\bar{c}$ fluctuate.

First we look at the conditional characteristics of the *p*-chart and then at those of the *c*-chart.

3.2.3.1 Conditional characteristics of the *p*-chart

Once we observed a value u of the random variable U we can calculate the conditional probability of a signal. The Phase II conditional run-length distribution is then completely known (see e.g. Table 3.3).

Tables 3.5 and 3.6 illustrate the exact steps to calculate the conditional probability of a no-signal, the conditional probability of a signal or the conditional false alarm rate (*CFAR*), the conditional average run-length (*CARL*) and the conditional standard deviation of the run-length (*CSDRL*) for the *p*-chart. These are all conditional Phase II properties as they all depend on an observed value from Phase I.

For illustration purposes we assume a total of $T = mn = 20$ individual Phase I observations is used to estimate p using $\bar{p} = U/mn$ as point estimate and that $p_1 = p = 0.5$. The latter assumption implies that the process operated at a fraction nonconforming of $p = 0.5$ during Phase I and that in Phase II the process continues to operate at this same level so that $p_1 = 0.5$; this is the same as saying that the process is in-control in Phase II. However, note that, because of sampling variation the observed value of $\bar{p}$ may of course not be equal to p (see e.g. Remark 4(i)).

The calculations of Table 3.5 are based on the assumption that $m = 4$ independent Phase I reference samples each of size $n = 5$ are used whereas the computations of Table 3.6 are based on $m = 1$ with $n = 20$.

In particular, column 1 lists all the values of U (the total number of possible nonconforming items in the entire Phase I reference sample) that can possibly be attained. This ranges from a minimum of zero to a maximum of twenty. Column 2 converts the observed value u of U into a point estimate of the unknown true fraction of nonconforming items, that is, we calculate $\bar{p} = u/20 = \bar{p}_{\text{obs}}$ which estimates p. Because each row entry in each of the succeeding columns (i.e. columns 3 to 12) is computed by conditioning on a row entry from column 1 (or, equivalently, from column 2) we start calculating the conditional properties in columns 1 and/or 2 and sequentially proceed to the right-hand side of the tables. Thus, given a value u or $\bar{p}_{\text{obs}}$ the lower and the upper control limits are estimated in columns 3 and 4 using (3-26). These estimated limits are then used to compute the two constants $\hat{a}$ and $\hat{b}$ defined in (3-31), which are shown in columns 5 and 6, respectively. Finally, columns 7

through 10 list the probability of a no-signal, the *FAR*, the in-control *ARL* and the in-control *SDRL* given the observed value u from column 1, respectively. These properties are labeled $\Pr(\text{No Signal} \mid U, p)$, *CFAR*, $CARL_0$ and $CSDRL_0$, and calculated using (3-30) and the expressions in Table 3.3. Columns 11 and 12 show the values of the probability mass function (p.m.f) and the cumulative distribution function (c.d.f) of the random variable $U \mid p = 0.5 \sim Bin(20,0.5)$, that is,

$$\Pr(U = u \mid p = 0.5) = \binom{20}{u} 0.5^{20} \quad \text{and} \quad \Pr(U \le u \mid p = 0.5) = \sum_{j=0}^{u} \binom{20}{j} 0.5^{20} \quad \text{for} \quad u = 0,1,2,...,20.$$

Both these probability functions are useful when interpreting the characteristics of the conditional run-length distribution. The former shows the exact probability of obtaining a particular value u of U whereas the latter can be used to find the percentiles of the distribution of U.

T = 20 with m = 4 and n = 5

Consider Table 3.5 which uses a total of $T = 20$ individual in-control Phase I reference observations from $m = 4$ independent samples each of size $n = 5$.

There are two unique scenarios. The first takes place when $U = 0$ (the minimum value possible) and the second occurs when $U = 4 \times 5 = 20$ (the maximum value). In both these cases the probability of a no-signal is zero by definition and the chart signals once the first Phase II sample is observed. As a result the conditional in-control average run-length is $CARL_0 = 1$. In the former situation the estimated control limits are $L\hat{C}L_p = U\hat{C}L_p = 0$ and in the latter the limits are $L\hat{C}L_p = U\hat{C}L_p = 1$. In both these situations the constants $\hat{a}$ and $\hat{b}$ need not be calculated; this is indicated by NA (read as "not applicable") in columns 5 and 6, respectively (see e.g. (3-30) and Remark 5(ii)).

The probability that none or all of the Phase I reference observations are nonconforming is of course rather small. The probabilities of these two events are $P(U = 0 \mid 0.5) = P(U = 20 \mid 0.5) = 0.5^{20}$ which are zero when rounded to four decimal places (see e.g. column 11). For all other values of $U \neq 0$ and $U \neq mn = 20$, that is, when $U \in \{1,2,...,19\}$, we proceed with the calculation of the conditional characteristics as follows.

Table 3.5: Conditional probability of a no-signal, the conditional false alarm rate (*CFAR*), the in-control conditional average run-length ($CARL_0$) and the in-control conditional standard deviation of the run-length ($CSDRL_0$) of the *p*-chart in Case U for $m = 4$ and $n = 5$, assuming that $p_1 = p = 0.5$

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
u	$\bar{p}_{obs}$	$L\hat{C}L_p$	$U\hat{C}L_p$	$\hat{a}$	$\hat{b}$	Pr(No Signal | *U*, *p*)	*CFAR*	$CARL_0$	$CSDRL_0$	Pr(*U*=*u*| *p*)	Pr(*U*<=*u*| *p*)
0	0.00	0.00	0.00	NA	NA	0.0000	1.0000	1.00	0.00	0.0000	0.0000
1	0.05	-0.24	0.34	NA	1	0.1875	0.8125	1.23	0.53	0.0000	0.0000
2	0.10	-0.30	0.50	NA	2	0.5000	0.5000	2.00	1.41	0.0002	0.0002
3	0.15	-0.33	0.63	NA	3	0.8125	0.1875	5.33	4.81	0.0011	0.0013
4	0.20	-0.34	0.74	NA	3	0.8125	0.1875	5.33	4.81	0.0046	0.0059
5	0.25	-0.33	0.83	NA	4	0.9688	0.0313	32.00	31.50	0.0148	0.0207
6	0.30	-0.31	0.91	NA	4	0.9688	0.0313	32.00	31.50	0.0370	0.0577
7	0.35	-0.29	0.99	NA	4	0.9688	0.0313	32.00	31.50	0.0739	0.1316
8	0.40	-0.26	1.06	NA	5	1.0000	0.0000	∞	∞	0.1201	0.2517
9	0.45	-0.22	1.12	NA	5	1.0000	0.0000	∞	∞	0.1602	0.4119
10	0.50	-0.17	1.17	NA	5	1.0000	0.0000	∞	∞	0.1762	0.5881
11	0.55	-0.12	1.22	NA	5	1.0000	0.0000	∞	∞	0.1602	0.7483
12	0.60	-0.06	1.26	NA	5	1.0000	0.0000	∞	∞	0.1201	0.8684
13	0.65	0.01	1.29	0	5	0.9688	0.0313	32.00	31.50	0.0739	0.9423
14	0.70	0.09	1.31	0	5	0.9688	0.0313	32.00	31.50	0.0370	0.9793
15	0.75	0.17	1.33	0	5	0.9688	0.0313	32.00	31.50	0.0148	0.9941
16	0.80	0.26	1.34	1	5	0.8125	0.1875	5.33	4.81	0.0046	0.9987
17	0.85	0.37	1.33	1	5	0.8125	0.1875	5.33	4.81	0.0011	0.9998
18	0.90	0.50	1.30	2	5	0.5000	0.5000	2.00	1.41	0.0002	1.0000
19	0.95	0.66	1.24	3	5	0.1875	0.8125	1.23	0.53	0.0000	1.0000
20	1.00	1.00	1.00	NA	NA	0.0000	1.0000	1.00	0.00	0.0000	1.0000

Suppose, for instance, that we observe seven nonconforming items out of the possible twenty in the entire Phase I reference sample. Our chance to find exactly seven nonconforming items is approximately 0.0739 (which is relatively high, see e.g. column 11); the probability to find less than seven nonconforming items is $P(U < 7 \mid 0.5) \approx 0.0577$ (see e.g. column 12).

A value of $U = 7$ gives a point estimate for p of $\bar{p} = 7/20 = 0.35$ so that (3-26) yields an estimated upper control limit and an estimated lower control limit of

$$U\hat{C}L_p = 0.35 + 3\sqrt{0.35(0.65)/5} = 0.99 \quad \text{and} \quad L\hat{C}L_p = 0.35 - 3\sqrt{0.35(0.65)/5} = -0.29$$

respectively .

Because $nL\hat{C}L_p = (5)(-0.29) = -1.45$ is less than zero the chart has no lower control limit. We therefore do not calculate a value for $\hat{a}$ in this case. The constant $\hat{b}$, on the other hand, is found to be

$$\hat{b} = \min\{[nU\hat{C}L_p],5\} = \min\{[(5)(0.99)],5\} = \min\{[4.95],5\} = 4.$$

Finally, after substituting $\hat{b}$ in (3-30) we calculate the conditional probability of a no-signal and then also the *CFAR*, the $CARL_0$ and the $CSDRL_0$ using expressions (3-36), (3-37) and (3-38) in Table 3.3.

The conditional probability of a no-signal is

$$\hat{\beta}(p_1 = 0.5, m = 4, n = 5 \mid U = 7, p = 0.5) = \hat{\beta}(p_1 = 0.5, m = 4, n = 5 \mid \bar{p} = 0.35, p = 0.5) = 0.9688,$$

so that the conditional false alarm rate is

$$CFAR(p_1 = 0.5,4,5 \mid U = 7, p = 0.5) = 1 - 0.9688 = 0.0313.$$

The Phase II p-chart then has an in-control conditional *ARL* of

$$CARL_0(p_1 = 0.5,4,5 \mid U = 7, p = 0.5) = 1/0.0313 = 32.00$$

and an in-control conditional *SDRL* of

$$SDRL_0(p_1 = 0.5,4,5 \mid U = 7, p = 0.5) = \sqrt{0.9688}/0.0313 = 31.50.$$

If the process remains to operate at $p_1 = 0.5$ (i.e. the process stays in-control) we expect that the chart would, on average, give a false alarm or erroneous signal on every 32nd sample. This is more often that what we would nominally expect from a 3-sigma Shewhart-type control chart, which typically has an in-control *ARL* of 370.4. We also see that the conditional false alarm rate (*CFAR*), particularly for $U = 7$, is much higher than the nominally expected 0.0027 even though the point estimate $\bar{p} = 0.35$ is $\left|\sqrt{5}(0.35 - 0.50)/\sqrt{0.35(1-0.35)}\right| = 0.70$ standard deviation units from the supposedly known value of $p = 0.5$.

For values of U from 8 to 12 the *CFAR* is equal to zero and as a result the moments of the run-length distribution, such as the $CARL_0$ and the $CSDRL_0$, are all undefined; this implies that, in practice, the conditional Phase II chart will not signal and that the $CARL_0$ and the $CSDRL_0$ are both infinite. Although we typically want a high in-control *ARL*, an *ARL* of infinity is not practical. Thus, $m = 4$ subgroups each of size $n = 5$ is not adequate to control the false alarm rate (*FAR*) at a small yet practically desirable level, and at the same time ensure that a high in-control *ARL* is achieved. This suggests that one needs more reference data and that n needs to be larger relative to m in order to achieve any reasonable probability of a false alarm with attributes data.

T = 20 with m = 1 and n = 20

To study the effect of choosing a larger value of n relative to m suppose that a total of $T = 20$ in-control Phase I reference observations are available but in one sample of twenty observations, that is, $m = 1$ and $n = 20$. Calculations for this situation are shown in Table 3.6.

We observe that the conditional probability of a no-signal i.e.

$$\Pr(\text{No Signal} \mid U, p) = \hat{\beta}(p_1 = 0.5, m = 1, n = 20 \mid U = u, p = 0.5)$$

is non-zero for all values of $U = 0,1,...,20$. As a result none of the *CFAR*'s values are zero and therefore all the moments (such as the in-control *ARL*, the in-control *SDRL* etc.) of the conditional run-length distribution are defined and finite. This suggests the need for a very careful choice of the number of reference samples m and the size n of each of the samples before a *p*-chart with an unknown value of p is implemented in practice.

Table 3.6: Conditional probability of a no-signal, the conditional false alarm rate (*CFAR*), the in-control conditional average run-length ($CARL_0$) and the in-control conditional standard deviation of the run-length ($CSDRL_0$) of the *p*-chart in Case U for $m = 1$ and $n = 20$, assuming that $p_1 = p = 0.5$

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
u	$\bar{p}_{obs}$	$L\hat{C}L_p$	$U\hat{C}L_p$	$\hat{a}$	$\hat{b}$	Pr(No Signal | *U*, *p*)	*CFAR*	$CARL_0$	$CSDRL_0$	Pr(*U*=*u*| *p*)	Pr(*U*<=*u*| *p*)
0	0.00	0.00	0.00	NA	NA	0.0000	1.0000	1.00	0.00	0.0000	0.0000
1	0.05	-0.10	0.20	NA	3	0.0013	0.9987	1.00	0.04	0.0000	0.0000
2	0.10	-0.10	0.30	NA	6	0.0577	0.9423	1.06	0.25	0.0002	0.0002
3	0.15	-0.09	0.39	NA	7	0.1316	0.8684	1.15	0.42	0.0011	0.0013
4	0.20	-0.07	0.47	NA	9	0.4119	0.5881	1.70	1.09	0.0046	0.0059
5	0.25	-0.04	0.54	NA	10	0.5881	0.4119	2.43	1.86	0.0148	0.0207
6	0.30	-0.01	0.61	NA	12	0.8684	0.1316	7.60	7.08	0.0370	0.0577
7	0.35	0.03	0.67	0	13	0.9423	0.0577	17.34	16.84	0.0739	0.1316
8	0.40	0.07	0.73	1	14	0.9793	0.0207	48.27	47.77	0.1201	0.2517
9	0.45	0.12	0.78	2	15	0.9939	0.0061	163.66	163.16	0.1602	0.4119
10	0.50	0.16	0.84	3	16	0.9974	0.0026	388.07	387.57	0.1762	0.5881
11	0.55	0.22	0.88	4	17	0.9939	0.0061	163.66	163.16	0.1602	0.7483
12	0.60	0.27	0.93	5	18	0.9793	0.0207	48.27	47.77	0.1201	0.8684
13	0.65	0.33	0.97	6	19	0.9423	0.0577	17.34	16.84	0.0739	0.9423
14	0.70	0.39	1.01	7	20	0.8684	0.1316	7.60	7.08	0.0370	0.9793
15	0.75	0.46	1.04	9	20	0.5881	0.4119	2.43	1.86	0.0148	0.9941
16	0.80	0.53	1.07	10	20	0.4119	0.5881	1.70	1.09	0.0046	0.9987
17	0.85	0.61	1.09	12	20	0.1316	0.8684	1.15	0.42	0.0011	0.9998
18	0.90	0.70	1.10	13	20	0.0577	0.9423	1.06	0.25	0.0002	1.0000
19	0.95	0.80	1.10	16	20	0.0013	0.9987	1.00	0.04	0.0000	1.0000
20	1.00	1.00	1.00	NA	NA	0.0000	1.0000	1.00	0.00	0.0000	1.0000

The conditional false alarm rate

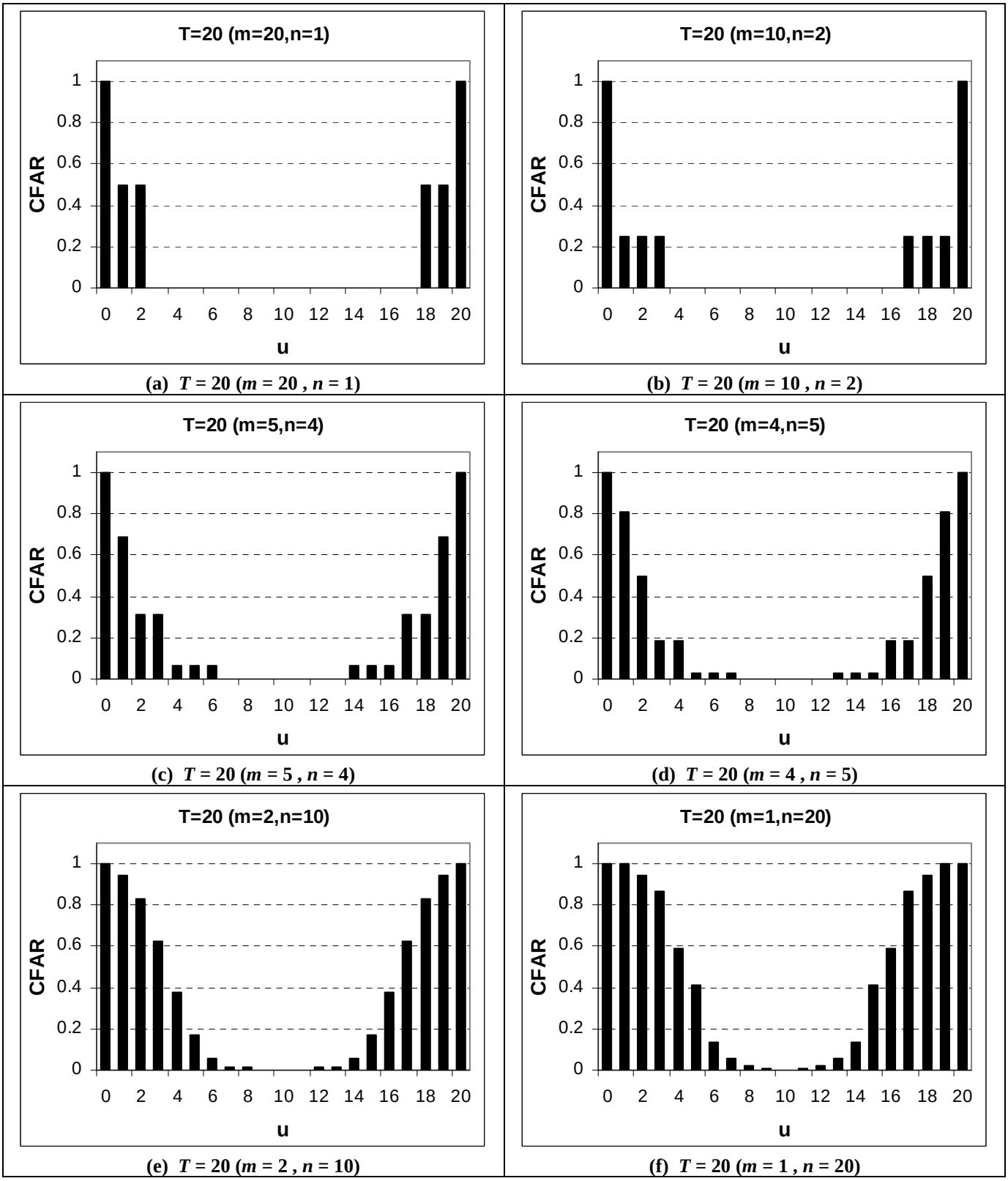
Panels (a) to (f) of Figures 3.1 and 3.2 display the conditional false alarm rate (*CFAR*)

$$1-\hat{\beta}(p_1 = 0.5, m, n \mid U = u, p = 0.5) \quad \text{as a function of} \quad u = 0,1,\ldots,mn$$

for various combinations of *m* and *n* when a total of $T = 20$ and a total of $T = 50$ individual Phase I reference observations are used to estimate p. For illustration purposes we assume that $p_1 = p = 0.50$.

The impact of the actual number of nonconforming items u in the entire Phase I reference sample is easily noticed. The distribution of the *CFAR* is seen to be U-shaped and symmetric at the point $mn/2$; this is the mean value of *U*. For values of *U* near the two tails the *CFAR* can be very high, sometimes close to 1 or 100%, which obviously means many false alarms. Of course, this only happens at the rather extreme values of *U* that occur with very small probabilities (see e.g. columns 11 and 12 in Tables 3.5 and 3.6). However, even when U is not as extreme there can be a significantly high probability of a false alarm and it is seen that only when U takes on a value in the neighbourhood of its mean, will the *CFAR* be reasonably small. A potential problem is that for some combinations of *m* and *n* values, especially with smaller values of *n* relative to m, some of the *CFAR* values equal 0, which (as mentioned before) leads to an in-control average run-length that is undefined.

Note that, panels (d) and (f) of Figure 3.1 are in fact displaying the *CFAR*'s of column 8 in Tables 3.5 and 3.6, respectively.



(a) $T = 20$ ($m = 20$, $n = 1$)

(b) $T = 20$ ($m = 10$, $n = 2$)

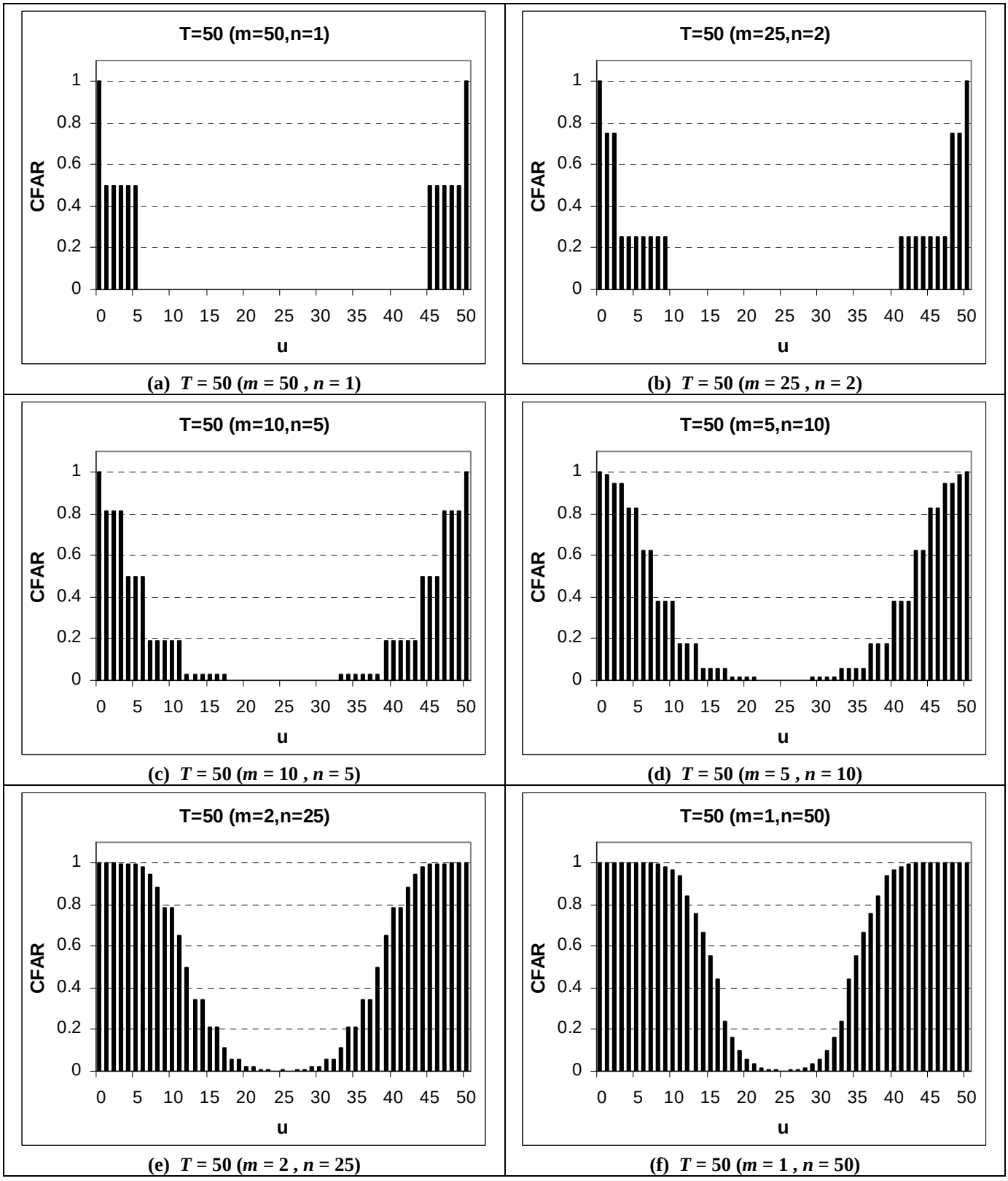
(c) $T = 20$ ($m = 5$, $n = 4$)

(d) $T = 20$ ($m = 4$, $n = 5$)

(e) $T = 20$ ($m = 2$, $n = 10$)

(f) $T = 20$ ($m = 1$, $n = 20$)

Figure 3.1: The conditional false alarm rate (*CFAR*) as a function of $u = 0,1,...,20$ for various combinations of m and n such that $T = mn = 20$



(a) $T = 50$ ($m = 50$, $n = 1$)

(b) $T = 50$ ($m = 25$, $n = 2$)

(c) $T = 50$ ($m = 10$, $n = 5$)

(d) $T = 50$ ($m = 5$, $n = 10$)

(e) $T = 50$ ($m = 2$, $n = 25$)

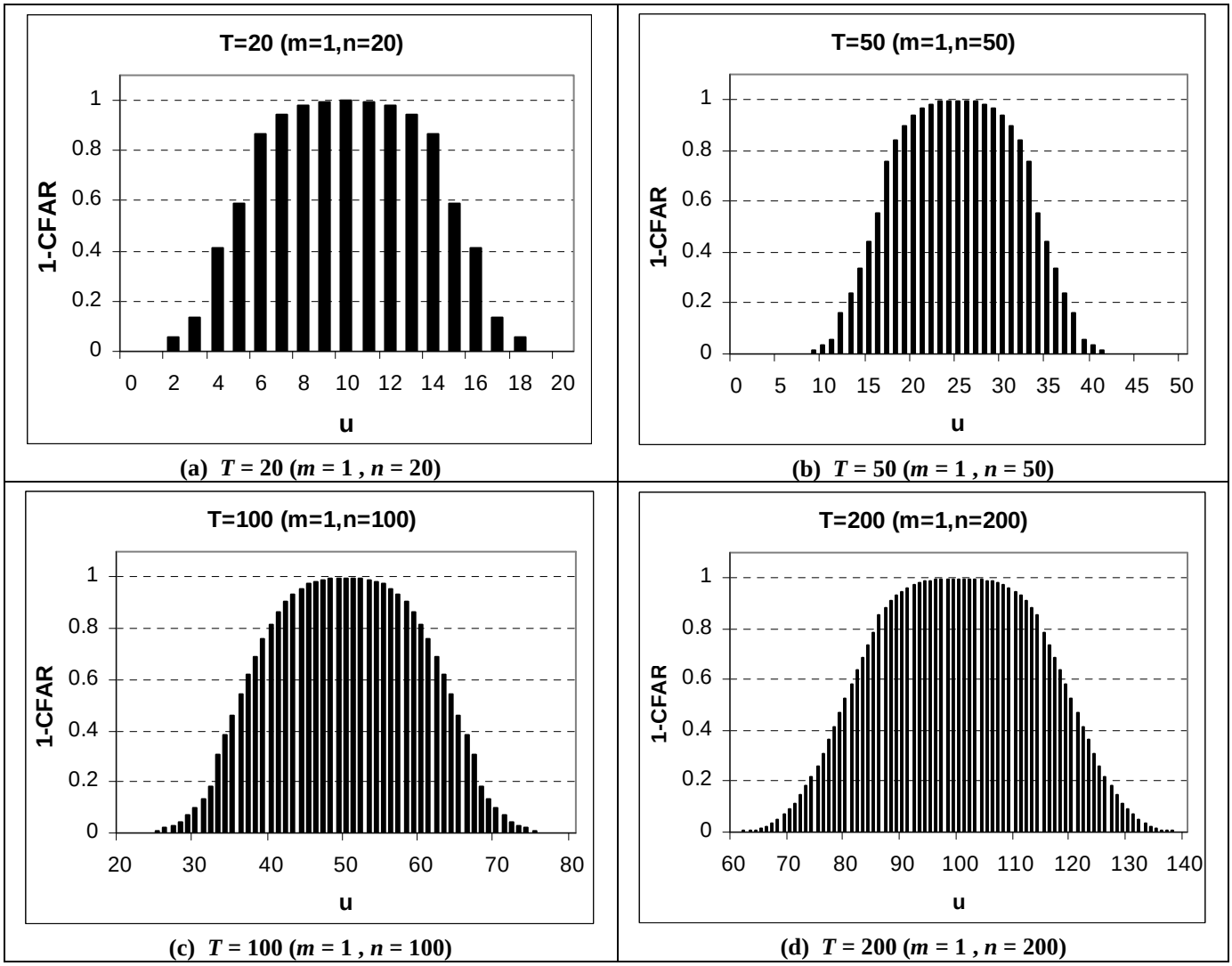
(f) $T = 50$ ($m = 1$, $n = 50$)

Figure 3.2: The conditional false alarm rate (*CFAR*) as a function of $u = 0,1,...,50$ for various combinations of m and n such that $T = mn = 50$

The conditional probability of a no-signal

The distribution of *1-CFAR*, which is the conditional probability of a no-signal when the process is in-control, is shown in panels (a) to (d) of Figure 3.3 for $T = 20$, 50, 100 and 200 when $m = 1$ and $n = T$ i.e. for large n relative to m.

It is seen that the distribution of *1-CFAR* is bell-shaped and symmetric; these two characteristics follow from that of *CFAR* shown in Figure 3.1 and 3.2.



(a) $T = 20$ ($m = 1$, $n = 20$) **(b) $T = 50$ ($m = 1$, $n = 50$)**

(c) $T = 100$ ($m = 1$, $n = 100$) **(d) $T = 200$ ($m = 1$, $n = 200$)**

Figure 3.3: The conditional probability of a no-signal when the process is in-control (*1-CFAR*) as a function of $u = 0,1,...,T$ for $m = 1$ and $n = T = 20$, 50, 100 and 200

The out-of-control conditional performance of the *p*-chart

The in-control performance of the Phase II *p*-chart (in theory) refers to the characteristics of the chart in the situation where the process operates at the same level in Phase II as what it did in Phase I; this is the scenario when $p_1 = p$. However, because p is unknown and estimated by $\bar{p}$, the observed value $\bar{p}_{obs}$ plays the role of p so that the conditional in-control performance (in practice) refers to the situation when $p_1 = \bar{p}_{obs}$ (see e.g. the earlier section labelled "Phase II implementation and operation"). The out-of-control performance (in practice) then refers to the characteristics of the *p*-chart when $p_1 \neq \bar{p}_{obs}$.

Taking into consideration the aforementioned, we can study the out-of-control performance of the Phase II *p*-chart by making use of the results from the previous section. In particular, by conditioning on a specific observed value $\bar{p}_{obs}$, the run-length distribution is affected in the same way it would be if the unknown true fraction nonconforming was to change from p (in Phase I) to p_1 (in Phase II). In other words, the out-of-control performance of the Phase II *p*-chart (i.e. when p has incurred either a downward or an upward shift to p_1 so that $p_1 \neq p$) is equivalent to the performance of the conditional *p*-chart when $p \neq \bar{p}_{obs}$ i.e. if p was either overestimated or underestimated (see e.g. Jones, Champ and Rigdon, (2004)); this correspondence allows us to examine the out-of-control performance of the *p*-chart by using the conditional statistical characteristics.

To this end, consider, for example, Table 3.7 which lists the false alarm rate (*CFAR*), the average run-length ($CARL_0$) and the standard deviation of the run-length ($CSDRL_0$) of the conditional run-length distribution for different combinations of m and n, provided that $T = mn = 20$ and $p_1 = p = 0.5$. In each case the run-length distribution is conditioned on an estimate of p through a particular realization u of the random variable U or, equivalently, on a specific realization $\bar{p}_{obs}$.

The values on which we condition are, for illustration proposes only, $U = 7$ (i.e. $\bar{p} = 7/20 = 0.35$), $U = 8$ (i.e. $\bar{p} = 8/20 = 0.40$) and $U = 10$ (i.e. $\bar{p} = 10/20 = 0.50$). These values correspond to the 10^{th}, the 25^{th} and the 50^{th} percentiles of the probability distribution of $U \sim Bin(mn = 20, p = 0.5)$, respectively; note that, because the $Bin(20,0.5)$ distribution is symmetric, conditioning on $U = 7$ and $U = 8$ are like conditioning on $U = 20 - 7 = 13$ (i.e. $\bar{p} = 13/20 = 0.65$) and $U = 20 - 8 = 12$ (i.e. $\bar{p} = 12/20 = 0.60$), which are the 90^{th} and the 75^{th} percentiles of the probability distribution of U, respectively.

In particular, by assuming that $p_1 = p = 0.5$ and then conditioning on $U = 7$ or $U = 8$ (i.e. $\bar{p} = 0.35$ or $\bar{p} = 0.40$) the performance of the Phase II p-chart are comparable to that of a process that has sustained a permanent step shift from 0.35 to 0.5 or encountered a lasting step shift from 0.4 to 0.5 i.e. an increase of either 43% or 25%, respectively. Similarly, if we assume that $p_1 = p = 0.5$ and then condition on $U = 13$ or $U = 12$ (i.e. $\bar{p} = 0.65$ or $\bar{p} = 0.60$) the performance of the Phase II p-chart is like that of a process that has sustained a permanent step shift from 0.65 to 0.5 (a decrease of 23%) or incurred a step shift from 0.6 to 0.5 (a decrease of 17%).

When $(m, n) = (1,20)$ and we condition on a value of $U = 8$ (or 12), which is the 25th (or the 75th) percentile of the distribution of $U \mid p = 0.5 \sim Bin(20,0.5)$, the *CFAR* is 0.0207 and the $CARL_0$ is 48.27. The *CFAR* is approximately $(0.0207/0.0160 - 1) \times 100\% \approx 29\%$ higher than the probability of a signal of $1 - \beta(p = 0.4 \text{ or } 0.6, p_0 = 0.5, n = 20) = 0.0160$ of Case K whereas the $CARL_0$ is roughly $(48.27/62.5 - 1) \times 100\% \approx 23\%$ lower than the out-of-control (OOC) *ARL* of Case K following a sustained shift from 0.4 or 0.6 to 0.5, which is equal to $ARL(p = 0.4 \text{ or } 0.6, p_0 = 0.5, n = 20) = 62.5$ (see e.g. Tables A3.4 and A3.5 in Appendix 3A).

This means that when $m = 1$ and $n = 20$, and p is either underestimated or overestimated by 25% (i.e. the process fraction nonconforming has endured either a 25% decrease or increase and is out-of-control), the *p*-chart of Case U would be better at detecting such a shift than the *p*-chart of Case K. However, note that, this superior performance is a side-effect of estimating p.

The same is true for other combinations of (m, n). For example, if our Phase I reference data consisted of $m = 2$ samples each of size $n = 10$ and we then condition on $U = 8$ (or 12), the conditional *FAR* is $CFAR = 0.0107$ and the in-control conditional *ARL* is $CARL_0 = 93.09$. These values are approximately 73% higher and 43% lower than the probability of a signal and the out-of-control *ARL* of 0.0062 and 162.6 if p had been known.

In contrast, it is noteworthy to see what happens if we condition on $U = 10$ (i.e. the 50th percentile of the distribution of U), which implies that our estimate of p is spot on, that is, the point estimate $\bar{p} = 0.5$ on which we condition is equal to p, so that we are in actual fact dealing with the in-control (IC) performance of the *p*-chart in Case U.

In this case, the *CFAR* and the in-control conditional *ARL* for both the scenarios $(m, n) = (1,20)$ and $(m, n) = (2,10)$, are exactly equal to the in-control performance of the *p*-chart in Case K with

$FAR = 0.0026$ & $ARL_0 = 388.07$ and $FAR = 0.0020$ & $ARL_0 = 512.00$, respectively (see e.g. Tables A3.4 and A3.5 in Appendix 3A). Furthermore, note that, as mentioned before, for some combinations of (m,n), especially when $m >> n$, it happens that for certain values of U the $CFAR$ equals zero which causes the $CARL_0$ and $CSDRL_0$ to be undefined, which is undesirable.

To summarize, when $T = 20$ and p is either underestimated or overestimated (i.e. the process is OOC), the Case U p-chart would do better than the Case K chart at detecting a shift, and only if our estimate $\bar{p}$ of p is on target (i.e. the process is IC) would the performance of the Case U and Case K charts be similar.

Table 3.7: The false alarm rate (*CFAR*), the average run-length (*CARL*) and the standard deviation of the run-length (*CSDRL*) of the conditional run-length distribution for different combinations of m and n, provided that $T = mn = 20$ and $p_1 = p = 0.5$

$T = 20$		$U = 7$ or 13 (OOC) ($\bar{p} = 0.35$ or 0.65)			$U = 8$ or 12 (OOC) ($\bar{p} = 0.4$ or 0.6)			$U = 10$ (IC) ($\bar{p} = 0.5$)		
		10th or 90th Percentile			25th or 75th Percentile			50th Percentile		
m	***n***	***CFAR***	$CARL_0$	$CSDRL_0$	***CFAR***	$CARL_0$	$CSDRL_0$	***CFAR***	$CARL_0$	$CSDRL_0$
1	**20**	0.0577	17.34	16.84	0.0207	48.27	47.77	0.0026	388.07	387.57
2	**10**	0.0107	93.09	92.59	0.0107	93.09	92.59	0.0020	512.00	511.50
4	**5**	0.0313	32.00	31.50	0.0	∞	∞	0.0	∞	∞
5,4 **20,1**	**10,2**	0.0	∞	∞	0.0	∞	∞	0.0	∞	∞

Calculations similar to those in Table 3.7 are shown in Tables 3.8, 3.9, 3.10, and 3.11 for a larger range of values for T; we specifically look at $T = 10$,15, 25, 30, 50, 75, 100, 200, 250, 300, 500, 750, 1000 and 1500.

For each value of T we look at all possible combinations of m and n such that $T = mn$ where both m and n are integers. We again condition on the 10th (or the 90th), the 25th (or the 75th), and the 50th percentiles of $U \mid p = 0.5 \sim Bin(T = mn, 0.5)$ so that the interpretation of these conditional characteristics is similar to those for $T = 20$ of Table 3.7. The values of the percentiles of U and the corresponding values of $\bar{p}$ are clearly indicated.

The characteristics that are highlighted in grey indicate those (m,n) combinations for which the Case U p-chart performs worse than the Case K p-chart; for all the other (m,n) combinations the Case U p-chart performs better or just as well as the Case K p-chart.

The conditional characteristics of Tables 3.8, 3.9, 3.10, and 3.11 are of great help to the practitioner as he/she gets an idea of the ramifications when (or if) p is underestimated or overestimated for his/her particular combination of m and n values at hand (even before any data is collected); this is similar to investigating the power of a test.

Table 3.8: The false alarm rate (*CFAR*), the average run-length (*CARL*) and the standard deviation of the run-length (*CSDRL*) of the conditional run-length distribution for different combinations of m and n, provided that $T = 10$, 15, 25 and 30 and $p_1 = p = 0.5$

		10^{th} or 90^{th} Percentile			25^{th} or 75^{th} Percentile			50^{th} Percentile		
m	***n***	***CFAR***	$CARL_0$	$CSDRL_0$	***CFAR***	$CARL_0$	$CSDRL_0$	***CFAR***	$CARL_0$	$CSDRL_0$
$T = 10$		$U = 3$ or 7 (OOC) ($\bar{p} = 0.3$ or 0.7)			$U = 4$ or 6 (OOC) ($\bar{p} = 0.4$ or 0.6)			$U = 5$ (IC) ($\bar{p} = 0.5$)		
1	**10**	0.0547	18.29	17.78	0.0107	93.09	92.59	0.0020	512.00	511.50
2	**5**	0.0313	32.00	31.50	0.0	∞	∞	0.0	∞	∞
5,2	**10,1**	0.0	∞	∞	0.0	∞	∞	0.0	∞	∞
$T = 15$		$U = 5$ or 10 (OOC) ($\bar{p} = 0.33\dot{}$ or $0.66\dot{}$)			$U = 6$ or 9 (OOC) ($\bar{p} = 0.4$ or 0.6)			$U = 7$ (IC) ($\bar{p} = 0.46\dot{}$)		
1	**15**	0.0592	16.88	16.37	0.0176	56.79	56.29	0.0042	239.18	238.68
3	**5**	0.0	32.00	31.50	0.0	∞	∞	0.0	∞	∞
5,3	**15,1**	0.0	∞	∞	0.0	∞	∞	0.0	∞	∞
$T = 25$		$U = 9$ or 16 (OOC) ($\bar{p} = 0.36$ or 0.64)			$U = 11$ or 14 (OOC) ($\bar{p} = 0.44$ or 0.56)			$U = 12$ (IC) ($\bar{p} = 0.48$)		
1	**25**	0.0539	18.56	18.05	0.0074	135.23	134.73	0.0025	400.98	400.48
5,5	**25,1**	0.0	∞	∞	0.0	∞	∞	0.0	∞	∞
$T = 30$		$U = 11$ or 19 (OOC) ($\bar{p} = 0.36\dot{}$ or $0.63\dot{}$)			$U = 13$ or 17 (OOC) ($\bar{p} = 0.43\dot{}$ or $0.56\dot{}$)			$U = 15$ (IC) ($\bar{p} = 0.5$)		
1	**30**	0.1002	9.98	9.46	0.0081	123.58	123.08	0.0014	698.86	698.36
2	**15**	0.0176	56.89	56.39	0.0037	268.59	268.09	0.0010	1024.00	1023.50
3	**10**	0.0107	93.09	92.59	0.0010	1024.00	1023.50	0.0020	512.00	511.50
5	**6**	0.0	64.00	63.50	0.0	∞	∞	0.0	∞	∞
6,5 **15,2**	**10,3** **30,1**	0.0	∞	∞	0.0	∞	∞	0.0	∞	∞

Table 3.9: The false alarm rate (*CFAR*), the average run-length (*CARL*) and the standard deviation of the run-length (*CSDRL*) of the conditional run-length distribution for different combinations of m and n, provided that $T = 50$, 75, 100, 200 and 250 and $p_1 = p = 0.5$

		10^{th} or 90^{th} Percentile			25^{th} or 75^{th} Percentile			50^{th} Percentile		
m	n	$CFAR$	$CARL_0$	$CSDRL_0$	$CFAR$	$CARL_0$	$CSDRL_0$	$CFAR$	$CARL_0$	$CSDRL_0$
$T = 50$		$U = 20$ or 30 (OOC) ($\bar{p} = 0.4$ or 0.6)			$U = 23$ or 27 (OOC) ($\bar{p} = 0.46$ or 0.54)			$U = 25$ (IC) ($\bar{p} = 0.5$)		
1	50	0.0595	16.82	16.31	0.0078	127.77	127.27	0.0026	384.29	383.79
2	25	0.0217	46.18	45.68	0.0078	128.67	128.17	0.0041	245.26	244.76
5	10	0.0107	93.09	92.59	0.0010	1024.00	1023.50	0.0020	512.00	511.50
10,5 50,1	25,2	0.0	∞	∞	0.0	∞	∞	0.0	∞	∞
$T = 75$		$U = 32$ or 43 (OOC) ($\bar{p} = 0.426^{\cdot}$ or $0.573^{\cdot}$)			$U = 35$ or 40 (OOC) ($\bar{p} = 0.46^{\cdot}$ or $0.53^{\cdot}$)			$U = 37$ (IC) ($\bar{p} = 0.493^{\cdot}$)		
1	75	0.0527	18.98	18.47	0.0104	96.39	95.89	0.0038	260.67	260.17
3	25	0.0074	135.23	134.73	0.0025	400.98	400.48	0.0025	400.98	400.48
5	15	0.0037	268.59	268.09	0.0042	239.18	238.68	0.0010	1024.00	1023.50
15,5 75,1	25,3	0.0	∞	∞	0.0	∞	∞	0.0	∞	∞
$T = 100$		$U = 44$ or 56 (OOC) ($\bar{p} = 0.44$ or 0.56)			$U = 47$ or 53 (OOC) ($\bar{p} = 0.47$ or 0.53)			$U = 50$ (IC) ($\bar{p} = 0.5$)		
1	100	0.0443	22.56	22.05	0.0107	93.51	93.01	0.0035	284.28	283.78
2	50	0.0165	60.74	60.23	0.0035	289.59	289.09	0.0026	384.29	383.79
4	25	0.0074	135.23	134.73	0.0025	400.98	400.48	0.0041	245.26	244.76
5	20	0.0061	163.66	163.16	0.0015	671.30	670.80	0.0026	388.07	387.57
10	10	0.0010	1024.00	1023.50	0.0010	1024.00	1023.50	0.0020	512.00	511.50
20,5 50,2	25,4 100,1	0.0	∞	∞	0.0	∞	∞	0.0	∞	∞
$T = 200$		$U = 91$ or 109 (OOC) ($\bar{p} = 0.455$ or 0.545)			$U = 95$ or 105 (OOC) ($\bar{p} = 0.475$ or 0.525)			$U = 100$ (IC) ($\bar{p} = 0.5$)		
1	200	0.0384	26.02	25.52	0.0098	102.24	101.74	0.0023	438.70	438.20
2	100	0.0176	56.69	56.19	0.0062	160.75	160.25	0.0035	284.28	283.78
4	50	0.0078	127.77	127.27	0.0038	265.37	264.87	0.0026	384.29	383.79
5	40	0.0084	119.25	118.75	0.0036	281.45	280.95	0.0022	450.16	449.66
8	25	0.0074	135.23	134.73	0.0025	400.98	400.48	0.0041	245.26	244.76
10	20	0.0061	163.66	163.16	0.0015	671.30	670.80	0.0026	388.07	387.57
20	10	0.0010	1024.00	1023.50	0.0020	512.00	511.50	0.0020	512.00	511.50
25	8	0.0039	256.00	255.50	0.0	∞	∞	0.0	∞	∞
40,5 100,2	50,4 200,1	0.0	∞	∞	0.0	∞	∞	0.0	∞	∞
$T = 250$		$U = 115$ or 135 (OOC) ($\bar{p} = 0.46$ or 0.54)			$U = 120$ or 130 (OOC) ($\bar{p} = 0.48$ or 0.52)			$U = 125$ (IC) ($\bar{p} = 0.5$)		
1	250	0.0438	22.85	22.35	0.0097	103.13	102.63	0.0029	347.38	346.88
2	125	0.0157	63.53	63.03	0.0063	159.02	158.52	0.0022	449.14	448.64
5	50	0.0078	127.77	127.27	0.0038	265.37	264.87	0.0026	384.29	383.79
10	25	0.0078	128.67	128.17	0.0025	400.98	400.48	0.0041	245.26	244.76
25	10	0.0010	1024.00	1023.50	0.0020	512.00	511.50	0.0020	512.00	511.50
50,5 250,1	125,2	0.0	∞	∞	0.0	∞	∞	0.0	∞	∞

Table 3.10: The false alarm rate (*CFAR*), the average run-length (*CARL*) and the standard deviation of the run-length (*CSDRL*) of the conditional run-length distribution for different combinations of m and n, provided that $T = 300$, 500 and 750 and $p_1 = p = 0.5$

		10^{th} or 90^{th} Percentile			25^{th} or 75^{th} Percentile			50^{th} Percentile		
m	n	$CFAR$	$CARL_0$	$CSDRL_0$	$CFAR$	$CARL_0$	$CSDRL_0$	$CFAR$	$CARL_0$	$CSDRL_0$
$T = 300$		$U = 139$ or 161 (OOC) ($\bar{p} = 0.463^{\cdot}$ or $0.536^{\cdot}$)			$U = 144$ or 156 (OOC) ($\bar{p} = 0.48$ or 0.52)			$U = 150$ (IC) ($\bar{p} = 0.5$)		
1	300	0.0470	21.29	20.79	0.0122	81.82	81.32	0.0032	315.53	315.03
2	150	0.0205	48.81	48.30	0.0058	173.44	172.94	0.0024	415.71	415.21
3	100	0.0106	94.51	94.01	0.0065	154.96	154.46	0.0035	284.28	283.78
4	75	0.0102	97.68	97.18	0.0058	171.50	171.00	0.0024	409.13	408.63
5	60	0.0069	144.06	143.56	0.0036	274.60	274.10	0.0027	374.47	373.97
6	50	0.0078	127.77	127.27	0.0038	265.37	264.87	0.0026	384.29	383.79
10	30	0.0028	360.50	360.00	0.0033	300.58	300.08	0.0014	698.86	698.36
12	25	0.0025	400.98	400.48	0.0025	400.98	400.48	0.0041	245.26	244.76
15	20	0.0061	163.66	163.16	0.0015	671.30	670.80	0.0026	388.07	387.57
20	15	0.0042	239.18	238.68	0.0010	1024.00	1023.50	0.0010	1024.00	1023.50
25	12	0.0034	292.57	292.07	0.0034	292.57	292.07	0.0005	2048.00	2047.50
30	10	0.0010	1024.00	1023.50	0.0020	512.00	511.50	0.0020	512.00	511.50
50,6 75,4 150,2	60,5 100,3 300,1	0.0	∞	∞	0.0	∞	∞	0.0	∞	∞
$T = 500$		$U = 236$ or 264 (OOC) ($\bar{p} = 0.472$ or 0.528)			$U = 242$ or 258 (OOC) ($\bar{p} = 0.484$ or 0.516)			$U = 250$ (IC) ($\bar{p} = 0.5$)		
1	500	0.0405	24.68	24.17	0.0113	88.24	87.74	0.0027	370.81	370.31
2	250	0.0184	54.39	53.88	0.0070	143.25	142.75	0.0029	347.38	346.88
4	125	0.0100	100.06	99.56	0.0038	260.91	260.41	0.0022	449.14	448.64
5	100	0.0062	160.75	160.25	0.0038	266.28	265.78	0.0035	284.28	283.78
10	50	0.0038	265.37	264.87	0.0038	265.37	264.87	0.0026	384.29	383.79
20	25	0.0025	400.98	400.48	0.0025	400.98	400.48	0.0041	245.26	244.76
25	20	0.0015	671.30	670.80	0.0015	671.30	670.80	0.0026	388.07	387.57
50	10	0.0010	1024.00	1023.50	0.0020	512.00	511.50	0.0020	512.00	511.50
100,5 250,2	125,4 500,1	0.0	∞	∞	0.0	∞	∞	0.0	∞	∞
$T = 750$		$U = 357$ or 393 (OOC) ($\bar{p} = 0.476$ or 0.524)			$U = 366$ or 384 (OOC) ($\bar{p} = 0.488$ or 0.512)			$U = 375$ (IC) ($\bar{p} = 0.5$)		
1	750	0.0430	23.24	22.73	0.0089	112.45	111.95	0.0024	413.68	413.18
2	375	0.0194	51.54	51.03	0.0051	196.82	196.32	0.0027	370.96	370.46
3	250	0.0134	74.53	74.02	0.0051	197.15	196.65	0.0029	347.38	346.88
5	150	0.0090	111.07	110.56	0.0038	262.77	262.27	0.0024	415.71	415.21
6	125	0.0061	163.01	162.51	0.0041	242.72	242.22	0.0022	449.14	448.64
10	75	0.0055	181.29	180.79	0.0032	317.07	316.57	0.0024	409.13	408.63
15	50	0.0038	265.37	264.87	0.0018	565.23	564.73	0.0026	384.29	383.79
25	30	0.0033	300.58	300.08	0.0033	300.58	300.08	0.0014	698.86	698.36
30	25	0.0025	400.98	400.48	0.0025	400.98	400.48	0.0041	245.26	244.76
50	15	0.0042	239.18	238.68	0.0010	1024.00	1023.50	0.0010	1024.00	1023.50
75	10	0.0020	512.00	511.50	0.0020	512.00	511.50	0.0020	512.00	511.50
125,6 250,3 750,1	150,5 375,2	0.0	∞	∞	0.0	∞	∞	0.0	∞	∞

Table 3.11: The false alarm rate (*FAR*), the average run-length (*ARL*) and the standard deviation of the run-length (*SDRL*) of the conditional run-length distribution for different combinations of m and n, provided that $T = 1000$ and 1500 and $p_1 = p = 0.5$

		10^{th} or 90^{th} Percentile			25^{th} or 75^{th} Percentile			50^{th} Percentile		
m	n	*CFAR*	*CARL*	*CSDRL*	*CFAR*	*CARL*	*CSDRL*	*CFAR*	*CARL*	*CSDRL*
$T = 1000$		$U = 480$ or 520 (OOC) ($\bar{p} = 0.48$ or 0.52)			$U = 489$ or 511 (OOC) ($\bar{p} = 0.489$ or 0.511)			$U = 500$ (IC) ($\bar{p} = 0.5$)		
1	1000	0.0410	24.40	23.90	0.0106	94.61	94.11	0.0026	378.00	377.50
2	500	0.0178	56.25	55.75	0.0056	179.73	179.23	0.0027	370.81	370.31
4	250	0.0097	103.13	102.63	0.0051	197.15	196.65	0.0029	347.38	346.88
5	200	0.0067	149.04	148.53	0.0033	306.27	305.77	0.0023	438.70	438.20
8	125	0.0063	159.02	158.52	0.0041	242.72	242.22	0.0022	449.14	448.64
10	100	0.0065	154.96	154.46	0.0038	266.28	265.78	0.0035	284.28	283.78
20	50	0.0038	265.37	264.87	0.0018	565.23	564.73	0.0026	384.29	383.79
25	40	0.0036	281.45	280.95	0.0022	450.16	449.66	0.0022	450.16	449.66
40	25	0.0025	400.98	400.48	0.0025	400.98	400.48	0.0041	245.26	244.76
50	20	0.0015	671.30	670.80	0.0026	388.07	387.57	0.0026	388.07	387.57
100	10	0.0020	512.00	511.50	0.0020	512.00	511.50	0.0020	512.00	511.50
125,8 250,4 1000,1	200,5 500,2	0.0	∞	∞	0.0	∞	∞	0.0	∞	∞
$T = 1500$		$U = 725$ or 775 (OOC) ($\bar{p} = 0.483^{\cdot}$ or $0.516^{\cdot}$)			$U = 737$ or 763 (OOC) ($\bar{p} = 0.4913^{\cdot}$ or $0.5086^{\cdot}$)			$U = 750$ (IC) ($\bar{p} = 0.5$)		
1	1500	0.0418	23.92	23.41	0.0095	105.36	104.85	0.0025	398.62	398.12
2	750	0.0187	53.44	52.94	0.0061	163.72	163.22	0.0024	413.68	413.18
3	500	0.0113	88.24	87.74	0.0045	221.21	220.71	0.0027	370.81	370.31
4	375	0.0089	112.99	112.48	0.0040	247.16	246.66	0.0027	370.96	370.46
5	300	0.0091	109.95	109.45	0.0038	265.71	265.21	0.0032	315.53	315.03
6	250	0.0070	143.25	142.75	0.0038	261.95	261.45	0.0029	347.38	346.88
10	150	0.0059	168.33	167.83	0.0027	364.75	364.25	0.0024	415.71	415.21
12	125	0.0038	260.91	260.41	0.0026	383.35	382.84	0.0022	449.14	448.64
15	100	0.0038	266.28	265.78	0.0027	376.82	376.32	0.0035	284.28	283.78
20	75	0.0032	317.07	316.57	0.0032	317.07	316.57	0.0024	409.13	408.63
25	60	0.0036	274.60	274.10	0.0019	535.30	534.80	0.0027	374.47	373.97
30	50	0.0038	265.37	264.87	0.0018	565.23	564.73	0.0026	384.29	383.79
50	30	0.0033	300.58	300.08	0.0033	300.58	300.08	0.0014	698.86	698.36
60	25	0.0025	400.98	400.48	0.0025	400.98	400.48	0.0041	245.26	244.76
75	20	0.0015	671.30	670.80	0.0026	388.07	387.57	0.0026	388.07	387.57
100	15	0.0010	1024.00	1023.50	0.0010	1024.00	1023.50	0.0010	1024.00	1023.50
125	12	0.0034	292.57	292.07	0.0005	2048.00	2047.50	0.0005	2048.00	2047.50
150	10	0.0020	512.00	511.50	0.0020	512.00	511.50	0.0020	512.00	511.50
250,6 375,4 750,2	300,5 500,3 1500,1	0.0	∞	∞	0.0	∞	∞	0.0	∞	∞

3.2.3.2 Conditional characteristics of the *c*-chart

Like the *p*-chart, once we observe a value v of the random variable V we can calculate the conditional probability of a no-signal of the *c*-chart so that the Phase II conditional run-length distribution and its associated conditional characteristics are completely known (see e.g. Table 3.4). To this end, Table 3.12 illustrates the steps to calculate the conditional probability of a no-signal, the conditional false alarm rate (*CFAR*), the conditional average run-length (*CARL*) and the conditional standard deviation of the run-length (*CSDRL*) of the *c*-chart.

For illustration purposes we assume that $c_1 = c = 20$; this implies that the process operated at a level of twenty nonconformities (on average) in an inspection unit during Phase I and that in Phase II the process continues to operate at this same level. In addition, we assume that $m = 100$ Phase I inspection units are available to estimate c using $\bar{c} = V/m = \bar{c}_{\text{obs}}$, which (because of sampling variation) may or may not be equal to c.

In particular, column 1 lists some values of $V = 0(200)6000$, which (in theory) can be any integer greater than or equal to zero. Column 2 converts the observed value v of V of column 1 into a point estimate of c by calculating $\bar{c}_{\text{obs}} = v/100$. Because each row entry in each of the succeeding columns (i.e. columns 3 to 10) is computed by conditioning on a row entry from column 1 or column 2, we start calculating the conditional properties in column 1 or 2 and sequentially proceed to the right-hand side of the table. So, given a value v or $\bar{c}_{\text{obs}}$ the lower and the upper control limits are estimated in columns 3 and 4 using (3-27) and then used to compute the two constants $\hat{d}$ and $\hat{f}$ defined in (3-33), which are shown in columns 5 and 6, respectively. Finally, columns 7 through 10 list the probability of a no-signal, the *FAR*, the *ARL* and the *SDRL* conditioned on the observed value v from column 1, respectively. These properties are labeled $\Pr(\text{No Signal} \mid V, c)$, *CFAR*, $CARL_0$ and $CSDRL_0$, and calculated using (3-32) and the expressions in Table 3.4.

An examination of Table 3.12 reveals one special scenario i.e. when $V = 0$ (the minimum possible value). In this particular case the estimated control limits are $L\hat{C}L_c = U\hat{C}L_c = 0$ so that the constants $\hat{d}$ and $\hat{f}$ need not be calculated (see e.g. expression (3-32) and Remark 6(ii)); as a result, the probability of a no-signal is defined to be zero so that the *c*-chart signals with probability one once the first Phase II inspection unit is sampled i.e. both the conditional *FAR* and the conditional *ARL* are one (as shown in columns 8 and 9, respectively). For values of $V \neq 0$ we proceed as follows.

Table 3.12: Conditional Probability of a no-signal, the conditional false alarm rate (*CFAR*), the conditional average run-length (*CARL*) and the conditional standard deviation of the run-length (*CSDRL*) of the *c*-chart in Case U for $m = 100$ and assuming that $c_1 = c = 20$

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
v	$\bar{c}_{obs}$	$L\hat{C}L_c$	$U\hat{C}L_c$	$\hat{d}$	$\hat{f}$	Pr(No Signal | *V*, c)	*CFAR*	$CARL_0$	$CSDRL_0$
0	0	0.00	0.00	NA	NA	0.0000	1.0000	1.00	0.00
200	2	-2.24	6.24	0	6	0.0003	0.9997	1.00	0.02
400	4	-2.00	10.00	0	9	0.0050	0.9950	1.01	0.07
600	6	-1.35	13.35	0	13	0.0661	0.9339	1.07	0.28
800	8	-0.49	16.49	0	16	0.2211	0.7789	1.28	0.60
1000	10	0.51	19.49	0	19	0.4703	0.5297	1.89	1.29
1200	12	1.61	22.39	1	22	0.7206	0.2794	3.58	3.04
1400	14	2.78	25.22	2	25	0.8878	0.1122	8.91	8.40
1600	16	4.00	28.00	4	27	0.9475	0.0525	19.05	18.54
1800	18	5.27	30.73	5	30	0.9865	0.0135	73.82	73.32
2000	20	6.58	33.42	6	33	0.9971	0.0029	339.72	339.22
2200	22	7.93	36.07	7	36	0.9988	0.0012	832.30	831.80
2400	24	9.30	38.70	9	38	0.9949	0.0051	195.92	195.42
2600	26	10.70	41.30	10	41	0.9892	0.0108	92.39	91.89
2800	28	12.13	43.87	12	43	0.9610	0.0390	25.63	25.13
3000	30	13.57	46.43	13	46	0.9339	0.0661	15.12	14.61
3200	32	15.03	48.97	15	48	0.8435	0.1565	6.39	5.87
3400	34	16.51	51.49	16	51	0.7789	0.2211	4.52	3.99
3600	36	18.00	54.00	18	53	0.6186	0.3814	2.62	2.06
3800	38	19.51	56.49	19	56	0.5297	0.4703	2.13	1.55
4000	40	21.03	58.97	21	58	0.3563	0.6437	1.55	0.93
4200	42	22.56	61.44	22	61	0.2794	0.7206	1.39	0.73
4400	44	24.10	63.90	24	63	0.1568	0.8432	1.19	0.47
4600	46	25.65	66.35	25	66	0.1122	0.8878	1.13	0.38
4800	48	27.22	68.78	27	68	0.0525	0.9475	1.06	0.24
5000	50	28.79	71.21	28	71	0.0343	0.9657	1.04	0.19
5200	52	30.37	73.63	30	73	0.0135	0.9865	1.01	0.12
5400	54	31.95	76.05	31	76	0.0081	0.9919	1.01	0.09
5600	56	33.55	78.45	33	78	0.0027	0.9973	1.00	0.05
5800	58	35.15	80.85	35	80	0.0008	0.9992	1.00	0.03
6000	60	36.76	83.24	36	83	0.0004	0.9996	1.00	0.02
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Suppose, for example, that we observe two thousand four hundred nonconformities in the entire Phase I reference sample. The value of $V = 2400$ gives an observed value of the point estimate for c of $\bar{c}_{obs} = 2400/100 = 24$ so that (3-27) yields an estimated upper control limit and an estimated lower control of

$$U\hat{C}L_c = 24 + 3\sqrt{24} = 38.70 \qquad \text{and} \qquad L\hat{C}L_c = 24 - 3\sqrt{24} = 9.30$$

respectively.

The constants $\hat{d}$ and $\hat{f}$ are thus found to be

$$\hat{d} = \max\{0,[L\hat{C}L_c]\} = \max\{0,[9.3]\} = 9 \qquad \text{and} \qquad \hat{f} = [U\hat{C}L_c - 1] = [38.70] = 38$$

so that upon substituting $\hat{d}$ and $\hat{f}$ in (3-32) we calculate the conditional probability of a no-signal and then also the *CFAR*, the $CARL_0$ and the $CSDRL_0$ using expressions (3-42), (3-43) and (3-44) in Table 3.4.

The conditional probability of a no-signal, in particular, is

$$\hat{\beta}(c_1 = 20, m = 100 \,|\, V = 2400, c = 20) = \hat{\beta}(c_1 = 20, m = 100 \,|\, \bar{c} = 24, c = 20) = 0.9949$$

so that the conditional false alarm rate is

$$CFAR(c_1 = 20{,}100 \,|\, V = 2400, c = 20) = 1 - 0.9949 = 0.0051.$$

The Phase II c-chart then has a conditional in-control *ARL* of

$$CARL_0(c_1 = 20{,}100 \,|\, V = 2400, c = 20) = 1/0.0051 = 195.92$$

and a conditional in-control *SDRL* of

$$SDRL_0(c_1 = 20{,}100 \,|\, V = 2400, c = 20) = \sqrt{0.9949}/0.0051 = 195.42.$$

The conditional false alarm rate

Figure 3.4 displays the conditional false alarm rate (*CFAR*), that is, $1 - \hat{\beta}(c_1 = 20, m \,|\, V = v, c = 20)$ as a function of $v = 0{,}1{,}2{,}...$ when m = 50 or 75 or 100 individual Phase I inspection units are used to estimate c; the curve labeled $m = 100$ corresponds to the *CFAR* 's of column 8 in Table 3.12.

The impact of the actual observed number of nonconformities v in the entire Phase I reference sample is easily noticed. The distribution function of the *CFAR* is seen to be slightly negatively U-shaped. For values of *V* near the two tails (i.e. the extreme left and right) the *CFAR* can be very high, sometimes close to 1 or 100%, which obviously means many false alarms.

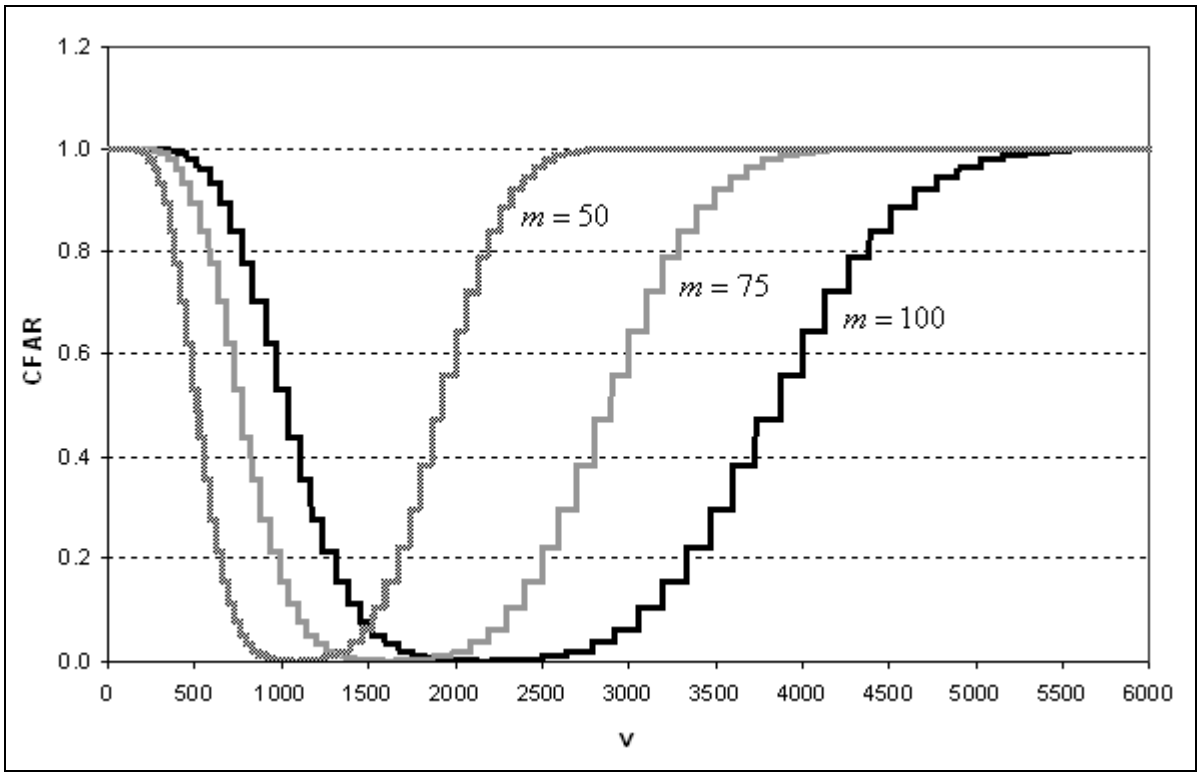


Figure 3.4: The conditional false alarm rate (*CFAR*) as a function of $v = 0,1,2,...$ for $m = 50,75$ and 100

However, even when V is not near the two tails there can be a significantly high probability of a false alarm; this is more easily seen from Figure 3.5, which (for illustration purposes) displays values of $1-\hat{\beta}(c_1 = 20, m = 100 \mid V = v, c = 20)$ for values of v between 1800 and 2600 only.

It is seen that only when V takes on a value in the neighbourhood of its mean i.e. $E(V \mid c = 20) = mc = 100 \times 20 = 2000$ (or, equivalently, when $\bar{c}$ is close to the true average number of nonconformities, which is 20 in this case) will the *CFAR* be reasonably small and close to its Case K value of 0.0029 (see e.g. Table A3.12 in Appendix 3A).

However even though the *CFAR* may be small, it is (for most values of v) still far from the typical or nominal expected value of 0.0027 of a Shewhart X-bar chart with 3-sigma limits. Thus, the performance of the *c*-chart, as measured by the false alarm rate, is considerably degraded and unfavourably affected by a poor point estimate $\bar{c}$.

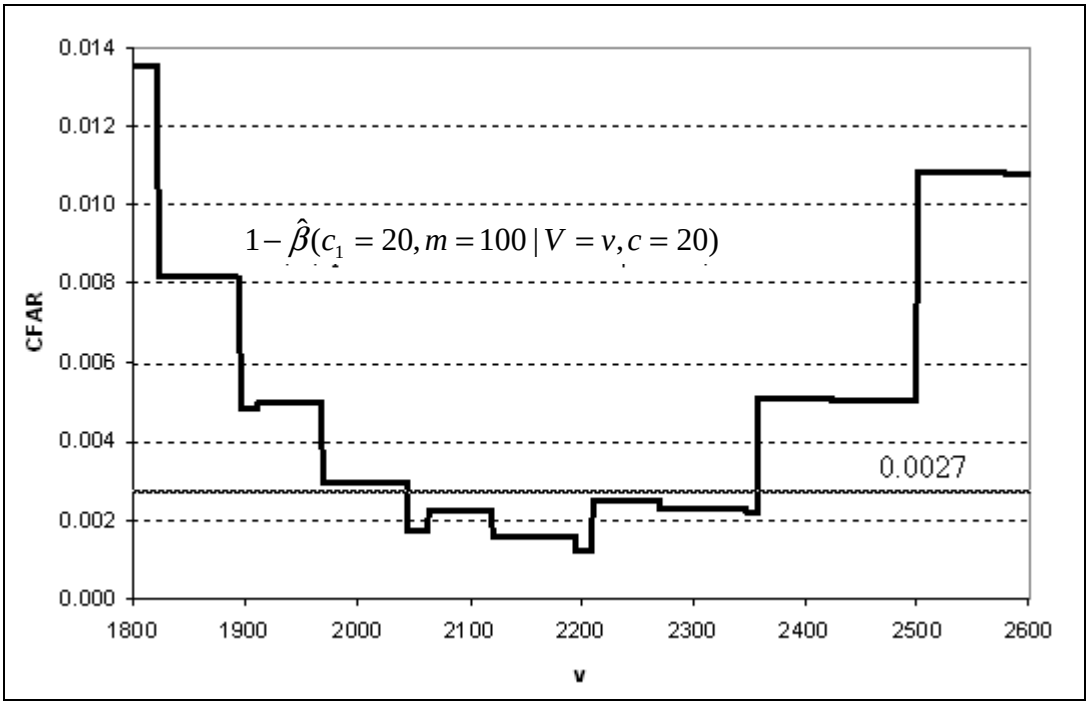


Figure 3.5: The conditional false alarm rate (*CFAR*) as a function of $v = 1800,...,2600$ when $m = 100$ in relation to the nominal *FAR* of 0.0027

The out-of-control conditional performance of the *c*-chart

The in-control performance of the Phase II *c*-chart (in theory) refers to the characteristics of the chart in the situation where the process operates at the same level in Phase II than what it did in Phase I; this is the scenario when $c_1 = c$. But, because c is unknown and estimated by $\bar{c}$, the observed value $\bar{c}_{\text{obs}}$ plays the role of c so that the conditional in-control performance (in practice) refers to the situation when $c_1 = \bar{c}_{\text{obs}}$ (see e.g. the earlier section labelled "Phase II implementation and operation"). The out-of-control performance (in practice) then refers to the characteristics of the *c*-chart when $c_1 \neq \bar{c}_{\text{obs}}$.

In view of the abovementioned, we can study the out-of-control performance of the Phase II *c*-chart by making use of the results from the previous section concerning the conditional characteristics of the Phase II *c*-chart. In particular, note that, by conditioning on a specific observed value $\bar{c}_{\text{obs}}$ the run-length distribution of the Phase II *c*-chart is affected in the same way it would be if the unknown true average number of nonconformities in an inspection was to change from c (in Phase I) to c_1 (in Phase II). In other words, the out-of-control performance of the Phase II *c*-chart (i.e. when c has incurred either a downward or an upward shift to c_1 so that $c_1 \neq c$) is equivalent to the performance of the *c*-chart when $c \neq \bar{c}_{\text{obs}}$ i.e. if c was either overestimated or underestimated (see e.g. Jones, Champ and Rigdon, (2004)); this correspondence allows us to examine the out-of-control performance of the *c*-chart by using the conditional statistical characteristics of the Phase II *c*-chart.

Consider, for example, Tables 3.13, 3.14 and 3.15 which list the false alarm rate (*CFAR*), the average run-length (*CARL*) and the standard deviation of the run-length (*CSDRL*) of the conditional run-length distribution assuming that $c = 5, 10, 15, 20 \text{ and } 30$ with $m = 10, 15, 20, 25, 50, 75, 100, 150$ and 200. For each combination of (m, c)-values the run-length distribution is conditioned on (for illustration purposes only) the 10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th} percentiles of the distribution of $V \mid c \sim Poi(mc)$.

In particular, suppose that $(m, c) = (20, 20)$ and we observed $V = 400$ so that our estimate $\bar{c} = 20$ is spot on. In this case, Table 3.14 shows that the conditional false alarm rate is

$$CFAR(c_1 = 20, m = 20 \mid V = 400, c = 20) = 0.0029,$$

the conditional average run-length is

$$CARL_0(c_1 = 20, m = 20 \mid V = 400, c = 20) = 1/0.0029 = 339.72$$

and the conditional standard deviation of the run-length is

$$CSDRL_0(c_1 = 20, m = 20 \mid V = 400, c = 20) = \sqrt{1 - 0.0029} / 0.0029 = 339.22.$$

These conditional characteristics i.e. conditioned on the 50th percentile of $V \mid c = 20 \sim Poi(400)$, are identical to the in-control characteristics of the Case K *c*-chart, that is, $1 - \beta(c = 20, c_0 = 20) = 0.0029$, $ARL(c = 20, c_0 = 20) = 339.72$ and $SDRL(c = 20, c_0 = 20) = 339.22$, which can be found from Table A3.12 in Appendix 3A. To illustrate the out-of-control (OOC) performance of the Case U *c*-chart we should condition on a percentile of $V \mid c \sim Poi(mc)$ other than the 50th percentile. To this end, consider again, for example, the situation when $(m, c) = (20, 20)$, but conditioning on the 25th percentile of $V \mid c = 20 \sim Poi(mc = 400)$, that is, $V = 386$ or $\bar{c} = 386 / 20 = 19.30$; this implies that c is underestimated by approximately $(20 / 19.3 - 1)100\% \approx 4\%$ or, equivalently, that the average number of nonconformities in an inspection unit has increased by 4%.

Table 3.14 shows that the $CFAR = 0.0050$, the $CARL_0 = 200.70$ and the $CSDRL_0 = 200.20$. Compared to the probability of a signal of $1 - \beta(c = 19.30, c_0 = 20) = 0.0020$, the OOC average run-length of $ARL(c = 19.30, c_0 = 20) = 507.85$ and the OOC standard deviation of the run-length of $SDRL(c = 19.30, c_0 = 20) = 507.35$ of Case K (which are found by evaluating expressions (3-7), (3-21) and (3-22), respectively) we observe that the Case U *c*-chart would detect an increase from 19.30 to 20 quicker than the *c*-chart of Case K. However, this is (as mentioned earlier in case of the *p*-chart) a side-effect of estimating c and not due to improved performance.

On the other hand, when $(m, c) = (20, 20)$, and we condition on the 90th percentile of $V \mid c = 20 \sim Poi(mc = 400)$, that is, $V = 426$ or $\bar{c} = 426 / 20 = 21.30$, which implies that c is overestimated by $(21.3 / 20 - 1)100\% = 6.5\%$ (or, equivalently, that the average number of nonconformities in an inspection unit has decreased by 6.5%), Table 3.14 shows that the $CFAR = 0.0016$, the $CARL_0 = 632.01$ and the $CSDRL_0 = 631.51$ which implies that the Case U *c*-chart performs worse than the Case K *c*-chart with probability of a signal of $1 - \beta(c = 21.3, c_0 = 20) = 0.0068$, an out-of-control *ARL* of $ARL(c = 21.30, c_0 = 20) = 146.15$ and an out-of-control *SDRL* of $SDRL(c = 21.30, c_0 = 20) = 145.65$.

Note that, when conditioning on a particular percentile of V, the OOC performance of the Case U *c*-chart is the same for two or more (m, c) combinations and thus the overlap of certain of the cells as

seen in Tables 3.13, 3.14 and 3.15. For example, the OOC performance of the Case U c-chart when (i) $(m,c)=(20,20)$ and conditioning on the 90^{th} percentile of $V \mid c=20 \sim Poi(mc=400)$, and (ii) $(m,c)=(15,20)$ and we condition on the 90^{th} percentile of $V \mid c=20 \sim Poi(15\times 20=300)$ i.e. $V=322$ so that $\bar{c}=21.47$ (which corresponds to an decrease of 7.35% in c from 21.47 to 20), are similar.

Table 3.13: The false alarm rate (*FAR*), the average run-length (*ARL*) and the standard deviation of the run-length (*SDRL*) of the conditional run-length distribution of the *c*-chart for *m* = 10, 15, 20, 25, 50, 75, 100, 150, 200 when *c* = 5 and 10

	c = 5									*c* = 10								
Percentile	***m* = 10**	**15**	**20**	**25**	**50**	**75**	**100**	**150**	**200**	**10**	**15**	**20**	**25**	**50**	**75**	**100**	**150**	**200**
10^{th} (OOC)	0.0204 48.94 48.44				0.0122 82.03 81.53					0.0143 69.82 69.32		0.0072 138.28 137.78						0.0035 285.74 285.23
25^{th} (OOC)	0.0204 48.94 48.44			0.0122 82.03 81.53						0.0072 138.28 137.78				0.0035 285.74 285.23				
50^{th} (IC)	*CFAR* = 0.0122 *CARL* = 82.03 *CSDRL* = 81.53									*CFAR* = 0.0035 *CARL* = 285.74 *CSDRL* = 285.23								
75^{th} (OOC)	0.0088 114.20 113.70					0.0122 82.03 81.53				0.0016 612.12 611.62				0.0035 285.74 285.23				
90^{th} (OOC)	0.0074 134.48 133.98			0.0088 114.20 113.70						0.0012 833.99 833.49		0.0016 612.12 611.62					0.0035 285.74 285.23	

Table 3.14: The false alarm rate (*FAR*), the average run-length (*ARL*) and the standard deviation of the run-length (*SDRL*) of the conditional run-length distribution of the *c*-chart for *m* = 10, 15, 20, 25, 50, 75, 100, 150, 200 when *c* = 15 and 20

	c = 15									*c* = 20								
Percentile	***m* = 10**	**15**	**20**	**25**	**50**	**75**	**100**	**150**	**200**	**10**	**15**	**20**	**25**	**50**	**75**	**100**	**150**	**200**
10^{th} (OOC)	0.0112 89.25 88.75			0.0062 160.66 160.16		0.0064 156.34 155.84		0.0035 283.83 283.33		0.0135 73.82 73.32		0.0082 122.49 121.99			0.0050 200.70 200.20			
25^{th} (OOC)	0.0062 160.66 160.16			0.0064 156.34 155.84		0.0035 283.83 283.33				0.0048 208.36 207.86		0.0050 200.70 200.20				0.0029 339.72 339.22		
50^{th} (IC)	*CFAR* = 0.0035 *CARL* = 283.83 *CSDRL* = 283.33									*CFAR* = 0.0029 *CARL* = 339.72 *CSDRL* = 339.22								
75^{th} (OOC)	0.0019 518.90 518.40							0.0035 283.83 283.33		0.0023 440.99 440.49		0.0017 573.34 572.84			0.0029 339.72 339.22			
90^{th} (OOC)	0.0017 582.29 581.79			0.0026 388.74 388.24		0.0019 518.90 518.40				0.0016 632.01 631.51		0.0023 440.99 440.49				0.0017 573.34 572.84		0.0029 339.72 339.22

Table 3.15: The false alarm rate (*FAR*), the average run-length (*ARL*) and the standard deviation of the run-length (*SDRL*) of the conditional run-length distribution of the *c*-chart for *m* = 10, 15, 20, 25, 50, 75, 100, 150, 200 when *c* = 30

	c = 30								
Percentile	***m* = 10**	**15**	**20**	**25**	**50**	**75**	**100**	**150**	**200**
10th (OOC)	0.0098 102.05 101.55	0.0064 155.37 154.87			0.0041 242.41 241.91		0.0044 229.10 228.60		
25th (OOC)	0.0064 155.37 154.87	0.0041 242.41 241.91		0.0044 229.10 228.60				0.0029 349.94 349.44	
50th (IC)	*CFAR* = 0.0029 *CARL* = 349.94 *CSDRL* = 349.44								
75th (OOC)	0.0024 415.11 414.61				0.0019 527.54 527.04		0.0029 349.94 349.44		
90th (OOC)	0.0025 405.45 404.95	0.0018 553.19 552.69			0.0024 415.11 414.61			0.0019 527.54 527.04	

3.2.4 Unconditional Phase II run-length distributions and characteristics

The conditional run-length distribution and the associated characteristics of the conditional run-length distribution present the performance of a chart only for one particular realization of the point estimator and a supposed value for the parameter. For each individual realization of $\bar{p} = U/mn$ or $\bar{c} = V/m$ and the true p or c value the performance of the chart will be different – some charts performing acceptable and others poorly.

In case of the *p*-chart the variable U can take on any value between and including 0 and mn i.e. $U \in \{0,1,...,mn\}$, so that there is a finite number $mn+1$ possible values on which we can condition. For the *c*-chart the variable V can be any positive integer greater or equal to zero i.e. $V \in \{0,1,2,...\}$, and so the number of possible values on which we can condition is infinite.

To avoid calculating the conditional performance of the charts for each realization of the point estimator and to asses the overall performance of the charts, the influence of a single realization should ideally be removed. The unconditional run-length distribution and its associated characteristics serve this purpose and better represent the overall performance of the charts when the parameters are estimated and let one see the bigger picture.

The unconditional characteristics of the charts can be found from the conditional run-length distribution by averaging over the distributions of U and V respectively, and allow us to look at the marginal (or the unconditional) run-length distribution. This incorporates the additional variation introduced to the run-length through the estimation of p and c by taking into account all possible realizations of the random variables on which we condition. In particular, we derive expressions for the:

(i) unconditional run-length distribution,
(ii) unconditional average run-length, and
(iii) unconditional variance of the run-length

of each chart.

Unconditional run-length distribution: *p*-chart and *c*-chart (Case U)

Because:

(i) the observations in the Phase I reference sample are assumed to be independent and identically distributed, that is, $X_i \sim iidBin(n, p)$ and $Y_i \sim iidPoi(c)$ for $i = 1,2,...,m$, and

(ii) we assume that the Phase I X_i's and Y_i's are independent from the Phase II observations i.e. $X_i \sim iidBin(n, p_1)$ and $Y_i \sim iidPoi(c_1)$ for $i = m+1, m+2,...$,

the joint probability distribution of

(i) the Phase I point estimator $U = mn\bar{p}$ and the Phase II run-length random variable N_p, and

(ii) the Phase I point estimator $V = m\bar{c}$ and the Phase II run-length random variable N_c

can straightforwardly be obtained (see e.g. Definition 4.2.1 in Casella and Berger, (2002) p. 148) as

$$\Pr(N_p = j, U = u; p_1, m, n \mid p) = \Pr(N_p = j; p_1, m, n \mid U = u, p).\Pr(U = u \mid p) \qquad (3\text{-}46)$$

and

$$\Pr(N_c = j, V = v; c_1, m \mid c) = \Pr(N_c = j; c_1, m \mid V = v, c).\Pr(V = v \mid c) \qquad (3\text{-}47)$$

for $j = 1,2,...$, $u = 0,1,...,mn$ and $v = 0,1,2,...$ where

$$\Pr(N_p = j; p_1, m, n \mid U = u, p) \quad \text{and} \quad \Pr(N_c = j; c_1, m \mid V = v, c)$$

are the conditional run-length distributions of the *p*-chart and the *c*-chart given in Tables 3.3 and 3.4, respectively, and

$$\Pr(U = u \mid p) = \binom{mn}{u} p^u (1-p)^{mn-u} \quad \text{for} \quad u = 0,1,2,...,mn$$

and

$$\Pr(V = v \mid c) = \frac{e^{-mc}(mc)^v}{v!} \quad \text{for} \quad v = 0,1,2,...$$

are the probability distributions of the estimators U and V, which depend on the unknown parameters p and c, respectively.

The marginal or unconditional run-length distributions are then found from the joint probability distributions and given by

$$\begin{aligned}\Pr(N_p = j; p_1, m, n \mid p) &= \sum_{u=0}^{mn} \Pr(N_p = j, U = u; p_1, m, n \mid p) \\ &= \sum_{u=0}^{mn} \Pr(N_p = j; p_1, m, n \mid U = u, p).\Pr(U = u \mid p) \\ &= \sum_{u=0}^{mn} \hat{\beta}(p_1, m, n \mid u, p)^{j-1}[1 - \hat{\beta}(p_1, m, n \mid u, p)].\binom{mn}{u} p^u (1-p)^{mn-u}\end{aligned} \tag{3-48}$$

and

$$\begin{aligned}\Pr(N_c = j; c_1, m \mid c) &= \sum_{v=0}^{\infty} \Pr(N_c = j, V = v; c_1, m \mid c) \\ &= \sum_{v=0}^{\infty} \Pr(N_c = j; c_1, m \mid V = v, c).\Pr(V = v \mid c) \\ &= \sum_{v=0}^{\infty} \hat{\beta}(c_1, m \mid v, c)^{j-1}[1 - \hat{\beta}(c_1, m \mid v, c)].\frac{e^{-mc}(mc)^v}{v!}\end{aligned} \tag{3-49}$$

for $j = 1,2,...$ (see e.g. Definition 4.2.1 in Casella and Berger, (2002) p. 148).

One can think of these unconditional distributions as weighted averages i.e. the conditional distributions averaged over all possible values of the parameter estimators, where a weight is the probability of obtaining a particular realization of the point estimator which is given by $\Pr(U = u \mid p)$ or $\Pr(V = v \mid p)$.

It is important to note that the unconditional run-length distributions in (3-48) and (3-49) are unconditional only with respect to the random variables U and V; the unconditional run-length distributions still depend on the parameters p and c. This means that when we evaluate the unconditional run-length distributions and the associated characteristics of the unconditional run-length distributions, the results apply only for those particular values of p and c that are used.

The unconditional average run-length and the unconditional variance of the run-length distributions

Apart from the unconditional run-length distributions we can also compute higher order moments of the unconditional run-length distribution.

The unconditional k^{th} non-central moments, for example, are

$$E(N_p^k \mid p) = E_U(E(N_p^k \mid U, p)) \quad \text{and} \quad E(N_c^k \mid c) = E_V(E(N_c^k \mid V, c))$$

where

$$E(N_p^k \mid U, p) \quad \text{and} \quad E(N_c^k \mid V, c)$$

are the k^{th} non-central moments of the conditional run-length distributions of the *p*-chart and *c*-chart, respectively (see e.g. Theorem 5.4.4 in Bain and Engelhardt, (1992) p. 183).

In particular, when $k = 1$ we have that the unconditional average run-length, denoted by *UARL*, which are

$$UARL_p = E(N_p \mid p) = E_U(E(N_p \mid U, p)) \quad \text{and} \quad UARL_c = E(N_c \mid c) = E_V(E(N_c \mid V, c))$$

where

$$E(N_p \mid U, p) = (1 - \hat{\beta}(p_1, m, n \mid U, p))^{-1} \quad \text{and} \quad E(N_c \mid V, c) = ((1 - \hat{\beta}(c_1, m \mid V, c))^{-1}$$

are the conditional *ARL*'s (conditioned on particular observations of the random variables U and V), respectively.

Hence, it follows that

$$\begin{aligned} UARL(p_1, m, n \mid p) &= \sum_{u=0}^{mn} (1 - \hat{\beta}(p_1, m, n \mid u, p))^{-1} \Pr(U = u \mid p) \\ &= \sum_{u=0}^{mn} (1 - \hat{\beta}(p_1, m, n \mid u, p))^{-1} \binom{mn}{u} p^u (1-p)^{mn-u} \end{aligned} \tag{3-50}$$

and

$$\begin{aligned} UARL(c_1, m \mid c) &= \sum_{v=0}^{\infty} (1 - \hat{\beta}(c_1, m \mid v, c))^{-1} \Pr(V = v \mid c) \\ &= \sum_{v=0}^{\infty} (1 - \hat{\beta}(c_1, m \mid v, c))^{-1} \frac{e^{-mc}(mc)^v}{v!}. \end{aligned} \tag{3-51}$$

Similarly, the unconditional variance of the run-length, denoted by *UVARL*, can be found using

(i) the conditional variance of the run-length (*CVAR*),
(ii) the conditional average run-length (*CARL*), and
(iii) the unconditional average run-length (*UARL*),

and is given by

$$UVARL = E_Z(CVARL) + E_Z(CARL^2) - UARL^2 . \qquad (3\text{-}52a)$$

where Z plays the role of U and/or V .

Result (3-52a) follows from the fact that, in general, the unconditional variance can be obtained from the expected value of the conditional variance and the variance of the conditional expected value i.e.

$$\begin{aligned} \text{var}(N) &= E_Z(\text{var}(N \mid Z)) + \text{var}_Z(E(N \mid Z)) \\ &= E_Z(\text{var}(N \mid Z)) + \{E_Z[(E(N \mid Z))^2] - [E_Z(E(N \mid Z))]^2\} \qquad (3\text{-}52b) \\ UVARL &= E_Z(CVARL) + E_Z(CARL^2) - UARL^2 \end{aligned}$$

where $\text{var}(N)$ is the unconditional variance of the run-length ,

$$CVARL = \text{var}(N \mid Z) = \hat{\beta}/(1-\hat{\beta})^2$$

denotes the conditional variance of the run-length,

$$CARL = E(N \mid Z) = 1/(1-\hat{\beta})$$

denotes the conditional average run-length, $\hat{\beta}$ denotes (in general) the conditional probability of a no-signal and Z plays the role of U and/or V , which is the random variable on which we condition in the particular case (see e.g. Theorem 5.4.3 in Bain and Engelhardt, (1992) p. 182).

In case of the *p*-chart, using (3-52a), the unconditional variance of the run-length is

$$\begin{aligned}
UVARL_p &= E_U\left(CVARL\right)+\left\{E_U\left(CARL^2\right)-[UARL]^2\right\} \\
&= E_U\left(\frac{\hat{\beta}(p_1,m,n\,|\,U,p)}{(1-\hat{\beta}(p_1,m,n\,|\,U,p))^2}\right)+\left\{E_U\left(\frac{1}{(1-\hat{\beta}(p_1,m,n\,|\,U,p))^2}\right)-\left[E_U\left(\frac{1}{1-\hat{\beta}(p_1,m,n\,|\,U,p)}\right)\right]^2\right\} \\
&= E_U\left(\frac{1+\hat{\beta}(p_1,m,n\,|\,U,p)}{(1-\hat{\beta}(p_1,m,n\,|\,U,p))^2}\right)-\left[E_U\left(\frac{1}{1-\hat{\beta}(p_1,m,n\,|\,U,p)}\right)\right]^2 \\
&= \sum_{u=0}^{mn}\left(\frac{1+\hat{\beta}(p_1,m,n\,|\,u,p)}{(1-\hat{\beta}(p_1,m,n\,|\,u,p))^2}\right)\binom{mn}{u}p^u(1-p)^{mn-u}-\left[\sum_{u=0}^{mn}\left(\frac{1}{1-\hat{\beta}(p_1,m,n\,|\,u,p)}\right)\binom{mn}{u}p^u(1-p)^{mn-u}\right]^2
\end{aligned}$$

whilst for the *c*-chart we have

$$\begin{aligned}
UVARL_c &= E_V\left(CVARL\right)+\left\{E_V\left(CARL^2\right)-[UARL]^2\right\} \\
&= E_V\left(\frac{\hat{\beta}(c_1,m\,|\,V,c)}{(1-\hat{\beta}(c_1,m\,|\,V,c))^2}\right)+\left\{E_V\left(\frac{1}{(1-\hat{\beta}(c,c_1,m\,|\,V,c))^2}\right)-\left[E_V\left(\frac{1}{1-\hat{\beta}(c,c_1,m\,|\,V,c)}\right)\right]^2\right\} \\
&= E_V\left(\frac{1+\hat{\beta}(c_1,m\,|\,V,c)}{(1-\hat{\beta}(c_1,m\,|\,V,c))^2}\right)-\left[E_V\left(\frac{1}{1-\hat{\beta}(c_1,m\,|\,V,c)}\right)\right]^2 \\
&= \sum_{v=0}^{\infty}\left(\frac{1+\hat{\beta}(c_1,m\,|\,v,c)}{(1-\hat{\beta}(c_1,m\,|\,v,c))^2}\right)\frac{e^{-mc}(mc)^v}{v!}-\left[\sum_{v=0}^{\infty}\left(\frac{1}{1-\hat{\beta}(c_1,m\,|\,v,c)}\right)\frac{e^{-mc}(mc)^v}{v!}\right]^2.
\end{aligned}$$

The unconditional standard deviation of the run-length follows by taking the square root of the unconditional variance of the run-length i.e. $USDRL=\sqrt{UVARL}$.

The unconditional probability mass function (u.p.m.f), the unconditional cumulative distribution function (u.c.d.f), the unconditional false alarm rate (*UFAR*), the unconditional average run-length (*UARL*), and the unconditional variance of the run-length (*UVARL*) for the *p*-chart and the *c*-chart are summarized in Tables 3.16 and 3.17, respectively.

These characteristics, as mentioned before, are important as they help us understand the full impact of estimating the unknown parameters on the performance of the charts. Note, however, that when evaluating the unconditional distributions and the unconditional characteristics in Tables 3.16 and 3.17

one still has to select values for p and c; hence, the results are only applicable to the particular values of p and c that are selected.

Table 3.16: The unconditional probability mass function (u.p.m.f), the unconditional cumulative distribution function (u.c.d.f), the unconditional false alarm rate (*UFAR*), the unconditional average run-length (*UARL*) and the unconditional variance of the run-length (*UVARL*) of the *p*-chart in Case U

u.p.m.f	$\Pr(N_p = j; p_1, m, n \mid p) = \sum_{u=0}^{mn} (\hat{\beta}(p_1, m, n \mid u, p)^{j-1}(1-\hat{\beta}(p_1, m, n \mid u, p)) \binom{mn}{u} p^u (1-p)^{mn-u}$	(3-53)
u.c.d.f	$\Pr(N_p \le j; p_1, m, n \mid p) = \sum_{u=0}^{mn} (1-(\hat{\beta}(p_1, m, n \mid u, p))^j) \binom{mn}{u} p^u (1-p)^{mn-u} \quad j = 1,2,....$	(3-54)
UFAR	$UFAR(p_1, m, n \mid p = p_1) = \sum_{u=0}^{mn} (1-\hat{\beta}(p_1, m, n \mid u, p = p_1)) \binom{mn}{u} p_1^u (1-p_1)^{mn-u}$	(3-55)
UARL	$UARL(p_1, m, n \mid p) = \sum_{u=0}^{mn} (1-\hat{\beta}(p_1, m, n \mid u, p))^{-1} \binom{mn}{u} p^u (1-p)^{mn-u}$	(3-56)
UVARL	$UVARL_p = \sum_{u=0}^{mn} \left(\frac{1+\hat{\beta}(p_1, m, n \mid u, p)}{(1-\hat{\beta}(p_1, m, n \mid u, p))^2} \right) \binom{mn}{u} p^u (1-p)^{mn-u} - \left[\sum_{u=0}^{mn} \left(\frac{1}{1-\hat{\beta}(p_1, m, n \mid u, p)} \right) \binom{mn}{u} p^u (1-p)^{mn-u} \right]^2$	(3-57)

Table 3.17: The unconditional probability mass function (u.p.m.f), the unconditional cumulative distribution function (u.c.d.f.), the unconditional false alarm rate (*UFAR*), the unconditional average run-length (*UARL*) and the unconditional variance of the run-length (*UVARL*) of the *c*-chart in Case U

u.p.m.f	$\Pr(N_c = j; c_1, m \mid c) = \sum_{v=0}^{\infty} \hat{\beta}(c_1, m \mid v, c)^{j-1}(1-\hat{\beta}(c_1, m \mid v, c)) \frac{e^{-mc}(mc)^v}{v!} \quad j = 1,2,....$	(3-58)
u.c.d.f	$\Pr(N_c \le j; c_1, m \mid c) = \sum_{v=0}^{\infty} 1-(\hat{\beta}(c_1, m \mid v, c))^j \frac{e^{-mc}(mc)^v}{v!} \quad j = 1,2,....$	(3-59)
UFAR	$UFAR(c_1, m \mid c = c_1) = \sum_{v=0}^{\infty} 1-\hat{\beta}(c_1, m \mid v, c = c_1) \frac{e^{-mc_*}(mc_1)^v}{v!}$	(3-60)
UARL	$UARL(c_1, m \mid c) = \sum_{v=0}^{\infty} (1-\hat{\beta}(c_1, m \mid v, c))^{-1} \frac{e^{-mc}(mc)^v}{v!}$	(3-61)
UVARL	$UVARL_c = \sum_{v=0}^{\infty} \left(\frac{1+\hat{\beta}(c_1, m \mid v, c)}{(1-\hat{\beta}(c_1, m \mid v, c))^2} \right) \frac{e^{-mc}(mc)^v}{v!} - \left[\sum_{v=0}^{\infty} \left(\frac{1}{1-\hat{\beta}(c_1, m \mid v, c)} \right) \frac{e^{-mc}(mc)^v}{v!} \right]^2$	(3-62)

3.2.4.1 Unconditional characteristics of the *p*-chart

The necessary steps and calculations to obtain a numerical value for a particular unconditional characteristic of the Phase II run-length distribution of the *p*-chart are explained via the examples shown in Tables 3.18 and 3.19; these tables are essentially the same as Tables 3.5 and 3.6 that we used to illustrate the mechanics for calculating the *FAR* , the *ARL* and the *SDRL* of the conditional run-length distribution. However, here, we go a step further and calculate the unconditional characteristics of the run-length distribution, that is, the unconditional *FAR* (*UFAR*), the unconditional *ARL* (*UARL*) and the unconditional *SDRL* (*USDRL*). In addition, note that, although we still assume that $p_1 = p = 0.5$ we now assume that $T = mn = 15$ with $(m,n) = (1,15)$ and $(m,n) = (3,5)$ individual Phase I reference observations are used to estimate p .

T = 15 with m = 1 and n = 15

First consider Table 3.18 which assumes that $(m,n) = (1,15)$. Recall that to calculate the conditional properties we begin in column 1 and sequentially move to the right-hand side of the table up to column 9. To illustrate the concept once more, assume that we observe nine nonconforming items from the entire fifteen reference observations i.e. suppose that $U = 9$, so that we get a point estimate of $\bar{p}_{\text{obs}} = 9/15 = 0.6$ for the unknown true fraction nonconforming $0 < p < 1$ in column 2. Thus, using (3-26), we find that the estimated control limits are $L\hat{C}L_p = 0.22$ and $U\hat{C}L_p = 0.98$; these values are listed in columns 3 and 4, respectively. Then, making use of (3-31) we find that the charting constants are $\hat{a} = 3$ and $\hat{b} = 14$ (which are listed in columns 5 and 6, respectively) so that (3-36) yields a conditional false alarm rate of $CFAR(p_1 = 0.5, m = 1, n = 15 \mid U = 9, p = 0.5) = 0.0176$ which leads to a conditional average run-length and a conditional variance of the run-length (found from (3-37) and (3-38)) of

$$CARL(p_1 = 0.5, m = 1, n = 15 \mid U = 9, p = 0.5) = 56.79$$

and

$$CVARL = \left[CSDRL(p_1 = 0.5, m = 1, n = 15 \mid U = 9, p = 0.5)\right]^2 = 3168.35;$$

these values are displayed in columns 8 and 9, respectively.

To calculate the unconditional properties of the *p*-chart we calculate a weighted average of all the values (rows) for each of columns 7, 8 and 9, respectively. The weights are found from the probability

distribution of the random variable $U \mid p = 0.5 \sim Bin(15,0.5)$ which is given in column 10 and calculated from evaluating $\Pr(U = u \mid p = 0.5) = \binom{15}{u} 0.5^{15}$ for $u = 0,1,...,15$.

Table 3.18: The conditional and unconditional characteristics of the run-length distribution for $m = 1$ and $n = 15$ when $p = p_1 = 0.5$

Phase I						Phase II : Conditional Properties				Phase II : Unconditional Properties			
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)=(7)x(10)	(12)=(8)x(10)	(13)=(8)^{2}x(10)	(14)=(9)x(10)
U	$\bar{p}_{obs}$	$L\hat{C}L_p$	$U\hat{C}L_p$	$\hat{a}$	$\hat{b}$	$CFAR$	$CARL$	$CVARL$	$\Pr(U=u\|p)$	$CFAR.\Pr(U=u\|p)$	$CARL.\Pr(U=u\|p)$	$CARL^2.\Pr(U=u\|p)$	$CVARL.\Pr(U=u\|p)$
0	0.00	0.00	0.00	NA	NA	1.0000	1.00	0.00	0.0000	0.00003	0.00003	0.0000	0.0000
1	0.07	-0.13	0.26	NA	3	0.9824	1.02	0.02	0.0005	0.00045	0.00047	0.0005	0.0000
2	0.13	-0.13	0.40	NA	5	0.8491	1.18	0.21	0.0032	0.00272	0.00377	0.0044	0.0007
3	0.20	-0.11	0.51	NA	7	0.5000	2.00	2.00	0.0139	0.00694	0.02777	0.0555	0.0278
4	0.27	-0.08	0.61	NA	9	0.1509	6.63	37.30	0.0417	0.00629	0.27609	1.8299	1.5538
5	0.33	-0.03	0.70	NA	10	0.0592	16.88	268.12	0.0916	0.00543	1.54714	26.1189	24.5717
6	0.40	0.02	0.78	0	11	0.0176	56.79	3168.35	0.1527	0.00269	8.67418	492.60905	483.9349
7	0.47	0.08	0.85	1	12	0.0042	239.18	56969.08	0.1964	0.00082	46.97080	11234.5932	11187.6224
8	0.53	0.15	0.92	2	13	0.0042	239.18	56969.08	0.1964	0.00082	46.97080	11234.5932	11187.6224
9	0.60	0.22	0.98	3	14	0.0176	56.79	3168.35	0.1527	0.00269	8.67418	492.6091	483.9349
10	0.67	0.30	1.03	4	15	0.0592	16.88	268.12	0.0916	0.00543	1.54714	26.1189	24.5717
11	0.73	0.39	1.08	5	15	0.1509	6.63	37.30	0.0417	0.00629	0.27609	1.8299	1.5538
12	0.80	0.49	1.11	7	15	0.5000	2.00	2.00	0.0139	0.00694	0.02777	0.0555	0.0278
13	0.87	0.60	1.13	9	15	0.8491	1.18	0.21	0.0032	0.00272	0.00377	0.0044	0.0007
14	0.93	0.74	1.13	11	15	0.9824	1.02	0.02	0.0005	0.00045	0.00047	0.0005	0.0000
15	1.00	1.00	1.00	NA	NA	1.0000	1.00	0.00	0.0000	0.00003	0.00003	0.0000	0.0000
										0.05074	**115.00**	**23510.42**	**23395.42**
										UFAR	***UARL***	***USDRL* = 183.52**	

Unconditional false alarm rate

To obtain the unconditional false alarm rate (*UFAR*), we need the conditional false alarm rate and the related probability $\Pr(U = u \mid p = 0.5)$ for $u = 0,1,...,15$, which are listed in columns 7 and 10, respectively. Multiplying corresponding row entries of column 7 and column 10, we end up with column 11, that is,

$$CFAR(p_1 = 0.5,1,15 \mid U = u, p = 0.5) \times \Pr(U = u \mid p = 0.5) = (1 - \hat{\beta}(0.5,1,15 \mid U, p)) \times \Pr(U = u \mid p = 0.5)$$

for $u = 0,1,...,15$ so that adding up all the entries in column 11 yields the unconditional false alarm rate i.e.

$$UFAR(p_1 = 0.5,1,15 \mid p = 0.5) = \sum_{u=0}^{15} CFAR(0.5,1,15 \mid U = u, p = 0.5) \Pr(U = u \mid p = 0.5) = 0.05074$$

(see e.g. (3-55) in Table 3.16). The unconditional *FAR* value implies that the probability of a signal on any new incoming Phase II sample, for any practitioner, while the process is in-control at a fraction nonconforming of 0.5, is expected to be 0.05074.

Unconditional average run-length

Like the unconditional *FAR*, the unconditional *ARL* is found by multiplying each of the conditional average run-length values listed in column 8 with the corresponding probability $\Pr(U = u \mid p = 0.5)$ listed in column 10 and then adding up all the resultant products.

To this end, column 12 lists all the values of

$$CARL(p_1 = 0.5,1,15 \mid U = u, p = 0.5) \times \Pr(U = u \mid p = 0.5) \qquad \text{for} \qquad u = 0,1,...,15$$

so that by totalling the values of column 12 we find the unconditional average run-length to be

$$UARL(p_1 = 0.5,1,15 \mid p = 0.5) = \sum_{u=0}^{15} CARL(p_1 = 0.5,1,15 \mid U = u, p = 0.5) \Pr(U = u \mid p = 0.5) = 115.00$$

(see Table 3.16, (3-56)).

An unconditional *ARL* of 115.00 means that a practitioner that estimates p using $\bar{p} = U / mn$, (which is based on a Phase I reference sample that consists of a total of $T = 15$ individual observations from 1 sample of size 15) can expect that his Phase II *p*-chart would, on average, signal on the 115th sample when the process remains in-control at a fraction nonconforming of 0.5.

Unconditional variance of the run-length

Using expression (3-52a) to calculate the unconditional variance of the run-length we note that, $E_U(CVARL) = 23395.42$ (listed in column 14), $E_U(CARL^2) = 23510.42$ (listed in column 13) so that the unconditional standard deviation of the run-length is found to be

$$USDRL = \sqrt{E_U(CVARL) + E_U(CARL^2) - UARL^2} = \sqrt{23395.42 + 23510.42 - (115.00)^2} = 183.52 .$$

The unconditional standard deviation is the same for all the users and measures the overall variation in the run-length distribution.

Remark 7

In particular, note that, for $T = 15$ where $(m, n) = (1, 15)$:

(i) The unconditional average run-length is not equal to the reciprocal of the unconditional false alarm rate i.e. $UARL \neq (UFAR)^{-1}$. The reason is that the unconditional run-length distribution is not geometric (see e.g. expression (3-53) in Table 3.16).

This is unlike in Case K where $ARL = (FAR)^{-1}$ (see e.g. expression (3-12) in Table 3.1), which makes both the average run-length and false alarm rate popular measures of a control chart's performance.

(ii) The unconditional average run-length is smaller than the unconditional standard deviation of the run-length; this is not the situation in Case K where $ARL > SDRL = \sqrt{ARL(ARL-1)}$ (see e.g. Appendix 3A, section 3.4.2.2) and is due to extra variation introduced to the run-length distribution when estimating p.

(iii) The unconditional FAR is greater than the FAR of 0.0010 of Case K whilst the unconditional ARL and the unconditional $SDRL$ is less than the ARL of 1024.00 and $SDRL$ of 1023.50 of Case K, respectively.

This implies that a Phase II p-chart in Case U, based on an estimate of p using $T = 15$ observations, will signal more often than the Case K p-chart with a known standard.

$T = 15$ with $m = 3$ and $n = 5$

To study the effect of choosing a smaller value of n relative to m (i.e. changing the composition of the reference sample while keeping the total number of Phase I observation the same) on the unconditional characteristics of the run-length distribution, Table 3.19 shows the calculations necessary to obtain the unconditional *FAR*, the unconditional *ARL* and the unconditional *SDRL* when $T = 15$ with $m = 3$ and $n = 5$.

Although the steps in calculating the values in Table 3.19 are similar to that of Table 3.18, we note that the finer points where the $CFAR = 0$, are somewhat lost when we look at the unconditional FAR, which is found by averaging the conditional FAR (given in column (7)) over all fifteen values of U and their associated probabilities (as given in column (10)). For example, from column (11) in Table 3.19 an unconditional *FAR* equal to 0.01726 is found, which is more than six times the nominal false alarm rate of 0.0027; in spite of this, the unconditional ARL and the unconditional $SDRL$ are still undefined. One can therefore deduce that three subgroups each consisting of five in-control observations do not work satisfactorily in practice.

Table 3.19: The conditional and unconditional characteristics of the run-length distribution for $m = 3$ and $n = 5$ when $p = p_1 = 0.5$

Phase I						Phase II : Conditional Properties				Phase II : Unconditional Properties			
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)=(7)x(10)	(12)=(8)x(10)	(13)=(8)^{2}x(10)	(14)=(9)x(10)
U	$\bar{p}_{obs}$	$L\hat{C}L_p$	$U\hat{C}L_p$	$\hat{a}$	$\hat{b}$	$CFAR$	$CARL$	$CVARL$	$\Pr(U=u\|p)$	$CFAR.\Pr(U=u\|p)$	$CARL.\Pr(U=u\|p)$	$CARL^2.\Pr(U=u\|p)$	$CVARL.\Pr(U=u\|p)$
0	0.00	0	0	NA	NA	1.0000	1.00	0.00	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.07	-0.27	0.40	NA	2	0.5000	2.00	2.00	0.0005	0.0002	0.0009	0.0018	0.0009
2	0.13	-0.32	0.59	NA	2	0.5000	2.00	2.00	0.0032	0.0016	0.0064	0.0128	0.0064
3	0.20	-0.34	0.74	NA	3	0.1875	5.33	23.11	0.0139	0.0026	0.0741	0.3950	0.3209
4	0.27	-0.33	0.86	NA	4	0.0313	32.00	992.00	0.0417	0.0013	1.3330	42.6563	41.3232
5	0.33	-0.30	0.97	NA	4	0.0313	32.00	992.00	0.0916	0.0029	2.9326	93.8438	90.9111
6	0.40	-0.26	1.06	NA	5	0.0000	∞	∞	0.1527	0.0000	∞	∞	∞
7	0.47	-0.20	1.14	NA	5	0.0000	∞	∞	0.1964	0.0000	∞	∞	∞
8	0.53	-0.14	1.20	NA	5	0.0000	∞	∞	0.1964	0.0000	∞	∞	∞
9	0.60	-0.06	1.26	NA	5	0.0000	∞	∞	0.1527	0.0000	∞	∞	∞
10	0.67	0.03	1.30	0	5	0.0313	32.00	992.00	0.0916	0.0029	2.9326	93.8438	90.9111
11	0.73	0.14	1.33	0	5	0.0313	32.00	992.00	0.0417	0.0013	1.3330	42.6563	41.3232
12	0.80	0.26	1.34	1	5	0.1875	5.33	23.11	0.0139	0.0026	0.0741	0.3950	0.3209
13	0.87	0.41	1.32	2	5	0.5000	2.00	2.00	0.0032	0.0016	0.0064	0.0128	0.0064
14	0.93	0.60	1.27	2	5	0.5000	2.00	2.00	0.0005	0.0002	0.0009	0.0018	0.0009
15	1.00	1.00	1.00	NA	NA	1.0000	1.00	0.00	0.0000	0.0000	0.0000	0.0000	0.0000
										0.01726	∞	∞	∞
										UFAR	***UARL***	***USDRL*** = ∞	

To illustrate and help understand the overall effects of parameter estimation on the properties of the *p*-chart in more detail, some results (similar to those in Tables 3.18 and 3.19) are presented in Tables 3.20, 3.21 and 3.22 for $T = 10, 20, 25, 30, 50, 75, 100, 200, 250, 300, 500, 750, 1000$ and 1500, each time considering several combinations of m and n values so that $T = mn$. Thus, we look at what happens to the unconditional characteristics (in particular the *UFAR* and the *UARL*) when:

(a) T increases, and
(b) when the composition of the Phase I sample changes i.e. varying m and n.

The resulting unconditional *FAR*'s and the unconditional in-control *ARL*'s are listed under *UFAR* and $UARL_0$, respectively. Also shown is the percentage difference of the unconditional *FAR* and in-control unconditional *ARL* of Case U versus

(a) the *FAR* and *ARL* of Case K (see e.g. Tables A3.4 and A3.5 in Appendix 3A), and
(b) the nominal *FAR* of 0.0027 and the nominal *ARL* of 370.

Several interesting facts emerge from an examination of the results in Tables 3.20, 3.21 and 3.22:

(i) A lot of reference data is needed before the *UFAR* is anywhere near the nominal value of 0.0027 implicitly expected in a typical application of the *p*-chart. In addition, the choice of the number of subgroups m and the subgroup size n are both seen to be important.

For example, the calculations show that unlike in the case with variables data, when studying attributes data the subgroup size n needs to be much larger than the number of subgroups m, to ensure that the *UFAR* is reasonably close to the nominal value and (at the same time) ensure that the *UARL* is not undefined (see e.g. Table 3.21 where $T = 300$ with $m = 10$ and $n = 30$).

(ii) There is great variation in the *UFAR* values and it could be hundreds of percents off from its nominal value of 0.0027 and/or its Case K value for many combinations of m and n that are typically used in practice.

For example, when $T = 100$, we find that

(a) with $m = 4$ and $n = 25$ the *UFAR* is 191.5% above the nominal value and 92% above its Case K value of 0.0041, and
(b) with $m = 20$ and $n = 5$ the *UFAR* is 95.9% lower than the nominal value but close to its Case K value of zero.

(iii) Unless one is careful about the choice of *m* and *n*, the unconditional in-control average run-length of the chart can be undefined particularly when $m >> n$, which is undesirable in practice. This is due to the fact that the conditional probability of a false alarm can be zero for certain values of m and n, since although U can take on any integer value between 0 and mn (including both) with a non-zero probability, the binomial distribution (for the number of nonconforming items within each monitored group) assigns zero probability to any value greater than n.

(iv) The effect of the discreteness of the binomial distribution is also seen to be substantial on both the *FAR* and *ARL* values. For example, unlike in the variables case, with attributes data, only a certain number of ARL_0 values are attainable depending on the combination of values of m, n and p the user has at hand.

(v) As mentioned before, unlike in Case K, the unconditional *ARL* is not equal to the reciprocal of the unconditional *FAR* nor is it smaller than the unconditional *SDRL* (not listed here); this is an important effect of estimating the unknown parameter p.

(vi) For the (m,n) combinations where $n \leq 8$ the Case K *FAR* is zero and the associated Case K *ARL* is undefined (see e.g. Tables A3.4 and A3.5 in Appendix 3A).

In these cases, it is not practical to calculate the percentage difference and therefore indicated by an asterisk. In addition, for those (m,n) combinations where the *UFAR* is zero and/or the *UARL* is undefined it is impractical to calculate the percentage difference from the nominal values and thus indicated by the hash sign.

The aforementioned results suggest that there is a need for a large amount of reference data, with a larger amount of data in each subgroup than the number of subgroups i.e. $n >> m$. For example, when $T = 200$ with $m = 8$ and $n = 25$, the *UFAR* is 0.00447 which is 65.5% above the nominal value, whereas when $T = 500$, both $(m,n) = (25,20)$ and $(m,n) = (20,25)$ lead to an unconditional false alarm rate close to the nominal. This suggests one would need at least 400-500 in-control reference data points to achieve any meaningful control of the false alarm rate near the nominal 0.0027. An examination of the *UARL* values also lead to similar conclusions, in the sense that the combination of the number of subgroups and the size of the subgroup play an important role in dictating the (stable) properties of the *p*-chart.

Table 3.20: The unconditional false alarm rate ($UFAR$) and the unconditional in-control average run-length ($UARL_0$) values for the p-chart for various values of m and n such that $T = mn$ when $p = p_1 = 0.5$

	m	n	$UFAR$	$UARL_0$	% difference from Case K FAR[1]	% difference from Case K ARL[2]	% difference from nominal FAR=0.0027[3]	% difference from nominal ARL=370[4]
T = 10	1	10	0.06896	168.73	3348.1	203.5	2454.2	-54.4
	2	5	0.03552	∞	*	*	1215.6	#
	5	2	0.00684	∞	*	*	153.2	#
	10	1	0.01172	∞	*	*	334.0	#
T = 20	1	20	0.04553	135.62	1651.0	186.1	1586.2	-63.3
	2	10	0.01913	455.94	856.4	12.3	608.5	23.2
	4	5	0.01021	∞	*	*	278.1	#
	5	4	0.00787	∞	*	*	191.4	#
	10	2	0.00065	∞	*	*	-76.1	#
	20	1	0.00020	∞	*	*	-92.5	#
T = 25	1	25	0.04567	171.89	1014.0	42.7	1591.7	-53.5
	5	5	0.00405	∞	*	*	50.1	#
	25	1	0.00001	∞	*	*	-99.6	#
T = 30	1	30	0.04287	235.16	2962.0	197.2	1487.7	-36.4
	2	15	0.01765	288.78	1665.1	254.6	553.7	-22.0
	3	10	0.01119	605.83	459.4	-15.5	314.3	63.7
	5	6	0.00724	∞	*	*	168.2	#
	6	5	0.00333	∞	*	*	23.2	#
	10	3	0.00066	∞	*	*	-75.7	#
	15	2	0.00008	∞	*	*	-97.0	#
	30	1	0	∞	*	*	#	#
T = 50	1	50	0.03686	140.47	1317.6	173.6	1265.1	-62.0
	2	25	0.01838	171.32	348.2	43.2	580.6	-53.7
	5	10	0.00600	553.53	200.1	-7.5	122.3	49.6
	10	5	0.00104	∞	*	*	-61.5	#
	25,2	50,1	0	∞	*	*	#	#
T = 75	1	75	0.04094	105.69	1606.0	287.1	1416.4	-71.4
	3	25	0.01022	254.50	149.2	-3.6	278.5	-31.2
	5	15	0.00612	492.82	512.3	107.8	126.8	33.2
	15	5	0.00033	∞	*	*	-87.7	#
	25,3	75,1	0	∞	*	*	#	#
T = 100	1	100	0.04006	108.45	1044.5	162.1	1383.6	-70.7
	2	50	0.01475	234.72	467.3	63.7	446.3	-36.6
	4	25	0.00787	246.68	92.0	-0.6	191.5	-33.3
	5	20	0.00577	348.72	122.0	11.3	113.7	-5.8
	10	10	0.00332	647.93	65.9	-21.0	22.9	75.1
	20	5	0.00011	∞	*	*	-95.9	#
	25,4 100,1	50,2	0	∞	*	*	#	#

[1] $\%\ \text{deviation} = 100(UFAR/FAR_{\text{Case K}} - 1)$; [2] $\%\ \text{deviation} = 100(UARL_0/ARL_{\text{Case K}} - 1)$;

[3] $\%\ \text{deviation} = 100(UFAR/0.0027 - 1)$; [4] $\%\ \text{deviation} = 100(UARL_0/370 - 1)$

Table 3.21: The unconditional false alarm rate (*UFAR*) and the unconditional in-control average run-length ($UARL_0$) values for the *p*-chart for various values of m and n such that $T = mn$ when $p = p_1 = 0.5$

	m	n	*UFAR*	$UARL_0$	% difference from Case K *FAR*[1]	% difference from Case K *ARL*[2]	% difference from nominal *FAR*=0.0027[3]	% difference from nominal *ARL*=370[4]
$T = 200$	**1**	**200**	0.03328	162.34	1347.0	170.2	1132.6	-56.1
	2	**100**	0.01587	164.72	353.3	72.6	487.6	-55.5
	4	**50**	0.00708	259.93	172.2	47.8	162.1	-29.7
	5	**40**	0.00593	275.48	169.4	63.4	119.5	-25.5
	8	**25**	0.00447	312.51	9.0	-21.5	65.5	-15.5
	10	**20**	0.00374	409.26	43.8	-5.2	38.5	10.6
	20	**10**	0.00207	683.52	3.7	-25.1	-23.2	84.7
	25	**8**	0.00171	∞	*	*	-36.7	#
	40,5 100,2	**50,4 200,1**	0	∞	*	*	#	#
$T = 250$	**1**	**250**	0.03546	131.33	1122.7	164.5	1213.3	-64.5
	2	**125**	0.01440	187.77	554.7	139.2	433.4	-49.3
	5	**50**	0.00616	285.03	137.0	34.8	128.3	-23.0
	10	**25**	0.00392	330.65	-4.4	-25.8	45.1	-10.6
	25	**10**	0.00181	697.63	-9.4	46.8	-32.9	88.5
	50,5 250,1	**125,2**	0	∞	*	*	#	#
$T = 300$	**1**	**300**	0.03725	120.40	1064.0	162.1	1279.6	-67.5
	2	**150**	0.01488	177.10	520.1	134.7	451.2	-52.1
	3	**100**	0.01006	197.37	187.3	44.0	272.5	-46.7
	4	**75**	0.00765	223.40	218.9	83.1	183.5	-39.6
	5	**60**	0.00602	282.98	122.8	32.3	122.8	-23.5
	6	**50**	0.00532	276.87	104.7	38.8	97.2	-25.2
	10	**30**	0.00387	369.27	176.1	89.3	43.2	-0.2
	12	**25**	0.00362	343.82	-11.7	-28.7	34.0	-7.1
	15	**20**	0.00303	444.15	16.4	-12.6	12.1	20.0
	20	**15**	0.00245	670.29	144.9	52.8	-9.3	81.2
	25	**12**	0.00228	988.67	355.3	107.1	-15.7	167.2
	30	**10**	0.00175	693.34	-12.5	-26.2	-35.2	87.4
	50	**6**	0.00001	∞	*	*	-99.6	#
	60,5 100,3 300,1	**75,4 150,2**	0	∞	*	*	#	#
$T = 500$	**1**	**500**	0.03406	139.83	1161.6	165.2	1161.6	-62.2
	2	**250**	0.01416	187.50	388.3	85.3	424.4	-49.3
	4	**125**	0.00735	245.93	234.2	82.6	172.3	-33.5
	5	**100**	0.00638	236.94	82.3	20.0	136.3	-36.0
	10	**50**	0.00398	328.92	53.0	16.8	47.3	-11.1
	20	**25**	0.00296	373.74	-27.8	-34.4	9.7	1.0
	25	**20**	0.00258	470.72	-0.9	-17.6	-4.6	27.2
	50	**10**	0.00175	626.47	-12.5	-18.3	-35.2	69.3
	100,5 250,2	**125,4 500,1**	0	∞	*	*	#	#

[1] $\%\ \text{deviation} = 100(UFAR/FAR_{\text{Case K}} - 1)$; [2] $\%\ \text{deviation} = 100(UARL_0/ARL_{\text{Case K}} - 1)$;

[3] $\%\ \text{deviation} = 100(UFAR/0.0027 - 1)$; [4] $\%\ \text{deviation} = 100(UARL_0/370 - 1)$

Table 3.22: The unconditional in-control average run-length ($UARL_0$) values of m and n such that $T = mn$ when $p = p_1 = 0.5$

	m	n	$UFAR$	$UARL_0$	% difference from Case K FAR[1]	% difference from Case K ARL[2]	% difference from nominal FAR=0.0027[3]	% difference from nominal ARL=370[4]
T = 750	1	750	0.03373	154.68	1305.5	167.4	1149.3	-58.2
	2	375	0.01429	195.58	429.2	89.7	429.2	-47.1
	3	250	0.00927	220.66	219.5	57.4	243.2	-40.4
	5	150	0.00628	239.35	161.6	73.7	132.6	-35.3
	6	125	0.00543	259.56	147.0	73.0	101.3	-29.8
	10	75	0.00417	291.37	73.6	40.4	54.3	-21.3
	15	50	0.00339	351.05	30.3	9.5	25.5	-5.1
	25	30	0.00282	426.66	101.5	63.8	4.5	15.3
	30	25	0.00270	388.13	-34.3	-36.8	-0.2	4.9
	50	15	0.00180	822.08	79.8	24.6	-33.4	122.2
	75	10	0.00180	590.93	-9.8	-13.4	-33.2	59.7
	125,6 250,3 750,1	150,5 375,2	0	∞	*	*	#	#
T = 1000	1	1000	0.03362	142.52	1193.2	165.2	1145.3	-61.5
	2	500	0.01434	193.35	431.0	91.8	431.0	-47.7
	4	250	0.00746	221.21	157.4	57.0	176.5	-40.2
	5	200	0.00613	253.58	166.6	73.0	127.1	-31.5
	8	125	0.00456	286.10	107.5	57.0	69.1	-22.7
	10	100	0.00422	285.01	20.6	-0.3	56.3	-23.0
	20	50	0.00312	364.70	20.0	5.4	15.6	-1.4
	25	40	0.00290	374.32	31.7	20.3	7.3	1.2
	40	25	0.00260	393.72	-36.7	-37.7	-3.8	6.4
	50	20	0.00232	475.03	-10.7	-18.3	-14.0	28.4
	100	10	0.00186	559.96	-6.9	-8.6	-31.0	51.3
	125	8	0.00024	∞	*	*	-91.0	#
	200,5 500,2	250,4 1000,1	0	∞	*	*	#	#
T = 1500	1	1500	0.03315	149.71	1226.1	166.3	1127.9	-59.5
	2	750	0.01401	189.14	483.9	118.7	419.0	-48.9
	3	500	0.00943	205.42	249.4	80.5	249.4	-44.5
	4	375	0.00716	233.25	165.1	59.0	165.1	-37.0
	5	300	0.00629	235.43	96.4	34.0	132.8	-36.4
	6	250	0.00538	258.84	85.4	34.2	99.1	-30.0
	10	150	0.00418	287.19	74.3	44.8	55.0	-22.4
	12	125	0.00382	307.31	73.8	46.2	41.6	-16.9
	15	100	0.00361	308.79	3.3	-7.9	33.9	-16.5
	20	75	0.00327	331.32	36.3	23.5	21.1	-10.5
	25	60	0.00303	360.64	12.4	3.8	12.4	-2.5
	30	50	0.00287	380.90	10.5	0.9	6.4	2.9
	50	30	0.00253	466.77	81.0	49.7	-6.2	26.2
	60	25	0.00254	397.09	-38.1	-38.2	-6.0	7.3
	75	20	0.00235	453.58	-9.8	-14.4	-13.1	22.6
	100	15	0.00135	933.58	34.6	9.7	-50.2	152.3
	125	12	0.00109	1686.74	118.2	21.4	-59.6	355.9
	150	10	0.00191	533.16	-4.4	-4.0	-29.2	44.1
	250,6 375,4 750,2	300,5 500,3 1500,1	0	∞	*	*	#	#

[1] $\%\,\text{deviation} = 100(UFAR/FAR_{\text{Case K}} - 1)$; [2] $\%\,\text{deviation} = 100(UARL_0/ARL_{\text{Case K}} - 1)$;

[3] $\%\,\text{deviation} = 100(UFAR/0.0027 - 1)$; [4] $\%\,\text{deviation} = 100(UARL_0/370 - 1)$

3.2.4.2 Unconditional characteristics of the *c*-chart

The unconditional characteristics of the *c*-chart can be calculated in the same manner as that of the *p*-chart. To this end, the necessary steps are shown in Table 3.23 where we assume that $c = c_1 = 1$ and $m = 5$ individual and independent Phase I inspection units are used to estimate c.

First, we calculate the conditional characteristics in columns 7, 8 and 9 (based on the observed value u or $\bar{c}_{obs}$ and the estimated control limits and resulting chart constants listed in columns 1 to 6) and then we calculate the unconditional properties of the run-length distribution (in particular, the *UFAR*, the *UARL* and the *USDRL* using expressions (3-59), (3-61) and (3-52)) by means of the results of columns 11 to 14. Note, however, that although theoretically $V \in \{0,1,2,...\}$, Table 3.23 only shows the conditional properties for $V \in \{0,1,2,...,20\}$ in order to save space.

Table 3.23: The conditional and unconditional characteristics of the run-length distribution for $m = 5$ when $c = 1$

Phase I						Phase II : Conditional Properties				Phase II : Unconditional Properties			
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)=(7)x(10)	(12)=(8)x(10)	(13)=(8)^{2}x(10)	(14)=(9)x(10)
v	$\bar{c}_{obs}$	$L\hat{C}L_c$	$U\hat{C}L_c$	$\hat{d}$	$\hat{f}$	*CFAR*	*CARL*	*CVARL*	Pr(V=v|c)	*CFAR*.Pr(V=v|c)	*CARL*.Pr(V=v|c)	*CARL*^{2}.Pr(V=v|c)	*CVARL*.Pr(V=v|c)
0	0.0	0.00	0.00	0	0	1.0000	1.0000	0.0000	0.00674	0.00674	0.00674	0.0067	0.0000
1	0.2	-1.14	1.54	0	1	0.6321	1.5820	0.9207	0.03369	0.02130	0.05330	0.0843	0.0310
2	0.4	-1.50	2.30	0	2	0.4482	2.2312	2.7472	0.08422	0.03775	0.18792	0.4193	0.2314
3	0.6	-1.72	2.92	0	2	0.4482	2.2312	2.7472	0.14037	0.06291	0.31321	0.6988	0.3856
4	0.8	-1.88	3.48	0	3	0.3869	2.5849	4.0967	0.17547	0.06788	0.45356	1.1724	0.7188
5	1.0	-2.00	4.00	0	3	0.3869	2.5849	4.0967	0.17547	0.06788	0.45356	1.1724	0.7188
6	1.2	-2.09	4.49	0	4	0.3715	2.6915	4.5527	0.14622	0.05433	0.39356	1.0593	0.6657
7	1.4	-2.15	4.95	0	4	0.3715	2.6915	4.5527	0.10444	0.03881	0.28111	0.7566	0.4755
8	1.6	-2.19	5.39	0	5	0.3685	2.7139	4.6513	0.06528	0.02405	0.17716	0.4808	0.3036
9	1.8	-2.22	5.82	0	5	0.3685	2.7139	4.6513	0.03627	0.01336	0.09842	0.2671	0.1687
10	2.0	-2.24	6.24	0	6	0.3680	2.7177	4.6680	0.01813	0.00667	0.04928	0.1339	0.0846
11	2.2	-2.25	6.65	0	6	0.3680	2.7177	4.6680	0.00824	0.00303	0.02240	0.0609	0.0385
12	2.4	-2.25	7.05	0	7	0.3679	2.7182	4.6704	0.00343	0.00126	0.00933	0.0254	0.0160
13	2.6	-2.24	7.44	0	7	0.3679	2.7182	4.6704	0.00132	0.00049	0.00359	0.0098	0.0062
14	2.8	-2.22	7.82	0	7	0.3679	2.7182	4.6704	0.00047	0.00017	0.00128	0.0035	0.0022
15	3.0	-2.20	8.20	0	8	0.3679	2.7183	4.6707	0.00016	0.00006	0.00043	0.0012	0.0007
16	3.2	-2.17	8.57	0	8	0.3679	2.7183	4.6707	0.00005	0.00002	0.00013	0.0004	0.0002
17	3.4	-2.13	8.93	0	8	0.3679	2.7183	4.6707	0.00001	0.00001	0.00004	0.0001	0.0001
18	3.6	-2.09	9.29	0	9	0.3679	2.7183	4.6708	0.00000	0.00000	0.00001	0.0000	0.0000
19	3.8	-2.05	9.65	0	9	0.3679	2.7183	4.6708	0.00000	0.00000	0.00000	0.0000	0.0000
20	4.0	-2.00	10.00	0	9	0.3679	2.7183	4.6708	0.00000	0.00000	0.00000	0.0000	0.0000
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
										0.40672	**2.51**	**6.35**	**3.85**
										UFAR	***UARL***	***USDRL* = 1.98**	

To obtain the unconditional false alarm rate, for instance, we need the conditional false alarm rate and the related probability $\Pr(V = v \mid c = 1)$ for $v = 0,1,2,...$ listed in columns 7 and 10, respectively. Multiplying the corresponding row entries of columns 7 and 10, we end up with column 11, that is,

$$CFAR(c_1 = 1, m = 5 \mid V = v, c = 1) \times \Pr(V = v \mid c = 1) = (1 - \hat{\beta}(c_1 = 1\ ,\ m = 5 \mid V = v, c = 1)) \times \Pr(V = v \mid c = 1)$$

$$\text{for} \quad v = 0,1,2,...$$

so that summing the entries in column 11 yields the unconditional false alarm rate i.e.

$$UFAR(c_1 = 1, m = 5 \mid c = 1) = \sum_{v=0}^{\infty} CFAR(c_1 = 1, m = 5 \mid V = v, c = 1) \Pr(V = v \mid c = 1) = 0.40672$$

(see e.g. (3-61) in Table 3.17). Similarly, we find an unconditional *ARL* and unconditional *SDRL* of 2.51 and 1.98, respectively. Note that, in the calculation of the unconditional characteristics in Table 3.23 the summation was done until $P(V = v \mid c = 1) \approx 0$.

Compared to the Case K *FAR*, *ARL* and *SDRL* of 0.3869, 2.58 and 2.02, respectively (see e.g. Table A3.12 in Appendix 3A) we see that the unconditional values are not far off. However, the unconditional values do not measure up to the nominal *FAR*, *ARL* and *SDRL* values of 0.0027, 370 and 369 typically expected from a 3-sigma control chart; the reason for this big discrepancy is twofold:

(i) the normal approximation to the $Poi(c)$, for small c, is inaccurate so that the charting formula (mean $\pm$ 3 standard deviations) may be inaccurate, and

(ii) due to the discrete nature of the Poisson distribution only certain (conditional) *FAR*, *ARL* and *SDRL* values can be attained.

Note that, from Table 3.23 it is clear that, unlike in case of the *p*-chart, none of the *CFAR* values of the *c*-chart are zero and thus none of the moments, such as the *UFAR* and the *USDRL*, are undefined.

To illustrate the effect of parameter estimation on the overall performance of the *c*-chart, Table 3.24 displays the *UFAR*, the *UARL* and the *USDRL* for various values of m when $c = c_1 = 1, 2, 4, 6, 8, 10, 20 \text{ and } 30$. Also shown are the *FAR*, the *ARL* and the *SDRL* for Case K and the nominal values – given in the last two rows of the table.

We observe that, in general:

(i) As the size of the Phase I reference sample m becomes larger, the unconditional properties gets closer to the Case K values, regardless the value of c.

For instance, when $c = c_1 = 8$ and $m = 20$, the $UFAR = 0.0054$, the $UARL = 315.32$ and the $USDRL = 468.24$ but, when the Phase I sample increase to $m = 500$ inspection units, the $UFAR = 0.0041$, the $UARL = 246.81$ and the $USDRL = 247.68$, which is close to the *FAR*, the *ARL* and the *SDRL* of Case K i.e. 0.0041, 246.70 and 246.20, respectively;

(ii) Unless c and m are both large the *UFAR*, the *UARL* and the *USDRL* are nowhere near the nominally expected values of 0.0027, 370.0 and 369.9.

For instance, when $c = 6$ and $m = 25$, the $UFAR = 0.0079$, the $UARL = 156.49$ and the $USDRL = 181.41$ but, when $c = 20$ and $m = 200$ the $UFAR = 0.0032$, the $UARL = 333.40$ and the $USDRL = 352.01$ gets closer but still not equal to the nominal values. Although this could be a reason for concern for the practitioner, the nominal values are not entirely appropriate given the fact that the Poisson distribution is discrete ;

(iii) The unconditional *ARL* is not equal to the reciprocal of the unconditional *FAR* nor is it smaller than the unconditional *SDRL* (for all combinations of m and c).

This is unlike the situation of the *c*-chart in Case K and is a result of estimating the unknown parameter c; this was also observed in the case of the p-chart with an unknown standard.

Table 3.24: The unconditional false alarm rate (*UFAR*), the unconditional in-control average run-length ($UARL_0$) and the unconditional in-control standard deviation of the run-length ($USDRL_0$) values for the *c*-chart for various values of m when $c = c_1 = 1, 2, 4, 6, 8, 10, 20$ and 30 [1]

	c							
m	**1**	**2**	**4**	**6**	**8**	**10**	**20**	**30**
5	0.4067	0.1603	0.0325	0.0136	0.0104	0.0095	0.0078	0.0072
	2.51	6.54	38.49	166.91	436.17	399.00	303.41	269.39
	1.98	6.22	42.31	226.85	855.24	664.34	420.94	345.01
10	0.3901	0.1485	0.0272	0.0097	0.0069	0.0062	0.0052	0.0048
	2.58	6.82	40.34	162.21	370.41	378.91	330.91	307.82
	2.03	6.37	42.30	205.76	653.10	577.22	427.50	369.61
15	0.3845	0.1463	0.0259	0.0087	0.0060	0.0053	0.0045	0.0042
	2.61	6.88	41.04	159.53	326.93	356.59	333.40	321.49
	2.05	6.40	42.33	194.19	525.47	520.67	416.14	376.15
20	0.3824	0.1448	0.0252	0.0082	0.0054	0.0048	0.0041	0.0038
	2.62	6.93	41.48	157.90	315.32	353.51	338.79	328.11
	2.06	6.44	42.35	187.07	468.24	489.44	412.20	377.23
25	0.3813	0.1446	0.0248	0.0079	0.0052	0.0045	0.0039	0.0037
	2.63	6.94	41.74	156.49	298.67	343.85	336.93	330.19
	2.07	6.44	42.37	181.41	425.57	460.87	403.04	373.39
30	0.3807	0.1439	0.0247	0.0077	0.0050	0.0044	0.0038	0.0036
	2.63	6.96	41.78	155.76	290.10	333.52	334.53	332.60
	2.07	6.46	42.32	177.97	395.40	438.50	392.87	372.15
50	0.3799	0.1434	0.0241	0.0073	0.0047	0.0040	0.0035	0.0033
	2.63	6.99	42.22	154.09	276.24	322.48	335.16	336.25
	2.08	6.48	42.42	169.72	344.81	401.00	379.88	366.80
100	0.3796	0.1424	0.0239	0.0070	0.0044	0.0038	0.0033	0.0032
	2.64	7.03	42.40	154.12	261.79	308.18	334.20	339.49
	2.08	6.52	42.44	162.69	300.06	360.11	363.95	360.06
200	0.3795	0.1413	0.0239	0.0066	0.0042	0.0037	0.0032	0.0031
	2.64	7.08	42.45	156.83	252.11	295.09	333.40	341.43
	2.08	6.57	42.48	160.31	268.28	320.34	352.01	355.01
300	0.3794	0.1407	0.0238	0.0064	0.0041	0.0036	0.0031	0.0030
	2.64	7.11	42.47	159.17	248.61	289.82	332.90	342.34
	2.08	6.59	42.49	160.92	255.61	302.17	344.89	352.47
500	0.3794	0.1402	0.0238	0.0062	0.0041	0.0035	0.0031	0.0030
	2.64	7.13	42.48	161.91	246.81	285.81	331.23	339.38
	2.08	6.61	42.51	162.24	247.68	287.77	334.07	342.30
Case K	0.3869	0.1399	0.0264	0.0061	0.0041	0.0035	0.0029	0.0029
	2.58	7.15	37.81	163.74	246.70	285.74	339.72	349.94
	2.02	6.63	37.31	163.24	246.20	285.23	339.22	349.44
Nominal	0.0027 , 370.0 , 369.9							

[1]The three rows of each cell shows the *UFAR*, the $UARL_0$ and the $USDRL_0$, respectively

Example 1 continued: A Phase II *p*-chart

Phase I and Phase II (conditional)

Recall that the final Phase I data consisted of $m = 28$ samples each of size $n = 50$ (see pages 160-161). Based on these data, it was found that $\bar{p} = 301/1400 = 0.215$ so that the estimated Phase II control limits were set at $U\hat{C}L_p = 0.3893$ and $L\hat{C}L_p = 0.0407$. Given the particular Phase I data, it was shown that the resultant Phase II p-chart has a conditional false alarm rate of $CFAR = 0.002218$ and a conditional average run-length of $CARL_0 = 450.89$.

To get an idea of the general performance of a Phase II p-chart based on $m = 28$ samples each of size $n = 50$ (even prior to collecting the data) one has to look at the unconditional properties of the Phase II p-chart; the unconditional properties takes into account all the possible realizations of $\bar{p} = \frac{U}{mn} \in \{0, \frac{1}{mn}, \frac{2}{mn}, ..., \frac{mn-1}{mn}, 1\}$.

Phase II (unconditional)

Using (3-56) and averaging over all $mn + 1 = 28 \times 50 + 1 = 1401$ possible values and the corresponding binomial probabilities of U, the in-control unconditional *ARL* is found to be

$$UARL_0(p_1 = 0.2, m = 28, n = 50 \mid p = 0.2) = \sum_{u=0}^{1400} (1 - \hat{\beta}(0.2, 28, 50 \mid u, 0.2))^{-1} \binom{1400}{u} 0.2^u (0.8)^{1400-u} = 401.51$$

which is about 11% smaller than the in-control conditional *ARL* for the given data,

$$CARL_0(p_1 = 0.2, 28, 50 \mid U = 301, p = 0.2) = 450.89.$$

Perhaps more importantly, it is seen that when p is estimated from Phase I data, the in-control unconditional *ARL* is 8.5% higher than the corresponding in-control *ARL* of 370 as obtained in the standard known case. Thus, when p is estimated, the in-control *ARL* can be much larger than the nominal value.

Example 2 continued: A Phase II *c*-chart

Phase I and Phase II (conditional)

The final Phase I data consisted of $m = 24$ inspection units each of 100 individual items of product; this resulted in a point estimate $\bar{c} = 472/24 = 19.67$ so that the estimated Phase II control limits were set at $U\hat{C}L_c = 32.97$ and $L\hat{C}L_c = 6.36$ (see pages 162 – 163). Given the particular Phase I data, it was shown that the resultant Phase II c-chart has a conditional false alarm rate of $CFAR = 0.004983$; it follows that the conditional average run-length is $CARL_0 = 1/0.004983 = 200.68$.

Like in the case of the Phase II p-chart, one can get an idea of the general performance of a Phase II c-chart based on $m = 24$ inspection units each of 100 individual items of product (even prior to collecting the data) by looking at the unconditional properties of the Phase II c-chart; the unconditional properties take into account all the possible realizations of $\bar{c} = \frac{V}{m} \in \{0, \frac{1}{m}, \frac{2}{m}, ...\}$.

Phase II (unconditional)

Using (3-60) and (3-61), and averaging over all the possible values and the corresponding probabilities of $V \mid c = 20 \sim Poi(mc = 480)$, the unconditional false alarm rate ($UFAR$) is found to be 0.0039 and the in-control unconditional average run length ($UARL_0$) is found to be 335.30.

The $UFAR$ is 20% less than the $CFAR$ of 0.004983 and the $UARL_0$ is 67% larger than the $CARL$ of 200.68; both these conditional properties are based on an observed value of V equal to 472.

Note that, with regards to the unconditional chart properties, the in-control unconditional average run-length ($UARL_0$) is 1.3% less than the in-control average run-length of 339.72 one would have obtained in Case K for $c_0 = 20$ and the unconditional false alarm rate $UFAR$ is 34.5% larger than the FAR of 0.0029 obtainable in Case K (see e.g. Table A3.12 in Appendix 3A); we can thus expect more false alarms (given the Phase I data at hand) than what would be the case if it is known that $c = 20$.

3.3 Concluding remarks: Summary and recommendations

The false alarm rate (FAR) and the in-control average run length (ARL_0) of the *p*-chart and the *c*-chart are substantially affected by the estimation of the unknown true fraction of nonconforming items p and/or the unknown true average number of nonconformities in an inspection unit c. Calculations show that when p and c are estimated:

(i) The unconditional FAR, unlike in Case K, is not equal (not even close) to the reciprocal of the unconditional ARL_0 and vice versa;

(ii) The unconditional ARL_0 is, unlike in Case K, smaller than the unconditional $SDRL_0$;

(iii) Unless *m* and/or *n* are rather large, the unconditional false alarm rates and the in-control unconditional average run-lengths can be substantially different from the nominal values of 0.0027 and 370;

(iv) Even if more Phase I data is available, neither the $UARL_0$ nor the $UFAR$ will necessarily be exactly equal to the commonly used nominally expected values (primarily due to the discreteness of the underlying distributions);

(v) The typical recommendation of using between $m = 10$ and 25 subgroups of size 5 appears to be inadequate and can be very problematic with attributes charts with regard to a true FAR or true ARL_0; and

(vi) Since one deals with a discrete (binomial or Poisson) distribution in the case of attributes charts, it is rather unlikely to be able to guarantee an exact false alarm rate as is typical for a variables control chart.

For the *p*-chart, in particular, even with a large amount of reference data, if *m* is (much) larger than *n* (as is the case in a typical variables charting situation) the false alarm rate can be too small and the in-control average run-length can be undefined, which are, of course, undesirable. In practice, at least $T \geq 200$ reference data points are recommended, in 10 subgroups of 20 observations each; a general rule is $n/m \geq 0.5$. To this end, Table 3.20, 3.21 and 3.22 can provide valuable guidance in the

process of choosing *m* and *n*. Similarly, for the *c*-chart, Table 3.24 can be used as a guide to select an appropriate number m of Phase I inspection units.

If the necessary amount of reference data is not available in a given situation, the user can calculate the exact unconditional false alarm rates and the exact in-control unconditional *ARL* values using the formulas given in this chapter for the specific *m* and/or *n* values at hand and get an idea of the ramifications of estimating p and/or c.

Another alternative would be to adjust the control limits by finding the value of the charting constant $k > 0$ (which is equal to 3 in routine applications) so that the unconditional *FAR* equals a specified FAR^* or the unconditional *ARL* equals a particular ARL_0^*, say. This would mean either expanding or contracting the control limits and entails, for example, in case of the Phase II *p*-chart, solving for k from

$$UARL_0(p_1 = p_*, m, n, k \mid p = p_*) = \sum_{u=0}^{mn} (1 - \hat{\beta}(p_1 = p_*, m, n \mid u, p = p_*))^{-1} \binom{mn}{u} p_*^u (1 - p_*)^{mn-u} = ARL_0^*$$

where m, n and $p_1 = p = p_*$ for some $0 < p_* < 1$ and

$$\hat{\beta}(p_1, m, n, k \mid u, p) = \Pr(L\hat{C}L_p < p_i < U\hat{C}L_p \mid \bar{p}, p)$$

$$= \Pr(\bar{p} - k\sqrt{\bar{p}(1-\bar{p})/n} < p_i < \bar{p} + k\sqrt{\bar{p}(1-\bar{p})/n} \mid \bar{p}, p)$$

$$= \Pr(\tfrac{u}{mn} - k\sqrt{\tfrac{u}{mn}(1-\tfrac{u}{mn})/n} < p_i < \tfrac{u}{mn} + k\sqrt{\tfrac{u}{mn}(1-\tfrac{u}{mn})/n} \mid u, p)$$

(see e.g. expression (3-30) where $k = 3$).

However, note that, in solving the above equation the user has to, as in the preceding examples, specify a value of p - the same parameter that is unknown! This implies that the practitioner has to know the process that is monitored quite well because the charting constant(s) found from solving the above equation would only be appropriate for the particular p that is selected.

To overcome the predicament of choosing a specific value for p (denoted p_*) one can, for instance, make use of the idea of mixture distributions and assume a particular distribution for p, say $f(p; \underline{\boldsymbol{\theta}})$ for $0 < p < 1$ where $\underline{\boldsymbol{\theta}}$ are the (known or specified) parameters of the distribution (which

handles our uncertainty about the parameter p by treating it as a random variable rather than a fixed value) and then solve for k from

$$\int_0^1 \left(\sum_{u=0}^{mn} (1-\hat{\beta}(p_1 = p, m, n, k \mid u, p))^{-1} \binom{mn}{u} p^u (1-p)^{mn-u} \right) f(p;\underline{\boldsymbol{\theta}}) dp = ARL_0^*.$$

Again, the exact equations given in this chapter can be helpful in this regard, but the practitioner still needs to select and substantiate, from a practical point of view, a distribution $f(p;\underline{\boldsymbol{\theta}})$ and provide the parameter(s) for this distribution.

If the idea of mixture distributions is to be followed, we suggest that $f(p;\underline{\boldsymbol{\theta}})$ and its parameters be chosen in such a way, that best conveys the practitioner's believe about the unknown true fraction nonconforming. For example, because we know that $0 < p < 1$, one possibility is to use the type I (or standard) Beta distribution with parameters (α, β), which has the interval (0,1) as support, as a prior distribution. But which beta distribution should we use? If it is believed that p is in the neighborhood of 0.25 (say) we may, for instance, choose a Beta$(\alpha = 1, \beta = 3)$ distribution which has a mean of 0.25 and a variance of 0.0375. Other options are certainly also available.

A third approach that one can use to obtain the appropriate Phase II control limits is a Bayesian procedure. As an example, we briefly outline the Bayes approach for the Phase II *p*-chart. According to Bayes' theorem the posterior distribution, g, is proportional to the likelihood function, L, times the prior distribution, f.

For the *p*-chart the likelihood is

$$L(p \mid \text{Phase I data}) \propto p^{\sum_{i=1}^{m} x_i} (1-p)^{mn - \sum_{i=1}^{m} x_i}$$

where the Phase I data are the observed values of X_i, $i = 1,2,...,m$ and denoted by x_i, $i = 1,2,...,m$. The Jeffreys' prior (which is the best noninformative prior for the unknown parameter p) is given by

$$f(p) \propto p^{-\frac{1}{2}} (1-p)^{-\frac{1}{2}}.$$

From the likelihood and the prior it follows that the posterior distribution of p is a beta distribution i.e.

$$g(p \mid \text{Phase I data}) = \left[B\left(\sum_{i=1}^{m} x_i + \frac{1}{2}, mn - \sum_{i=1}^{m} x_i + \frac{1}{2} \right) \right]^{-1} p^{\sum_{i=1}^{m} x_i - \frac{1}{2}} (1-p)^{mn - \sum_{i=1}^{m} x_i - \frac{1}{2}} \quad , \quad 0 < p < 1.$$

If the process remains in-control during Phase II monitoring, the control limits for a Phase II sample of n independent Bernoulli trials (units), which results in Y_i, $i = m+1, m+2, \ldots$ successes (nonconforming units), can be derived using a predictive distribution, h. The conditional distribution of Y_i, $i = m+1, m+2, \ldots$, given the sample size n and p, is binomial(n, p) and the unconditional predictive distribution of Y_i is a beta-binomial distribution i.e.

$$\begin{aligned} h(y \mid \text{Phase I data}) &= \int_0^1 \binom{n}{y} p^y (1-p)^{n-y} g(p \mid \text{Phase I data}) dp \\ &= \binom{n}{y} \frac{B\left(\sum_{i=1}^{m} x_i + \frac{1}{2} + y, mn - \sum_{i=1}^{m} x_i + \frac{1}{2} + n - y \right)}{B\left(\sum_{i=1}^{m} x_i + \frac{1}{2}, mn - \sum_{i=1}^{m} x_i + \frac{1}{2} \right)} \end{aligned}$$

where $y = 0,1,2,\ldots,n$ and $i = m+1, m+2, \ldots$.

The Phase II control limits, via a Bayes approach, are then derived (using the unconditional predictive distribution) from the resulting rejection region of size α, that is, $R(\alpha)$, which is defined as

$$\alpha = \sum_{R(\alpha)} h(y \mid \text{Phase I data}).$$

Because there is currently no evidence to suggest that the one approach (i.e. either assuming that p is deterministic and unknown and therefore specifying a value for p or using a Bayes approach) is superior and none of the approaches is without any obstacles, more research is needed to find suitable charting constants for the Phase II attributes charts.

3.4 Appendix 3A: Characteristics of the *p*-chart and the *c*-chart in Case K

The characteristics of the *p*-chart and the *c*-chart in Case K are important because it

(i) helps us understand the operation and the performance of the charts in the simplest of cases (when the parameters are known), and

(ii) provide us with benchmark values that we can use to determine the effect of estimating the parameters on the operation and the performance of the charts in Case U (when the parameters are unknown).

We compute and examine the characteristics of the *p*-chart and that of the *c*-chart in two different sections. For each chart we give an example that shows

(i) the calculations that are needed to implement the chart, and

(ii) how to determine the chart's characteristics via its run-length distribution.

Each example is followed by a general discussion of the results which were obtained from an analysis of the chart's in-control (IC) and out-of-control (OOC) characteristics listed in Tables 3.1 and 3.2, respectively.

To the author's knowledge none of the standard textbooks and/or articles currently available in the literature give such a detailed elucidation of the *p*-chart's or the *c*-chart's characteristics as is done here.

3.4.1 The *p*-chart in Case K: An example

We first look at an example of a *p*-chart in Case K in order to illustrate the typical application of the chart. In other words, we investigate the properties of the chart for a specific combination of p_0 (the specified value of p) and n while varying $0 < p < 1$ (the true fraction nonconforming). In later sections the performance of the chart is then further studied by considering multiple (various) combinations of p_0 and n.

Example A1: A Case K *p*-chart

Assume that the sample size $n = 50$ and suppose that the true fraction nonconforming p is known or specified to be $p_0 = 0.2$. The upper control limit, the centerline and the lower control limit are then set at

$$UCL_p = 0.2 + 3\sqrt{0.2(0.8)/50} = 0.3697 \qquad CL_p = 0.20 \qquad LCL_p = 0.2 - 3\sqrt{0.2(0.8)/50} = 0.0303.$$

Twelve X_i values (or counts) that were simulated from a $Bin(50,0.2)$ distribution are shown in Table A3.1. Without any loss of generality these counts may be regarded as the number of nonconforming items in twelve independent random samples each of size 50 from a process with a true fraction nonconforming of 0.2. The corresponding sample fraction nonconforming $p_i = X_i / 50$ for each sample is also shown; these are the charting statistics of our *p*-chart.

The *p*-chart is shown in Figure A3.1. The chart displays the two control limits (UCL and LCL), the centerline (CL), and the sample fraction nonconforming p_i from each sample. Because none of the points plot outside the limits we continue to monitor the process. Once a point does plot outside the limits the charting procedure will stop and a search for assignable causes (i.e. additional and/or unwanted sources of variation) will begin.

Table A3.1: Data for the *p*-chart in Case K

Sample number / Time: i	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Counts: X_i	12	8	6	14	8	12	9	7	13	16	11	8
Sample fraction nonconforming: $p_i = X_i / 50$	0.24	0.16	0.12	0.28	0.16	0.24	0.18	0.14	0.26	0.32	0.22	0.16

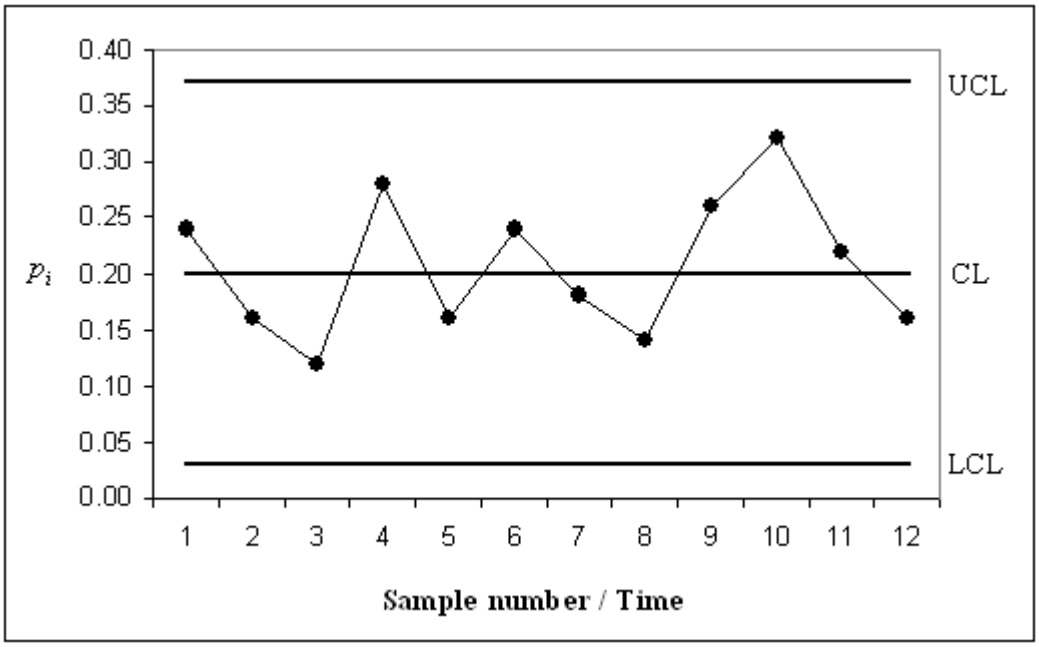


Figure A3.1: A *p*-chart in Case K

Given the operation of the chart it is natural to ask 'How long before the chart signals?' or 'What is the probability for a point to plot between or outside the control limits?' etc. These performance based questions are relevant while the process remains in-control and even more so when a shift occurs.

To answer these questions we study the run-length distribution of the chart. The run-length distribution, as mentioned earlier in section 3.1.1, is characterized entirely by the probability of a no-signal $\beta(p, p_0, n)$ or, equivalently, the probability of a signal $1-\beta(p, p_0, n)$ (see e.g. Table 3.1).

Our starting point when analyzing the performance of the chart is therefore to find the probability of a no-signal. Once we have the probability of a no-signal both the in-control and the out-of-control properties of the *p*-chart in Case K are easily found.

Performance of the *p*-chart

For our particular combination of the parameters i.e. $n = 50$ and $p_0 = 0.2$, the control limits are set at $UCL_p = 0.3697$ and $LCL_p = 0.0303$, so that we proceed as follows to find the probability of a no-signal:

First, we calculate the two charting constants a and b defined in (3-5), which gives

$$a = [nLCL_p] = [(50)(0.0303)] = [1.52] = 1 \quad \text{and} \quad b = \min\{[nUCL_p], n\} = \min\{[18.49], 50\} = 18.$$

Using (3-4) shows that the probability of a no-signal is

$$\beta(p, p_0 = 0.2, 50) = I_p(1,48) - I_p(18,31) \qquad \text{for} \qquad 0 < p < 1$$

so that substituting values for p allow us to study the in-control (when $p = 0.2$) and the out-of-control (when $p \neq 0.2$) properties and performance of the chart.

In-control properties

While the true fraction nonconforming p remains constant and equal to $p_0 = 0.2$ we have the in-control scenario. The probability of a no-signal is then

$$\beta(p = 0.2, 0.2, 50) = I_{0.2}(1,48) - I_{0.2}(18,31) = 0.9973$$

and the probability of a signal, or the false alarm rate, is

$$FAR(0.2,50) = 1 - \beta(0.2, 0.2, 50) = 0.0027.$$

The in-control run-length distribution is therefore geometric with probability of success 0.0027, which we write as $N_0 \sim Geo(0.0027)$.

Expressions (3-15) and (3-16) in Table 3.1 show that the in-control ARL and the in-control $SDRL$ can be calculated as

$$ARL_0 = 1/(0.0027) = 370.4 \qquad \text{and} \qquad SDRL_0 = \sqrt{0.9973}/(0.0027) = 369.9$$

respectively.

An in-control *ARL* of 370.4 means that while the process remains in-control we could expect the chart to issue a false alarm or an erroneous signal (on average) every 370th sample. However, with the large standard deviation of 369.9 we could expect a phase (or cycle) during which the chart signals frequently i.e. many false signals occurring one after the other within a relatively short period of time, which is then followed by a phase where the chart hardly ever signals.

Out-of-control properties

When the true fraction nonconforming changes it implies that p is no longer equal to its specified value of $p_0 = 0.2$ and then we deal with the out-of-control case. Since p can increase or decrease we consider both situations.

Increase in p : Upward shift

Suppose that p increases by 12.5% from 0.2 to 0.225. The probability of a no-signal of 0.9973 then becomes

$$\beta(p = 0.225, 0.2, 50) = I_{0.225}(1,48) - I_{0.225}(18,31) = 0.0097$$

so that the probability of a signal becomes $1 - \beta(0.225, 0.2, 50) = 0.9903$.

Assuming that the change in p is permanent so that all future samples that we collect come from a process with a fraction nonconforming equal to $p = 0.225$, the out-of-control run-length distribution of the *p*-chart is $N_1 \sim Geo(0.9903)$. The out-of-control average run-length is then calculated using (3-15) as

$$ARL_1 = 1/(0.9903) = 1.01$$

and implies that (on average) we could expect the chart to signal on approximately the 1st sample following an increase from 0.2 to 0.225. The out-of-control *SDRL* of the run-length distribution is

$$SDRL_1 = \sqrt{0.0097}/(0.9903) = 0.01$$

and is calculated using (3-16).

Decrease in p : Downward shift

Suppose that the true fraction nonconforming permanently decreased by 25% from 0.2 to 0.15. The probability of a no-signal then changes from its in-control value of 0.9973 to

$$\beta(p = 0.15, 0.2, 50) = I_{0.15}(1,48) - I_{0.15}(18,31) = 0.003$$

so that the out-of-control run-length distribution is geometric with probability of success equal to the probability of a signal $1 - \beta(0.15, 0.2, 50) = 0.997$. We could thus expect the chart to signal (on average) on the 1st sample following the change (decrease) in p.

The OC-curve

The OC-curve is the probability of a no-signal $\beta(p, p_0, n)$ plotted as a function of p for a known (specified) value of p_0 and a given (selected) sample size *n*.

The OC-curve for $n = 50$ and $p_0 = 0.2$, that is, $\beta(p, p_0 = 0.2, n = 50)$ for $0 < p \leq 0.55$ is displayed in Figure A3.2. The probability of a signal $1 - \beta(p, p_0 = 0.2, n = 50)$ as a function of p is also shown. These two probabilities are plotted on the vertical axis for a given value of p on the horizontal axis. Table A3.2 displays values of the OC and the probability of a signal for selected values of $p = 0.025(0.025)0.550$; it also shows the average run-length and the standard deviation of the run-length associated with each value of p.

Examining Figure A3.2 we begin at the in-control value of $p = 0.2$ where the probability of a no-signal is $\beta(0.2,0.2,50) = 0.9973$ and the probability of a signal is equal to $1 - \beta(0.2,0.2,50) = 0.0027$; these two points are indicated on the graphs. We observe that:

(i) As we move in either direction away from $p = 0.2$ (i.e. either to the left or to the right) the probability of a no-signal, in general, decreases whereas the probability of a signal, in general, increases.

This indicates that as p changes (moves away) from the known or specified value of 0.2 the likelihood of a signal that the process is out-of-control increases. We can therefore expect the chart to signal more often (sooner) when the process is out-of-control than when the process is in-control; which is good and confirms that using a control chart is an effective tool in detecting changes in a process.

(ii) The values of $\beta(p,0.2,50)$ and $1 - \beta(p,0.2,50)$ vary between zero and one, and happens since both functions compute a probability.

In particular, as the process moves further out-of-control the probability of a no-signal approaches zero whereas the probability of a signal approaches one.

(iii) Neither the probability of a no-signal nor the probability of a signal is symmetric about $p = 0.2$.

This implies, for example, that the rate at which $\beta(p,0.2,50)$ changes as p decreases or increases (i.e. moves to the left or to the right away from 0.2) is not the same. A decrease and an increase of 10% (say) in p from 0.2 to 0.18 and 0.22 (respectively) would therefore not result in the same decrease in $\beta(p,0.2,50)$. The same is true for the probability of a signal and happens since the binomial (50,0.2) distribution is not symmetric.

(iv) As p decreases from 0.2 the probability of a no-signal increases slightly until it reaches a maximum and then decreases (as mentioned in (i)). Similarly, the probability of a signal first decreases a little as p decreases until it reaches a minimum and then it increases again.

This tendency is also seen in Table A3.2. For instance, at $p = 0.2$ we have $\beta(p = 0.2,0.2,50) = 0.9973$ which is less than the probability of a no-signal at $p = 0.175$ of $\beta(p = 0.175,0.2,50) = 0.9988$. For a detailed discussion on this phenomenon see e.g. Acosta-Mejia (1999).

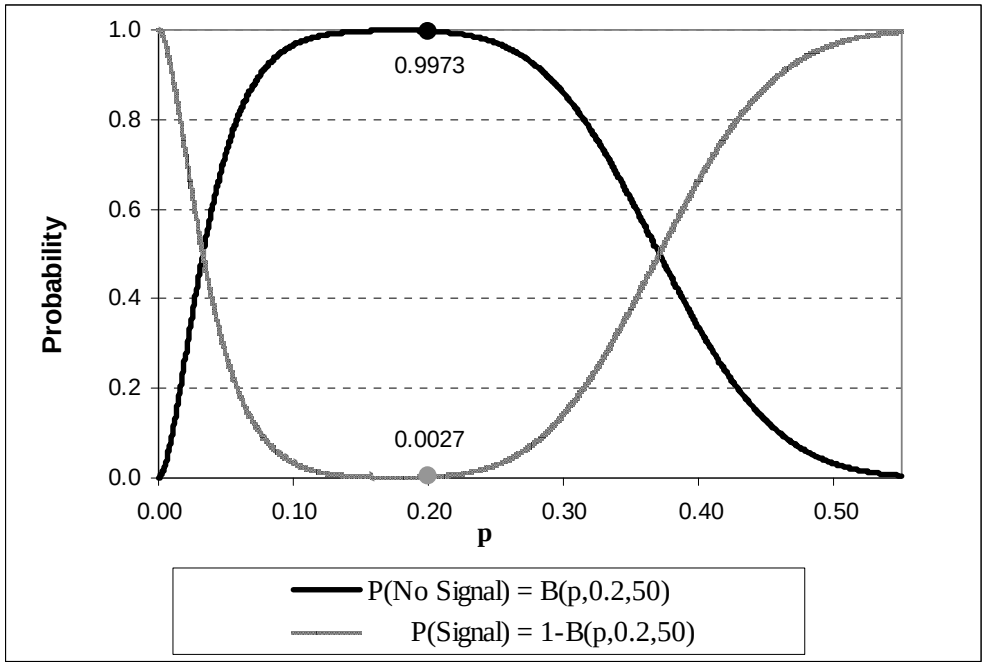


Figure A3.2: The OC-curve and the probability of a signal as a function of p when $p_0 = 0.2$ and $n = 50$

Table A3.2: The Probability of a no-signal, the Probability of a signal, the *ARL* and the *SDRL* for $p = 0.025(0.025)0.550$ when $p_0 = 0.2$ and $n = 50$

p	Pr(No Signal | process OOC)	Pr(Signal | process OOC)	*ARL*	*SDRL*
0.025	0.3565	0.6435	1.55	0.93
0.050	0.7206	0.2794	3.58	3.04
0.075	0.8975	0.1025	9.76	9.24
0.100	0.9662	0.0338	29.60	29.09
0.125	0.9897	0.0103	97.42	96.92
0.150	0.9970	0.0030	337.26	336.76
0.175	0.9988	0.0012	802.13	801.63
0.200	0.9973	0.0027	369.84	369.34
0.225	0.9903	0.0097	103.13	102.63
0.250	0.9713	0.0287	34.79	34.29
0.275	0.9306	0.0694	14.42	13.91
0.300	0.8594	0.1406	7.11	6.60
0.325	0.7544	0.2456	4.07	3.54
0.350	0.6216	0.3784	2.64	2.08
0.375	0.4758	0.5242	1.91	1.32
0.400	0.3356	0.6644	1.51	0.87
0.425	0.2167	0.7833	1.28	0.59
0.450	0.1273	0.8727	1.15	0.41
0.475	0.0678	0.9322	1.07	0.28
0.500	0.0325	0.9675	1.03	0.19
0.525	0.0139	0.9861	1.01	0.12
0.550	0.0053	0.9947	1.01	0.07

Average run-length

The average run-length is the expected number of samples that must be collected before the chart signals.

To quickly detect changes in a process it is desirable that the average run-length

$$ARL(p, p_0, n) = 1/(1-\beta(p, p_0, n))$$

is at its maximum when the process is in-control i.e. when $p = p_0$. This is not always the case for the *p*-chart. For a *p*-chart based on a charting statistic that has a (positively) skewed distribution such as the $Bin(50,0.2)$ distribution the value of $ARL(p, p_0 = 0.2, n = 50)$ increases initially as p decreases; this causes the *p*-chart to have poor performance in detecting small to moderate decreases in p.

Figure A3.3 displays the average run-length $ARL(p,0.2,50)$ as a function of p for $0.05 \leq p \leq 0.3$. The value of $ARL(p,0.2,50)$ is plotted on the vertical axis for a specific value of p on the horizontal axis. The average run-length is much higher for values of p slightly less than 0.2 than at 0.2 i.e. the point that indicates the in-control average run-length of 369.84. In particular, at $p = 0.175$ the average run-length is 802.13 (see e.g. Table A3.2).

This phenomenon, as mentioned before, is caused by the skewness of the binomial distribution and the smaller the value of p the greater the skewness and the larger the problem. For a detailed discussion on this phenomenon see e.g. Acosta-Mejia (1999).

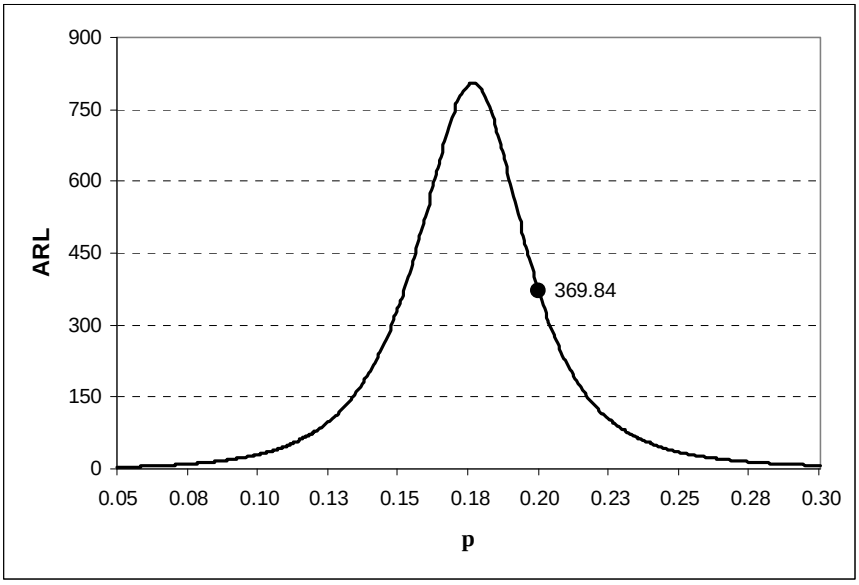


Figure A3.3: The average run-length (*ARL*) as a function of p when $n = 50$ and $p_0 = 0.2$

Run-length distributions

In Case K the in-control and the out-of-control run-length distributions are both geometric (see e.g. Table 3.1).

A graphical display of the in-control and the out-of-control run-length distributions is useful since it helps us (better) see the effect of a change in the process parameter on the entire run-length distribution.

We consider two types of displays: Boxplot-like graphs and probability mass functions (p.m.f's). The former (visually) reveals more about the change in the run-length distribution than do the p.m.f's.

Boxplot-like graphs

Figure A3.4 shows boxplot-like graphs (i.e. the minimum value is replaced by the 1st percentile of the run-length distribution and the maximum value is replaced by the 99th percentile of the run-length distribution) of the in-control as well as the out-of-control run-length distributions. Figure A3.4 is accompanied by Table A3.3 which summarizes some of the properties of the in-control and the out-of-control run-length distributions.

Studying Figure A3.4 and Table A3.3 we note that:

(i) The run-length distributions are severely positively skewed i.e. the spread (variation) in the upper 25% of the distribution between the 75th percentile (or Q_3) and the 99th percentile, is much larger than the spread in the lower 25% of the distribution between the 1st percentile and the 25th percentile (or Q_1).

The skewness of the run-length distribution is confirmed by the fact that the average run-length (indicated by the diamond symbol) is larger than the median run-length (indicated by the circle) in all three the distributions. The exact numerical values of the average run-lengths and the median run-lengths are also indicated. The skewness follows from the fact that the run-length distributions are geometric.

(ii) The run-length distribution is considerably altered following a process change.

Compare, for example, the boxplot-like graph associated with the run-length distribution of the out-of-control process (when $p = 0.225$) to the boxplot-like graph of the in-control run-length distribution (when $p = 0.2$). In particular we see that both the average run-length of 103.1 and the median run-length of 72 of the out-of-control run-length distribution is far less than the average run-length of 369.8 and the median run-length of 257 associated with the in-control run-length distribution. A comparison of the percentiles and the standard deviation of the run-length leads to the same conclusion.

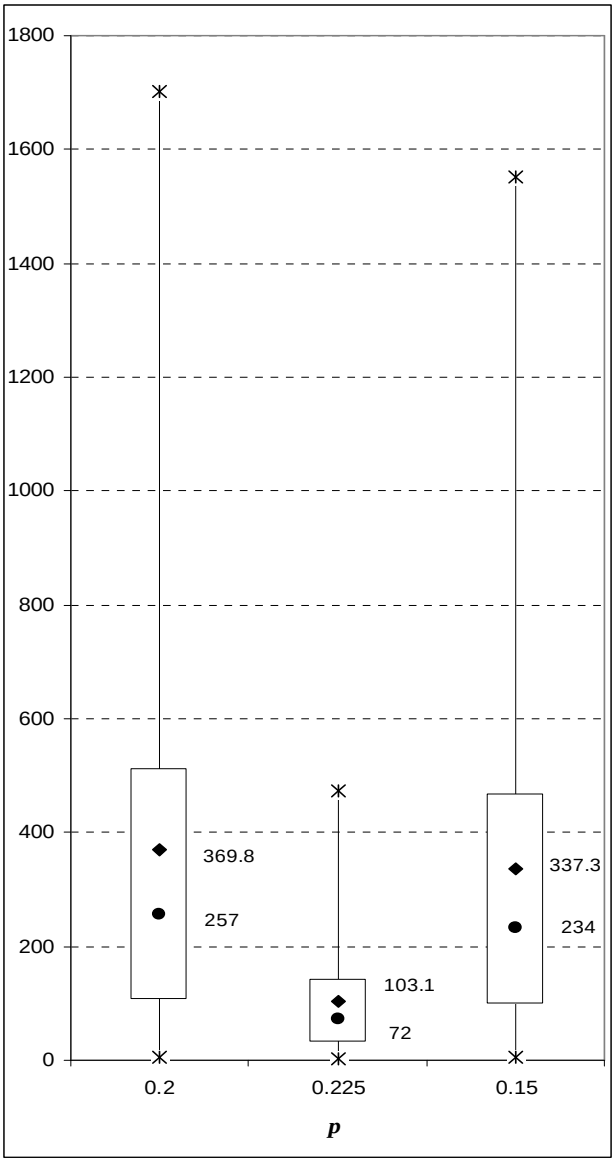


Figure A3.4: Boxplot-like graphs of the in-control and the out-of-control run-length distributions of the *p*-chart in Case K

Table A3.3: Summary measures of the in-control (IC) and the out-of-control (OOC) run-length distributions of the *p*-chart when $n = 50$ and $p_0 = 0.2$ in Case K

	IC	OOC (increase in *p*)	OOC (decrease in *p*)
p	**0.2**	**0.225**	**0.15**
Pr(No Signal)	0.9973	0.9903	0.997
Pr(Signal)	0.0027	0.0097	0.003
ARL	369.84	103.13	337.26
SDRL	369.34	102.63	336.76
1st percentile	4	2	4
5th percentile	19	6	18
10th percentile	39	11	36
25th (Q_1)	107	30	97
50th (*MDRL*)	257	72	234
75th (Q_3)	513	143	467
90th percentile	851	237	776
95th percentile	1107	308	1009
99th percentile	1701	473	1551

Probability mass functions

Studying the p.m.f's of the run-length distributions is another way to look at the effect of a change in the process on the performance of the chart.

Figure A3.5 displays the p.m.f's of the in-control and the out-of-control run-length distributions, that is,

$$\Pr(N_0 = j;0.2,0.2,50) = 0.9973^{j-1}0.0027 \qquad \text{and} \qquad \Pr(N_1 = j;0.15,0.2,50) = 0.003^{j-1}0.997$$

for $j = 1,2,\ldots.$. The former is the in-control p.m.f and the latter the out-of-control p.m.f which corresponds to a decrease by 25% in the fraction of nonconforming p from 0.2 to 0.15.

For values of j less than approximately 370 the likelihood of obtaining these shorter run-lengths is larger following a decrease in the fraction non-conforming. We can write this as $\Pr(N_1 = j) > \Pr(N_0 = j)$ for $j < 370$. The converse also holds, that is, for values of j larger than approximately 370 we see that $\Pr(N_1 = j) < \Pr(N_0 = j)$. This means that the p-chart will signal sooner when the process moves out-of-control than when it is in-control; which is good.

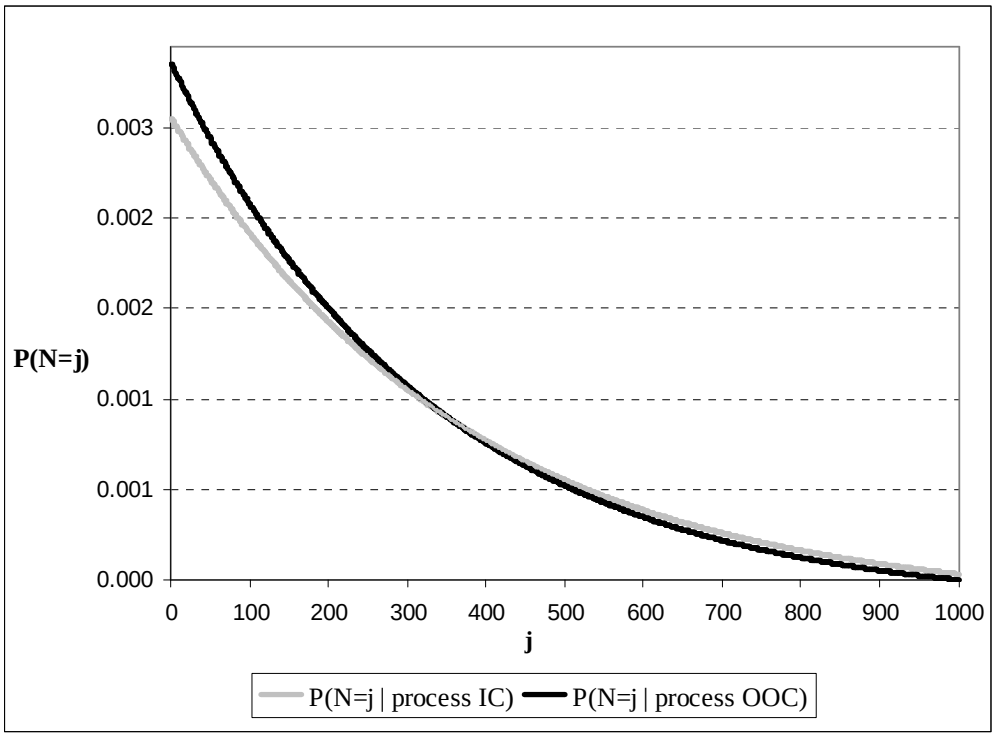


Figure A3.5: The probability distributions of N_0 (when $p = p_0 = 0.2$) and N_1 (when $p = 0.2$ with $p_0 = 0.15$)[1]

[1] *Note*: Instead of displaying the usual histograms, the tops of the bars of the histograms have been joined to better display the shapes of these distributions, and the bars of the histograms have been deleted.

3.4.2 The *p*-chart in Case K: Characteristics of the in-control run-length distribution

The preceding example focused on only one particular combination of n and p_0 i.e. $n = 50$ and $p_0 = 0.2$. Other combinations of n and p_0 are also of interest and gives us an idea of the *p*-chart's performance over a wider range of the parameters.

The false alarm rate and the average run-length are two well-known characteristics of the run-length distribution and most often used to measure a chart's performance. More recently other characteristics of the run-length distribution, such as the standard deviation and the percentiles (quartiles), have also been used and supplemented the false alarm rate and the average run-length.

We study all the abovementioned performance measures for the *p*-chart.

3.4.2.1 False alarm rate

The false alarm rate (FAR) is the probability of a signal when the process is truly in-control and is given by $1-\beta(p = p_0, p_0, n)$ where $\beta(p = p_0, p_0, n)$ is found from (3-4). We can calculate the FAR by substituting different combinations of values for n and p_0 into $1-\beta(p = p_0, p_0, n)$.

Table A3.4 lists the FAR-values (rounded to 4 decimal places) for $p_0 = 0.01$, 0.025, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.40 and 0.50 when the sample size $n =$ 1(1)10, 12, 15(5)30, 40, 50, 75, 100, 125, 150(50)300, 375, 500, 750, 1000 and 1500.

For some combinations of n and p_0 (especially for small values of n and large values of p_0) we observe that the false alarm rate is zero. Although we typically expect (desire) a small false alarm rate, zero is not practical since all moments (such as the average and the standard deviation) of the run-length distribution will be undefined (see e.g. Tables A3.5 and A3.6).

Table A3.4: The false alarm rate (*FAR*) of the p-chart as a function of the sample size n and the known or the specified true fraction nonconforming p_0 in Case K

Sample size n	The known or the specified true fraction nonconforming p_0									
	0.01	**0.025**	**0.05**	**0.10**	**0.15**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**
1	0.0100	0.0250	0.0500	0.1000	0.0	0.0	0.0	0.0	0. 0	0.0
2	0.0199	0.0494	0.0025	0.0100	0.0225	0.0	0.0	0.0	0.0	0.0
3	0.0297	0.0731	0.0073	0.0280	0.0034	0.0080	0.0156	0.0	0.0	0.0
4	0.0394	0.0036	0.0140	0.0037	0.0120	0.0016	0.0039	0.0081	0.0	0.0
5	0.0490	0.0059	0.0226	0.0086	0.0022	0.0067	0.0010	0.0024	0.0	0.0
6	0.0585	0.0088	0.0328	0.0158	0.0059	0.0016	0.0046	0.0007	0.0041	0.0
7	0.0679	0.0121	0.0038	0.0027	0.0121	0.0047	0.0013	0.0038	0.0016	0.0
8	0.0773	0.0158	0.0058	0.0050	0.0029	0.0104	0.0042	0.0013	0.0007	0.0
9	0.0865	0.0200	0.0084	0.0083	0.0056	0.0031	0.0013	0.0043	0.0003	0.0039
10	0.0043	0.0246	0.0115	0.0128	0.0099	0.0064	0.0035	0.0016	0.0017	0.0020
12	0.0062	0.0349	0.0196	0.0043	0.0046	0.0039	0.0028	0.0017	0.0028	0.0005
15	0.0096	0.0057	0.0055	0.0127	0.0036	0.0042	0.0042	0.0037	0.0024	0.0010
20	0.0169	0.0130	0.0159	0.0024	0.0059	0.0026	0.0039	0.0013	0.0021	0.0026
25	0.0258	0.0238	0.0072	0.0095	0.0021	0.0056	0.0034	0.0019	0.0016	0.0041
30	0.0361	0.0064	0.0033	0.0078	0.0029	0.0031	0.0029	0.0024	0.0012	0.0014
40	0.0075	0.0174	0.0034	0.0051	0.0043	0.0031	0.0019	0.0030	0.0018	0.0022
50	0.0138	0.0081	0.0032	0.0032	0.0019	0.0027	0.0031	0.0031	0.0021	0.0026
60	0.0224	0.0039	0.0028	0.0057	0.0024	0.0022	0.0017	0.0029	0.0022	0.0027
75	0.0069	0.0113	0.0041	0.0027	0.0028	0.0025	0.0036	0.0024	0.0030	0.0024
100	0.0184	0.0037	0.0043	0.0049	0.0034	0.0040	0.0038	0.0031	0.0029	0.0035
125	0.0087	0.0043	0.0040	0.0032	0.0031	0.0026	0.0029	0.0033	0.0025	0.0022
150	0.0042	0.0047	0.0036	0.0020	0.0030	0.0031	0.0025	0.0032	0.0034	0.0024
200	0.0043	0.0048	0.0027	0.0034	0.0022	0.0035	0.0025	0.0026	0.0030	0.0023
250	0.0040	0.0046	0.0042	0.0024	0.0027	0.0034	0.0021	0.0030	0.0024	0.0029
300	0.0036	0.0041	0.0027	0.0030	0.0028	0.0031	0.0027	0.0030	0.0026	0.0032
375	0.0051	0.0034	0.0031	0.0035	0.0032	0.0024	0.0029	0.0023	0.0026	0.0027
500	0.0052	0.0047	0.0032	0.0023	0.0033	0.0030	0.0023	0.0029	0.0030	0.0027
750	0.0044	0.0031	0.0027	0.0029	0.0030	0.0030	0.0024	0.0028	0.0025	0.0024
1000	0.0033	0.0036	0.0030	0.0027	0.0030	0.0030	0.0024	0.0027	0.0027	0.0026
1500	0.0034	0.0031	0.0026	0.0030	0.0027	0.0027	0.0026	0.0026	0.0029	0.0025

Figure A3.6 displays the *FAR*-values for $n = 10, 25$ and 50 on the vertical axis for selected values of p_0 on the horizontal axis. Also shown is the nominal *FAR* of 0.0027, which is the *FAR* on a 3-sigma Shewhart X-bar control chart when the charting statistics follow a normal distribution.

Figure A3.6 shows that for small values of p_0 the *FAR* of the *p*-chart is considerably larger than the nominal value of 0.0027. For larger values of p_0 (or, values nearer to 0.5) the *FAR* is closer to the nominal of 0.0027 but still not equal. This illustrates that even for known values of the true fraction nonconforming there is no guarantee that the *FAR* of the *p*-chart will be equal to the nominal 0.0027.

There are two reasons for these discrepancies:

(i) when p is small the normal approximation to the binomial distribution is poor so both the charting constant $k = 3$ and the charting formula (mean $\pm$ 3 standard deviations) may be inaccurate, and

(ii) due to the discrete nature of the binomial distribution only certain *FAR* values can be attained.

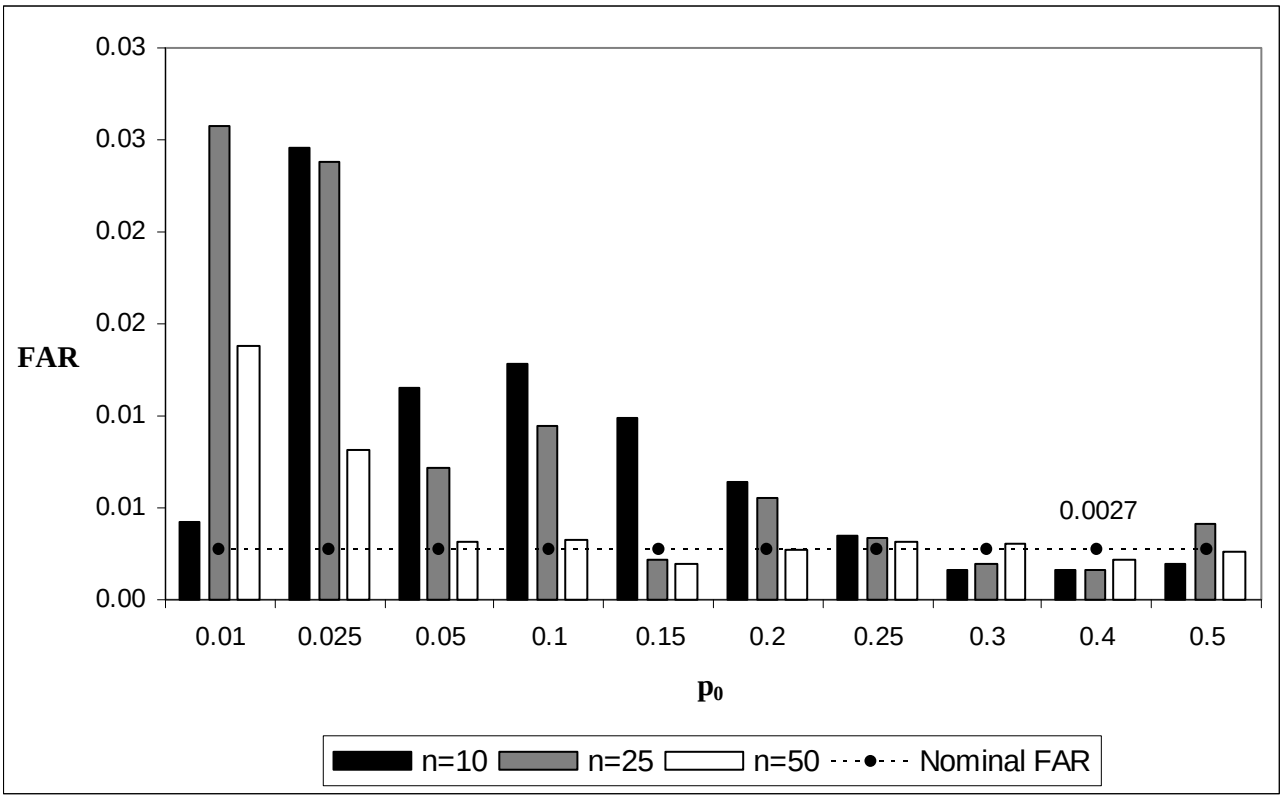


Figure A3.6: The false alarm rate (*FAR*) of the *p*-chart for $n = 10, 25$ and 50 when $p_0 = 0.01, 0.025, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4$ and 0.5 in Case K compared to the nominal *FAR* of 0.0027

3.4.2.2 Average run-length and standard deviation of the run-length

The average run-length (ARL) is the expected value or the mean of the run-length distribution and is equal to the reciprocal of the probability of a signal, that is,

$$ARL(p, p_0, n) = 1/(1-\beta(p, p_0, n)).$$

The in-control ARL is found by replacing p with p_0 in $ARL(p, p_0, n)$ and is the reciprocal of the false alarm rate, that is,

$$ARL_0 = ARL(p_0, p_0, n) = 1/(1-\beta(p_0, p_0, n)) = 1/FAR.$$

The ARL is a measure of how fast (or slow) the control chart signals and is therefore typically used for out-of-control performance comparisons of the charts.

Since the geometric distribution is (severely) positively skewed the ARL becomes questionable as the sole metric for a chart's performance and we therefore need to look at the standard deviation of the run-length ($SDRL$) too.

The $SDRL$ measures the variation or the spread in the run-length distribution and is given by

$$SDRL(p, p_0, n) = \sqrt{\beta(p, p_0, n)}/(1-\beta(p, p_0, n)).$$

The in-control $SDRL$ is found by substituting p_0 for p in $SDRL(p, p_0, n)$ which gives

$$SDRL_0 = SDRL(p_0, p_0, n) = \sqrt{\beta(p_0, p_0, n)}/(1-\beta(p_0, p_0, n)) = \sqrt{1-FAR}/FAR;$$

this shows that the $SDRL_0$ is (like the ARL_0) a function of the FAR.

The values of the ARL_0 and the $SDRL_0$ that correspond to the FAR-values of Table A3.4 are shown in Tables A3.5 and A3.6 (rounded to 2 decimal places), respectively. We can also calculate the ARL_0 and the $SDRL_0$ for different combinations of n and p_0 not shown in Tables A3.5 and Table A3.6 and is carried out by direct evaluation of expressions (3-15) and (3-16).

For example, to find the in-control ARL and the in-control $SDRL$ when $p = 0.25$, $p_0 = 0.25$, and $n = 11$ we proceed as follow:

We first calculate the control limits. These are given by (3-1) as

$$UCL_p = 0.25 + 3\sqrt{0.25(0.75)/11} = 0.6417 \quad \text{and} \quad LCL_p = 0.25 - 3\sqrt{0.25(0.75)/11} = -0.1417.$$

Then (3-5) shows that $b = \min\{[7.0584],11\} = 7$. The constant a need not be calculated since the lower control limits is negative, that is, $LCL_p < 0$. Using (3-4) we find that the probability of a no-signal is $\beta(0.25,0.25,11) = 1 - I_{0.25}(7,3) = 0.9988$ so that the false alarm rate is

$$FAR(0.25,0.25,7) = 1 - \beta(0.25,0.25,11) = I_{0.25}(7,3) = 0.0012.$$

The in-control ARL is therefore $ARL_0 = (1-0.9988)^{-1} = 841.6$ and the in-control standard deviation is $SDRL_0 = \sqrt{0.9988}/(0.0012) = 846.1$.

The calculations for the out-of-control ARL and the out-of-control $SDRL$ are similar; we simply replace p in $\beta(p,0.25,11) = 1 - I_p(7,3)$ with a value other than $p_0 = 0.25$ and proceed along the same lines.

Table A3.5: The in-control average run-length (ARL_0) of the p-chart as a function of the sample size n and the known or the specified true fraction nonconforming p_0 in Case K

Sample size n	The known (specified) true fraction nonconforming p_0									
	0.01	**0.025**	**0.05**	**0.10**	**0.15**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**
1	100.00	40.00	20.00	10.00	∞	∞	∞	∞	∞	∞
2	50.25	20.25	400.00	100.00	44.44	∞	∞	∞	∞	∞
3	33.67	13.67	137.93	35.71	296.30	125.00	64.00	∞	∞	∞
4	25.38	275.77	71.33	270.27	83.46	625.00	256.00	123.46	∞	∞
5	20.40	168.26	44.26	116.82	448.93	148.81	1024.00	411.52	∞	∞
6	17.09	114.06	30.51	63.09	169.92	625.00	215.58	1371.74	244.14	∞
7	14.72	82.84	266.17	366.57	82.62	214.04	744.73	263.80	610.35	∞
8	12.94	63.17	172.76	199.03	350.40	96.09	236.59	775.00	1525.88	∞
9	11.56	49.96	119.60	120.03	177.66	326.12	744.73	233.05	3814.70	256.00
10	234.40	40.63	86.93	78.15	101.28	157.00	285.25	628.78	596.05	512.00
12	161.96	28.63	51.10	230.98	215.44	256.20	359.52	591.14	355.85	2048.00
15	103.84	176.24	182.91	78.61	277.35	235.86	238.49	273.78	417.02	1024.00
20	59.31	77.19	62.89	419.10	168.89	385.38	253.67	781.93	468.26	388.07
25	38.82	41.96	139.57	105.53	467.01	180.02	296.70	522.87	611.72	245.26
30	27.66	157.04	304.65	128.47	339.86	321.44	341.52	410.34	854.91	698.86
40	133.38	57.31	294.82	197.51	231.84	325.83	539.81	331.42	550.59	450.16
50	72.37	122.96	313.64	310.57	512.93	369.84	320.92	323.37	469.25	384.29
60	44.60	259.52	351.05	176.03	411.27	446.91	585.24	347.13	457.45	374.47
75	144.51	88.38	242.82	368.47	351.24	404.72	280.73	424.38	336.52	409.13
100	54.42	270.11	233.96	203.98	294.90	250.93	265.00	324.31	344.84	284.28
125	114.61	230.59	248.37	312.50	322.82	392.14	349.00	303.11	405.93	449.14
150	237.46	212.87	277.54	488.03	329.49	325.75	398.29	313.45	293.42	415.71
200	232.80	206.23	370.42	294.04	449.57	284.28	401.99	389.85	333.58	438.70
250	248.43	219.07	240.23	415.64	376.32	291.56	467.00	338.68	424.89	347.38
300	277.57	244.39	365.86	335.28	354.65	324.53	373.71	330.57	384.63	315.53
375	197.63	296.17	325.93	284.05	314.43	413.51	343.76	431.32	381.72	370.96
500	192.01	213.20	316.36	429.94	306.11	328.38	434.37	345.98	336.29	370.81
750	227.35	323.23	367.35	343.32	329.48	332.59	418.21	358.28	397.20	413.68
1000	300.16	279.22	327.92	370.18	331.16	330.18	410.94	374.21	374.59	378.00
1500	297.89	323.23	384.88	332.36	370.33	372.32	385.02	389.48	345.82	398.62

It is straightforward to show using (3-15) and (3-16) that $SDRL = \sqrt{ARL(ARL-1)}$ and implies that the standard deviation is always less than the average run-length i.e. $SDRL < ARL$, and holds whether the process is in-control or out-of-control.

This relationship between the $SDRL$ and the ARL is clearly visible from Tables A3.5 and A3.6. For example, for $n = 5$ and $p_0 = 0.025$ the in-control ARL equals 168.26 whereas the in-control $SDRL$ equals 167.67. We also looked at this relationship between the $SDRL$ and the ARL of the run-length distribution in Case U when the process parameters are unknown.

Note that, as mentioned before, the in-control average run-length in Table A3.5 and the in-control standard deviation of the run-length in Table A3.6 are undefined for the same combinations of n and p_0 for which the false alarm rate in Table A3.4 is zero. This is undesirable and shows that for some combinations of n and p_0 the *p*-chart would not perform satisfactorily in practice.

Table A3.6: The in-control standard deviation of the run-length ($SDRL_0$) of the p-chart as a function of the sample size n and the known or the specified fraction nonconforming p_0 in Case K

Sample size n	The known or the specified fraction nonconforming p_0									
	0.01	**0.025**	**0.05**	**0.10**	**0.15**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**
1	99.50	39.50	19.49	9.49	∞	∞	∞	∞	∞	∞
2	49.75	19.75	399.50	99.50	43.94	∞	∞	∞	∞	∞
3	33.17	13.16	137.43	35.21	295.80	124.50	63.50	∞	∞	∞
4	24.87	275.27	70.83	269.77	82.96	624.50	255.50	122.96	∞	∞
5	19.90	167.76	43.76	116.32	448.43	148.31	1023.50	411.02	∞	∞
6	16.58	113.56	30.01	62.59	169.42	624.50	215.08	1371.24	243.64	∞
7	14.21	82.34	265.67	366.07	82.12	213.54	744.23	263.30	609.85	∞
8	12.43	62.67	172.26	198.53	349.90	95.59	236.09	774.50	1525.38	∞
9	11.05	49.45	119.10	119.53	177.16	325.62	744.23	232.55	3814.20	255.50
10	233.90	40.13	86.43	77.65	100.77	156.50	284.75	628.28	595.55	511.50
12	161.45	28.13	50.60	230.48	214.94	255.70	359.02	590.64	355.35	2047.50
15	103.34	175.74	182.41	78.11	276.85	235.36	237.99	273.28	416.52	1023.50
20	58.81	76.69	62.39	418.60	168.39	384.88	253.17	781.43	467.76	387.57
25	38.32	41.45	139.07	105.02	466.51	179.52	296.20	522.37	611.22	244.76
30	27.16	156.53	304.15	127.97	339.36	320.93	341.02	409.84	854.41	698.36
40	132.88	56.81	294.32	197.01	231.34	325.33	539.31	330.92	550.09	449.66
50	71.87	122.46	313.14	310.07	512.43	369.34	320.42	322.87	468.75	383.79
60	44.10	259.02	350.55	175.52	410.77	446.41	584.74	346.63	456.95	373.97
75	144.01	87.88	242.32	367.97	350.74	404.22	280.23	423.88	336.02	408.63
100	53.92	269.61	233.46	203.48	294.40	250.43	264.50	323.81	344.34	283.78
125	114.11	230.09	247.87	312.00	322.32	391.64	348.50	302.61	405.43	448.64
150	236.96	212.37	277.03	487.53	328.99	325.25	397.79	312.95	292.92	415.21
200	232.30	205.73	369.92	293.54	449.07	283.78	401.49	389.35	333.08	438.20
250	247.93	218.57	239.72	415.14	375.82	291.06	466.50	338.18	424.39	346.88
300	277.07	243.89	365.36	334.78	354.15	324.03	373.21	330.07	384.13	315.03
375	197.13	295.67	325.43	283.55	313.92	413.01	343.26	430.82	381.22	370.46
500	191.51	212.70	315.86	429.44	305.61	327.88	433.87	345.48	335.79	370.31
750	226.85	322.73	366.85	342.82	328.98	332.09	417.71	357.78	396.70	413.18
1000	299.66	278.72	327.42	369.68	330.66	329.68	410.44	373.71	374.09	377.50
1500	297.39	322.73	384.38	331.86	369.83	371.82	384.52	388.98	345.32	398.12

3.4.2.3 Run-length distribution

Figure A3.6 showed the discrepancy between the false alarm rate (*FAR*) of the *p*-chart in Case K and the nominal *FAR* of 0.0027 i.e. the *FAR* associated with a 3-sigma X-bar chart for a normal process. Because the run-length distribution holds more information than the *FAR* it is instructive to also look at graphs of the run-length distribution of the *p*-chart in Case K compared to the run-length distribution of the 3-sigma Shewhart X-bar chart.

Figure A3.7 displays boxplot-like graphs of the run-length distributions of the *p*-chart in Case K for $n = 10, 25 \text{ and } 50$ when $p_0 = 0.05, 0.1, 0.2, 0.3 \text{ and } 0.5$. Also shown in Figure A3.7 is the boxplot-like graph of the 3-sigma Shewhart X-bar chart, which has a *FAR* of 0.0027, an in-control *ARL* of 370.4 and an in-control *SDRL* of 369.9.

The properties of the 3-sigma Shewhart X-bar chart are the nominally expected values for a 3-sigma chart such as the *p*-chart. We therefore typically use the performance characteristics of the X-bar chart as benchmark values for that of the *p*-chart (or any other Shewhart-type chart) in Case K.

Table A3.7 accompanies Figure A3.7 and shows the false alarm rate (*FAR*), the average run-length (*ARL*), the standard deviation of the run-length (*SDRL*) as well as the 1st, the 5th, the 10th, the 25th, the 50th, the 75th, the 95th and the 99th percentiles of all the run-length distributions displayed in Figure A3.7. The 25th percentile is the 1st quartile (typically denoted by Q_1), the 50th percentile is the 2nd quartile (also denoted by Q_2 and called the median run-length, or simply the *MDRL*), whereas the 75th percentile is the 3rd quartile (in some cases denoted by Q_3). These percentiles are all important descriptive statistics. For example, the inter-quartile range (*IQR*) is calculated as the difference between the 3rd and 1st quartiles, that is, $IQR = Q_3 - Q_1$. The *IQR* measures the spread of the middle 50% in the run-length distribution. The median run-length (*MDRL*) is a robust measure of the central tendency (location) of the run-length distribution and sometimes preferred instead of the average run-length.

All the abovementioned characteristics of the *p*-chart were computed using expressions (3-12) through (3-17) in Table 3.1. The properties of the 3-sigma Shewhart X-bar chart were calculated using expressions available in the literature (see e.g. Chakraborti, (2000)).

We assume that the 1st percentile is the minimum possible run-length and that the 99th percentile is the maximum achievable run-length and therefore compute the range (*R*) of the run-length distribution as the

difference between the 99th percentile and the 1st percentile, that is, $R = \max - \min =$ 99th percentile – 1st percentile.

Figure A3.7 shows that for $n = 10$ and $n = 25$ the run-length distribution of the *p*-chart is much different from that of the X-bar chart. For example, for $p_0 = 0.05$, 0.1 and 0.2 the *ARL* and the *SDRL* are both far less than the *ARL* of 370.4 and the *SDRL* of 369.9 of the X-bar chart (see e.g. Table A3.7). The range of the run-length distributions are also less. For $p_0 = 0.3$ and 0.5 the converse holds. In other words, the *ARL*, the *SDRL* and the range of the run-length distribution of the *p*-chart are all larger than what we would nominally expect from a 3-sigma Shewhart-type control chart.

For $n = 50$, the run-length distribution is more like that of the X-bar chart in that the *ARL* is approximately equal to 370.4, the *SDRL* is almost 369.9 and the range of the run-length distribution is close to being between 4 (the 1st percentile of the X-bar chart) and 1704 (the 99th percentile of the X-bar chart). However, the run-length distribution is still not exactly the same. This shows that even if the true fraction nonconforming is specified (known) and *n* is large, the *p*-chart still does not perform as (nominally) expected.

Table A3.7: Properties of the in-control (IC) run-length distribution of the *p*-chart for $n = 10, 25$ and 50 when $p_0 = 0.05, 0.1, 0.2, 0.3$ and 0.5 in Case K, and that of the 3-sigma Shewhart X-bar chart

					Percentiles / Quartiles								
n	***p*₀**	***FAR***	***ARL***	***SDRL***	**1st**	**5th**	**10th**	**25th (*Q*₁)**	**50th (*MDRL*)**	**75th (*Q*₃)**	**90th**	**95th**	**99th**
	0.05	0.0115	86.9	86.4	1	5	10	25	60	120	200	259	399
	0.10	0.0128	78.2	77.7	1	4	9	23	54	108	179	233	358
***n* = 10**	**0.20**	0.0064	157.0	156.5	2	9	17	46	109	217	361	469	721
	0.30	0.0016	628.8	628.3	7	33	67	181	436	871	1447	1883	2894
	0.50	0.0020	512.0	511.5	6	27	54	148	355	710	1178	1533	2356
	0.05	0.0072	139.6	139.1	2	8	15	41	97	193	321	417	641
	0.10	0.0095	105.5	105.0	2	6	12	31	73	146	242	315	484
***n* = 25**	**0.20**	0.0056	180.0	179.5	2	10	19	52	125	249	414	538	827
	0.30	0.0019	522.9	522.4	6	27	56	151	363	725	1203	1565	2406
	0.50	0.0041	245.3	244.8	3	13	26	71	170	340	564	734	1128
	0.05	0.0032	313.6	313.1	4	17	33	91	218	435	722	939	1443
	0.10	0.0032	310.6	310.1	4	16	33	90	215	430	714	929	1428
***n* = 50**	**0.20**	0.0027	369.8	369.3	4	19	39	107	257	513	851	1107	1701
	0.30	0.0031	323.4	322.9	4	17	35	93	224	448	744	968	1487
	0.50	0.0026	384.3	383.8	4	20	41	111	267	533	884	1150	1768
3-sigma X-bar		0.0027	370.4	369.9	4	19	39	107	257	513	852	1109	1704

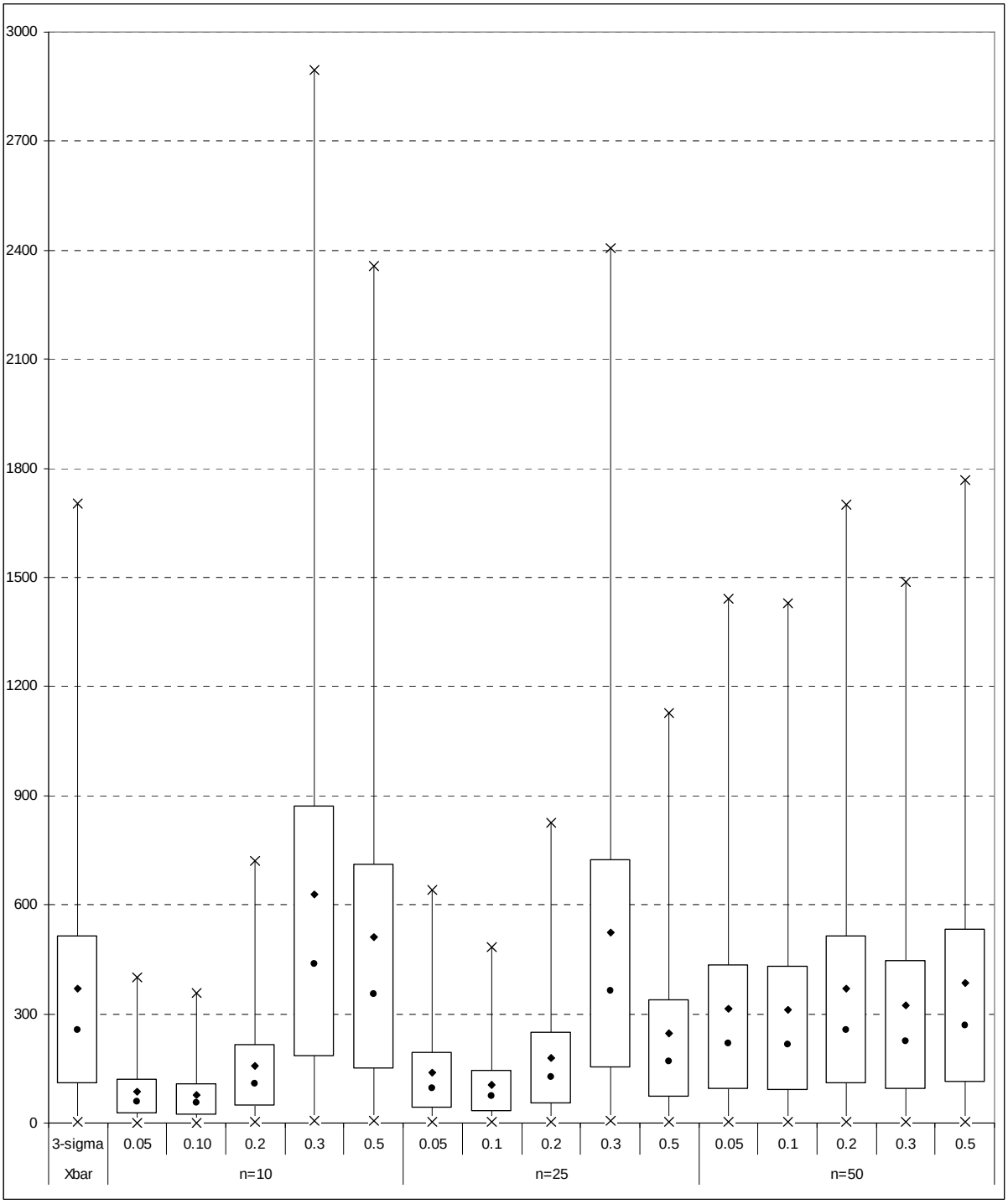


Figure A3.7: Boxplot-like graphs of the in-control (IC) run-length distribution of the *p*-chart for $n = 10, 25$ and 50 when $p_0 = 0.05, 0.1, 0.2, 0.3$ and 0.5 in Case K compared to the run-length distribution of the 3-sigma Shewhart X-bar chart

The foregoing discussion focused on the performance of the *p*-chart as measured by the false alarm rate, the average run-length, the standard deviation of the run-length and the percentiles of the run-length distribution and compared the *p*-chart's performance to that of the well-known 3-sigma Shewhart X-bar chart. It is also useful and important to know how to design a *p*-chart. The design of the *p*-chart in Case K is therefore looked at next.

3.4.2.4 The OC-curves and *ARL* curves

When designing a *p*-chart in Case K we need to choose a sample size n and while doing so keep in mind the size of the shift we are interested in detecting i.e. by how much the true fraction nonconforming p will differ from its specified value p_0 once a shift occurs.

Choosing the appropriate sample size is typically carried out by looking at a family of OC-curves or a family of *ARL*-curves, which are obtained by plotting multiple (at least two) OC-curves or multiple *ARL*-curves on the same set of axis.

Recall that an OC-curve is a graph (plot) of the probability of a no-signal $\beta(p, p_0, n)$ on the vertical axis for some values of $0 < p < 1$ on the horizontal axis. Hence, a family of two OC-curves is obtained by plotting $\beta(p, p_0, n = n_1)$ and $\beta(p, p_0, n = n_2)$, where n_1 and n_2 denote two different sample sizes, on the same set of axes; hence, each OC-curve corresponds to a specific sample size (in this case n_1 or n_2) but the value of p_0 is the same for each curve. Similarly, an *ARL*-curve is a graph (plot) of the average run-length $ARL(p, p_0, n)$ on the vertical axis for some values of $0 < p < 1$ on the horizontal axis so that family of two *ARL*-curves is obtained by plotting $ARL(p, p_0, n = n_1)$ and $ARL(p, p_0, n = n_2)$ on the same set of axes.

Suppose that we would like to compare and decide between two control charting plans to monitor the specified fraction nonconforming of $p_0 = 0.5$. Further, assume that both plans use a *p*-chart with 3-sigma control limits; the first plan uses $n = 25$ items per sample whereas the second plan uses double that i.e. $n = 50$; the question is then what the effect of sampling twice as many items is.

To assist us with our choice between the two control charting plans Figure A3.8 shows the OC-curve of each of the control charting plans. In other words, Figure A3.8 shows a family of two OC-curves where $\beta(p, p_0 = 0.5, n = 25)$ and $\beta(p, p_0 = 0.5, n = 50)$ are plotted on the vertical axis versus $0 < p < 1$ on the

horizontal axis. In addition, Table A3.8 lists some values of $\beta(p,0.5,25)$ and $\beta(p,0.5,50)$ for values of $p = 0.05(0.05)0.95$.

Figure A3.8 shows that the plan that uses $n = 50$ items per sample has a consistently lower β-risk or OC. Thus, if the objective is to detect a shift in the fraction nonconforming as soon as possible and we can afford the extra cost of sampling, this plan will be preferred. In the language of hypothesis testing, this shows that with all other things being equal, the power of test to detect a shift is higher for a larger sample size.

Figure A3.9 displays a family of two *ARL*-curves which corresponds to the OC-curves of Figure A3.8, that is, Figure A3.9 shows the average run-lengths $ARL(p, p_0 = 0.5, n = 25)$ and $ARL(p, p_0 = 0.5, n = 50)$ as functions of $0 < p < 1$. A decision based on the OC-curves of Figure A3.8 and a decision based on the *ARL*-curves of Figure A3.9 will therefore be exactly the same; this is so since the relationship between the average run-length and the probability of a no-signal is one-to-one and given by $ARL(p, p_0, n) = (1 - \beta(p, p_0, n))^{-1}$. Table A3.8 also shows the exact numerical values of $ARL(p,0.5,n)$ for $n = 50$ and 25 at $p = 0.05(0.05)0.95$.

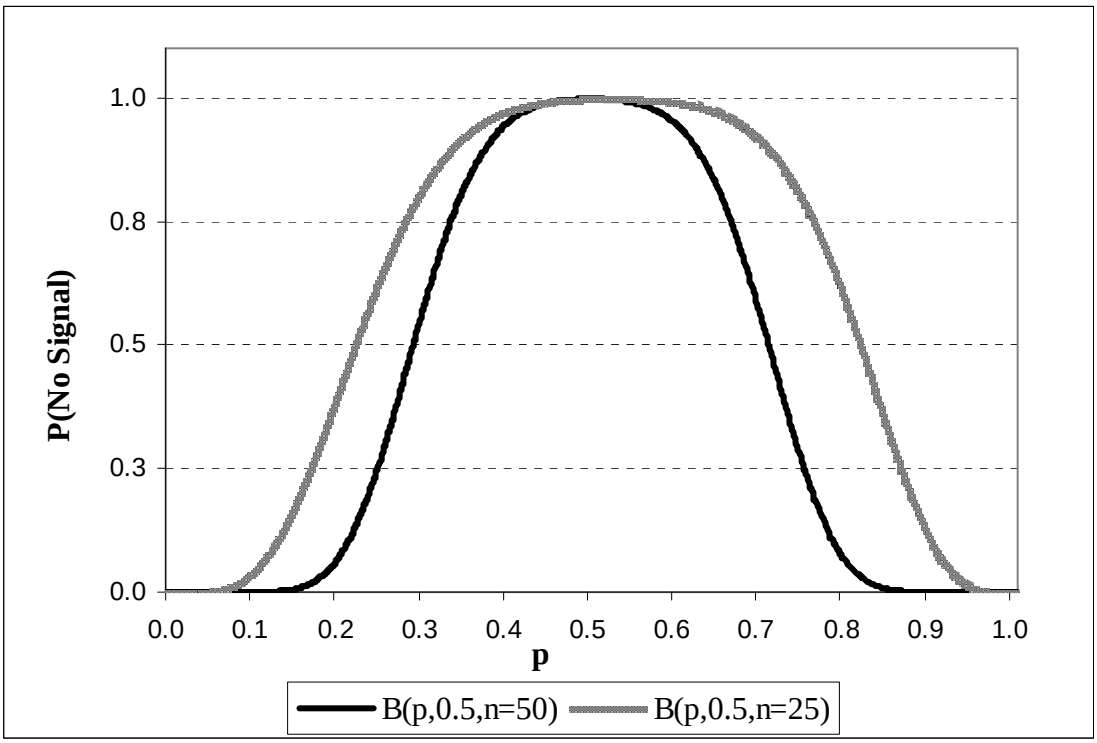


Figure A3.8: Family of Operating Characteristic (OC) Curves for the *p*-chart for a specified fraction nonconforming of $p_0 = 0.5$ when $n = 50$ and 25

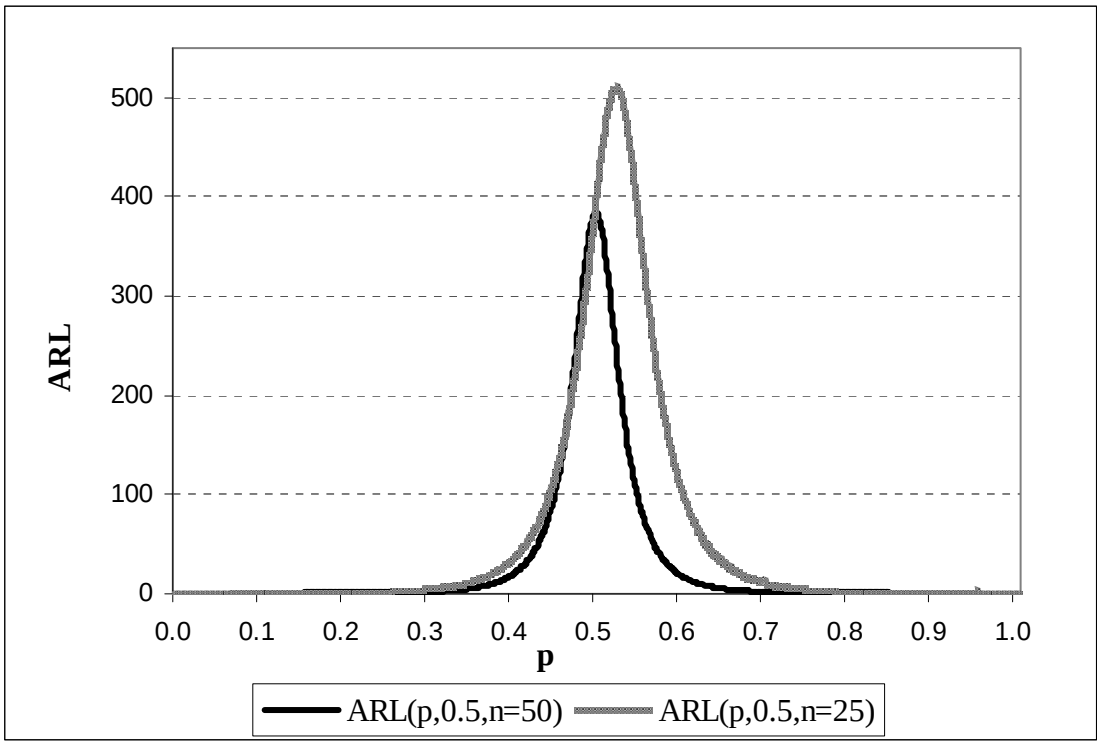


Figure A3.9: Family of Average Run-Length (*ARL*) Curves for the *p*-chart for a specified fraction nonconforming of $p_0 = 0.5$ when $n = 50$ and 25

Table A3.8: The values of the OC and the *ARL* of a *p*-chart n Case K when $p_0 = 0.5$ and $n = 50$ and 25

	Probability of a no-signal / Operating Characteristic (OC)		Average Run-Length (*ARL*)	
p	$\beta(p,0.5,n=50)$	$\beta(p,0.5,n=25)$	$ARL(p,0.5,n=50)$	$ARL(p,0.5,n=25)$
0.05	0.0000	0.0012	1.00	1.00
0.10	0.0001	0.0334	1.00	1.03
0.15	0.0053	0.1615	1.01	1.19
0.20	0.0607	0.3833	1.06	1.62
0.25	0.2519	0.6217	1.34	2.64
0.30	0.5532	0.8065	2.24	5.17
0.35	0.8122	0.9174	5.33	12.10
0.40	0.9460	0.9706	18.53	34.05
0.45	0.9895	0.9913	95.37	115.35
0.50	0.9974	0.9975	384.29	400.98
0.55	0.9895	0.9973	95.37	371.83
0.60	0.9460	0.9905	18.53	104.99
0.65	0.8122	0.9679	5.33	31.20
0.70	0.5532	0.9095	2.24	11.05
0.75	0.2519	0.7863	1.34	4.68
0.80	0.0607	0.5793	1.06	2.38
0.85	0.0053	0.3179	1.01	1.47
0.90	0.0001	0.0980	1.00	1.11
0.95	0.0000	0.0072	1.00	1.01

Summary

The *p*-chart is well-known, easy to use and its' applications is based on the implicit assumption that the binomial distribution is well approximated by the normal distribution, which, as one might expect, is not always the case. For example, as the preceding discussion shows, in some cases (especially for small values of n) the FAR is zero which implies that the ARL, the $SDRL$ and other moments are undefined. Moreover, the performance of the *p*-chart with a known or given or specified value for p might not be anything like that of the 3-sigma X-bar chart.

The *p*-chart is used to monitor the fraction nonconforming in a sample. The *c*-chart on the other hand is used to monitor the number of nonconformities in an inspection unit and is based on the Poisson distribution. We study the Case K *c*-chart in the next sections.

3.4.3 The *c*-chart in Case K: An example

We first look at an example of a *c*-chart in Case K to illustrate the typical application of the chart and investigate the characteristics of the chart for a specific value of c_0 (the specified value of c) while varying $c > 0$ (true average number of nonconformities in an inspection). The performance of chart is then further studied in subsequent sections by considering multiple (various) values of c_0.

Example A2: A Case K *c*-chart

Suppose that the true average number of nonconformities in an inspection unit c is known or specified to be $c_0 = 14$. The 3-sigma control limits for the *c*-chart are

$$UCL_c = 14 + 3\sqrt{14} = 25.22 \qquad CL_c = 14 \qquad LCL_c = 14 - 3\sqrt{14} = 2.78$$

and are calculated using (3-2).

Table A3.9 shows ten values simulated from a $Poi(14)$ distribution. We can assume without loss of generality that the values (counts) are the charting statistics of the *c*-chart; we therefore denote them by Y_i for $i = 1,2,...,10$. The *c*-chart is shown in Figure A3.10. The chart displays the upper control limit (*UCL*), the center line (*CL*), the lower control (*LCL*) and the Y_i's from each inspection unit plotted on the vertical axis versus the inspection unit number (time) on the horizontal axis. We see from Figure A3.10 that none of the 10 points plot out-of-control.

As long as no point plots outside the control limits we continue to monitor the process; this involves obtaining independent successive inspection units, calculating the charting statistic (i.e. the number of nonconformities) for each new inspection unit, and then plotting these one at a time on the chart. Once a point plots outside the limits it is taken as evidence that c is no longer equal to its specified value of $c_0 = 14$. A search for assignable causes is then started.

Table A3.9: Data for the *c*-chart in Case K

Inspection unit number / Time: i	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Counts: Y_i	17	9	17	12	16	16	9	21	15	11

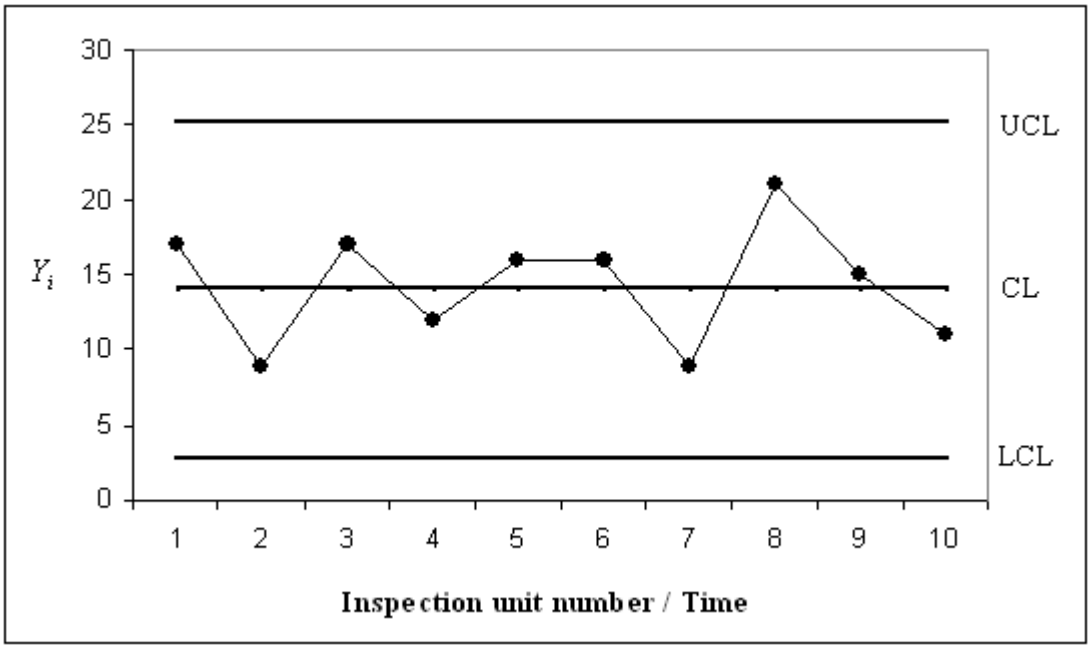


Figure A3.10: A *c*-chart in Case K

Performance of the *c*-chart

To study the performance of the aforementioned *c*-chart we analyze its in-control and out-of-control properties for which we need the probability of a no-signal, or equivalently, the probability of a signal. The probability of a no-signal completely characterizes the run-length distribution of the chart.

For $c_0 = 14$ it was shown that the upper control limit is $UCL_c = 25.22$ and the lower control limit is $LCL_c = 2.78$. Expression (3-8) shows that $d = [2.78] = 2$ and $f = [25.22] = 25$; these constants are needed to calculate the probability of a no-signal. We can study the in-control and the out-of-control performance of the chart by substituting values for c in the probability of a no-signal which is $\beta(c,14) = \Gamma_{26}(c) - \Gamma_3(c)$ and is found using (3-7).

In-control properties

As long as the true average number of nonconformities c remains unchanged and equal to its specified value of $c_0 = 14$ we deal with an in-control process. The probability of a no signal is then

$$\beta(c = 14,14) = \Gamma_{26}(14) - \Gamma_3(14) = 0.9973$$

so that the false alarm rate is $FAR = 1 - \beta(14,14) = 0.0027$.

The in-control run-length distribution is therefore geometric with probability of success equal to 0.0027, which we write as $N_0 \sim Geo(0.0027)$.

Out-of-control properties

When the true average number of nonconformities in an inspection unit changes, c is no longer equal to $c_0 = 14$ and implies that we have the out-of-control scenario. We look at the scenario when c increases; a decrease in c can be handled in a similar fashion.

Increase in c: Upward shift

Suppose c increases from 14 to 15; this is approximately a 7.14% increase in c. The probability of a no-signal decreases from 0.9973 (when the process was in-control) to

$$\beta(c = 15,14) = \Gamma_{26}(15) - \Gamma_3(15) = 0.9938$$

whereas the probability of a signal increases from 0.0027 to $1 - \beta(c = 15,14) = 0.0062$. The increase in the probability of a signal is good since the likelihood of detecting the shift increases.

The out-of-control run-length distribution is geometric with probability of success equal to 0.0062. Expression (3-21) shows that the out-of-control average run-length is $ARL_1 = 1/0.0062 = 160.66$. So, if it happens that c increases from 14 to 15 (and stays fixed at 15) one would expect the chart to detect such a shift (and signal) on approximately the 161st sample following the shift.

The OC-curve

The OC-curve and the probability of a signal as functions of c for $0 < c \leq 42$ are shown in Figure A3.11. In addition, Table A3.10 shows values of the probability of a no-signal $\beta(c,14) = \Gamma_{26}(c) - \Gamma_3(c)$ and the probability of a signal $1 - \beta(c,14) = 1 - \Gamma_{26}(c) + \Gamma_3(c)$ for values of $c = 2(2)42$.

Studying the OC-curve and the probability of a signal as function of c helps us see what the performance of our *c*-chart would be when a shift occurs. For example, if c was to decrease from $c = 14$ to $c = 8$ (which may be interpreted as an improvement in the process as approximately 42.9% less nonconformities (on average) in an inspection unit will in future be observed) we see from Table A3.10 that $1 - \beta(c = 8, c_0 = 14) = 0.0138$ so that the $ARL = 72.70$ and the $SDRL = 72.20$.

Note that, the two curves of Figure A3.11 are very similar to that of the *p*-chart considered earlier (see e.g. Figure A3.2); this is so because the values of n and p_0 (for the *p*-chart) and c_0 (in case of the *c*-chart) is such that the false alarm rate (*FAR*) of both the charts are 0.0027, and so, the IC run-length distributions of these charts and all other performance measures (including the OOC performance measures) are roughly the same.

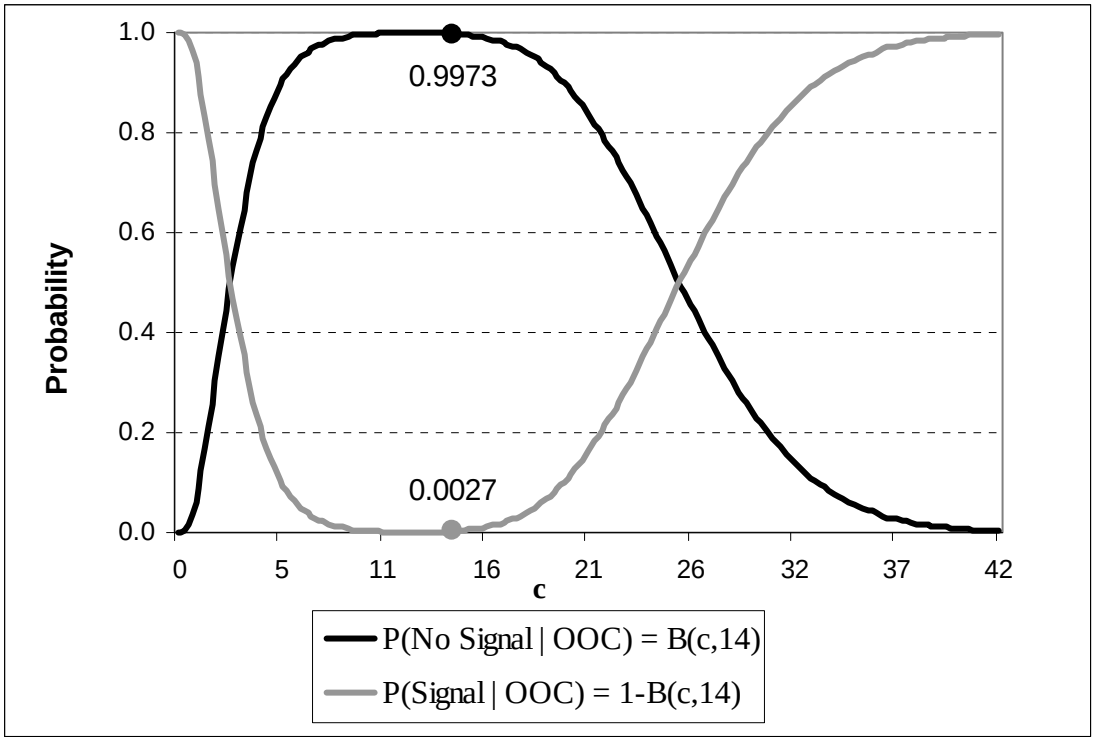


Figure A3.11: The OC-curve and the probability of a signal as a function of c when $c_0 = 14$

Table A3.10: The Probability of a no-signal and the Probability of a signal for $c = 2(2)42$ when $c_0 = 14$

c	P(No Signal | process OOC)	P(Signal | process OOC)	*ARL*	*SDRL*
2	0.3233	0.6767	1.48	0.84
4	0.7619	0.2381	4.20	3.67
6	0.9380	0.0620	16.14	15.63
8	0.9862	0.0138	72.70	72.20
10	0.9972	0.0028	358.80	358.30
12	0.9992	0.0008	1204.80	1204.30
14	0.9973	0.0027	370.16	369.66
16	0.9869	0.0131	76.13	75.63
18	0.9554	0.0446	22.42	21.91
20	0.8878	0.1122	8.91	8.40
22	0.7771	0.2229	4.49	3.95
24	0.6319	0.3681	2.72	2.16
26	0.4739	0.5261	1.90	1.31
28	0.3272	0.6728	1.49	0.85
30	0.2084	0.7916	1.26	0.58
32	0.1228	0.8772	1.14	0.40
34	0.0674	0.9326	1.07	0.28
36	0.0345	0.9655	1.04	0.19
38	0.0166	0.9834	1.02	0.13
40	0.0076	0.9924	1.01	0.09
42	0.0033	0.9967	1.00	0.06

Run-length distributions

Figure A3.12 displays boxplot-like graphs of the in-control and the out-of-control run-length distributions of the *c*-chart with the average run-lengths (*ARL*'s) and the median run-lengths (*MDRL*'s) indicated (the former by diamond symbols and the latter by circles). The exact numerical values of the *ARL*'s and the *MDRL*'s are also shown in Figure A3.12 and listed in Table A3.11 together with the probability of a no-signal, the probability of a signal and some percentiles (quartiles) of the in-control and the out-of-control run-length distributions.

The *ARL* and the *MDRL* measures the central tendency (location) of the run-length distribution. The *MDRL* however is more robust and outlier resistant than the *ARL*. In both the in-control and the out-of-control run-length distributions the *ARL* is larger than the *MDRL* and indicates that the in-control and the out-of-control run-length distributions are non-normal and positively skewed. The skewness of the run-length distributions is also observed by comparing the upper and the lower tails of each of the

distributions, that is, the distance between the 99th percentile and the 75th percentile to the distance between the 25th percentile and the 1st percentile; this comparison between the upper and lower tails is done separately for each distribution.

For example, for the in-control run-length distribution (when $c = c_0 = 14$) Table A3.11 shows that the distance between the 99th percentile and the 75th percentile is $1703 - 513 = 1190$ whereas the distance between the 25th and the 1st percentiles of the in-control run-length distribution is $107 - 4 = 103$. The latter is much larger (approximately 1190/103 = 11.5 times) than the former and shows, as mentioned before, that the in-control run-length distribution is positively skewed.

Most importantly however Figure A3.12 shows the overall difference between the in-control ($c = 14$) and the out-of-control ($c = 15$) run-length distributions. For example, the out-of-control average run-length is 160.7 compared to the in-control average run-length of 370.2. Similarly, the out-of-control median run-length is 112 versus the in-control median run-length of 257. Furthermore, the range (*R*) and the inter-quartile range (*IQR*) of the in-control and the out-of-control run-length distributions differ somewhat. Both the range and the inter-quartile range measure the spread (variation) in the run-length distributions. The range measures the overall spread and is the distance between the 99th percentile (maximum) and the 1st percentile (minimum). The *IQR*, on the other hand, is the distance between the 3rd and the 1st quartile, that is, $IQR = Q_3 - Q_1$ and measures the variation in the middle 50% of the distribution. For the in-control run-length distribution Table A3.11 shows that the range of the in-control run-length distribution is $R_0 = 1703 - 4 = 1699$ and that the inter-quartile range of the in-control run-length distribution is $IQR_0 = 513 - 107 = 406$. The values of R_0 and IQR_0 are both larger than that of the out-of-control run-length distribution. For the out-of-control run-length distribution we have that $R_1 = 738 - 2 = 736$ and $IQR_1 = 223 - 47 = 176$. This big discrepancy between the range and the inter-quartile range of the in-control and the out-of-control run-length distributions emphasizes that once a shift occurs, the run-length distribution is severely altered in that we can expect the chart to signal (detect the shift) sooner, which is of course good.

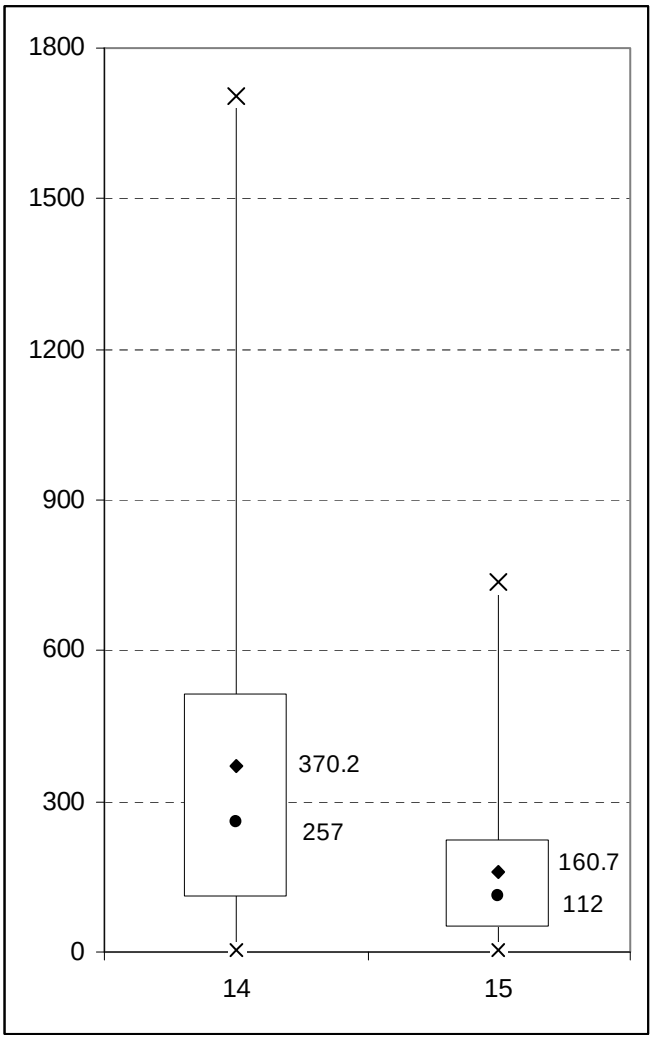


Figure A3.12: Boxplot-like graphs of the in-control (IC) and the out-of-control (OOC) run-length distributions of the *c*-chart in Case K

Table A3.11: Summary measures of the in-control (IC) and the out-of-control (OOC) run-length distributions of the *c*-chart when $c_0 = 14$ in Case K

	IC	OOC (increase in *c*)
c	**14**	**15**
Pr(No Signal)	0.9973	0.9938
Pr(Signal)	0.0027	0.0062
ARL	370.16	160.66
SDRL	369.66	160.16
1st percentile	4	2
5th percentile	19	9
10th percentile	39	17
25th (Q_1)	107	47
50th (*MDRL*)	257	112
75th (Q_3)	513	223
90th percentile	852	369
95th percentile	1108	480
99th percentile	1703	738

The preceding discussion focused on the properties of the *c*-chart for one particular value of c i.e. $c_0 = 14$. Other values of c_0 are also of interest in order to get an idea of the overall performance of the *c*-chart and can only be obtained by studying the characteristics of the *c*-chart for a wider range of values for c_0.

3.4.4 The *c*-chart in Case K: Characteristics of the in-control run-length distribution

To get a better idea of the overall performance of the *c*-chart in Case K we look at the run-length distribution and its characteristics for a range of values for c_0. We consider both small and large values of c_0.

To this end, Table A3.12 shows the control limits (LCL_c and UCL_c), the charting constants d and f, the probability of a no signal when the process is in-control, the false alarm rate (*FAR*), the in-control average run-length (ARL_0), and the in-control standard deviation of the run-length ($SDRL_0$) when $c_0 = 1(1)10(5)50$, 75 and 100, respectively.

Table A3.12 is accompanied by Table A3.13 which shows the percentiles of the run-length distributions of the *c*-chart for the same values of c_0. The values in columns (2) through (9) of Table A3.12 were computed using expressions (3-2), (3-7), (3-8) and the expressions in Table 3.2. The percentiles were calculated using expression (3-23) in Table 3.2.

For illustration purposes, consider the *c*-chart with $c_0 = 35$. Table A3.12 shows that

$$LCL_c = 35 - 3\sqrt{35} = 17.25 \quad \text{and} \quad UCL_c = 35 + 3\sqrt{35} = 52.75$$

so that $d = [17.25] = 17$ and $f = [52.75] = 52$. It thus follows that $\beta(c = 35, c_0 = 35) = 0.9967$, $ARL_0(c = 35, c_0 = 35) = 301.42$ and $SDRL_0(c = 35, c_0 = 35) = 300.92$. In addition Table A3.13 shows that the $MDRL_0 = 209$ and that the $Q_3 = 418$ and the $Q_1 = 87$ so that the $IQR = Q_3 - Q_1 = 402$.

However, note that, since the *FAR* and the *ARL* are most often used in OOC performance comparisons we primarily focus on the *FAR* and the in-control *ARL* in our discussion on the performance of the *c*-chart in Case K. In particular, we compare the *FAR* and the *ARL* of the *c*-chart in Case K to that of the well-known 3-sigma X-bar chart.

Table A3.12: Characteristics of the in-control run-length distribution of the *c*-chart for $c_0 = 1(1)10(5)50, 75$ and 100

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
c_0	LCL_c	UCL_c	d	f	Pr(No Signal | IC)	FAR	ARL_0	$SDRL_0$
1	-2.00	4.00	0	3	0.6131	0.3869	2.58	2.02
2	-2.24	6.24	0	6	0.8601	0.1399	7.15	6.63
3	-2.20	8.20	0	8	0.9464	0.0536	18.66	18.15
4	-2.00	10.00	0	9	0.9736	0.0264	37.81	37.31
5	-1.71	11.71	0	11	0.9878	0.0122	82.03	81.53
6	-1.35	13.35	0	13	0.9939	0.0061	163.74	163.24
7	-0.94	14.94	0	14	0.9934	0.0066	150.85	150.35
8	-0.49	16.49	0	16	0.9959	0.0041	246.70	246.20
9	0.00	18.00	0	17	0.9946	0.0054	183.72	183.22
10	0.51	19.49	0	19	0.9965	0.0035	285.74	285.23
15	3.38	26.62	3	26	0.9965	0.0035	283.83	283.33
20	6.58	33.42	6	33	0.9971	0.0029	339.72	339.22
25	10.00	40.00	10	39	0.9960	0.0040	248.14	247.64
30	13.57	46.43	13	46	0.9971	0.0029	349.94	349.44
35	17.25	52.75	17	52	0.9967	0.0033	301.42	300.92
40	21.03	58.97	21	58	0.9964	0.0036	275.36	274.86
45	24.88	65.12	24	65	0.9976	0.0024	413.04	412.54
50	28.79	71.21	28	71	0.9975	0.0025	396.70	396.20
75	49.02	100.98	49	100	0.9967	0.0033	299.77	299.27
100	70.00	130.00	70	129	0.9967	0.0033	307.36	306.86

Table A3.13: Percentiles of the in-control run-length distribution of the *c*-chart for $c_0 = 1(1)10(5)50, 75$ and 100

	Percentiles of the run-length distribution								
c_0	1^{st}	5^{th}	10^{th}	25^{th} (Q_1)	50^{th} ($MDRL$)	75^{th} (Q_3)	90^{th}	95^{th}	99^{th}
1	2	2	2	2	2	3	5	7	10
2	2	2	2	2	5	10	16	20	31
3	2	2	2	6	13	26	42	55	84
4	2	2	4	11	26	52	86	112	172
5	5	5	9	24	57	114	188	245	376
6	2	9	18	47	114	227	376	490	752
7	2	8	16	44	105	209	347	451	693
8	3	13	26	71	171	342	567	738	1134
9	2	10	20	53	128	255	422	549	844
10	3	15	31	83	198	396	657	855	1314
15	3	15	30	82	197	393	653	849	1305
20	4	18	36	98	236	471	782	1017	1563
25	3	13	27	72	172	344	571	742	1141
30	4	18	37	101	243	485	805	1047	1610
35	4	16	32	87	209	418	693	902	1386
40	3	15	29	80	191	382	633	824	1266
45	5	22	44	119	286	572	950	1236	1900
50	4	21	42	114	275	550	913	1187	1825
75	4	16	32	87	208	415	690	897	1379
100	4	16	33	89	213	426	707	920	1414

3.4.4.1 False alarm rate and average run-length

Figure A3.13 shows the percentage difference between the false alarm rate (*FAR*) of the *c*-chart in Case K and the nominal *FAR* of 0.0027 i.e. the *FAR* of a 3-sigma Shewhart X-bar chart. The percentage difference is seen to be mostly positive; only for $c_0 = 45$ and 50 is the percentage difference negative. It is also clear that, in general, the *FAR* is far from 0.0027; especially for small values of c_0 i.e. less than or equal to 10, say.

In particular, Figure A3.13 shows, in general, that (i) the *FAR* is hundreds of percents larger than 0.0027, and (ii) as c_0 increases the percentage difference gets smaller. A *c*-chart based on $c_0 = 6$, for instance, has a *FAR* of 0.0061, which is 126% larger than 0.0027 whereas a *c*-chart based on $c_0 = 10$ has a *FAR* of 0.0035 (which is 30% larger than 0.0027) and a *c*-chart based on $c_0 = 35$ has a *FAR* equal to 0.0033 (which is only 23% larger).

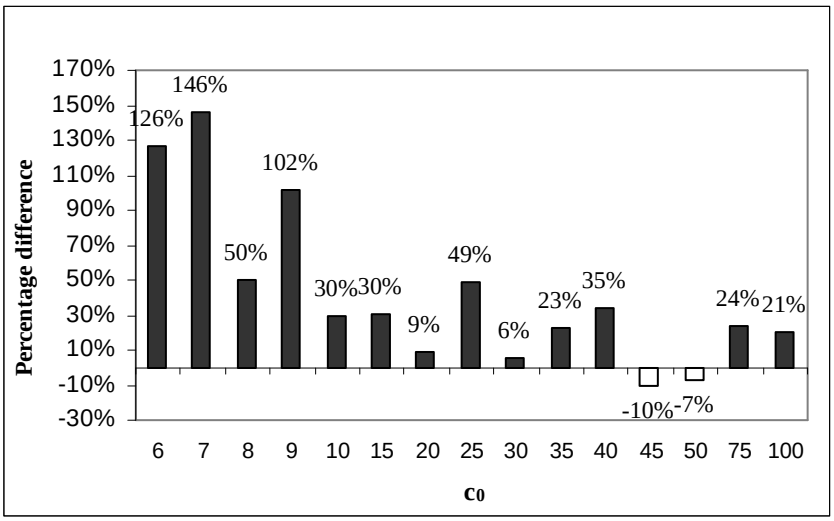


Figure A3.13: Percentage difference between the false alarm rate (*FAR*) of the *c*-chart and the nominal *FAR* of 0.0027 for $c_0 = 6(1)10(5)50$, 75 and 100

Figure A3.14 shows the percentage difference between the average run-length (*ARL*) of the *c*-chart in Case K and that of the nominal *ARL* of 370.4, which is the *ARL* of a 3-sigma Shewhart X-bar chart. The percentage difference is seen to be mostly negative and implies shorter in-control *ARL*'s than nominally expected from a 3-sigma chart like the *c*-chart. Thus, we can deduce that, unless the specified value c_0 of c is reasonably large, the *c*-chart will erroneously signal more often than what is nominally expected from a 3-sigma chart.

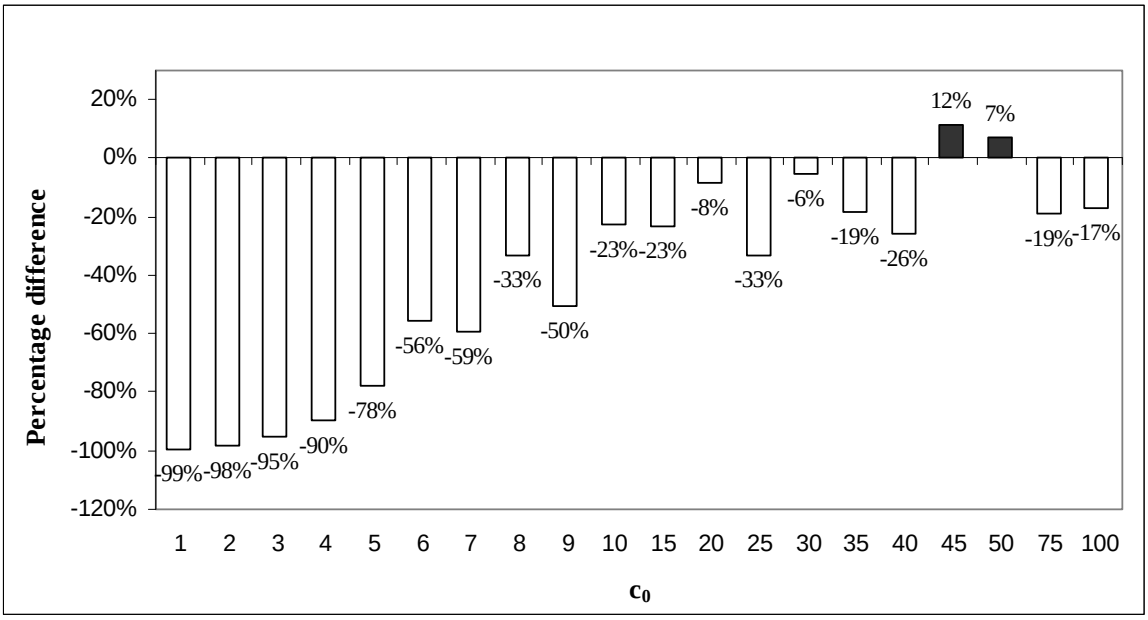


Figure A3.14: Percentage difference between the average run-length (*ARL*) of the *c*-chart and the nominal *ARL* of 370.4 for $c_0 = 1(1)10(5)50$, 75 and 100.

3.4.4.2 The run-length distribution

It is good to make a visual comparison of the run-length distributions since it gives us an overall idea of just how different (or similar) the run-length distribution of the *c*-chart is to that of the 3-sigma X-bar chart. Figure A3.15 displays boxplot-like graphs of the run-length distribution of the *c*-chart when $c_0 = 6(1)10(5)50$ and also shows a boxplot-like graph of the run-length distribution of the 3-sigma X-bar chart.

We see that, in general, for small values of c_0 the run-length distribution of the c-chart differs substantially from that of the 3-sigma X-bar chart in that the ARL_0 and the $MDRL_0$ are considerably smaller and the spread (as measured by the range *R*) in the run-length distribution of the *c*-chart is noticeably less than that of the X-bar chart. Only for larger values of c_0 does the run-length distribution of the c-chart become more like that of the 3-sigma X-bar chart.

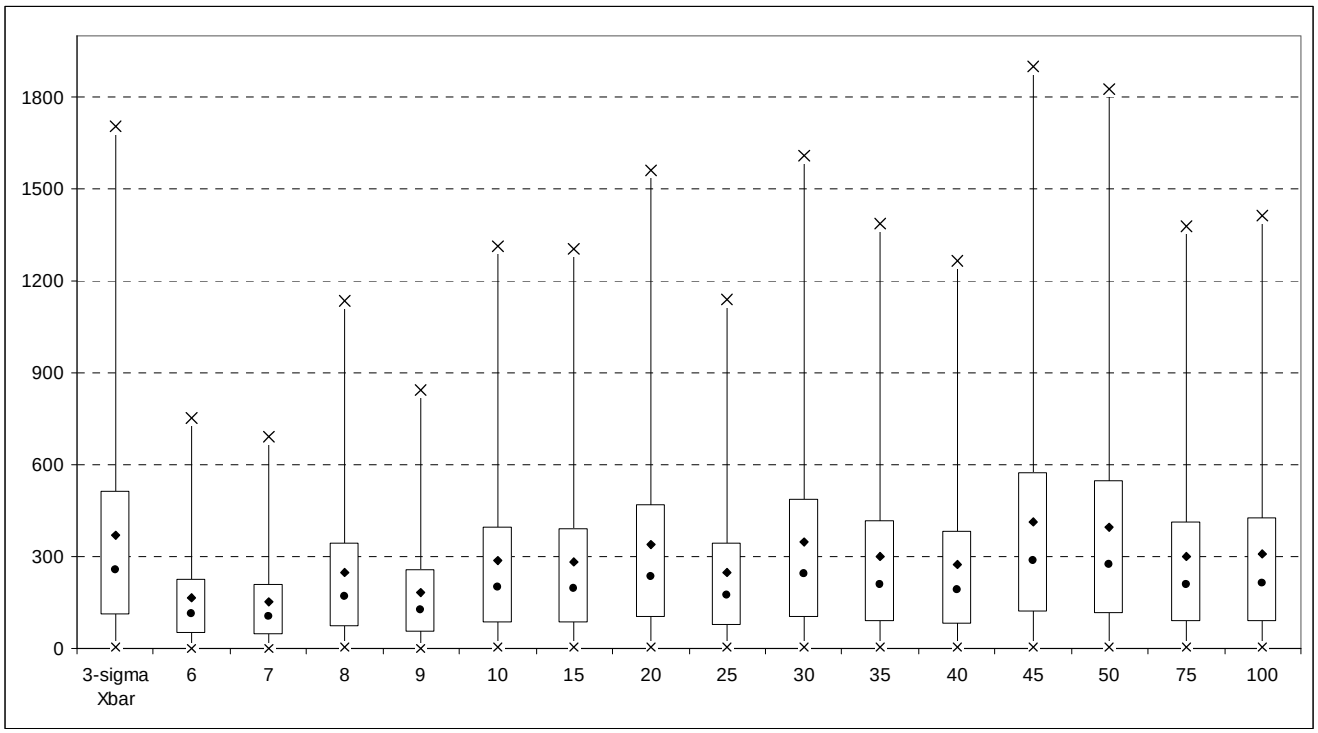


Figure A3.15: Boxplot-like graphs of the in-control (IC) run-length distribution of the *c*-chart for $c_0 = 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 75$ and 100 in Case K compared to the run-length distribution of the 3-sigma Shewhart X-bar chart

Summary

Like the *p*-chart, the *c*-chart is well-known and easy to apply but, even in Case K, the *c*-chart does not perform anything like the 3-sigma Shewhart X-bar chart. The discrepancy is due to the facts that

(i) when c is small the normal approximation to the Poisson distribution is poor so both the charting constant $k = 3$ and the charting formula (mean $\pm$ 3 standard deviations) may be inaccurate, and

(ii) due to the discrete nature of the Poisson distribution only certain *FAR* values can be attained.

Chapter 4

Nonparametric Shewhart-type control charts with runs-type signaling rules: Case K and Case U

4.0 Chapter overview

Introduction

The commonly used control charts in the statistical process control (SPC) environment with data that can be measured on a continuous numerical scale are generally designed and used with a specific parametric distribution, such as the normal distribution, in mind. It is well-known that if the underlying process is not as assumed, the performance of the parametric charts can be significantly degraded. In this context, one key problem is the lack of in-control robustness of some of the well-known parametric charts; this, for example, implies that there can be too many false alarms than what is typically expected and obviously this could mean considerable loss of time and resources (see e.g. Chakraborti et al., (2004)). Thus it is desirable, from a practical point of view, to develop and apply a set of control charts that are not designed under the assumption of normality (or any other parametric distribution). Such charts can expected to be more flexible in that they require no or little assumption regarding the underlying process distribution and would be useful in some applications.

To this end, nonparametric control charts are helpful. Nonparametric control charts have received considerable attention over the last few years. See, for example, Bakir (2004), Albers and Kallenberg (2004), Chakraborti et al. (2004) and Albers et al. (2006), where various nonparametric alternatives to the classical Shewhart-type charts are proposed, by adapting (for example) the corresponding nonparametric test for the process parameter, and have been shown to outperform the Shewhart $\overline{X}$ chart (and some other well-known charts) in terms of in-control robustness and efficiency, particularly

for heavy-tailed distributions. A thorough review of the literature on nonparametric charts can be found in Chakraborti et al. (2001, 2007).

The main advantage of the nonparametric charts is that they have known in-control properties that remain unchanged for all continuous distributions. Thus, for example, while the false alarm rate (FAR) of a Shewhart or Cumulative Sum (CUSUM) or Exponentially Weighted Moving Average (EWMA) chart for the mean will fluctuate depending on the underlying distribution, the FAR of a nonparametric chart can be calculated exactly and will remain the same (for in-control conditions) no matter what the distribution; this is a very useful feature for the practitioner.

A formal definition of a nonparametric or distribution-free control chart is given in terms of its run-length distribution. The number of samples that needs to be collected before the first out-of-control signal is given by a chart is a random variable called the run-length; the probability distribution of the run-length is referred to as the run-length distribution. If the in-control run-length distribution is the same for every continuous probability distribution the chart is called distribution-free or nonparametric (see e.g. Chakraborti et al., (2004)).

Note that, the term nonparametric is not intended to imply that there are no parameters involved, quite to the contrary. While the term distribution-free seems to be a better description of what one expects these charts to accomplish, nonparametric is perhaps the term more often used. In this chapter, both terms (distribution-free and nonparametric) are used to emphasize the fact that they mean the same.

Motivation

To construct a nonparametric control chart for the specified (or known or target) median of a distribution that is continuous and symmetric Bakir (2004) considered a Shewhart-type chart based on the Wilcoxon signed-rank (SR) test statistic. This chart, referred to as the *1-of-1* SR chart, signals when a single charting statistic falls outside the control limits, was shown to compete well with the Shewhart $\overline{X}$ chart in the case of the normal distribution and it performed better in the case of a heavier-tailed (than the normal) distribution such as the double exponential and the Cauchy. However, the false alarm rates for the *1-of-1* SR chart were just too high (i.e. the in-control average run-lengths were too short) for an application in practice, unless the subgroup size n was in the neighborhood of 20, which is not typical in SPC.

Having realized the potential and yet the practical shortcoming of the *1-of-1* SR chart, Chakraborti and Eryilmaz (2007) extended the idea of Bakir (2004) and considered an alternative class of nonparametric charts using the same SR statistic as charting statistic but incorporating some signaling rules based on runs; these charts are called runs-rule enhanced signed-rank charts. The new SR charts were shown to be more appealing from a practitioner's point of view in that they have much larger attainable in-control average run-lengths (ARL_0 's), much smaller attainable *FAR*'s and have better out-of-control (OOC) performance than the *1-of-1* SR chart.

Although the SR charts are useful, the requirement that the underlying distribution be symmetric is an additional assumption to be verified and may in fact not be satisfied in some situations in practice. If not much knowledge is available about the shape of the distribution, an alternative nonparametric test called the sign test can be used to make inference about any percentile including the median, whereas the SR test applies only to the median. The sign test does not require the assumption of symmetry of the underlying continuous population distribution (see e.g. Gibbons and Chakraborti, (2003)) and is easy to apply. Another advantage is that one does not require the actual measurements to be available to apply the sign test; all one needs to know is how many of the observations within each sample are larger (or smaller) than the specified value of the parameter (percentile) of interest. Thus the sign test can be applied with binary data, when the only information available, for each unit tested, is whether or not the measurement was higher (or lower) than the target value of the percentile of interest. Neither the normal theory chart nor the SR charts can be applied with just the dichotomized data.

A further requirement for applying the SR charts (and charts based on the sign test) is that the in-control process median (or percentile) must be specified; a situation commonly referred to as Case K. This may not be the case in some applications and could limit the application of the charts, with or without signaling rules. For example, when a new product is being developed not much information or expert knowledge may be initially available to specify the distribution and/or the in-control value of the percentile of interest. Hence there is a need to also develop control charts when the in-control process percentile of interest (or, in general, the location) is unknown. This is the scenario where the process distribution is continuous and unknown (no symmetry necessary) and the in-control percentile (or the location parameter) is unknown or unspecified (unlike in Case K); this situation is referred to as Case U.

Our objective is to overcome the drawbacks of the SR charts by studying and developing a new class of nonparametric control charts with runs-type signaling rules for the scenario where the

percentile (or location parameter) of interest of the process distribution is known and then, second, when it is unknown, without having to assume symmetry of the underlying process distribution. In the former situation (or Case K) the control charts are based on the well-known sign test statistic while in the latter scenario (or Case U) the charts are based on the two-sample median test statistic.

It will be seen that the charts we consider provide a new class of flexible, yet powerful, nonparametric charts to be used in practice.

Although one can consider other types of nonparametric charts such as the CUSUM and the EWMA (see e.g. the reviews by Chakraborti et al. (2001, 2007)), in this chapter, we keep the discussion focused on the Shewhart-type charts because of their inherent practical appeal and global effectiveness (see e.g. Montgomery, (2005) p. 385). The development of nonparametric CUSUM and EWMA charts will be a topic for future research.

Methodology

We use a Markov chain approach (see e.g. Fu and Lou, (2003)) to derive the necessary results (such as the run-length distributions, average run-lengths etc.) for our runs-rule enhanced charts because this approach provides a more compact and unified view of the derivations, and as stated by Balakrishnan and Koutras (2002), p.14, “The Markov techniques possess a great advantage (over the classical combinatory methods) as they are easily adjustable to many run-related problems; they often simplify the solutions to specific problems they are applied on and remain valid even for cases involving non-identical or dependent trials”. In some cases, however, we draw on the results of the geometric distribution of order k (see e.g. Balakrishnan and Koutras, (2002), Chapter 2) to obtain closed form and explicit expressions for the run-length distributions and/or their associated performance characteristics.

In Case U, like in Chapter 3, we use a two-step approach to derive the run-length distributions which involve the method of conditioning (see e.g. Chakraborti, (2000)). First we derive the conditional run-length distributions i.e. conditioned on two order statistics (control limits), which lets one focus on specific values of the control limits. Second we derive the unconditional (or marginal) run-length distributions by averaging over the joint distribution of the two order statistics. The unconditional run-length distributions reflect the bigger picture and reveal the overall performance of the charts taking into account that the control limits are estimated.

Layout of Chapter 4

In Section 4.1 we define and describe in detail (using graphs) the runs-type signaling rules for the one-sided and two-sided charts. In Section 4.2 we derive the run-length distributions of our new nonparametric control charts with signaling rules for Case K and then, in Section 4.3, we derive the run-length distributions of the charts for Case U. Section 4.4 gives a summary and some concluding remarks.

4.1 Runs-type signaling rules

Introduction

We consider a class of nonparametric Shewhart-type control charts for monitoring the percentile (or location parameter) of a process of which the distribution is assumed to be continuous but not necessarily symmetric. First we study the scenario where the π^{th} percentile of the process distribution is known and then, second, when it is unknown. In the former situation (or Case K), studied in Section 4.2, the control charts are based on the well-known sign test statistic while in the latter scenario (or Case U), which is looked at in Section 4.3, the charts are based on the two-sample median test statistic.

Signaling rules

The new control charts are "runs-rule enhanced" charts in which a process is declared out-of-control (OOC) when either

(i) A single charting statistic plots outside the control limits (*1-of-1* chart), or

(ii) k consecutive charting statistics all plot outside the control limits (*k-of-k* chart), or

(iii) exactly k of the last w charting statistics plot outside the control limits (*k-of-w* chart).

It is clear that rules (i) and (ii) are special cases of rule (iii). Rule (i) is the simplest and most frequently used whilst rules (ii) and (iii) have been used in the context of the parametric Shewhart-type charts such as the well-known $\overline{X}$ chart (see e.g. Derman and Ross, (1997) and Klein, (2000)).

One-sided and two-sided charts

We consider one-sided and two-sided control charts. Amongst the one-sided charts two situations can arise: (i) when just upward shifts are of interest so that an upper control limit is adequate, and (ii) when only detecting downward shifts are of interest so that a lower control limit is sufficient. The former is called the positive-sided (or upper one-sided) chart whereas the latter is labeled the negative-sided (or lower one-sided) chart. We study both the upper and the lower one-sided charts.

When a shift in any direction (up or down) is of concern a two-sided chart is used which has an upper and a lower control limit.

Charting statistic and control limits

We denote the charting statistic for the i^{th} subgroup, in general, by Q_i for $i = 1,2,3,...$ and the upper and the lower control limits by UCL and LCL, respectively; this allows us to simultaneously deal with Case K and Case U when we define and describe the runs-type signaling rules. Later, when we individually discuss the control charts of Case K and Case U we define and replace Q_i, UCL and LCL by their appropriate and representative counterparts.

Signaling indicators

Let

$$\xi_i^+ = I(Q_i \geq UCL) = \begin{cases} 1 & \text{if} \quad Q_i \geq UCL \\ 0 & \text{if} \quad Q_i < UCL \end{cases} \tag{4-1}$$

and

$$\xi_i^- = I(Q_i \leq LCL) = \begin{cases} 1 & \text{if} \quad Q_i \leq LCL \\ 0 & \text{if} \quad Q_i > LCL \end{cases} \tag{4-2}$$

for $i = 1,2,3,...$ denote the indicator functions for the one-sided charts corresponding to the events $\{Q_i \geq UCL\}$ and $\{Q_i \leq LCL\}$, respectively. In other words, ξ_i^+ (ξ_i^-) denotes the signaling indicator for the event when Q_i plots on or above (below) the upper (lower) control limit of the positive-sided (negative-sided) chart. Likewise, we let

$$\xi_i = \begin{cases} 1 & \text{if} \quad Q_i \geq UCL \\ 0 & \text{if} \quad LCL < Q_i < UCL \\ 2 & \text{if} \quad Q_i \leq LCL \end{cases} \tag{4-3}$$

denote the signaling indicator for the two-sided chart so that ξ_i indicates whether Q_i plots on or above the UCL (in which case $\xi_i = 1$), between the LCL and the UCL (so that $\xi_i = 0$), or on or below the LCL ($\xi_i = 2$).

Remark 1

Not only do the signaling indicators in (4-1), (4-2) and (4-3) allow us to clearly define and describe signaling rules (i), (ii) and (iii), but their statistical properties (e.g. whether they are independent, their "success" probabilities such as $\Pr(\xi_i^+ = 1)$, $\Pr(\xi_i^- = 1)$ and $\Pr(\xi_i^\pm = 1)$ etc.) are also important since they play a key role in deriving the run-length distributions of the new class of proposed runs-rule enhanced charts.

4.1.1 The *1-of-1* charts

The *1-of-1* charts are the least complicated and most frequently used charts. The signaling rules for the ***1-of-1*** charts, as defined by the signaling indicators in (4-1), (4-2) and (4-3), are given by:

The ***1-of-1* upper one-sided** chart signals when event A_1 occurs where:

$$A_1 : \{Q_i \geq UCL\} \Leftrightarrow \{\xi_i^+ = 1\}^*,$$

The ***1-of-1* lower one-sided** chart signals when event A_2 occurs where:

$$A_2 : \{Q_i \leq LCL\} \Leftrightarrow \{\xi_i^- = 1\}, \text{ and}$$

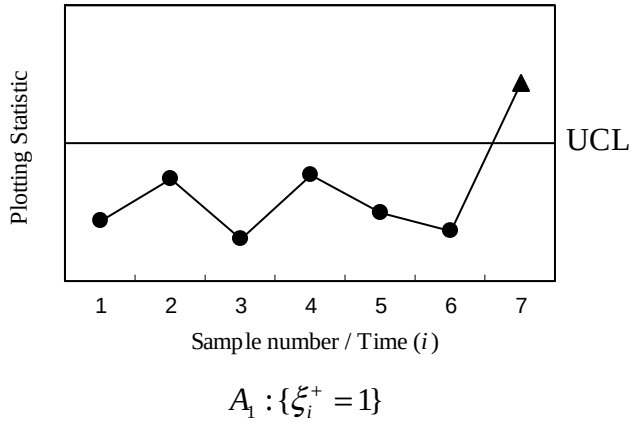
The ***1-of-1* two-sided** chart signals when event A occurs where:

$$A : \{Q_i \geq UCL \text{ or } Q_i \leq LCL\} \Leftrightarrow \{A_1 \text{ or } A_2\} \Leftrightarrow \{\xi_i^+ = 1 \text{ or } \xi_i^- = 1\} \Leftrightarrow \{\xi_i = 1 \text{ or } \xi_i = 2\}.$$

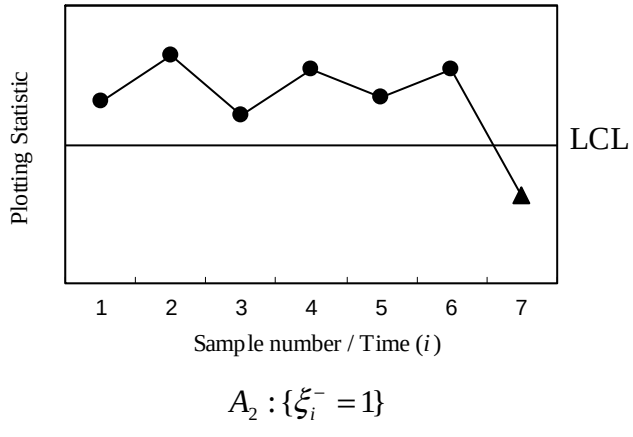
* The symbol "$\Leftrightarrow$" in an expression such as $P \Leftrightarrow Q$ is read as 'P is true if and only if Q is true'.

For illustration, panels (a) and (b) of Figure 4.1 show examples of the *1-of-1* upper and lower one-sided charts whilst Figure 4.2 displays examples of the *1-of-1* two-sided chart.

The *1-of-1* upper (lower) one-sided charts signals, for illustration only, at time $i = 7$ when Q_7 plots above (below) the UCL (LCL). The process is therefore declared OOC with the conclusion of an upward (downward) shift in the process location. Similarly, both of the *1-of-1* two-sided charts signal at time $i = 7$; the first chart signals when Q_7 plots above the upper control limit (indicative of an upward shift) whereas the second chart signals when Q_7 plots below the LCL (indicative of a downward shift).



(a) The *1-of-1* rule for the upper one-sided chart (upward shift detected)



(b) The *1-of-1* rule for the lower one-sided chart (downward shift detected)

Figure 4.1: The *1-of-1* rule for the upper and the lower one-sided charts

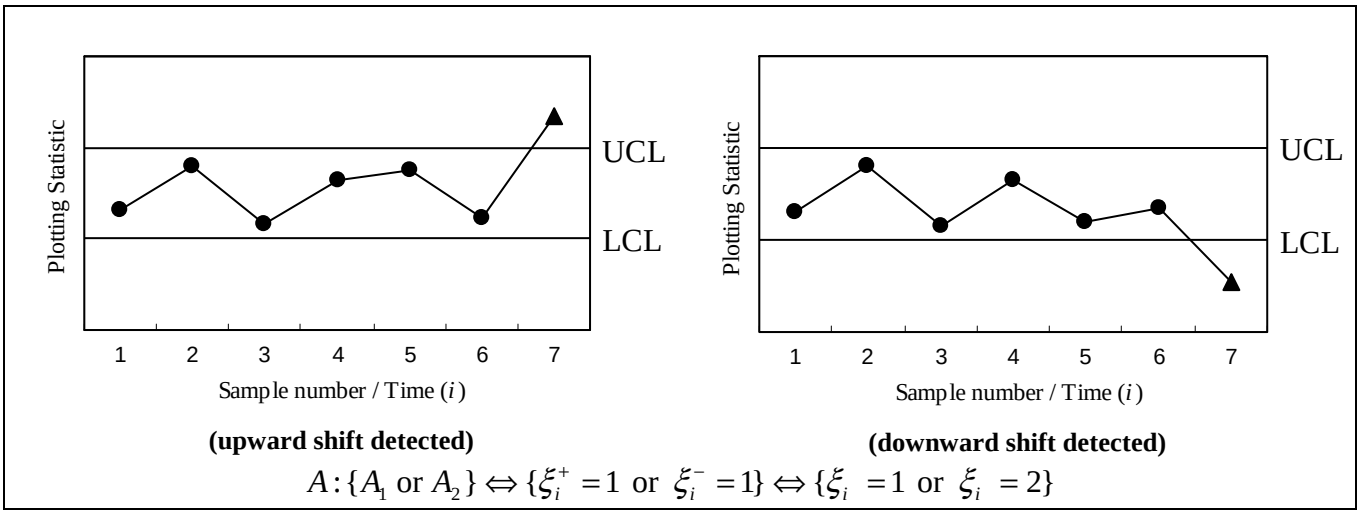


Figure 4.2: The *1-of-1* rule for the two-sided chart

Because the *1-of-1* charts are based on signaling rule (i) which uses only the information from the most recent (last) sample to make a decision whether or not the process is in-control (IC), one feels these charts can be improved upon by using rules (ii) and (iii); this idea is discussed in the next section.

4.1.2 The *k-of-k* and *k-of-w* charts

The runs-rule Shewhart-type charts we consider adopt a sequential approach and use the information from multiple samples including the most recent one to signal. Unlike the CUSUM and EWMA charts, the sequential approach we study is carried out by considering runs of the charting statistic outside the control limit(s) which includes the charting statistic from the current sample and one or more charting statistics from past samples. The resulting charts are easy to use and it will be seen that they offer the user greater practical flexibility in designing a chart so that more (practically) attractive, attainable, false alarm rates are available. Moreover, it will be shown that the new charts have higher efficiency (i.e. smaller OOC *ARL*'s) compared to the *1-of-1* charts.

According to the *k-of-k* $(k \geq 2)$ chart the control chart signals at any point in time when k consecutive charting statistics (from k consecutive samples), of which the last one is the most recent one, all plot outside the control limit(s). A generalization of the *k-of-k* chart is the *k-of-w* chart which signals when exactly k of the last w charting statistics all plot outside the limit(s), of which the last one plots outside the control limits.

It is clear that we can consider charts for any pair of positive integers k and w where $1 \leq k \leq w$ and $w \geq 2$. Although various values of k and w can be considered in theory, from a practical point of view, it is important that the resulting charts are easy to apply; so we focus on the *2-of-2* $(k = w = 2)$ and the *2-of-3* $(k = 2, w = 3)$ charts. A user is therefore required to keep track of only the last two or three of the most recent charting statistics.

4.1.2.1 One-sided *k-of-k* and *k-of-w* charts

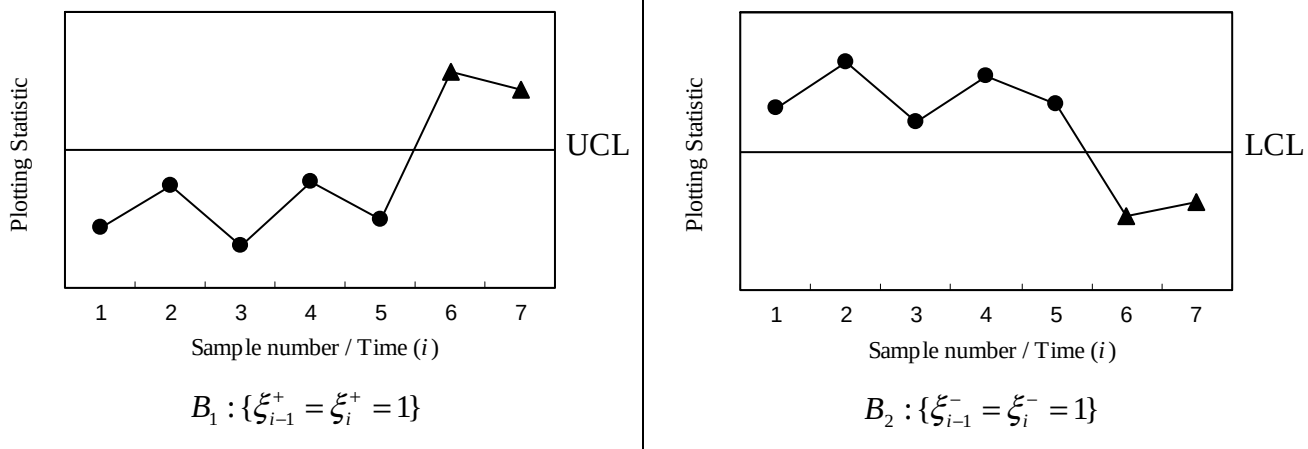
As noted earlier, the upper (lower) one-sided chart has only an upper (lower) control limit and is typically used when only an upward (downward) shift is of concern. We first describe the *2-of-2* upper (lower) one-sided chart and then the *2-of-3* upper (lower) one-sided chart.

One-sided *2-of-2* charts

The *2-of-2* chart requires the user to keep track of only the last two charting statistics Q_{i-1} and Q_i at any given point in time, $i \geq 2$. The upper one-sided *2-of-2* chart signals (declares the process OOC) if both Q_{i-1} and Q_i plot on or outside the upper control limit; otherwise no signal is given and we declare the process IC. Likewise, the lower one-sided *2-of-2* chart signals if both Q_{i-1} and Q_i plot on or outside the lower control limit. Thus the ***2-of-2*** one-sided charts are:

The ***2-of-2* upper one-sided** chart signals when the event B_1 occurs
where $B_1 : \{\xi_{i-1}^{+} = \xi_i^{+} = 1\}$, and
The ***2-of-2* lower one-sided** chart signals when the event B_2 occurs
where $B_2 : \{\xi_{i-1}^{-} = \xi_i^{-} = 1\}$.

For illustration, panels (a) and (b) of Figure 4.3 show examples of the *2-of-2* upper and the *2-of-2* lower one-sided charts. Both of the charts signal, again for illustration only, at time $i = 7$ i.e. on the first occurrence of a run of length two of the charting statistic above (below) the UCL (LCL) at time or sample number 7. Hence, the process is declared OOC with the conclusion of an upward (downward) shift in the process location.



(a) The *2-of-2* rule for the upper one-sided chart (upward shift detected)

(b) The *2-of-2* rule for the lower one-sided chart (downward shift detected)

Figure 4.3: The *2-of-2* rule for the upper and the lower one-sided charts

Remark 2

The *2-of-2* one-sided charts can be alternatively defined in terms of the minimum and/or the maximum of the statistics Q_{i-1} and Q_i. For example, the *2-of-2* upper one-sided chart signals if $\min(Q_{i-1}, Q_i)$ plots on or above the *UCL*. Similarly the *2-of-2* lower one-sided chart signals if $\max(Q_{i-1}, Q_i)$ plots on or below the *LCL*. We next consider a generalization of the *2-of-2* chart.

One-sided *2-of-3* charts

The *2-of-2* charts utilize moving (over time) blocks of only two charting statistics. It is therefore natural to investigate if there can be any sizeable gain in efficiency when moving blocks of three charting statistics are utilized. Thus we consider *2-of-3* charts for which at any time point $i \geq 3$ we need to keep track of Q_{i-2}, Q_{i-1} and Q_i; the upper (lower) one-sided *2-of-3* chart signals if two of these three statistics plot on or above (below) the upper (lower) control limit.

A similar chart was considered by Klein (2000) for the Shewhart $\overline{X}$ chart. However, note that although there are three ways for exactly two of the last three charting statistics to plot on or above (below) the upper (lower) control limit, we take only two of the three ways, namely where the last charting statistic plots on or above (below) the upper (lower) control limit to define a signal.

This is unlike the chart of Klein (2000) which can signal in any of the three ways; our charts eliminate the possibility that the process is declared OOC when both the first and the second charting statistics plot outside the control limits but the third (last) one plots between the control limits (see e.g. Figure 4.5), which we feel is somewhat undesirable in practice. Thus our ***2-of-3*** one-sided charts are:

The ***2-of-3* upper one-sided** chart signals when the event $C(+)=\{C_1 \text{ or } C_2\}$ occurs

where $C_1:\{Q_{i-2}<UCL \text{ and } Q_{i-1}\geq UCL \text{ and } Q_i \geq UCL\} \Leftrightarrow \{\xi^+_{i-2}=0, \xi^+_{i-1}=1, \xi^+_i=1\}$

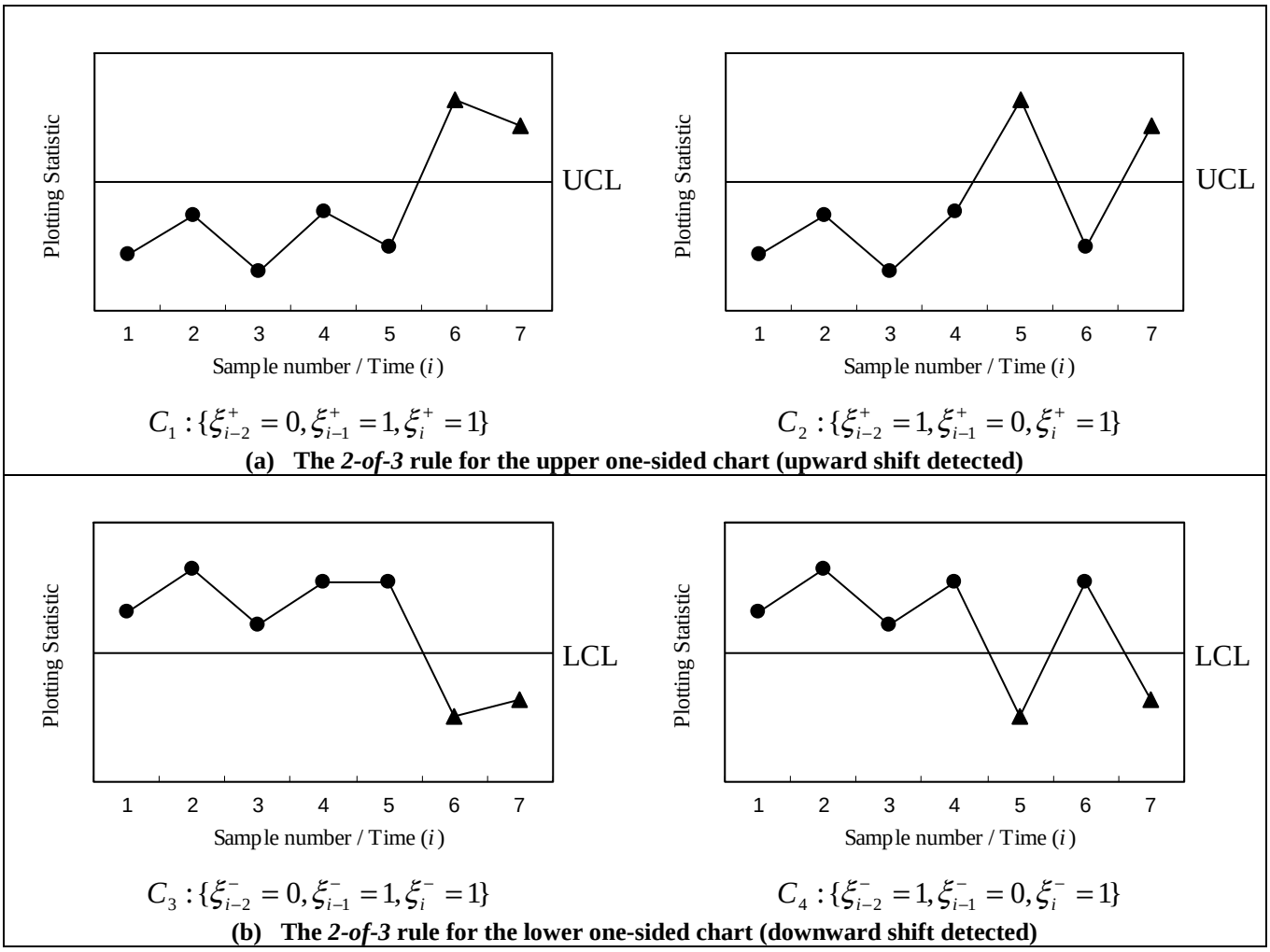
$C_2:\{Q_{i-2}\geq UCL \text{ and } Q_{i-1}<UCL \text{ and } Q_i \geq UCL\} \Leftrightarrow \{\xi^+_{i-2}=1, \xi^+_{i-1}=0, \xi^+_i=1\}$.

The ***2-of-3* lower one-sided** chart signals when the event $C(-)=\{C_3 \text{ or } C_4\}$ occurs

where $C_3:\{Q_{i-2}>LCL \text{ and } Q_{i-1}\leq LCL \text{ and } Q_i \leq LCL\} \Leftrightarrow \{\xi^-_{i-2}=0, \xi^-_{i-1}=1, \xi^-_i=1\}$

$C_4:\{Q_{i-2}\leq LCL \text{ and } Q_{i-1}>LCL \text{ and } Q_i \leq LCL\} \Leftrightarrow \{\xi^-_{i-2}=1, \xi^-_{i-1}=0, \xi^-_i=1\}$.

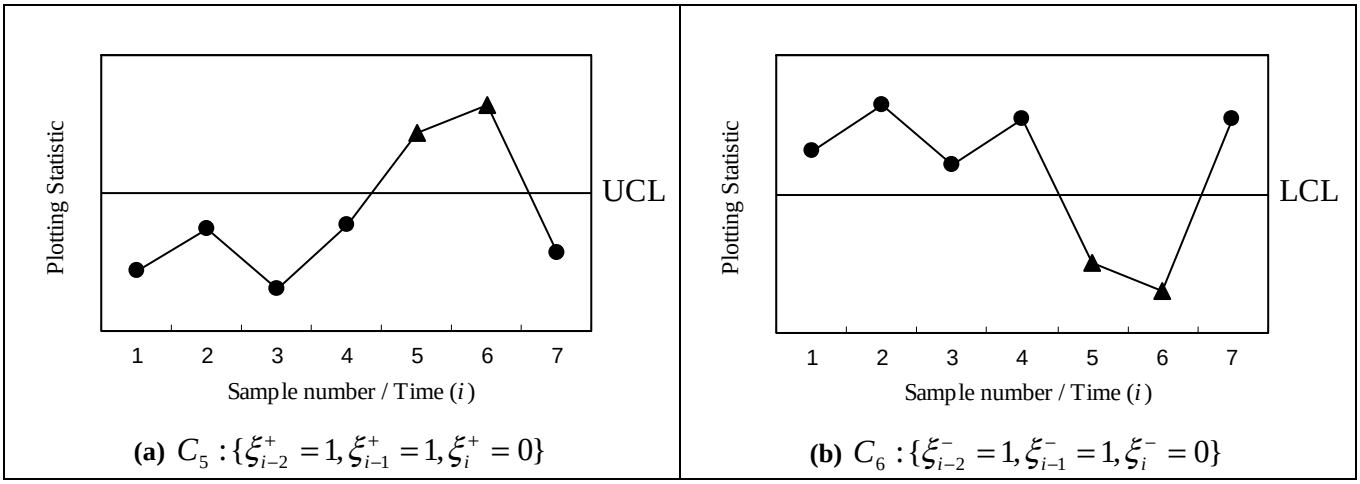
Panels (a) and (b) of Figure 4.4 show examples of what the signaling events C_1, C_2, C_3 and C_4 might look like in case of the *2-of-3* upper and lower one-sided charts. For example, both of the *2-of-3* upper one-sided charts, shown in panel (a), signal at time $i=7$ and the signals are interpreted to be indicative of an upward shift since both the charting statistics fall above the upper control limit.



(a) The *2-of-3* rule for the upper one-sided chart (upward shift detected)

(b) The *2-of-3* rule for the lower one-sided chart (downward shift detected)

Figure 4.4: The *2-of-3* rule for the upper and the lower one-sided charts



(a) $C_5 : \{\xi^+_{i-2} = 1, \xi^+_{i-1} = 1, \xi^+_i = 0\}$

(b) $C_6 : \{\xi^-_{i-2} = 1, \xi^-_{i-1} = 1, \xi^-_i = 0\}$

Figure 4.5: The *2-of-3* events excluded as signaling events for the upper and the lower one-sided charts

4.1.2.2 Two-sided *k-of-k* and *k-of-w* charts

Two-sided charts are typically used to detect either an upward or a downward shift and thus have both an upper and a lower control limit.

Two-sided *2-of-2* charts

Like the *2-of-2* one-sided charts, we also need to keep track of both Q_{i-1} and Q_i at any time point $i \geq 2$, but for the two-sided chart there are two control limits and so there are four ways for Q_{i-1} and Q_i to plot outside the limits. Any one (or more) of the four scenarios may be used to define a signal. We consider two *2-of-2* two-sided charts; both capable of detecting an upward or a downward shift in the location parameter.

The first *2-of-2* two-sided chart signals when any two successive charting statistics both plot on or outside the control limits. In other words, a signal is given when:

(i) both charting statistics plot on or above the UCL, or
(ii) both charting statistics plot on or below the LCL, or
(iii) the first charting statistic plots on or above the UCL and the second charting statistic plots on or below the LCL, or
(iv) the first charting statistic plots on or below the LCL and the second charting statistic plots on or above the UCL.

This signaling rule was proposed by Derman and Ross (1997) in the context of the Shewhart $\overline{X}$ chart; we refer to this chart as the ***2-of-2* DR** two-sided chart.

The second *2-of-2* two-sided chart signals when two successive charting statistics:

(i) both plot on or above the UCL, or
(ii) both plot on or below the LCL.

This signaling rule was considered by Klein (2000) in the context of the Shewhart $\overline{X}$ chart; we refer to this chart as the ***2-of-2* KL** two-sided chart.

More specifically, the ***2-of-2*** two-sided charts are:

The ***2-of-2*** **DR two-sided** chart signals when the event $D(DR):\{D_1 \text{ or } D_2 \text{ or } D_3 \text{ or } D_4\}$ occurs, and

The ***2-of-2*** **KL two-sided** chart signals when the event $D(KL)=\{D_1 \text{ or } D_2\}$ occurs

where

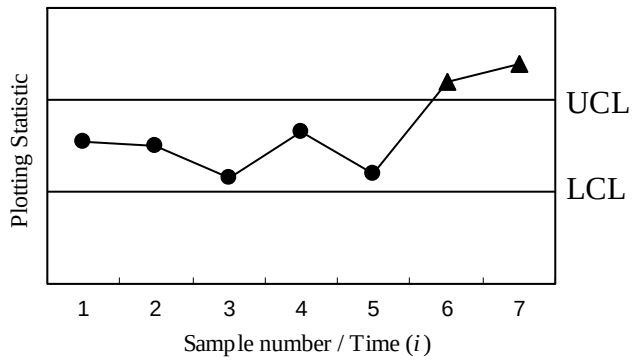
$$D_1:\{Q_{i-1}\geq UCL \text{ and } Q_i \geq UCL\} \Leftrightarrow \{\xi_{i-1}=1,\xi_i=1\},$$

$$D_2:\{Q_{i-1}\leq LCL \text{ and } Q_i \leq LCL\} \Leftrightarrow \{\xi_{i-1}=2,\xi_i=2\},$$

$$D_3:\{Q_{i-1}\geq UCL \text{ and } Q_i \leq LCL\} \Leftrightarrow \{\xi_{i-1}=1,\xi_i=2\}, \text{ and}$$

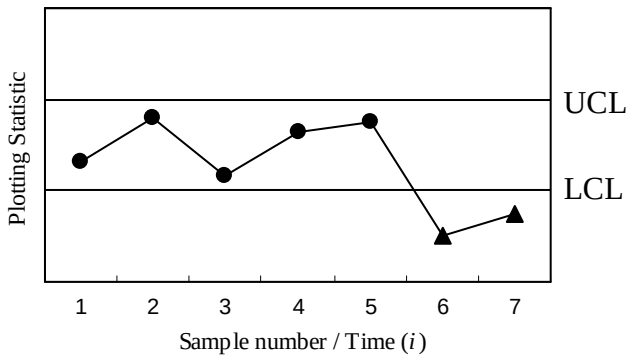
$$D_4:\{Q_{i-1}\leq LCL \text{ and } Q_i \geq UCL\} \Leftrightarrow \{\xi_{i-1}=2,\xi_i=1\}.$$

Figure 4.6 shows some examples of the events D_i, $i=1,2,3,4$. It is clear that the *2-of-2* DR chart signals on the seventh sample in each of the four panels of Figure 4.6 whereas the *2-of-2* KL chart signals only in panels (a) and (b). Thus, whenever the *2-of-2* KL chart signals so does the *2-of-2* DR chart, but the converse may not always happen. Furthermore, it seems that the *2-of-2* DR and KL charts are both suitable for detecting an upward or a downward shift, but the *2-of-2* DR chart can also detect a possible swing; this is when an upward shift is immediately followed by a downward shift or vice versa (Chakraborti and Eryilmaz, (2007)).



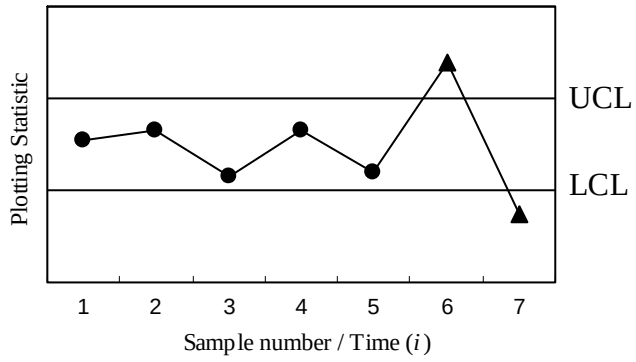
$D_1 : \{\xi_{i-1} = 1, \xi_i = 1\}$

(a) (signal above / upward shift detected)



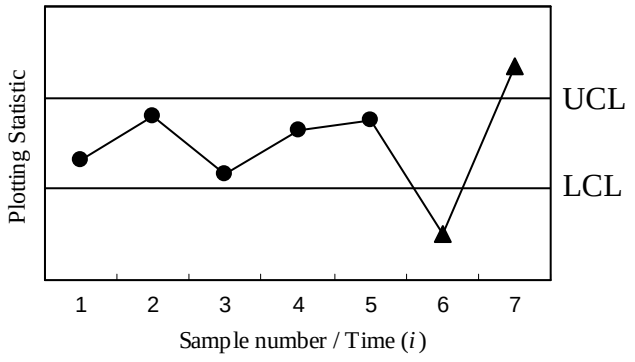
$D_2 : \{\xi_{i-1} = 2, \xi_i = 2\}$

(b) (signal below / downward shift detected)



$D_3 : \{\xi_{i-1} = 1, \xi_i = 2\}$

(c) (signal above followed by a signal below / possible swing detected)



$D_4 : \{\xi_{i-1} = 2, \xi_i = 1\}$

(d) (signal below followed by a signal above / possible swing detected)

Figure 4.6: The *2-of-2* rule for the two-sided charts

Two-sided *2-of-3* chart

Analogous to the *2-of-3* one-sided charts the *2-of-3* two-sided chart signals when:

(i) exactly two of the last three charting statistics both plot on or above the UCL, or

(ii) exactly two of the last three charting statistics both plot on or below the LCL.

Hence, the ***2-of-3*** two-sided chart defined in terms of the signaling indicator ξ_i is given by:

The ***2-of-3* two-sided** chart signals when event $E = \{E_1 \text{ or } E_2 \text{ or } E_3 \text{ or } E_4\}$ occurs where

$E_1 : \{LCL < Q_{i-2} < UCL \text{ and } Q_{i-1} \geq UCL \text{ and } Q_i \geq UCL\} \Leftrightarrow \{\xi_{i-2} = 0, \xi_{i-1} = 1, \xi_i = 1\}$,

$E_2 : \{Q_{i-2} \geq UCL \text{ and } LCL < Q_{i-1} < UCL \text{ and } Q_i \geq UCL\} \Leftrightarrow \{\xi_{i-2} = 1, \xi_{i-1} = 0, \xi_i = 1\}$,

$E_3 : \{LCL < Q_{i-2} < UCL \text{ and } Q_{i-1} \leq LCL \text{ and } Q_i \leq LCL\} \Leftrightarrow \{\xi_{i-2} = 0, \xi_{i-1} = 2, \xi_i = 2\}$, and

$E_4 : \{Q_{i-2} \leq LCL \text{ and } LCL < Q_{i-1} < UCL \text{ and } Q_i \leq LCL\} \Leftrightarrow \{\xi_{i-2} = 2, \xi_{i-1} = 0, \xi_i = 2\}$.

Figure 4.7 displays examples of events E_i for $i = 1,2,3,4$ and shows that when there is a signal, the proposed *2-of-3* two-sided chart offers a practical interpretation for the signal. For example, when either event E_1 or E_2 occurs (shown in panels (a) and (b)) the signal is interpreted as an upward shift. Similarly, if either event E_3 or E_4 occurs (displayed in panels (c) and (d)) a downward shift is inferred.

Remark 3

Apart from events E_1, E_2, E_3 and E_4 there are a further eight scenarios in case of the *2-of-3* two-sided chart where exactly two of the last three charting statistics can plot outside the control limits. We, however, exclude these events as signaling events when we calculate the statistical characteristics or properties of the *2-of-3* two-sided control charts; even though four of the events may possibly be linked to genuine or tangible changes in the process.

Figure 4.8 shows the eight events together with the practical interpretations (if any). For example, panels (a) and (b) show events E_5 and E_6 that could be considered a swing in the process, whereas panels (c) and (d) show events E_7 and E_8 that could be interpreted as trends (up or down) in the process. The events in panels (e), (f), (g) and (h) are excluded as signaling events because, as mentioned earlier, the last point plots between the control limits.

Most importantly, by excluding events E_5, E_6, …, E_{12} we are left with events E_1, E_2, E_3 and E_4, which makes the signaling events of the *2-of-3* one-sided charts and that of the *2-of-3* two-sided chart more alike (compare, for example, the signaling events shown in Figure 4.4 with that of Figure 4.7).

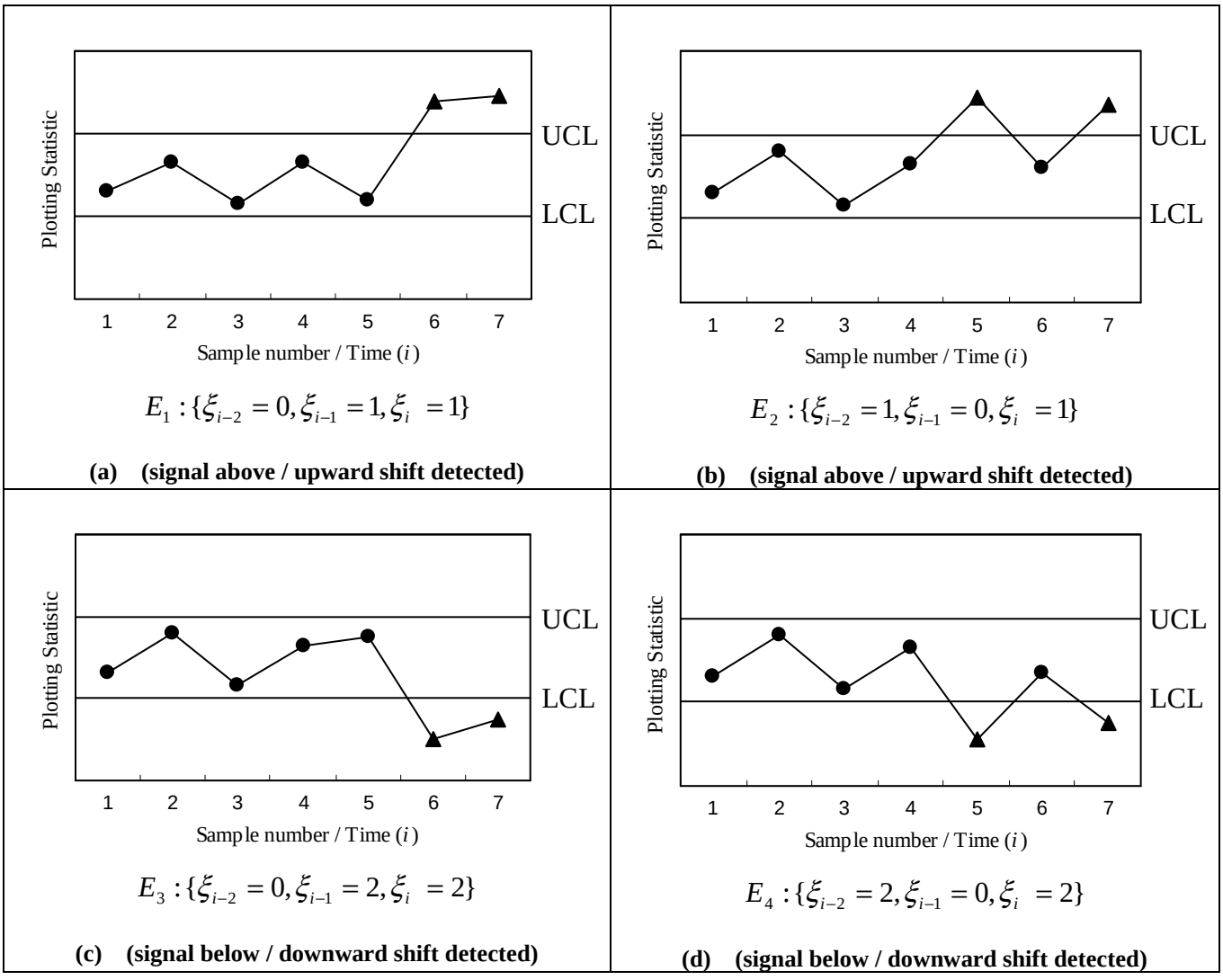
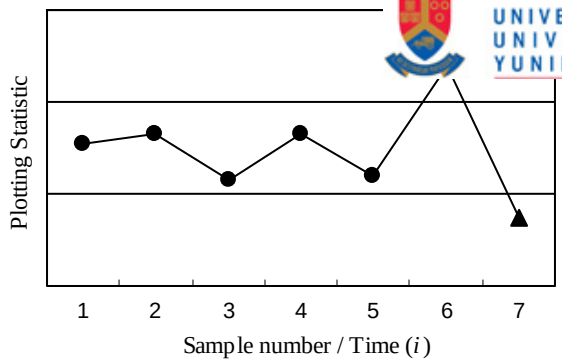
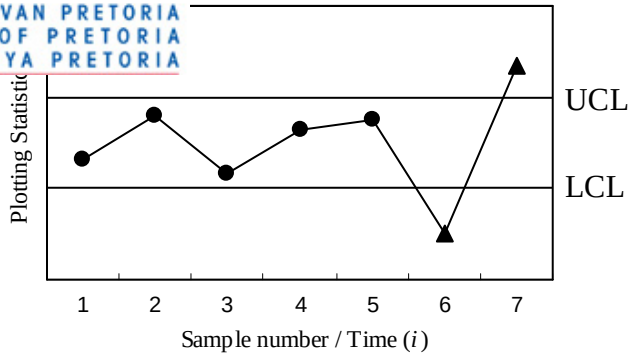


Figure 4.7: The *2-of-3* rule for the two-sided chart



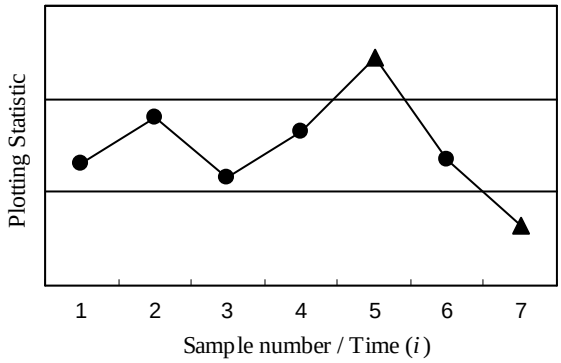
$E_5 : \{\xi_{i-2} = 0, \xi_{i-1} = 1, \xi_i = 2\}$

(a) (possible swing detected)



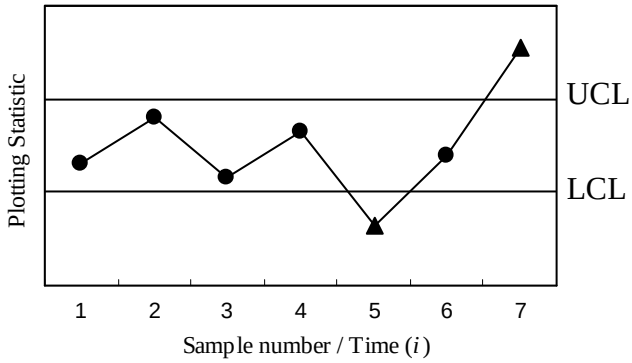
$E_6 : \{\xi_{i-2} = 0, \xi_{i-1} = 2, \xi_i = 1\}$

(b) (possible swing detected)



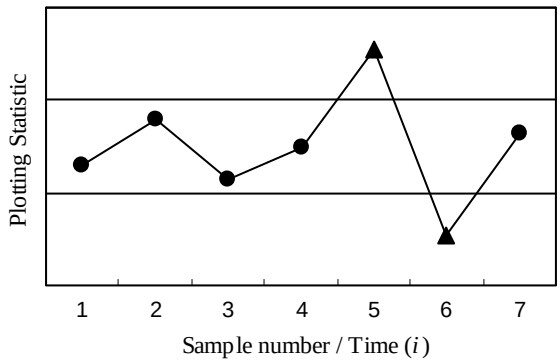
$E_7 : \{\xi_{i-2} = 1, \xi_{i-1} = 0, \xi_i = 2\}$

(c) (downward trend detected)



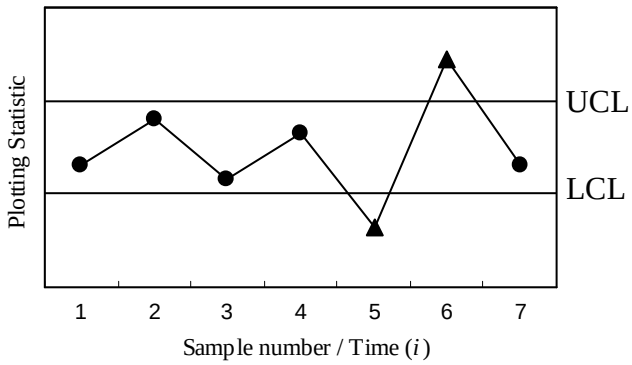
$E_8 : \{\xi_{i-2} = 2, \xi_{i-1} = 0, \xi_i = 1\}$

(d) (upward trend detected)



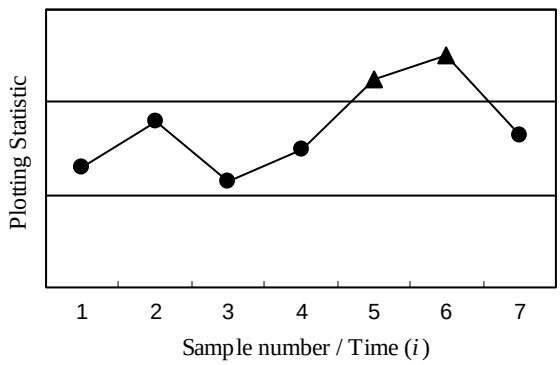
$E_9 : \{\xi_{i-2} = 1, \xi_{i-1} = 2, \xi_i = 0\}$

(e) (last point plots between *LCL* and *UCL*)



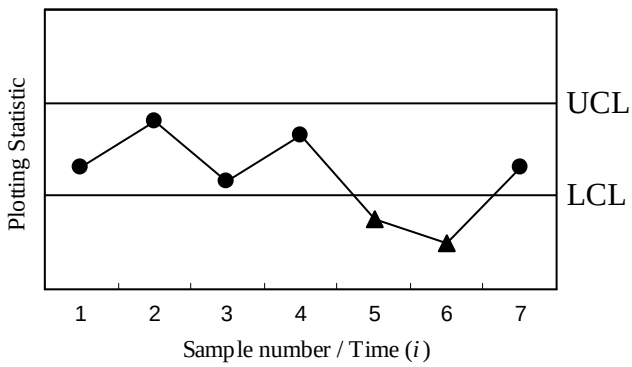
$E_{10} : \{\xi_{i-2} = 2, \xi_{i-1} = 1, \xi_i = 0\}$

(f) (last point plots between *LCL* and *UCL*)



$E_{11} : \{\xi_{i-2} = 1, \xi_{i-1} = 1, \xi_i = 0\}$

(g) (last point plots between *LCL* and *UCL*)



$E_{12} : \{\xi_{i-2} = 2, \xi_{i-1} = 2, \xi_i = 0\}$

(h) (last point plots between *LCL* and *UCL*)

Figure 4.8: The *2-of-3* events excluded as signaling events for the two-sided charts

Section 4.1 outlined, in general, the operation of the runs-rule enhanced charts. However, we are yet to define (choose) the charting statistic Q_i and the control limits (UCL and LCL). In Sections 4.2 and 4.3 we do just this and show, in particular, how to obtain the run-length distributions and how to design and implement the runs-rule enhanced charts in case the π^{th} percentile of the process distribution is known (Case K) and unknown (Case U). The performance of the charts is then further examined via properties of their run-length distributions such as the average run-length (ARL), the false alarm rate (FAR), the standard deviation of the run-length ($SDRL$) and some of the percentiles of the run-length distributions.

4.2 Sign charts for the known π^{th} quantile (Case K)

Introduction

Case K refers to the situation when the π^{th} quantile (or percentile) of the process distribution is known or specified. The new control charts we design for Case K are based on the well-known sign (test) statistic.

Assumptions

Let $(X_{i1}, X_{i2}, ..., X_{in})$ denote a random sample of size $n > 1$ taken at sampling stage (time) $i = 1,2,3,....$ Assume that the samples are independent and the observations come from a continuous distribution with cumulative distribution function (c.d.f) $F_X(x)$ with the unique $100\pi^{th}$ percentile denoted by $\theta = F_X^{-1}(\pi)$, $0 < \pi < 1$.

In many cases the percentile of interest is the median because it is a robust measure of central tendency so that $\pi = 0.5$ and $\theta = F_X^{-1}(0.5)$, however this is not necessary for our developments as the new sign charts can be applied for any percentile of interest.

Charting statistics

Amin et al. (1995) considered a *1-of-1* Shewhart-type sign chart for monitoring the median of a distribution based on the charting statistic

$$SN_i = \sum_{j=1}^{n} sign(X_{ij} - \theta_0) \quad \text{for} \quad i = 1,2,... \tag{4-4}$$

where $sign(x) = 1$ if $x > 0$, 0 if $x = 0$ and -1 if $x < 0$ and θ_0 denotes the specified value of the median.

We consider any percentile $\theta = F_X^{-1}(\pi)$ for $0 < \pi < 1$ and the charting statistic for our sign charts is the classical sign statistic

$$T_i = \sum_{j=1}^{n} I(X_{ij} > \theta_0) \quad \text{for} \quad i = 1,2,... \tag{4-5}$$

where θ_0 denotes the known (or the specified or the target) value of the percentile of interest and $I(X_{ij} > \theta_0)$ denotes the indicator function for the event $\{X_{ij} > \theta_0\}$. Thus T_i denotes the number of observations larger than θ_0 in the i^{th} sample and it is easily seen that T_i follows a binomial distribution with parameters n and probability of success $p = \Pr(X_{ij} > \theta_0)$. When $\theta = \theta_0$ the percentile of interest is equal to its specified value, the process is said to be in-control (IC) and then p is denoted by p_0 and equals

$$p_0 = \Pr(X_{ij} > \theta_0 \mid IC) = 1 - \pi . \qquad (4\text{-}6)$$

Thus, for example, when the percentile of interest is the median ($\pi = 0.5$), the process is IC when $\theta = \theta_0$ (the specified value of the median) and then $p = p_0 = 0.5$; similarly when θ is the first quartile ($\pi = 0.25$), the process is IC when $\theta = \theta_0$ (the specified value of the first quartile) and $p = p_0 = 0.75$ and so on.

Control limits

The upper and lower control limits of our sign charts are

$$UCL = n - b \qquad \text{and} \qquad LCL = a \qquad (4\text{-}7)$$

where the charting constants a and b are integers between (and including) zero and n, that is, $a, b \in \{0,1,2,...,n\}$, and selected so that the UCL is greater than the LCL; determination of a and b will be discussed later. Note that, the new sign charts do not have a centerline.

4.2.1 Run-length distributions of the sign charts

The run-length distribution and its associated characteristics (such as the mean, the standard deviation, the median etc.) reveal important information regarding the performance of a control chart (see e.g. Human and Graham, (2007)).

There are various approaches to finding the run-length distribution. We use (for the most part) a Markov chain approach (see e.g. Fu and Lou, (2003)) to derive the necessary results for our runs-rule enhanced charts because this approach provides a more compact and unified view of the derivations, and as stated by Balakrishnan and Koutras (2002), p.14, "The Markov techniques possess a great advantage (over the classical combinatory methods) as they are easily adjustable to many run-related problems; they often simplify the solutions to specific problems they are applied on and remain valid even for cases involving non-identical or dependent trials".

The Markov chain approach entails that we:

(a) classify each charting statistic T_i (based on its value) into one of two categories (for a one-sided chart) or into one of three categories (for a two-sided chart) depending on whether T_i plots on or above the UCL, on or below the LCL and/or between the LCL and UCL,

(b) define a new sequence of random variables $Y_1, Y_2, Y_3, \ldots$ (say) that keeps track of the classification of the $T_i's$, and then

(c) construct a Markov Chain $\{Z_i : i \geq 0\}$ to find the run-length distribution.

For example, consider the upper one-sided sign chart. Each T_i can be either on or above the UCL or below. Let $Y_i = 1$ (a success) in the former case and $Y_i = 0$ (a failure) in the latter case. Thus, corresponding to a sequence of $T_i's$ we get a sequence of $Y_i's$ that are all binary; for example, if $(T_1, T_2, T_3, T_4) = (4, 3, 8, 7)$ and $UCL = 5$, we get $(Y_1, Y_2, Y_3, Y_4) = (0, 0, 1, 1)$.

Thus, the run-length of the *1-of-1* upper one-sided chart i.e. the time when for the first time a T_i plots on or above the UCL, is "3" for our example and can be equivalently expressed as the time when for the first time we obtain a "1" (a success) among the four $Y_i's$.

Similarly, the run-length of the *2-of-2* upper one-sided chart i.e. the time when for the first time two consecutive $T_i's$ plot out-of-control, which equals "4" in our example, can be equivalently viewed as the time when for the first time we obtain two successive "1's" (or successes) among the four $Y_i's$.

Hence the run-length for the $T_i's$ can be equivalently defined as the waiting time for the first success (or, more generally, the first particular run or pattern of successes) among the $Y_i's$; it is this correspondence that makes the study of the statistical properties of the run-length random variable more amenable using results about the waiting time distributions in a sequence of Bernoulli (binary or two-state) and other types (three or more states) of random variables.

There is a rather vast literature on waiting time distributions. A detailed discussion about general results on the distribution theory of runs and patterns with various applications can be found in Balakrishnan and Koutras (2002) and Fu and Lou (2003). Some of these results pertain to the exact probability distribution of the waiting time for the first occurrence of a simple or a compound pattern in a sequence of i.i.d. (or homogeneous Markov dependent) 2-state (binary) or 3 or more-state trials (see e.g. Fu and Lou, (2003); Chapters 3, 4 and 5). The approach is to "properly imbed" (see e.g. Fu and Lou (2003), page 64; Definition 2.6) the random variable of interest (the run-length in our case) into a finite Markov chain which means constructing a "proper" Markov chain so that the probability that the run-length random variable N takes on some specific value is expressed in terms of the probability that the imbedded Markov chain $\{Z_i : i \geq 0\}$ resides in a specific subset S of the state space Ω.

The latter probability can be more easily computed using results about the transition probability matrix of the Markov chain. For example, given the $m \times m$ transition probability matrix of the Markov chain

$$\mathbf{M}_{m\times m} = \left[\begin{array}{c|c} \mathbf{Q}_{h\times h} & \mathbf{C}_{h\times(m-h)} \\ \hline \mathbf{0}_{(m-h)\times h} & \mathbf{I}_{(m-h)\times(m-h)} \end{array}\right]$$

(written in a partitioned form), the probability mass function (p.m.f), the expected value (ARL) and the variance ($VARL$) of the run-length random variable N can be directly obtained, using Theorems 5.2 and 7.4 of Fu and Lou (2003), as

$$P(N = j \mid n, a, b, \theta) = \boldsymbol{\xi}\mathbf{Q}^{j-1}(\mathbf{I} - \mathbf{Q})\mathbf{1} \qquad \text{for} \qquad j = 1,2,3,... \tag{4-8}$$

$$E(N \mid n, a, b, \theta) = \boldsymbol{\xi}(\mathbf{I} - \mathbf{Q})^{-1}\mathbf{1} \tag{4-9}$$

and

$$\text{var}(N \mid n, a, b, \theta) = \boldsymbol{\xi}(\mathbf{I} + \mathbf{Q})(\mathbf{I} - \mathbf{Q})^{-2}\mathbf{1} - (E(N))^2 \tag{4-10}$$

where the sub-matrix matrix $\mathbf{Q} = \mathbf{Q}_{h\times h}$ is called the essential transition probability sub-matrix, $\mathbf{I} = \mathbf{I}_{h\times h}$ (used in (4-8) and (4-10)) and $\mathbf{I}_{(m-h)\times(m-h)}$ (used in the definition of $\mathbf{M}_{m\times m}$) are identity matrices, $\xi_{1\times h} = (1, 0, 0, ..., 0)$ is the initial distribution, $\mathbf{1}_{h\times 1} = (1, 1, ..., 1)^{\mathrm{T}}$ is the unit vector, m denotes the number of states in the state space Ω and $m-h$ is the number of unique simple patterns that defines a signal; the (non-essential) matrix $\mathbf{C}_{h\times(m-h)}$ will be illustrated later.

The point is that we only need to construct the state space Ω and the essential transition probability sub-matrix $\mathbf{Q}_{h\times h}$ of the Markov chain in order to be capable to calculate the entire run-length distribution.

Signaling probabilities

Whilst the key to construct the state space Ω depends on the particular signaling rule and whether a one-sided or two-sided chart is looked at, the building blocks of the transition probability matrix are the one-step transition probabilities (i.e. the elements of the transition probability matrix).

The one-step transition probabilities are denoted by

$$p_{k,j} = \Pr(Z_i = j \mid Z_{i-1} = k)$$

and interpreted as the conditional probability given that at any specific time $i-1$ the system was in state k, the system will be in state j at time i for $i \geq 1$ and $j, k \in \Omega$. The transition probabilities $p_{k,j}$ are all functions of and depend on the signaling probabilities i.e. the probability for a single charting statistic to plot outside the control limit(s), and therefore play a key role in the derivation of the run-length distributions of the runs-rules enhanced charts. In case of the upper and lower one-sided charts the signaling probabilities are

$$p^{+}(n, b, \theta) = \Pr(T_i \geq UCL) = \Pr(T_i \geq n-b) = \Pr(\xi_i^{+} = 1) = I_p(n-b, b+1) \qquad (4\text{-}11)$$

and

$$p^{-}(n, a, \theta) = \Pr(T_i \leq LCL) = \Pr(T_i \leq a) = \Pr(\xi_i^{-} = 1) = 1 - I_p(a+1, n-a), \qquad (4\text{-}12)$$

respectively; for the two-sided chart the probability for any of the charting statistics to plot outside either the UCL or the LCL is

$$p^{\pm}(n, a, b, \theta) = 1 - \Pr(LCL < T_i < UCL) = 1 - I_p(a+1, n-a) + I_p(n-b, b+1) \qquad (4\text{-}13)$$

where $I_p(u,v) = [\beta(u,v)]^{-1}\int_0^p w^{u-1}(1-w)^{v-1}dw$ with $0 < p = \Pr(X_{ij} > \theta_0) < 1$, is the c.d.f of the $Beta(u,v)$ distribution, also known as the incomplete beta function, which helps us write various expressions in a more compact form.

Note that, for notational simplicity and brevity we denote the probabilities $p^+(n,b,\theta)$, $p^-(n,a,\theta)$, and $p^\pm(n,a,b,\theta)$ simply by p^+, p^- and $p^\pm$, respectively.

Remark 4

(i) If $\theta = \theta_0$ the signaling probabilities (and hence the distribution of T_i and the in-control run-length distributions and their associated characteristics) depend only on

a. the sample size n,
b. the charting constants a and/or b, and
c. the percentile of interest $\theta_0 = F_X^{-1}(\pi)$ where π is specified.

Any decision rule (signaling rule) based on the T_i's will therefore be distribution-free as long as the underlying distributions (at each point in time) are continuous and identical. It follows that the in-control run-length distributions of the runs-rules enhanced sign charts are distribution-free and therefore charts based on the T_i's will be distribution-free.

(ii) To obtain the in-control run-length distribution and its mean and variance one substitutes $\theta = \theta_0$ in expressions (4-8), (4-9) and (4-10); by substituting $\theta \neq \theta_0$ one obtains the corresponding results for the out-of-control situation which depends on the underlying process distribution.

4.2.2 Transition probability matrices of the sign charts

To illustrate the derivation of the transition probability matrices and the run-length distributions, we begin with the one-sided upper (lower) control charts and then proceed to study the two-sided charts.

In case of the one-sided and the two-sided charts, we first look at the run-length distribution of the *1-of-1* chart (which uses the least complicated signaling rule) before we study the run-length distributions of the run rules enhanced charts, that is, the *k-of-k* ($k \geq 2$) and the *k-of-w* ($1 \leq k \leq w$ and $w \geq 2$) sign charts.

For each chart the key is to imbed the run-length into a proper homogenous Markov chain and obtain the essential transition probability sub-matrix $\mathbf{Q}_{h\times h}$ associated with the particular Markov chain.

Note that, we discuss the derivation of the transition probability matrices of the one-sided and the two-sided sign charts in detail so that later, in Case U, we can merely make use of these results.

4.2.2.1 One-sided sign charts

For the upper one-sided sign chart we view the series of signaling indicators $\xi_1^+, \xi_2^+, \xi_3^+, \ldots$ associated with the charting statistics $T_1, T_2, T_3, \ldots$ and the *UCL* as a series of independent binary random variables, each being either "a success or 1" (T_i plotting on or above the *UCL*) or "a failure or 0" (T_i plotting below the *UCL*) with probabilities p^+ and $1-p^+$, respectively (see e.g. Figure 4.9).

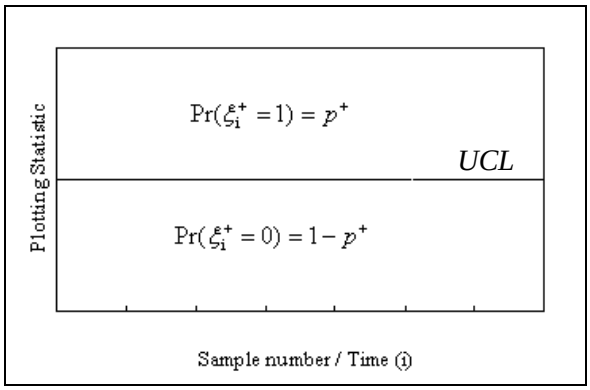


Figure 4.9: The two regions on the upper one-sided control chart ('0' and '1') and their associated probabilities used to classify the charting statistic

Upper one-sided *1-of-1* sign chart

The run-length N_{1of1}^{+} of the *1-of-1* upper one-sided chart is the waiting time until event A_1 (see panel (a) of Figure 4.1) occurs, which can be viewed as the waiting time for the first occurrence of a "1" (success) in the series of ξ_i^{+}'s (i.e. of 0's and 1's).

The imbedded Markov chain associated with N_{1of1}^{+} is a homogeneous Markov chain defined on the finite state space $\Omega = \{\phi, 0, 1\}$ with $m = 3$ states, where

(a) the state $\{1\}$ is called the "absorbing" state (when the process is declared out-of-control),

(b) the state $\{0\}$ is called the "transient" state (i.e. the process can remain in state $\{0\}$, which means that the process is IC and the charting procedure continues, or the process can move from state $\{0\}$ into state $\{1\}$, which implies that the process goes OOC and the charting procedure stops), and

(c) the state ϕ is the "dummy" state introduced for convenience. The dummy state ϕ is in fact added to the state space Ω so that with probability one the process is assumed to begin in-control with the intention that the corresponding initial probability distribution is taken as $\boldsymbol{\xi}_{1\times 2} = (1, 0)$.

The 3×3 transition probability matrix of $\{Z_i : i \geq 0\}$ associated with N_{1of1}^{+} is given by

$$\mathbf{M}_{3\times 3} = \left[\begin{array}{c|c} \mathbf{Q}_{2\times 2} & \mathbf{C}_{2\times 1} \\ \hline \mathbf{0}_{1\times 2} & \mathbf{I}_{1\times 1} \end{array}\right] = \left[\begin{array}{cc|c} p_{\phi,\phi} & p_{\phi,0} & p_{\phi,1} \\ p_{0,\phi} & p_{0,0} & p_{0,1} \\ \hline p_{1,\phi} & p_{1,0} & p_{1,1} \end{array}\right] = \left[\begin{array}{cc|c} 0 & 1-p^{+} & p^{+} \\ 0 & 1-p^{+} & p^{+} \\ \hline 0 & 0 & 1 \end{array}\right] \qquad (4\text{-}14)$$

where, for example, $p_{0,1}$ (the entry in the 2nd row and the 3rd column of $\mathbf{M}_{3\times 3}$) is the probability that the system goes from state $\{0\}$ (that is where T_{i-1} plots IC) at time $i-1$, to state $\{1\}$ (that is where T_i plot OOC) at time i; this probability is simply the probability that T_i plots at or above the *UCL* at time i, which is $\Pr(\xi_i^{+} = 1) = p^{+} = p^{+}(n, b, \theta)$. The rest of the elements of $\mathbf{M}_{3\times 3}$ in (4-14) can be calculated in a similar way.

Upper one-sided *2-of-2* sign chart

The run-length N^{+}_{2of2} of the *2-of-2* upper one-sided chart is the waiting time until event B_1 (see panel (a) of Figure 4.3) occurs, which can be viewed as the waiting time for the first occurrence of two consecutive 1's (i.e. successes) in the series of ξ_i^{+}'s (i.e. 0's and 1's).

The imbedded Markov chain associated with N^{+}_{2of2} is a homogeneous Markov chain defined on the finite state space $\Omega = \{\phi, 0, 1, 11\}$ with $m = 4$ states, where

(a) the last state $\{11\}$ is the absorbing state,

(b) the two states $\{0\}$ and $\{1\}$ are the transient (non-absorbent) states,

(c) and ϕ is the dummy state, which is again added to Ω so that (with probability one) the process is assumed to begin in-control and with the intention that the corresponding initial probability distribution is taken as $\boldsymbol{\xi}_{1\times3} = (1, 0, 0)$.

The transition probability matrix of $\{Z_i : i \geq 0\}$ associated with N^{+}_{2of2} is given by

$$\mathbf{M}_{4\times4} = \left[\begin{array}{c|c} \mathbf{Q}_{3\times3} & \mathbf{C}_{3\times1} \\ \hline \mathbf{0}_{1\times3} & \mathbf{I}_{1\times1} \end{array}\right] = \left[\begin{array}{ccc|c} p_{\phi,\phi} & p_{\phi,0} & p_{\phi,1} & p_{\phi,11} \\ p_{0,\phi} & p_{0,0} & p_{0,1} & p_{0,11} \\ p_{1,\phi} & p_{1,0} & p_{1,1} & p_{1,11} \\ \hline p_{11,\phi} & p_{11,0} & p_{11,1} & p_{11,11} \end{array}\right] = \left[\begin{array}{ccc|c} 0 & 1-p^{+} & p^{+} & 0 \\ 0 & 1-p^{+} & p^{+} & 0 \\ 0 & 1-p^{+} & 0 & p^{+} \\ \hline 0 & 0 & 0 & 1 \end{array}\right] \qquad (4\text{-}15)$$

where, for example, the probability that the system goes from state $\{1\}$ (where T_{i-1} plots OOC) at time $i-1$, to state $\{11\}$ (where both T_{i-1} and T_i plot OOC) at time i, denoted by $p_{1,11}$, is the entry in the 3rd row and the 4th column of $\mathbf{M}_{4\times4}$; this is simply the probability that T_i plots at or above the *UCL* at time *i*, which is, as earlier, $\Pr(\xi_i^{+} = 1) = p^{+} = p^{+}(n, b, \theta)$.

Remark 5

Because

(i) the signaling indicators ξ_i^+'s are a sequence of i.i.d. Bernoulli random variables each with probability of success p^+, and

(ii) the run-length random variable N_{1of1}^+ (N_{2of2}^+) is the waiting time for the first success (the first occurrence of two consecutive successes),

one can equivalently obtain the distribution (i.e. the p.m.f, the mean, the variance etc.) of N_{1of1}^+ (N_{2of2}^+) from the distribution of the variable T_k where $k=1$ (or 2).

The stopping time variable T_k ($k \geq 1$) is, in general, the waiting time to observe a sequence of k consecutive successes for the first time in a sequence of i.i.d. Bernoulli random variables with success probability α and should not be confused with the plotting statistic T_i defined in (4-5).

The distribution of T_k is known to be the geometric distribution of order k (see e.g. Chapter 2 of Balakrishnan and Koutras, (2002)) with p.m.f, expected value and variance given by

$$\Pr(T_k = j) = \begin{cases} 0 & \text{if} \quad 0 \leq j < k \\ \alpha^k & \text{if} \quad j = k \\ \sum_{i=1}^{[\frac{j+1}{k+1}]} (-1)^{i-1} \frac{\alpha^{ik}}{(1-\alpha)^{1-i}} \left\{ \binom{j-ik-1}{i-2} + (1-\alpha)\binom{j-ik-1}{i-1} \right\} & \text{if} \quad j \geq k+1 \end{cases} \tag{4-16}$$

$$E(T_k) = \frac{1-\alpha^k}{(1-\alpha)\alpha^k} \qquad \text{and} \qquad \text{var}(T_k) = \frac{1-(2k+1)(1-\alpha)\alpha^k - \alpha^{2k+1}}{(1-\alpha)^2 \alpha^{2k}} \tag{4-17}$$

respectively.

The equivalence between the distribution of T_k (i.e. the geometric distribution of order k) and the distribution of N^+_{kofk} (i.e. the waiting until for the first time k consecutive T_i's plot on or above the UCL) can be verified by substituting the essential transition probability sub-matrix $\mathbf{Q}_{h\times h}$ of $\{Z_i : i \geq 0\}$ associated with N^+_{kofk} in expressions (4-8), (4-9) and (4-10), and then simplifying symbolically (using, for example, computer software with matrix manipulations capabilities such as Scientific Workplace®); upon doing so one obtains explicit and closed form expressions, via the Markov chain approach, for the p.m.f, the ARL and the $VARL$ of the run-length random variable N^+_{kofk}.

For the upper one-sided *1-of-1* sign chart, for example, we substitute

$$\mathbf{Q}_{2\times 2} = \begin{bmatrix} 0 & 1-p^+ \\ 0 & 1-p^+ \end{bmatrix}$$

in expressions (4-8), (4-9) and (4-10) so that upon simplifying we obtain

$$\Pr(N^+_{1of1} = j \mid n,b,\theta) = (1-p^+)^{j-1}\,p^+ \quad \text{for} \quad j = 1,2,3,... \tag{4-18}$$

$$E(N^+_{1of1} \mid n,b,\theta) = 1/p^+ \quad \text{and} \quad \text{var}(N^+_{1of1} \mid n,b,\theta) = (1-p^+)/(p^+)^2. \tag{4-19}$$

Expressions (4-18) and (4-19) are identical to expressions (4-16) and (4-17) with $k = 1$ and $\alpha = p^+$, respectively.

Upper one-sided *2-of-3* sign chart

Like the upper one-sided *1-of-1* and *2-of-2* charts we can find the run-length distribution of the *2-of-3* upper one-sided chart using a Markov chain.

The run-length N^{+}_{2of3} of the *2-of-3* upper one-sided chart is the waiting time until event C_1 or C_2 (see panel (a) of Figure 4.4) occurs, which can be viewed as the waiting time for the first occurrence of the pattern $\Lambda = \{011 \text{ or } 101\}$ in the series of ξ_i^{+}'s (i.e. of 0's and 1's). The pattern Λ is called a "compound pattern" and written as: $\Lambda = \Lambda_1 \cup \Lambda_2$ where $\Lambda_1 = 011$ and $\Lambda_2 = 101$ are two so-called distinct "simple patterns".

The imbedded Markov chain associated with N^{+}_{2of3} is a homogeneous Markov chain defined on the state space $\Omega = \{\phi, 0, 1, 01, 10, \alpha_1, \alpha_2\}$ with $m = 7$ states, where

(a) the two states $\alpha_1 = \{011\}$ and $\alpha_2 = \{101\}$ are the absorbing states (when the process is declared OOC),

(b) the four states $\{0, 1, 01, 10\}$ are the transient states (i.e. the process can move from one of these states to another, which means that the charting procedure continues), and

(c) ϕ is the dummy state introduced for convenience.

The transient states are the sequential sub-patterns of $\Lambda_1 = 011$ and $\Lambda_2 = 101$, respectively. For example, the state $\{0\}$ is the sub-pattern of the state $\{01\}$, whereas the two states $\{0\}$ and $\{01\}$ are the sub-patterns of $\Lambda_1 = 011$, and the states $\{1\}$ and $\{10\}$ are the sub-patterns of $\Lambda_2 = 101$.

As earlier, the dummy state ϕ is again added to Ω so that (with probability one) the process is assumed to begin in-control, and the corresponding initial probability distribution is taken as $\xi_{1\times5} = (1, 0, 0, 0, 0)$.

The transition probability matrix of $\{\xi_i^+, i \geq 0\}$ associated with N_{2of3}^+ is given by

$$\mathbf{M}_{7\times7} = \left[\begin{array}{c:c} \mathbf{Q}_{5\times5} & \mathbf{C}_{5\times2} \\ \hdashline \mathbf{0}_{2\times5} & \mathbf{I}_{2\times2} \end{array}\right] = \left[\begin{array}{ccccc:cc} 0 & 1-p^+ & p^+ & 0 & 0 & 0 & 0 \\ 0 & 1-p^+ & 0 & p^+ & 0 & 0 & 0 \\ 0 & 0 & p^+ & 0 & 1-p^+ & 0 & 0 \\ 0 & 0 & 0 & 0 & 1-p^+ & p^+ & 0 \\ 0 & 1-p^+ & 0 & 0 & 0 & 0 & p^+ \\ \hdashline 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right] \quad \text{(4-20)}$$

where, for example, the probability that the system goes from state {01} (that is where T_{i-2} plots IC and T_{i-1} plots OOC) at time $i-1$, to state $\alpha_1 = \{011\}$ (that is where T_{i-2} plots IC and both T_{i-1} and T_i plot OOC) at time i is the entry in the 4th row and the 6th column of $\mathbf{M}_{7\times7}$.

Remark 6

A few comments concerning the application and the implementation of the upper one-sided *2-of-3* sign chart are in order:

(i) To declare a process out-of-control (OOC) we need at least three charting statistics (T_{i-2}, T_{i-1} and T_i, say) of which exactly one should plot in-control (IC) i.e. either T_{i-2} plots below the UCL with T_{i-1} and T_i plotting on or above the UCL or, T_{i-1} plots below the UCL with T_{i-2} and T_i plotting on or above the UCL (see e.g. events C_1 and C_2 in panel (a) of Figure 4.4). Thus, we can only declare the process OOC beginning from time $i \geq 3$ and, we need at least one charting statistic to plot below the upper control limit before we can declare the process OOC.

(ii) Because of these two build-in conditions of the upper one-sided *2-of-3* sign chart, the chart has a hitch at start-up: If $T_i \geq UCL$ for $i = 1,2,3,...,r$, that is, if all the charting statistics plot on or above the upper control limit from the time that the chart is implemented until time r, the chart would not immediately signal that the process is OOC even though the pattern of the points on the chart suggests otherwise. The chart would most likely give a "delayed" or a "late" OOC signal instead.

While this glitch is possible, we need to stress an important assumption:

> The design and the implementation of all the charts that are proposed in this chapter are based on an IC process at start-up as well as the trade-off between minimizing the probability that a charting statistic plots on or outside the control limit(s) when the process is actually IC and quickly detecting an OOC process.

This assumption means two things:

a. The process is IC at start-up; hence, we must ensure (to the extent that it is possible) that the process is IC *before* we start monitoring it.

b. We typically choose the UCL such that the probability that a T_i plots on or above the UCL when the process is IC i.e. $p_0^+ = \Pr(T_i \geq UCL \mid IC)$, is small, which automatically implies that the probability that a T_i plots below the UCL when the process is IC i.e. $1 - p_0^+ = \Pr(T_i < UCL \mid IC)$, is large.

The latter implies that the probability that all the charting statistics up to and including the r^{th} one plot on or above the *UCL* when the process is IC i.e. $(p_0^+)^r$, would decrease rapidly as r increases. But, most importantly, it also implies that as we continue to monitor the process, the probability that the r^{th} charting statistic plots below the *UCL* when the process is IC stays constant and equal to $1-p_0^+$; this is so because we assume that successive samples (or charting statistics) are independent.

Hence, what is of importance to the practitioner is to know what the risk is that this hitch occurs. This risk can be measured by calculating and studying the *odds* that a T_i plots below the *UCL* when the process is IC i.e. $(1-p_0^+)/(p_0^+)$, at any time $i=1,2,3,\ldots$.

To investigate the effect of p_0^+ on the above *odds*, Table 4.1 shows values of $(1-p_0^+)/(p_0^+)$ for values of $p_0^+ = 0.001(0.001)0.005$ and $0.01(0.01)0.20$. The values of p_0^+ that we use to construct Table 4.1 are representative of the typical values that one would consider when designing the proposed upper one-sided *2-of-3* chart (see e.g. Tables 4.6. and 4.7).

From Table 4.1 we observe that:

(i) The ratio $(1-p_0^+)/(p_0^+)$ is larger than or equal to 4.0 for all values of p_0^+ that we consider. This implies that, for a process that is IC at start-up (which is a fundamental assumption of our earlier theoretical developments and the reason for adding the dummy state, ϕ, to the state spaces off all the proposed charts) it is *at least* four times more likely for any new incoming T_i to plot below the *UCL* than for any new incoming T_i to plot on or above the *UCL*.

(ii) For $p_0^+ = 0.01$, which is a very reasonable choice considering all the values of p_0^+ in Tables 4.6 and 4.7, the ratio $(1-p_0^+)/(p_0^+)$ is equal to 99.0; this is relatively large.

(iii) The largest value for $(1-p_0^+)/(p_0^+)$ is 999.0 (when $p_0^+ = 0.001$) and will increase even further as p_0^+ decreases; this is good because smaller values of p_0^+ are typically preferred and also recommended in practice.

The above-mentioned observations are all relevant for the practitioner because, for a process that is IC at start-up, (which is a key assumption when implementing any of the charts that are proposed in this chapter) they show that the risk associated with the proposed *2-of-3* sign chart at start-up is: (a) almost negligible, and (b) decreases rapidly as we continue to monitor the process because the probability that all the charting statistics up to and including the r^{th} one plot on or above the *UCL* when the process is IC i.e. $(p_0^+)^r$, would decrease towards zero quickly as r increases. This should be reassuring for the practitioner.

Table 4.1: The ratio $(1-p_0^+)/(p_0^+)$ as a function of p_0^+

p_0^+	$(1-p_0^+)/(p_0^+)$	p_0^+	$(1-p_0^+)/(p_0^+)$
0.001	999.0	**0.09**	10.1
0.002	499.0	**0.10**	9.0
0.003	332.3	**0.11**	8.1
0.004	249.0	**0.12**	7.3
0.005	199.0	**0.13**	6.7
0.01	99.0	**0.14**	6.1
0.02	49.0	**0.15**	5.7
0.03	32.3	**0.16**	5.3
0.04	24.0	**0.17**	4.9
0.05	19.0	**0.18**	4.6
0.06	15.7	**0.19**	4.3
0.07	13.3	**0.20**	4.0
0.08	11.5		

To overcome the imperfection of the upper one-sided *2-of-3* sign chart at start-up, we could use the event $C_7 = \{T_1 \geq UCL, T_2 \geq UCL\} \Leftrightarrow \{\xi_1^+ = 1, \xi_2^+ = 1\}$, in addition to the events C_1 and C_2 shown in panel (a) of Figure 4.4, as a third signaling event. The event C_7 is special in two ways: (a) it prevents the hitch at start-up by enabling the chart to signal at time $i = 2$, and (b) it occurs if and only if the first two charting statistics, T_1 and T_2, both plot on or above the upper control limit; hence, event C_7 can not occur from time $i \geq 3$.

The resultant chart is an augmented upper one-sided *2-of-3* sign chart. Adding the extra event leads to an augmented state space i.e. $\Omega = \{\phi, 0, 1, 01, 10, 11, 011, 101\}$, where the three states $\{11\}$, $\{011\}$ and $\{101\}$ are the absorbent states and implies that the transition probability matrix in (4-20) be altered slightly to become

$$\mathbf{M}_{8\times 8} = \left[\begin{array}{c|c} \mathbf{Q}_{5\times 5} & \mathbf{C}_{5\times 3} \\ \hline \mathbf{0}_{3\times 5} & \mathbf{I}_{3\times 3} \end{array}\right] = \left[\begin{array}{ccccc|ccc} 0 & 1-p^+ & p^+ & 0 & 0 & 0 & 0 & 0 \\ 0 & 1-p^+ & 0 & p^+ & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1-p^+ & p^+ & 0 & 0 \\ 0 & 0 & 0 & 0 & 1-p^+ & 0 & p^+ & 0 \\ 0 & 1-p^+ & 0 & 0 & 0 & 0 & 0 & p^+ \\ \hline 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right].$$

To investigate the impact (i.e. gain or loss) of augmenting the transition probability matrix on the in-control performance of the chart, we calculated the in-control average run-lengths and the false alarm rates of the proposed upper one-sided *2-of-3* sign chart and that of the augmented upper one-sided *2-of-3* sign chart (when it is of interest to monitor the median of the process) for different combinations of the sample size, n, and the upper control limit, UCL.

The values of the in-control average run-lengths (denoted by ARL^{+}_{2of3} and ARL^{A+}_{2of3}, respectively) were calculated according to expression (4-9) using the transition probability matrix in (4-20) and the augmented transition probability matrix given above, respectively.

The false alarm rate of the proposed upper one-sided *2-of-3* sign chart (denoted FAR^{+}_{2of3}) and that of the augmented upper one-sided *2-of-3* sign chart (denoted FAR^{A+}_{2of3}) can be easily obtained from the definitions of the signaling events that are used by each chart and are given by

$$FAR^{+}_{2of3} = \begin{cases} 0 & \text{if} \quad i = 1 \text{ or } 2 \\ 2(1-p_0^+)(p_0^+)^2 & \text{if} \quad i \geq 3 \end{cases} \quad \text{and} \quad FAR^{A+}_{2of3} = \begin{cases} 0 & \text{if} \quad i = 1 \\ (p_0^+)^2 & \text{if} \quad i = 2 \\ 2(1-p_0^+)(p_0^+)^2 & \text{if} \quad i \geq 3 \end{cases}$$

respectively.

There is only a slight modification of the expression for FAR^{+}_{2of3} to obtain FAR^{A+}_{2of3}; this leads to the following similarities and/or differences in the false alarm rates of the charts:

(i) At time $i = 1$: $FAR^{+}_{2of3} = FAR^{A+}_{2of3} = 0$,

(ii) At time $i = 2$: $FAR^{+}_{2of3} = 0$ but $FAR^{A+}_{2of3} = (p_0^+)^2$, and

(iii) At time $i \geq 3$: $FAR^{+}_{2of3} = FAR^{A+}_{2of3} = 2(1-p_0^+)(p_0^+)^2$.

These similarities and/or differences are a direct consequence of the signaling events used by each chart i.e.

(i) Neither one of the charts can signal at time $i=1$ because, the proposed upper one-sided *2-of-3* sign chart needs at least three charting statistics to signal whereas the augmented upper one-sided *2-of-3* sign chart needs at least two charting statistics to signal.

(ii) It is only the augmented upper one-sided *2-of-3* sign chart that can give a false alarm at time $i=2$ and, it can do so if and only if event C_7 occurs.

(iii) From time $i \geq 3$, both the charts can signal if and only if event C_1 or event C_2 occurs. The event C_7, as mentioned earlier, can only occur at time $i=2$ and therefore does not influence the false alarm rate of the augmented upper one-sided *2-of-3* sign chart at or beyond time $i=3$.

Based on our calculations, we found that:

(i) The in-control average run-lengths of the two charts were almost identical; ARL^{+}_{2of3} is only slightly larger than ARL^{A+}_{2of3}.

(ii) Depending on the combination of n and UCL, the $FAR^{A+}_{2of3}=(p_0^{+})^2$ at time $i=2$ can be reasonably large, which might be a concern for the practitioner.

To further compare the impact of augmenting the proposed upper one-sided *2-of-3* sign chart, Table 4.2 shows values of the in-control probability mass functions (p.m.f's) and the in-control cumulative distribution functions (c.d.f's) of the run-length random variables, N^{+}_{2of3} and N^{A+}_{2of3}, associated with the proposed and the augmented charts; these values are denoted by $\Pr(N^{+}_{2of3}=i \mid IC)$, $\Pr(N^{A+}_{2of3}=i \mid IC)$, $\Pr(N^{+}_{2of3} \leq i \mid IC)$ and $\Pr(N^{A+}_{2of3} \leq i \mid IC)$, respectively and are calculated for values of $i=1,2,...,15$.

The calculations in Table 4.2 assume that we monitor the process median using samples of size $n = 5$ (which is a very popular choice in practice) and that the upper control limit is $UCL = 5$. For this particular combination of n and UCL, it is calculated that $p_0^+ = \Pr(T_i \geq 5 \mid IC) = 0.03125$ where $T_i \sim Bin(5, 0.5)$, and it was found that ($ARL_{2of3}^{+} = 552.65$; $FAR_{2of3}^{+} = 0.00189$ for $i \geq 3$) while ($ARL_{2of3}^{A+} = 552.13$; $FAR_{2of3}^{A+} = 0.00098$ at $i = 2$ and $FAR_{2of3}^{A+} = 0.00189$ for $i \geq 3$).

From Table 4.2 we see that there are two key differences with respect to the in-control characteristics and the in-control performance of the charts:

(i) The augmented upper one-sided *2-of-3* sign chart can signal incorrectly (with probability 0.00098) after having observed only two charting statistics whereas the proposed upper one-sided *2-of-3* sign chart cannot.

(ii) The ratio $\Pr(N_{2of3}^{A+} \leq i \mid IC) / \Pr(N_{2of3}^{+} \leq i \mid IC)$, decreases to 1 as i increases; this observation is supported by the fact that ARL_{2of3}^{A+} is only slightly less than ARL_{2of3}^{+} i.e. $ARL_{2of3}^{A+} / ARL_{2of3}^{+} \approx 1$. These observations imply that, from start-up (when the process is IC) the augmented upper one-sided *2-of-3* sign chart always has a higher cumulative probability for a shorter run-length than the proposed upper one-sided *2-of-3* sign chart..

Table 4.2: The in-control probability mass functions (p.m.f's) and the in-control cumulative distribution functions (c.d.f's) of the proposed *2-of-3* and the augmented *2-of-3* sign charts when $n = 5$ and $UCL = 5$

	***2-of-3* sign chart**		**Augmented *2-of-3* sign chart**	
i	$\Pr(N^{+}_{2of3} = i \mid IC)$	$\Pr(N^{+}_{2of3} \leq i \mid IC)$	$\Pr(N^{A+}_{2of3} = i \mid IC)$	$\Pr(N^{A+}_{2of3} \leq i \mid IC)$
1	0	0	0	0
2	0	0	0.00098	0.00098
3	0.00189	0.00189	0.00189	0.00287
4	0.00186	0.00375	0.00183	0.00470
5	0.00181	0.00556	0.00180	0.00651
6	0.00180	0.00736	0.00180	0.00831
7	0.00180	0.00917	0.00180	0.01011
8	0.00180	0.01097	0.00180	0.01191
9	0.00180	0.01276	0.00179	0.01370
10	0.00179	0.01455	0.00179	0.01549
11	0.00179	0.01634	0.00179	0.01728
12	0.00179	0.01813	0.00178	0.01906
13	0.00178	0.01991	0.00178	0.02085
14	0.00178	0.02169	0.00178	0.02262
15	0.00178	0.02347	0.00177	0.02440

To summarize the above discussion and our findings based on the analysis, we can state that:

(i) The proposed upper one-sided *2-of-3* sign chart has a hitch at start-up but, the odds that this problem occurs are typically small; this should be reassuring for the practitioner.

(ii) It is possible to fix the imperfection of the proposed upper one-sided *2-of-3* sign chart by adding a third signaling event but, even this modification has a drawback: the performance of the augmented chart is degraded at start-up i.e. its false alarm rate is nonzero at time $i = 2$ (unlike the proposed chart) and the cumulative probability for a shorter run-length is higher than that of the proposed chart.

(iii) Neither the proposed nor the augmented upper one-sided *2-of-3* sign chart can be implemented without taking a risk i.e. there is a trade-off between having a hitch at start-up and the possibility of a false alarm at time $i = 2$.

(iv) The inherent risk of each chart cannot be completely eliminated but, these risks can be minimized (or at least reduced) by ensuring that the process is IC at start-up and/or by choosing p_0^+ to be small.

(v) The in-control performance of the charts are almost identical: there is only a bit of a difference in their in-control ARL's and, at time $i = 2$ we have that $FAR_{2of3}^{A+} = (p_0^+)^2$ whereas $FAR_{2of3}^{+} = 0$.

(vi) If a shift/change in the process occurs *after* start-up i.e. from time $i \geq 3$, both the charts can signal only on the occurrence of events C_1 or C_2. So, the OOC performance of these two charts would be the same.

We recommend that practitioners use either the proposed or the augmented upper one-sided *2-of-3* sign chart but, we suggest that they familiarize themselves with the inherent risk associated with the selected chart. If the practitioner is not willing to accept the risk(s) associated with the *2-of-3* charts, he/she should use another chart e.g. the new proposed upper one-sided *2-of-2* sign chart or the original upper one-sided *1-of-1* sign chart.

Based on the above analysis and the fact that the augmented chart can signal after having observed only two charting statistics instead of the proposed three charting statistics (which implies that the augmented chart is not a "true" *2-of-3* chart), it was decided to focus on the proposed upper one-sided *2-of-3* sign chart and not to investigate the statistical properties of the augmented upper one-sided *2-of-3* sign chart any further in this thesis.

Furthermore, although the above discussion focussed specifically on the *upper* one-sided *2-of-3* sign chart, these comments also apply to the *lower* one-sided *2-of-3* chart and the *two-sided 2-of-3* chart. In fact, these comments are relevant for any one-sided or two-sided *k-of-w* chart whenever $k < w$. This is so, because we need at least w charting statistics before we can declare the process OOC and we need at least $w - k$ charting statistics to plot IC (i.e. below the UCL or, above the LCL or, between the LCL and UCL, depending on the chart that is used).

Remark 7

If the upper one-sided *2-of-3* sign chart were to signal upon any one of the three events in which two of the last three charting statistics can plot on or above the *UCL* i.e. the occurrence of either event C_1 or C_2 or C_5 (see e.g. panel (a) of Figure 4.4 and panel (a) of Figure 4.5), the p.m.f as well as the mean (ARL) and the variance ($VARL$) of N^{+}_{2of3} would be obtainable from the distribution and the associated characteristics of the random variable $T_k^{(w)}$.

The random variable $T_k^{(w)}$ is the waiting time for the first occurrence of a scan or run of type k/w, where the term scan or generalized run of type k/w refers to sub sequences $\xi_i^{+}, \xi_{i+1}^{+}, ..., \xi_{i+j-1}^{+}$ of length $j \le w$ such that the number of successes contained therein is at least k, that is, $\sum_{s=i}^{i+j-1} \xi_s^{+} \ge k$ (see e.g. Chapter 9 of Balakrishnan and Koutras, (2002)); the probability distribution of $T_k^{(w)}$ is known as the geometric distribution of order k/w and derived via combinatorial methods.

Because we exclude event C_5 as a signaling event in case of the upper one-sided *2-of-3* chart (because the possibility of declaring a process out-of-control when the first and second charting statistic plot OOC but the third one plots IC is undesirable in practice), we cannot make use of the p.m.f or the associated properties of the geometric distribution of order k/w, that is, $T_k^{(w)}$; this supports the statement in the beginning of section 4.2.1 that the Markov chain technique has a great advantage over the classical combinatory techniques for finding the distribution(s) of run-related problems.

Example 1

Consider the upper one-sided *2-of-3* sign chart for monitoring the median $\theta = F_X^{-1}(\pi = 0.5)$ and suppose that the subgroup size $n = 5$, the charting constant $b = 0$ (so that $UCL = 5$) and θ_0 denotes the target (IC) value for the median.

As noted earlier, when the process is in-control $p_0 = \Pr(X_{ij} > \theta_0 \mid IC) = 0.5$ and therefore

$$p_0^+ = p_0^+(n = 5, b = 0, \theta = \theta_0) = I_{p_0=0.5}(5,1) = 0.03125.$$

Substituting $p_0^+ = 0.03125$ for p^+ in (4-20) and using (4-9) we get

$$ARL_0 = E(N^+_{2of3,0} \mid n = 5, b = 0, \theta = \theta_0) = 552.65.$$

Similarly, using (4-10), the $SDRL_0 = 550.218$.

The in-control c.d.f of N^+_{2of3} can be obtained using the p.m.f in (4-8) and is given by

$$\Pr(N^+_{2of3,0} \le j) = \sum_{i=1}^{j} \boldsymbol{\xi} \mathbf{Q}^{i-1}_{5\times5,0}(\mathbf{I} - \mathbf{Q}_{5\times5,0})\mathbf{1}$$

where $N^+_{2of3,0}$ denotes the in-control run-length random variable and $\mathbf{Q}_{5\times5,0}$ is found from (4-20) by substituting p_0^+ for p^+. For illustration, we calculate and show the in-control p.m.f and the in-control c.d.f values for $j = 1,2,3,4,5,6$ in Table 4.3.

Table 4.3: The in-control probability mass function (p.m.f) and the in-control cumulative distribution function (c.d.f) for the upper one-sided *2-of-3* sign chart

j	**1**	**2**	**3**	**4**	**5**	**6**
$\Pr(N^+_{2of3,0} = j)$	0	0	0.00189	0.00186	0.00181	0.00180
$\Pr(N^+_{2of3,0} \le j)$	0	0	0.00189	0.00375	0.00556	0.00736

Given the c.d.f we can find the $100\pi^{\text{th}}$ percentile of the run-length distribution, which is the smallest integer j so that $\Pr(N^+_{2of3,0} \le j) \ge \pi$. For example, the second quartile (the median run-length, denoted $MDRL$) is found to be $Q_2 = 384$. The percentiles provide useful information regarding the efficacy of the control chart in addition to the moments such as the ARL_0 and the $SDRL_0$.

Lower one-sided *1-of-1*, *2-of-2* and *2-of-3* sign charts

By substituting $p^- = p^-(n,a,\theta)$, which is defined in (4-12), for $p^+ = p^+(n,b,\theta)$, which is defined in (4-11), in the transition probability matrices of (4-14), (4-15) and (4-20), the distributions of the run-length random variables N^-_{1of1}, N^-_{2of2} and N^-_{2of3} of the lower (negative) one-sided *1-of-1*, *2-of-2* and *2-of-3* charts, respectively can be straightforwardly obtained. This is so because each lower one-sided chart is a mirror image of the corresponding upper one-sided chart.

Also, note that, when we monitor the median, the in-control distribution of the plotting statistic is symmetric i.e. $T_i \sim Bin(n, 0.5)$. In this case, it makes practical sense to use symmetrically placed control limits and set $b = a$ so that $UCL = n - a$ and $LCL = a$; this implies that the control limits are equidistant from both ends. For this specific choice of the control limits we have that $\xi_i^+ \overset{d}{=} \xi_i^-$ i.e. the signaling indicators used to define the upper one-sided charts have the same distribution as the signaling indicators used to define the lower one-sided charts, and implies that the in-control performance of the lower and the upper one-sided sign charts, for monitoring the median, are identical. The performance of the upper and the lower one-sided sign charts will be further discussed in section 4.2.4 when we study their design.

Lastly, note that, the distributions of N^-_{1of1} and N^-_{2of2} can also be obtained from those of T_k (see Remark 5) by setting $\alpha = p^-$ and substituting $k = 1$ or $k = 2$ in (4-16) and (4-17), respectively.

Two-sided sign charts

The derivation of the transition probability matrices and the run-length distributions of the two-sided charts (via the Markov chain approach) parallel those of the one-sided charts.

For the two-sided charts, the signaling indicators ξ_i's are defined by a series of values 0, 1 or 2, depending on whether the corresponding charting statistic T_i plots between the two control limits, on or above the *UCL*, or on or below the *LCL*, respectively; the probabilities for these three events are $1-p^{\pm}=1-p^{+}-p^{-}$, p^{+} and p^{-}, respectively (see e.g. Figure 4.10 below).

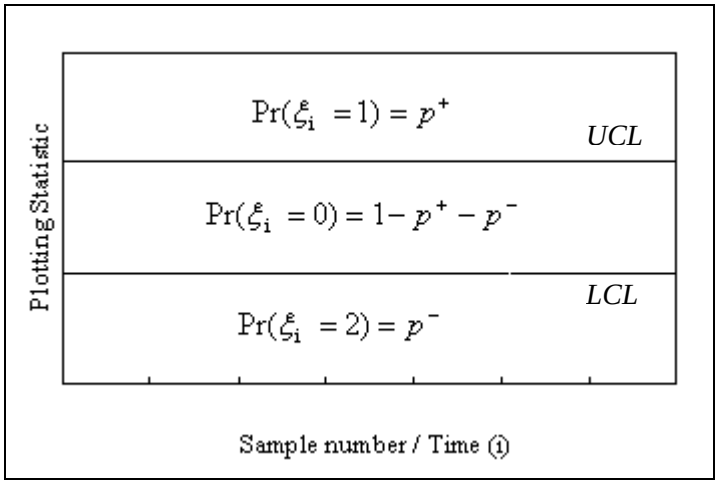


Figure 4.10: The three regions on the two-sided control chart ('0' , '1' and '2') and their associated probabilities used to classify the charting statistic

Two-sided *1-of-1* sign chart

For the *1-of-1* two-sided chart the run-length N_{1of1} is the waiting time for the first occurrence of the event $A = A_1 \cup A_2$ (see e.g. Figure 4.2), which is the first occurrence of the compound pattern $\Lambda = \Lambda_1 \cup \Lambda_2$ in the series of ξ_i's (i.e. among the 0's, 1's and 2's) where $\Lambda_1 = 1$ and $\Lambda_2 = 2$ are two distinct simple patterns in this situation.

The state space for the imbedded Markov chain associated with the variable N_{1of1} is $\Omega = \{\phi, 0, 1, 2\}$, which has $m = 4$ states. The absorbing states are $\{1\}$ and $\{2\}$ whereas $\{0\}$ is the transient state and $\{\phi\}$ is the dummy state.

The transition probability matrix $\mathbf{M}_{4\times4}$ is given by

$$\mathbf{M}_{4\times4} = \left[\begin{array}{c|c} \mathbf{Q}_{2\times2} & \mathbf{C}_{2\times2} \\ \hline \mathbf{0}_{2\times2} & \mathbf{I}_{2\times2} \end{array}\right] = \left[\begin{array}{cc|cc} p_{\phi,\phi} & p_{\phi,0} & p_{\phi,1} & p_{\phi,2} \\ p_{0,\phi} & p_{0,0} & p_{0,1} & p_{0,2} \\ \hline p_{1,\phi} & p_{1,0} & p_{1,1} & p_{1,2} \\ p_{2,\phi} & p_{2,0} & p_{2,1} & p_{2,2} \end{array}\right] = \left[\begin{array}{cc|cc} 0 & 1-p^+-p^- & p^+ & p^- \\ 0 & 1-p^+-p^- & p^+ & p^- \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right] \qquad (4\text{-}21)$$

where, for example, the entry in the 2nd row and 2nd column of $\mathbf{M}_{4\times4}$, denoted by $p_{0,0}$, is the probability that the system remains in state $\{0\}$, that is, where T_{i-1} plots IC at time $i-1$ and T_i plots IC at time i.

Two-sided *2-of-2* DR sign chart

The run-length N_{2of2}^{DR} of the *2-of-2* DR two-sided chart is the waiting time for the first occurrence of the event D_1 or D_2 or D_3 or D_4 (see e.g. Figure 4.6), which is the first occurrence of the compound pattern $\Lambda = \bigcup_{i=1}^{4} \Lambda_i$ in the series of ξ_i's, where $\Lambda_1 = 11$, $\Lambda_2 = 22$, $\Lambda_3 = 12$ and $\Lambda_4 = 21$ are the four distinct simple patterns.

The imbedded Markov chain, in this case, is defined on the state space $\Omega = \{\phi, 0, 1, 2, 11, 22, 12, 21\}$, where ϕ is the dummy state, the three states $\{0\}$, $\{1\}$ and $\{2\}$ are the transient states and the four states $\{11\}, \{22\}$, $\{12\}$ and $\{21\}$ are the absorbing states.

The transition probability matrix $\mathbf{M}_{8\times8}$ is given by

$$\mathbf{M}_{8\times8} = \left[\begin{array}{c:c} \mathbf{Q}_{4\times4} & \mathbf{C}_{4\times4} \\ \hdashline \mathbf{0}_{4\times4} & \mathbf{I}_{4\times4} \end{array}\right] = \left[\begin{array}{cccc:cccc} 0 & 1-p^+-p^- & p^+ & p^- & 0 & 0 & 0 & 0 \\ 0 & 1-p^+-p^- & p^+ & p^- & 0 & 0 & 0 & 0 \\ 0 & 1-p^+-p^- & 0 & 0 & p^+ & 0 & p^- & 0 \\ 0 & 1-p^+-p^- & 0 & 0 & 0 & p^- & 0 & p^+ \\ \hdashline 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right]. \qquad (4\text{-}22)$$

Remark 8

If we let

$$\xi_i^{\pm} = I(T_i \notin (LCL,UCL)) = \begin{cases} 1 & \text{if} \quad T_i \notin (LCL,UCL) \\ 0 & \text{if} \quad T_i \in (LCL,UCL) \end{cases}$$

where $I(T_i \notin (LCL,UCL))$ denotes the indicator function for the event $\{T_i \notin (LCL,UCL)\}$ then

$$\{\xi_i^{\pm} = 1\} \text{ if and only if } \{\xi_i = 1\} \cup \{\xi_i = 2\} \tag{4-23}$$

so that

$$\Pr(\xi_i^{\pm} = 1) = \Pr(\xi_i = 1) + \Pr(\xi_i = 2) \tag{4-24}$$

where the $\xi_i^{\pm}$'s is a sequence of i.i.d. Bernoulli random variables each with probability of success $\Pr(\xi_i^{\pm} = 1) = \Pr(T_i \notin (LCL,UCL)) = p^{\pm} = p^{+} + p^{-}$ and the ξ_i's is a sequence of i.i.d tri-variate random variables with probabilities $\Pr(\xi_i = 1) = p^{+}$, $\Pr(\xi_i = 0) = 1 - p^{\pm} = 1 - p^{+} - p^{-}$ and $\Pr(\xi_i = 2) = p^{-}$, respectively.

Expressions (4-23) and (4-24) permit us to define the signaling events and obtain the run-length distributions of the two-sided *1-of-1* and the two-sided *2-of-2* DR sign charts using the $\xi_i^{\pm}$'s instead of using the ξ_i's. This means that, instead of using the Markov chain approach, we can find the distributions of N_{1of1} and N_{2of2}^{DR} using the results (or properties) of the geometric distribution of order k.

In particular, it follows from (4-23) and (4-24) that the run-length N_{1of1} of the two-sided *1-of-1* chart, which is the waiting time for the first occurrence of the event $A = A_1 \cup A_2 = \{\xi_i = 1\} \cup \{\xi_i = 2\}$, is equivalent to the waiting time for the first success (i.e. 1) among the $\xi_i^{\pm}$'s, that is, $A = \{\xi_i^{\pm} = 1\}$. Likewise, the run-length N_{2of2}^{DR} of the two-sided 2-*of-2* chart, which is the waiting time for the first occurrence of the event D_1 or D_2 or D_3 or D_4 (see e.g. Figure 4.6), is the same as the waiting time for the first occurrence of two consecutive successes (two successive 1's) among the $\xi_i^{\pm}$'s, that is, $D(DR) = \{\xi_{i-1}^{\pm} = \xi_i^{\pm} = 1\}$ so that

$$\Pr(D(DR)) = \Pr(D_1 \cup D_2 \cup D_3 \cup D_4) = \Pr(\{\xi_{i-1}^{\pm} = \xi_i^{\pm} = 1\}) = (p^{\pm})^2.$$

The distributions of N_{1of1} and N^{DR}_{2of2} are therefore both geometric distributions of order k so that closed form expressions for the p.m.f's of N_{1of1} and N^{DR}_{2of2} can be conveniently obtained from (4-16) by setting

$$k=1 \quad \text{with} \quad \alpha = \Pr(A) = p^{\pm}$$

and

$$k=2 \quad \text{with} \quad \alpha = \Pr(D(DR)) = (p^{\pm})^2$$

instead of symbolically simplifying expression (4-8).

For example, upon substituting the essential transition probability sub-matrix of the two-sided *1-of-1* sign chart

$$\mathbf{Q}_{2\times 2} = \begin{bmatrix} 0 & 1-p^{+}-p^{-} \\ 0 & 1-p^{+}-p^{-} \end{bmatrix} = \begin{bmatrix} 0 & 1-p^{\pm} \\ 0 & 1-p^{\pm} \end{bmatrix}$$

into (4-8) and simplifying symbolically, we get an explicit formula for the p.m.f of N_{1of1} (via the Markov chain approach) that corresponds to the already available p.m.f one obtains after substituting $p^{\pm}$ and $k=1$ into (4-16) i.e.

$$\Pr(N_{1of1}=j \mid n,a,b,\theta) = \Pr(T_{k=1}=j \mid n,a,b,\theta) = (1-p^{\pm})^{j-1}p^{\pm} \quad \text{for} \quad j=1,2,3,\ldots. \qquad (4\text{-}25)$$

Two-sided *2-of-2* KL sign chart

The run-length N_{2of2}^{KL} of the *2-of-2* KL two-sided chart is the waiting time for the first occurrence of the event D_1 or D_2 (see e.g. panels (a) and (b) of Figure 4.6), which is the first occurrence of the compound pattern $\Lambda = \Lambda_1 \cup \Lambda_2$ in the series of ξ_i's, where $\Lambda_1 = 11$ and $\Lambda_2 = 22$ are the two distinct simple patterns in this case.

The imbedded Markov chain associated with the run-length variable N_{2of2}^{KL} is defined on the state space $\Omega = \{\phi, 0, 1, 2, 11, 22\}$, which has $m = 6$ states, where $\{11\}$ and $\{22\}$ are the two absorbing states.

The transition probability matrix $\mathbf{M}_{6\times6}$ of the Markov chain is given by

$$\mathbf{M}_{6\times6} = \left[\begin{array}{c|c} \mathbf{Q}_{4\times4} & \mathbf{C}_{4\times2} \\ \hline \mathbf{0}_{2\times4} & \mathbf{I}_{2\times2} \end{array}\right] = \left[\begin{array}{cccc|cc} 0 & 1-p^+-p^- & p^+ & p^- & 0 & 0 \\ 0 & 1-p^+-p^- & p^+ & p^- & 0 & 0 \\ 0 & 1-p^+-p^- & 0 & p^- & p^+ & 0 \\ 0 & 1-p^+-p^- & p^+ & 0 & 0 & p^- \\ \hline 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right]. \qquad (4\text{-}26)$$

Two-sided *2-of-3* sign chart

The run-length N_{2of3} of the *2-of-3* two-sided chart is the waiting time for the first occurrence of the event E_1 or E_2 or E_3 or E_4 (see e.g. Figure 4.7), which is the first occurrence of the compound pattern $\Lambda = \bigcup_{i=1}^{4} \Lambda_i$ in the series of ξ_i's (i.e. among the 0's, 1's and 2's), where $\Lambda_1 = 011$, $\Lambda_2 = 101$, $\Lambda_3 = 022$ and $\Lambda_4 = 202$ are the four distinct simple patterns.

The imbedded Markov chain, in this case, is defined on the finite state space $\Omega = \{\phi, 0, 1, 2, 01, 10, 02, 20, \alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ with $m = 12$ states, where the four states $\alpha_1 = \{011\}, \alpha_2 = \{101\}$, $\alpha_3 = \{022\}$ and $\alpha_4 = \{202\}$ are the absorbing states, ϕ is the dummy state, and the eight transient states are all the sequential sub-patterns of $\Lambda_1 = 011$, $\Lambda_2 = 101$, $\Lambda_3 = 022$ and $\Lambda_4 = 202$, respectively. In this case, the essential transition probability sub-matrix $\mathbf{Q}_{8\times8}$ of the transition probability matrix

$$\mathbf{M}_{12\times12} = \left[\begin{array}{c|c} \mathbf{Q}_{8\times8} & \mathbf{C}_{8\times4} \\ \hline \mathbf{0}_{4\times8} & \mathbf{I}_{4\times4} \end{array}\right]$$

is given by

$$\mathbf{Q}_{8\times8} = \begin{bmatrix} 0 & 1-p^+-p^- & p^+ & p^- & 0 & 0 & 0 & 0 \\ 0 & 1-p^+-p^- & 0 & 0 & p^+ & 0 & p^- & 0 \\ 0 & 0 & p^+ & p^- & 0 & 1-p^+-p^- & 0 & 0 \\ 0 & 0 & p^+ & p^- & 0 & 0 & 0 & 1-p^+-p^- \\ 0 & 0 & 0 & p^- & 0 & 1-p^+-p^- & 0 & 0 \\ 0 & 1-p^+-p^- & 0 & 0 & 0 & 0 & p^- & 0 \\ 0 & 0 & p^+ & 0 & 0 & 0 & 0 & 1-p^+-p^- \\ 0 & 1-p^+-p^- & 0 & 0 & p^+ & 0 & 0 & 0 \end{bmatrix} \quad (4\text{-}27)$$

whilst the non-essential transition probability sub-matrix $\mathbf{C}_{8\times4}$ is given by

$$\mathbf{C}_{4\times8} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & p^+ & 0 & 0 \\ 0 & 0 & 0 & 0 & p^+ & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & p^- \\ 0 & 0 & 0 & 0 & 0 & 0 & p^- & 0 \end{bmatrix}^T.$$

Remark 9

In general, if $F_1, F_2, ..., F_r$, $r \geq 1$ are the set of *all possible events* in which (a) k consecutive charting statistics, or (b) exactly k of the last w charting statistics) can plot OOC, one can design a chart that signals on the first occurrence of the event $F = \bigcup_{i=1}^{r} F_i$. The run-length of such a chart would be (a) geometric of order k, or (b) geometric of order k/w.

However, we prefer, due to practical considerations, to exclude some of the F_i's ; in doing so the distribution of the run-length is not necessarily geometric of order k or geometric of order k/w and we then use the Markov chain approach to find the run-length distribution.

For example, as mentioned earlier, because the two-sided *2-of-2* KL chart signals only if event D_1 or D_2 occurs for the first time and does not signal (unlike the two-sided *2-of-2* DR chart) in case event D_3 or event D_4 occurs (see Figure 4.6), the distribution of N_{2of2}^{KL}, in general, is not a geometric distribution of order $k = 2$.

Likewise, because the two-sided *2-of-3* sign chart signals only on the first occurrence of event E_1 or E_2 or E_3 or E_4 (see e.g. Figure 4.7) and excludes the remaining eight events in which exactly two of the last three charting statistics can plot on or outside the control limits i.e. events E_5, E_6, E_7, E_8, E_9, E_{10}, E_{11} and E_{12} (see e.g. Figure 4.8), as signaling events the distribution of N_{2of3}, in general, is not a geometric distribution of order $2/3$.

If, however, we were to design a two-sided *2-of-3* sign chart that signals on the first occurrence of either one of the events E_i for $i = 1,2,...,12$ the distribution of the run-length random variable associated with such a chart would be a geometric distribution of order $2/3$ with probability of success $\Pr(T_i \notin (LCL, UCL)) = p^{\pm} = p^{+} + p^{-}$ (also see Remark 7).

4.2.3 The in-control run-length characteristics of the one-sided and two-sided sign charts

The characteristics of the in-control (IC) run-length distributions are essential in the design of a control chart. Furthermore, for out-of-control (OOC) performance comparisons their in-control average run-length (ARL_0) and/or false alarm rate (FAR) should be equal or, at least, approximately so.

Tables 4.4 and 4.5 summarize the expressions for the ARL and the FAR of the various sign charts. The ARL expressions, in general, follow from having written the corresponding essential transition probability sub-matrix $\mathbf{Q}_{h\times h}$, substituting it in (4-9) and simplifying symbolically.

For example, for the *1-of-1* two-sided chart with state space $\Omega = \{\phi, 0, 1, 2\}$, it was shown that

$$\mathbf{Q}_{2\times 2} = \begin{bmatrix} 0 & 1-p^+-p^- \\ 0 & 1-p^+-p^- \end{bmatrix}$$

so that upon substitution in (4-9) and simplifying we get an explicit formula for the ARL given by

$$ARL_{1of1} = E(N_{1of1}) = \begin{bmatrix} 1 & 0 \end{bmatrix} \left[\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 1-p^+-p^- \\ 0 & 1-p^+-p^- \end{bmatrix} \right]^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{p^+ + p^-}.$$

Alternatively, in some cases (such as the upper and the lower one-sided *1-of-1* and *2-of-2* charts as well as the two-sided *1-of-1* and *2-of-2* DR charts) one can obtain closed form expressions by using available results of the geometric distribution of order k. For example, for the two-sided *2-of-2* DR chart, one obtains the ARL upon substituting $p^{\pm} = p^+ + p^-$ for α in (4-17); this gives

$$ARL^{DR}_{2of2} = E(N^{DR}_{2of2}) = E(T_2) = \frac{1-(p^+ + p^-)^2}{(1-p^+-p^-)(p^+ + p^-)^2} = \frac{p^+ + p^- + 1}{(p^+ + p^-)^2}.$$

Note that, for the in-control average run-length

$$p_0^+ = p_0^+(n, b, \theta = \theta_0) = I_{p_0}(n-b, b+1)$$

and

$$p_0^- = p_0^-(n, a, \theta = \theta_0) = 1 - I_{p_0}(a+1, n-a)$$

where p_0 is defined in (4-6), are to be substituted for p^+ and p^-, respectively, in the ARL expressions of Tables 4.4 and 4.5.

The expressions for the FAR can be obtained from the definitions of the charts in a straightforward manner. For example, for the *2-of-2* upper one-sided chart, the false alarm rate is

$$FAR^{+}_{2of2} = \Pr(B_1 \mid IC) = \Pr(T_{i-1} \geq UCL \mid IC) \times \Pr(T_i \geq UCL \mid IC) = (p_0^{+})^2$$

where B_1 is defined in panel (a) of Figure 4.3, whereas the false alarm rate for the *2-of-2* KL chart is

$$\begin{aligned} FAR^{KL}_{2of2} &= \Pr(D_1 \mid IC) + \Pr(D_2 \mid IC) \\ &= \Pr(T_{i-1} \geq UCL, T_i \geq UCL \mid IC) + \Pr(T_{i-1} \leq LCL, T_i \leq LCL \mid IC) \\ &= (p_0^{+})^2 + (p_0^{-})^2 \end{aligned}$$

where D_1 and D_2 are defined in panels (a) and (b) of Figure 4.6.

Table 4.4: Average run-lengths (*ARL*'s) and false alarm rates (*FAR*'s) of the upper one-sided sign charts

***1-of-1* upper**	***2-of-2* upper**	***2-of-3* upper**
$ARL^{+}_{1of1} = \dfrac{1}{p^{+}}$	$ARL^{+}_{2of2} = \dfrac{1+p^{+}}{(p^{+})^2}$	$ARL^{+}_{2of3} = \dfrac{(p^{+})^3 - 2(p^{+})^2 + p^{+} + 1}{(p^{+})^2[(p^{+})^2 - 3p^{+} + 2)]}$
$FAR^{+}_{1of1} = p_0^{+}$	$FAR^{+}_{2of2} = (p_0^{+})^2$	$FAR^{+}_{2of3} = 2(1-p_0^{+})(p_0^{+})^2$

Table 4.5: Average run-lengths (*ARL*'s) and false alarm rates (*FAR*'s) of the two-sided sign charts

1-of-1	***2-of-2* DR**	***2-of-2* KL**
$ARL_{1of1} = \dfrac{1}{p^{+}+p^{-}}$	$ARL^{DR}_{2of2} = \dfrac{p^{+}+p^{-}+1}{(p^{+}+p^{-})^2}$	$ARL^{KL}_{2of2} = \left(\dfrac{(p^{+})^2}{(p^{+}+1)} + \dfrac{(p^{-})^2}{(p^{-}+1)}\right)^{-1}$
$FAR_{1of1} = p_0^{+} + p_0^{-}$	$FAR^{DR}_{2of2} = (p_0^{+})^2 + (p_0^{-})^2 + 2(p_0^{+})(p_0^{-})$	$FAR^{KL}_{2of2} = (p_0^{+})^2 + (p_0^{-})^2$

2-of-3

$$ARL_{2of3} = \frac{\left[p^{-} + p^{+}(p^{-})^2 - p^{+}p^{-} - 2(p^{-})^2 + (p^{-})^3 + 1\right]^2 \left[p^{+} + (p^{+})^2 p^{-} - p^{+}p^{-} - 2(p^{+})^2 + (p^{+})^3 + 1\right]^2}{\left[(p^{+}+p^{-}-1)\begin{pmatrix} -2p^{+}(p^{-})^2 - 2(p^{+})^2 p^{-} + 3p^{+}(p^{-})^3 + 3(p^{+})^3 p^{-} - p^{+}(p^{-})^4 \\ -(p^{+})^4 p^{-} + 8(p^{+})^2(p^{-})^2 - 6(p^{+})^2(p^{-})^3 - 6(p^{+})^3(p^{-})^2 \\ +(p^{+})^2(p^{-})^4 + 2(p^{+})^3(p^{-})^3 + (p^{+})^4(p^{-})^2 - 2(p^{+})^2 + (p^{+})^3 \\ -2(p^{-})^2 + (p^{-})^3 \end{pmatrix}\right]^2}$$

$$= \frac{2q^3 - 3q^2 + q + 1}{2q^2(2q^2 - 5q + 2)} \quad \text{if} \quad p^{+} = p^{-} = q \text{ (say) i.e. symmetrically placed control limits (equidistant from both ends)}$$

$$FAR_{2of3} = 2(p_0^{+})^2(1 - p_0^{+} - p_0^{-}) + 2(p_0^{-})^2(1 - p_0^{+} - p_0^{-})$$

Remark 10

As mentioned earlier, the FAR and the ARL_0 of all the sign charts depend only on the probabilities p_0^+ and/or p_0^-, which in turn depend only on the sample size n and the charting constants a and/or b and not on the underlying process distribution. The in-control run-length distributions therefore remain the same for all continuous process distributions, and hence the proposed sign charts are nonparametric or distribution-free.

4.2.4 Design of the upper (lower) one-sided *1-of-1*, *2-of-2* and *2-of-3* sign charts

In order to design the proposed charts and assess their in-control performance the design parameters need to be chosen. The design parameters include

(i) the sample size n,

(ii) the charting constants a and b, and

(iii) the target value θ_0.

Because rational subgroups in SPC are small, we focus on $n = 4(1)15$, 20 and 25.

To monitor the center of a process, one typically chooses θ to be the process median and this is the case we study here; hence $\pi = 0.5$ so that $\theta = F_X^{-1}(0.5)$ and $p = p_0 = \Pr(X_{ij} > \theta_0 \mid IC) = 0.5$. However, other choices of θ might be desirable in some situations. For example, to monitor the 25th percentile of a processes' distribution we would set $\pi = 0.25$ so that $\theta = F_X^{-1}(0.25)$ and then $p_0 = \Pr(X_{ij} > \theta_0 \mid IC) = 0.75$. The sign charts are flexible enough to allow one to do that.

The charting constants a and b can be any integer between and including 0 and n. However, a and/or b are typically chosen so that the ARL_0 is reasonably large.

Tables 4.6 and 4.7 display the in-control characteristics of the *1-of-1* sign chart of Amin et al. (1995) and the new proposed runs-rule enhanced *2-of-2* and *2-of-3* one-sided sign charts.

Note that, Tables 4.6 and 4.7 apply to both the lower and the upper one-sided sign charts because in case of the median we have that $p = p_0 = 0.5$ when the process is in-control, which implies that the charting statistic T_i has a binomial (n, 0.5) distribution which is symmetric. Hence, when $b = a$ (as in Tables 4.6 and 4.7), the in-control performance of the lower and the upper one-sided sign charts are identical.

For example, if $n = 6$ and the LCL of the lower one-sided chart is $a = 1$, the UCL for the upper one-sided chart is simply $n - b = 6 - 1 = 5$, and both of these charts have an in-control ARL of 9.14 and a FAR of 0.10938.

An examination of the *ARL* and *FAR* values in Tables 4.6 and 4.7 reveal the advantages of the new sign charts

(i) They offer more practically attractive ARL_0's and *FAR*'s in that, for any particular combination of n and a, the attained ARL_0 values of the *2-of-2* and the *2-of-3* charts are much higher than those of the *1-of-1* chart with a corresponding decrease in the *FAR*.

For example, for $n = 5$ and $a = 0$ the ARL_0 of the *1-of-1* chart is 32.00 with a fairly large *FAR* of 0.03125, but for the *2-of-3* chart the ARL_0 increases to a more reasonable 552.65 and the *FAR* decreases to 0.00189, whereas for the *2-of-2* chart, the ARL_0 equals 1056.00 with a *FAR* of 0.00098.

(ii) Most importantly, when using the *1-of-1* chart the industry standard ARL_0 value of 370 and *FAR* of 0.0027 is far from being attainable, but with the proposed *2-of-2* chart, for example, when $n = 10$ and $a = 2$, we can be almost on target e.g. the ARL_0 and *FAR* values are 352.65 and 0.00299, respectively.

Thus, by carefully choosing the sample size n, the charting constants a and/or *b*, and the values of k and w, we can attain more familiar and recommended values for the ARL_0 and the *FAR* for the proposed nonparametric sign charts. Even for a sample size as small as $n = 4$, an ARL_0 of 272.00 with a *FAR* of 0.00391 is possible when the *2-of-2* chart is used with $a = 0$.

Amin et al. (1995) noted that the largest possible ARL_0 of their *1-of-1* one-sided sign chart for the median is 2^n. However, our runs-rules based sign charts provide a wider range of attainable ARL_0 values and false alarm rates. For instance, for $n = 15$ the sign charts can attain an ARL_0 (*FAR*) as low (high) as 6.00 (0.25) and an ARL_0 (*FAR*) as high (low) as 3293.23 (0.00031) with the *2-of-2* chart.

Table 4.6: The in-control characteristics (*ARL* and *FAR*) of the one-sided *1-of-1*, *2-of-2* and *2-of-3* sign charts for the median (for samples of size $n = 4(1)11$)*

Sample size	*LCL*	*UCL*	*1-of-1*		*2-of-2*		*2-of-3*	
n	***a***	***n-b=n-a***	ARL_0	***FAR***	ARL_0	***FAR***	ARL_0	***FAR***
4	**0**	**4**	16.00	0.06250	272.00	0.00391	148.68	0.00732
	1	**3**			13.44	0.09766	10.13	0.13428
5	**0**	**5**	32.00	0.03125	1056.00	0.00098	552.65	0.00189
	1	**4**	5.33	0.18750	33.78	0.03516	21.71	0.05713
	2	**3**			6.00	0.25000		
6	**0**	**6**	64.00	0.01563	4160.00	0.00024	2128.64	0.00048
	1	**5**	9.14	0.10938	92.73	0.01196	53.95	0.02131
	2	**4**			11.37	0.11816	8.94	0.15509
7	**0**	**7**	128.00	0.00781			8352.63	0.00012
	1	**6**	16.00	0.06250	272.00	0.00391	148.68	0.00732
	2	**5**			23.90	0.05133	16.13	0.07940
	3	**4**			6.00	0.25000		
8	**0**	**8**	256.00	0.00391				
	1	**7**	28.44	0.03516	837.53	0.00124	440.75	0.00239
	2	**6**	6.92	0.14453	54.79	0.02089	33.35	0.03574
	3	**5**			10.33	0.13197	8.34	0.16806
9	**0**	**9**	512.00	0.00195				
	1	**8**	51.20	0.01953	2672.64	0.00038	1375.36	0.00075
	2	**7**	11.13	0.08984	135.02	0.00807	76.56	0.01469
	3	**6**			19.45	0.06447	13.59	0.09620
	4	**5**			6.00	0.25000		
10	**0**	**10**	1024.00	0.00098				
	1	**9**	93.09	0.01074	8759.01	0.00012	4449.96	0.00023
	2	**8**	18.29	0.05469	352.65	0.00299	190.71	0.00565
	3	**7**	5.82	0.17188	39.67	0.02954	25.00	0.04893
	4	**6**			9.69	0.14209	7.98	0.17706
11	**0**	**11**	2048.00	0.00049				
	1	**10**	170.67	0.00586				
	2	**9**	30.57	0.03271	964.92	0.00107	506.04	0.00207
	3	**8**	8.83	0.11328	86.75	0.01283	50.73	0.02276
	4	**7**			16.92	0.07530	12.14	0.10928
	5	**6**			6.00	0.25000		

**Note*: Only ARL_0 values greater than 5 and less than 10 000 are shown.

Table 4.7: The in-control characteristics (*ARL* and *FAR*) of the one-sided *1-of-1*, *2-of-2* and *2-of-3* sign charts for the median (for samples of size $n = 12(1)15(5)25$)*

Sample size	*LCL*	*UCL*	*1-of-1*		*2-of-2*		*2-of-3*	
n	***a***	***n-b=n-a***	ARL_0	***FAR***	ARL_0	***FAR***	ARL_0	***FAR***
12	**0**	**12**	4096.00	0.00024				
	1	**11**	315.08	0.00317				
	2	**10**	51.85	0.01929	2740.07	0.00037	1409.56	0.00073
	3	**9**	13.70	0.07300	201.36	0.00533	111.64	0.00988
	4	**8**	5.16	0.19385	31.77	0.03758	20.58	0.06059
	5	**7**			9.25	0.14993	7.73	0.18375
13	**0**	**13**	8192.00	0.00012				
	1	**12**	585.14	0.00171				
	2	**11**	89.04	0.01123	8017.78	0.00013	4076.31	0.00025
	3	**10**	21.67	0.04614	491.35	0.00213	262.59	0.00406
	4	**9**	7.49	0.13342	63.67	0.01780	38.21	0.03085
	5	**8**			15.29	0.08441	11.20	0.11977
	6	**7**			6.00	0.25000		
14	**1**	**13**	1092.27	0.00092				
	2	**12**	154.57	0.00647				
	3	**11**	34.86	0.02869	1250.05	0.00082	651.82	0.00160
	4	**10**	11.14	0.08978	135.19	0.00806	76.66	0.01467
	5	**9**			26.97	0.04493	17.87	0.07082
	6	**8**			8.93	0.15623	7.55	0.18896
15	**1**	**14**	2048.00	0.00049				
	2	**13**	270.81	0.00369				
	3	**12**	56.89	0.01758	3293.23	0.00031	1689.92	0.00061
	4	**11**	16.88	0.05923	301.88	0.00351	164.28	0.00660
	5	**10**	6.63	0.15088	50.56	0.02276	31.02	0.03866
	6	**9**			14.14	0.09218	10.53	0.12839
	7	**8**			6.00	0.25000		
20	**2**	**18**	4969.55	0.00020				
	3	**17**	776.15	0.00129				
	4	**16**	169.23	0.00591				
	5	**15**	48.32	0.02069	2383.29	0.00043	1228.53	0.00084
	6	**14**	17.34	0.05766	318.13	0.00332	172.75	0.00627
	7	**13**	7.60	0.13159	65.35	0.01732	39.13	0.03007
	8	**12**			19.75	0.06336	13.76	0.09483
	9	**11**			8.32	0.16966	7.21	0.19956
25	**4**	**21**	2196.55	0.00046				
	5	**20**	490.52	0.00204				
	6	**19**	136.67	0.00732			9511.45	0.00011
	7	**18**	46.21	0.02164	2181.12	0.00047	1125.86	0.00092
	8	**17**	18.56	0.05388	363.08	0.00290	196.13	0.00549
	9	**16**	8.71	0.11476	84.64	0.01317	49.59	0.02332
	10	**15**			26.93	0.04502	17.85	0.07094
	11	**14**			11.30	0.11904	8.90	0.15594
	12	**13**			6.00	0.25000		

**Note*: Only ARL_0 values greater than 5 and less than 10 000 are shown.

4.2.5 Performance comparison of the one-sided sign charts

We compare the performance of the *1-of-1*, the *2-of-2* and the *2-of-3* sign charts to that of the competing *1-of-1* SR chart of Bakir (2004) and the *2-of-2* SR charts of Chakraborti and Eryilmaz (2007) under the normal, the double exponential (or Laplace) and the Cauchy distributions; for completeness, we also include the well-known Shewhart $\overline{X}$ chart.

The double exponential and the Cauchy distributions are normal like with different tail behavior (see e.g. Figure 4.11). For the double exponential distribution the scale parameter was set equal to $1/\sqrt{2}$ for a standard deviation of 1; for the Cauchy distribution the scale parameter was set equal to 0.2605 in order to achieve a tail probability of 0.05 above $\theta_0 + 1.645$ - the same as for a $N(\theta_0,1)$ distribution.

Without loss of generality, we take the in-control median to be $\theta_0 = 0$. All three distributions are symmetric and a shift refers to a shift in the mean (median). The amount of shift in the median was taken over the range $\delta = 0(0.2)1.2$.

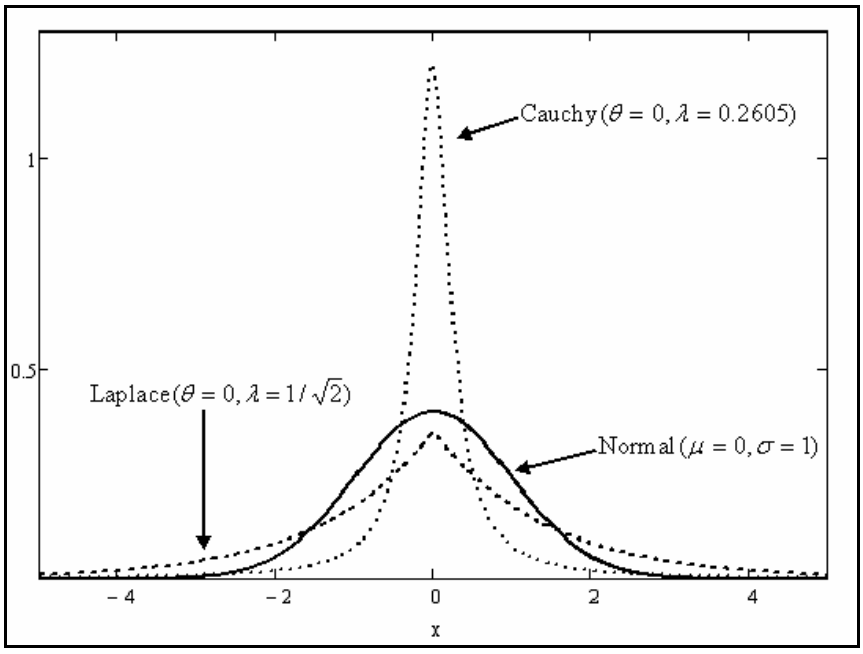


Figure 4.11: Probability distributions used for the performance comparison of the sign control charts

For comparison purposes the control charts are designed so that the ARL_0 values are high and are approximately equal. However, because the nonparametric charts are based on charting statistics that have discrete distributions, it is not possible to straightforwardly design the charts such that their ARL_0 values are all equal, and equal to some desired value such as 370 for a given sample size n (see e.g. Tables 4.6 and 4.7) .

Randomization was therefore used to ensure that the charts all have the same ARL_0 for a selected sample size. The technique is mainly used in the testing literature to compare the power of tests based on discrete test statistics so that they have identical nominal Type I error probability such as 0.05 (see e.g. Gibbons and Chakraborti, (2003)). We provide an example for illustration with the *1-of-1* sign chart; randomization for the other nonparametric charts can be handled in a similar way.

Example 2

Consider constructing a *1-of-1* upper one-sided sign chart with ARL_0 of 370 when samples of size $n = 10$ are used. From Table 4.6 we see that for this chart, exact in-control ARL values of 1024 (when $UCL = 10$) and 93.09 (when $UCL = 9$) are attainable that trap the target value 370.

The following randomized decision rule has an exact ARL_0 of 370:

"Declare the process OOC if $T_i \geq UCL = 10$ (with probability 1) and with probability q if $T_i = UCL - 1 = 9$, where $0 < q < 1$ is chosen such that $\Pr(T_i \geq 10 \mid IC) + q.\Pr(T_i = 9 \mid IC) = 1/370$".

Assuming that the median is the parameter of interest and the process is IC, $T_i \sim Bin(10,0.5)$ and therefore

$$q = [1/370 - \Pr(T_i \geq 10 \mid IC)]/\Pr(T_i = 9 \mid IC) = 0.1768 \approx 0.18 .$$

Hence, if we declare the process OOC every time the charting statistic is greater than or equal to 10 and declare the process OOC in 17.68% of the cases the charting statistic equals 9, we would have a *1-of-1* upper one-sided sign chart with an in-control ARL of 370.

In practice, we could use a random number generator to make a decision; for example, if the charting statistic equals 9, we could draw a random number between 1 and 100; if the drawn number is between 1 and 18, the process is declared OOC, otherwise it is not.

Note that, randomization is used to ensure that the in-control ARL values of the competing charts are equal so that their OOC performance can be fairly compared. The implementation and application of the charts remain as defined earlier and require no randomization.

The in-control (when $\delta = 0$) and the out-of-control (when $\delta \neq 0$) characteristics of the various charts for samples of size $n = 10$ are shown in Tables 4.8, 4.9 and 4.10 under the normal, the double exponential and the Cauchy distribution, respectively. The characteristics include the ARL, the $SDRL$ as well as the 5th, the 25th (the first quartile, Q_1), the 50th (the median run-length, $MDRL$), the 75th (the third quartile, Q_3) and the 95th percentiles of the run-length distribution.

Note that, since randomization was used and the out-of-control distribution for the SR statistic is unavailable for most distributions, we used simulations (100 000 samples each of $n = 10$) to estimate these characteristics in SAS®9.1; these programs can be found in Appendix 4A.

Table 4.8: In-control and out-of-control characteristics of the run-length distributions of the one-sided *1-of-1* sign, the *2-of-2* sign, the *2-of-3* sign, the *1-of-1* SR and the *2-of-2* SR chart for the median under the normal distribution

	***1-of-1* sign (*UCL*=10)**							***1-of-1* SR (*UCL*=53)**						
Shift	***ARL***	***SDRL***	**5th**	**Q_1**	***MDRL***	**Q_3**	**95th**	***ARL***	***SDRL***	**5th**	**Q_1**	***MDRL***	**Q_3**	**95th**
0.0	370	370.4	19	107	259	514	1107	370	369.7	19	106	255	509	1099
0.2	104	103.7	6	30	72	144	310	88.7	88.2	5	26	62	123	265
0.4	35.6	35.2	2	11	25	49	106	27.3	26.8	2	8	19	38	81
0.6	14.7	14.2	1	5	10	20	43	10.6	10.1	1	3	7	14	31
0.8	7.3	6.8	1	2	5	10	21	5	4.5	1	2	4	7	14
1.0	4.2	3.7	1	2	3	6	12	2.9	2.3	1	1	2	4	7
1.2	2.8	2.2	1	1	2	4	7	1.9	1.3	1	1	1	2	5
	***2-of-2* sign (*UCL*=9)**							***2-of-2* SR (*UCL*=33)**						
Shift	***ARL***	***SDRL***	**5th**	**Q_1**	***MDRL***	**Q_3**	**95th**	***ARL***	***SDRL***	**5th**	**Q_1**	***MDRL***	**Q_3**	**95th**
0.0	370	368.9	20	108	258	509	1107	370	371.9	21	108	256	514	1122
0.2	64	62.6	5	19	45	88	189	50.9	49.4	4	16	36	70	150
0.4	17.5	16.2	2	6	13	24	50	12.7	11.4	2	5	9	17	35
0.6	7.2	5.9	2	3	5	10	19	5.2	3.9	2	2	4	7	13
0.8	4	2.7	2	2	3	5	9	3.1	1.7	2	2	2	4	7
1.0	2.9	1.5	2	2	2	3	6	2.4	0.9	2	2	2	2	4
1.2	2.4	0.8	2	2	2	2	4	2.1	0.4	2	2	2	2	3
	***2-of-3* sign (*UCL*=9)**							**X-bar (*UCL* = 0.8797)**						
Shift	***ARL***	***SDRL***	**5th**	**Q_1**	***MDRL***	**Q_3**	**95th**	***ARL***	***SDRL***	**5th**	**Q_1**	***MDRL***	**Q_3**	**95th**
0.0	370	362.8	21	108	255	510	1094	370	370.3	20	107	257	513	1109
0.2	62.2	60.3	5	19	44	86	183	63.4	62.5	4	19	44	88	187
0.4	17.2	15.5	3	6	12	23	48	15.5	14.9	1	5	11	21	45
0.6	7.2	5.6	2	3	5	9	18	5.3	4.8	1	2	4	7	15
0.8	4.1	2.6	2	2	3	5	9	2.5	1.9	1	1	2	3	6
1.0	3	1.4	2	2	3	3	6	1.5	0.9	1	1	1	2	3
1.2	2.5	0.8	2	2	2	3	4	1.2	0.5	1	1	1	1	2

Table 4.9: In-control and out-of-control characteristics of the run-length distributions of the one-sided *1-of-1* sign, the *2-of-2* sign, the *2-of-3* sign, the *1-of-1* SR and the *2-of-2* SR chart for the median under the double exponential distribution

	***1-of-1* sign (*UCL*=10)**							***1-of-1* SR (*UCL*=53)**						
Shift	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**
0.0	370	371.2	19	107	257	514	1110	370	370.4	19	106	256	511	1111
0.2	54.6	53.8	3	16	38	75	162	48.7	48.2	3	14	34	68	145
0.4	16.6	16.1	1	5	12	23	49	13.3	12.8	1	4	9	18	39
0.6	7.5	7	1	3	5	10	21	5.7	5.1	1	2	4	8	16
0.8	4.3	3.8	1	2	3	6	12	3.2	2.6	1	1	2	4	8
1.0	3	2.4	1	1	2	4	8	2.2	1.6	1	1	2	3	5
1.2	2.2	1.7	1	1	2	3	6	1.7	1	1	1	1	2	4
	***2-of-2* sign (*UCL*=9)**							***2-of-2* SR (*UCL*=33)**						
Shift	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**
0.0	370	368.5	21	107	256	512	1099	370	369	21	107	255	511	1102
0.2	29	27.5	3	9	20	40	84	31.8	30.4	3	10	22	44	92
0.4	8	6.7	2	3	6	11	21	8.1	6.8	2	3	6	11	22
0.6	4.1	2.8	2	2	3	5	10	4	2.7	2	2	3	5	9
0.8	2.9	1.5	2	2	2	4	6	2.8	1.4	2	2	2	3	6
1.0	2.4	0.9	2	2	2	2	4	2.3	0.8	2	2	2	2	4
1.2	2.2	0.6	2	2	2	2	4	2.1	0.5	2	2	2	2	3
	***2-of-3* sign (*UCL*=9)**							**X-bar (*UCL*=0.9267)**						
Shift	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**
0.0	370	361	20	105	251	505	1089	370	368.1	20	107	257	510	1103
0.2	28.2	26.4	3	9	20	38	81	80.1	79.9	5	23	56	111	240
0.4	8	6.4	2	3	6	10	21	20.9	20.3	2	6	15	29	61
0.6	4.2	2.6	2	2	3	5	10	6.9	6.4	1	2	5	9	20
0.8	3	1.4	2	2	3	3	6	3	2.4	1	1	2	4	8
1.0	2.5	0.9	2	2	2	3	4	1.7	1.1	1	1	1	2	4
1.2	2.3	0.6	2	2	2	2	3	1.2	0.5	1	1	1	1	2

Table 4.10: In-control and out-of-control characteristics of the run-length distributions of the one-sided *1-of-1* sign, the *2-of-2* sign, the *2-of-3* sign, the *1-of-1* SR and the *2-of-2* SR chart for the median under the Cauchy distribution

	***1-of-1* sign (*UCL*=10)**							***1-of-1* SR (*UCL*=53)**						
Shift	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**
0.0	370	370.9	19	106	259	516	1113	370	367.7	19	105	255	511	1101
0.2	18.2	17.7	1	6	13	25	53	15.6	15.1	1	5	11	21	46
0.4	5.4	4.9	1	2	4	7	15	4.5	4	1	2	3	6	12
0.6	3.2	2.6	1	1	2	4	8	2.7	2.1	1	1	2	4	7
0.8	2.4	1.8	1	1	2	3	6	2.1	1.5	1	1	2	3	5
1.0	2	1.4	1	1	2	3	5	1.7	1.1	1	1	1	2	4
1.2	1.8	1.2	1	1	1	2	4	1.6	1	1	1	1	2	3
	***2-of-2* sign (*UCL*=9)**							***2-of-2* SR (*UCL*=33)**						
Shift	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**
0.0	370	368.4	21	107	257	511	1103	370	370.4	20	108	258	512	1111
0.2	8.8	7.4	2	3	6	12	24	11.4	10.1	2	4	8	15	31
0.4	3.3	1.9	2	2	2	4	7	4.1	2.8	2	2	3	5	10
0.6	2.5	1	2	2	2	3	5	2.9	1.5	2	2	2	4	6
0.8	2.2	0.7	2	2	2	2	4	2.5	1	2	2	2	3	5
1.0	2.1	0.5	2	2	2	2	3	2.3	0.8	2	2	2	2	4
1.2	2.1	0.4	2	2	2	2	3	2.2	0.6	2	2	2	2	4
	***2-of-3* sign (*UCL*=9)**							**X-bar (*UCL* = 30.6802)**						
Shift	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**
0.0	370	360.5	21	106	252	503	1085	370	368	19	107	258	513	1101
0.2	8.7	7.1	2	4	7	11	23	367	366.5	19	106	255	507	1096
0.4	3.4	1.8	2	2	3	4	7	367	367.2	19	106	254	507	1097
0.6	2.6	1	2	2	2	3	5	364	365.5	19	105	251	504	1090
0.8	2.3	0.7	2	2	2	3	4	361	361.7	19	104	248	500	1084
1.0	2.2	0.5	2	2	2	2	3	360	360	19	104	249	499	1077
1.2	2.2	0.4	2	2	2	2	3	355	353.8	19	103	247	492	1066

Table 4.11 summarizes our findings from Tables 4.8, 4.9 and 4.10 and ranks the charts (from the most to the least favorable) under each of the three distributions. The ranking was based primarily on their ARL (the current norm in the SPC literature), but since the run-length distributions are right (positive) skewed, we also looked at the median run-length ($MDRL$), the first and third quartiles (i.e. Q_1 and Q_3), as well as the 5^{th} and the 95^{th} percentiles.

Table 4.11: Ranking (from most to least favorable) of the one-sided nonparametric charts for the median under the normal, the double exponential and the Cauchy distributions based on out-of-control ARL and run-length percentiles. The $ARL_0 = 370$

Normal	Double Exponential	Cauchy
2-of-2 SR	*2-of-2* sign / *2-of-3* sign	*2-of-2* sign / *2-of-3* sign
2-of-2 sign / *2-of-3* sign	*2-of-2* SR	*2-of-2* SR
1-of-1 $\overline{X}$	*1-of-1* SR	*1-of-1* SR
1-of-1 SR	*1-of-1* sign	*1-of-1* sign
1-of-1 sign	*1-of-1* $\overline{X}$	*1-of-1* $\overline{X}$

Overall, it is concluded that the proposed sign charts

(i) have substantially better out-of-control performance (i.e. shorter ARL values) than the *1-of-1* sign chart of Amin et al. (1995),

(ii) compete well with the SR charts of Bakir (2004) and Chakraborti and Eryilmaz (2007), and

(iii) outperform the Shewhart $\overline{X}$ chart in case of the heavier tailed distributions.

4.2.6 Design of the two-sided *2-of-2* DR, the *2-of-2* KL and the *2-of-3* sign charts

The characteristics of the in-control run-length distribution are typically used in the design and/or the implementation of a chart. As noted before, the ARL_0 should be high so that the time and/or effort spent on searching for nonexistent out-of-control conditions is not wasted.

Tables 4.12 and 4.13 display the ARL_0 and the FAR values of the two-sided *1-of-1*, *2-of-2* DR, *2-of-2* KL and *2-of-3* sign charts, respectively. For simplicity we only consider symmetrically placed control limits for the median i.e. $LCL = a$ and $UCL = n - a$, so that $p_0^+ = \Pr(T_i \geq UCL \mid IC) = \Pr(T_i \leq LCL \mid IC) = p_0^-$. Asymmetric control limits may of course be used when necessary, say for monitoring percentiles other than the median.

To attain the desired ARL_0 and/or FAR (for any one of the four charting procedures) the practitioner may use Tables 4.12 and 4.13 to select the suitable charting constant a (hence the control limits) for the sample size n at hand. Note that, as pointed out by Amin et al. (1995), the largest possible in-control ARL for the two-sided *1-of-1* sign chart is 2^{n-1} when $p = 0.5$, and thus unless n is sufficiently large, it is not possible to get close (even approximately) to an ARL_0 such as 370; this makes the *1-of-1* charts somewhat unattractive from a practical point of view.

However, for any combination of n and a values the ARL_0 (or FAR) values of the *2-of-2* DR, the *2-of-2* KL and the *2-of-3* sign charts are higher (or smaller) than that of the *1-of-1* sign chart.

For example, if $n = 5$ and $a = 0$ the $LCL = 0$ and the $UCL = 5$; the *1-of-1* sign chart has an ARL_0 of 16.00 (with a FAR of 0.06250), whereas both the *2-of-2* DR and the *2-of-2* KL charts have much higher ARL_0 values, 272.00 and 528.00, respectively (and much smaller FAR values, 0.00391 and 0.00195, respectively). Therefore, the new two-sided sign charts with signaling rules are more useful to the practitioner.

Table 4.12: The in-control characteristics (*ARL* and *FAR*) of the two-sided *1-of-1*, the *2-of-2* DR, the *2-of-2* KL and the *2-of-3* sign charts for the median (for samples of size $n = 4(1)14$)*

Sample size	*LCL*	*UCL*	*1-of-1*		*2-of-2* DR		*2-of-2* KL		*2-of-3*	
n	***a***	***n-a***	ARL_0	*FAR*	ARL_0	*FAR*	ARL_0	*FAR*	ARL_0	*FAR*
4	**0**	**4**	8.00	0.12500	72.00	0.01563	136.00	0.00781	79.37	0.01367
	1	**3**					6.72	0.19531	8.74	0.14648
5	**0**	**5**	16.00	0.06250	272.00	0.00391	528.00	0.00195	285.27	0.00366
	1	**4**			9.78	0.14063	16.89	0.07031	13.75	0.08789
6	**0**	**6**	32.00	0.03125	1056.00	0.00098	2080.00	0.00049	1081.23	0.00095
	1	**5**			25.47	0.04785	46.37	0.02393	30.45	0.03738
	2	**4**					5.69	0.23633	8.75	0.14771
7	**0**	**7**	64.00	0.01563	4160.00	0.00024	8256.00	0.00012	4209.21	0.00024
	1	**6**	8.00	0.12500	72.00	0.01563	136.00	0.00781	79.37	0.01367
	2	**5**			7.08	0.20532	11.95	0.10266	11.01	0.11229
8	**0**	**8**	128.00	0.00781						
	1	**7**	14.22	0.07031	216.49	0.00494	418.77	0.00247	228.45	0.00460
	2	**6**			15.43	0.08356	27.40	0.04178	19.74	0.05940
	3	**5**					5.16	0.26395	9.00	0.14435
9	**0**	**9**	256.00	0.00391						
	1	**8**	25.60	0.03906	680.96	0.00153	1336.32	0.00076	701.4	0.00147
	2	**7**	5.57	0.17969	36.54	0.03229	67.51	0.01614	42.18	0.02649
	3	**6**			5.85	0.25787	9.72	0.12894	9.87	0.12692
10	**0**	**10**	512.00	0.00195						
	1	**9**	46.55	0.02148	2213.02	0.00046	4379.50	0.00023	2249.15	0.00045
	2	**8**	9.14	0.10938	92.73	0.01196	176.33	0.00598	100.94	0.01065
	3	**7**			11.37	0.11816	19.83	0.05908	15.43	0.07755
	4	**6**							9.32	0.13987
11	**0**	**11**	1024.00	0.00098						
	1	**10**	85.33	0.01172	7367.11	0.00014			7432.31	0.00014
	2	**9**	15.28	0.06543	248.87	0.00428	482.46	0.00214	261.61	0.00400
	3	**8**	4.41	0.22656	23.90	0.05133	43.38	0.02567	28.78	0.03970
	4	**7**			5.14	0.30121	8.46	0.15061	9.29	0.13590
12	**0**	**12**	2048.00	0.00049						
	1	**11**	157.54	0.00635						
	2	**10**	25.92	0.03857	697.98	0.00149	1370.04	0.00074	718.66	0.00143
	3	**9**	6.85	0.14600	53.77	0.02131	100.68	0.01066	60.31	0.01820
	4	**8**			9.23	0.15031	15.89	0.07515	13.18	0.09203
									9.66	0.13529
13	**0**	**13**	4096.00	0.00024						
	1	**12**	292.57	0.00342						
	2	**11**	44.52	0.02246	2026.71	0.00050	4008.89	0.00025	2061.32	0.00049
	3	**10**	10.84	0.09229	128.25	0.00852	245.67	0.00426	137.7	0.00773
	4	**9**			17.79	0.07121	31.83	0.03560	22.26	0.05221
	5	**8**					7.64	0.16881	8.99	0.14145
14	**0**	**14**	8192.00	0.00012						
	1	**13**	546.13	0.00183						
	2	**12**	77.28	0.01294	6049.95	0.00017			6109.11	0.00017
	3	**11**	17.43	0.05737	321.23	0.00329	625.02	0.00165	335.57	0.00310
	4	**10**	5.57	0.17957	36.58	0.03224	67.60	0.01612	42.23	0.02645
	5	**9**			7.92	0.17973	13.49	0.08987	11.84	0.10354
	6	**8**							10.00	0.13091

**Note*: Only ARL_0 values greater than 5 and less than 10 000 are shown.

Table 4.13: The in-control characteristics (*ARL* and *FAR*) of the two-sided *1-of-1*, the *2-of-2* DR, the *2-of-2* KL and the *2-of-3* sign charts for the median (for samples of size $n = 15(5)25$)[*]

Sample size	*LCL*	*UCL*	*1-of-1*		*2-of-2* DR		*2-of-2* KL		*2-of-3*	
n	***a***	***n-a***	ARL_0	***FAR***	ARL_0	***FAR***	ARL_0	***FAR***	ARL_0	***FAR***
15	**1**	**14**	1024.00	0.00098						
	2	**13**	135.40	0.00739						
	3	**12**	28.44	0.03516	837.53	0.00124	1646.62	0.00062	860.1	0.00119
	4	**11**	8.44	0.11847	79.69	0.01403	150.94	0.00702	87.38	0.01237
	5	**10**			14.30	0.09106	25.28	0.04553	18.53	0.06358
	6	**9**					7.07	0.18437	8.82	0.14483
20	**2**	**18**	2484.78	0.00040						
	3	**17**	388.07	0.00258						
	4	**16**	84.62	0.01182	7244.68	0.00014			7309.35	0.00014
	5	**15**	24.16	0.04139	607.90	0.00171	1191.64	0.00086	627.27	0.00164
	6	**14**	8.67	0.11532	83.87	0.01330	159.07	0.00665	91.73	0.01176
	7	**13**			18.24	0.06926	32.68	0.03463	22.74	0.05103
	8	**12**			5.93	0.25346	9.88	0.12673	9.94	0.12586
25	**3**	**22**	6388.89	0.00016						
	4	**21**	1098.27	0.00091						
	5	**20**	245.26	0.00408						
	6	**19**	68.34	0.01463	4738.32	0.00021	9408.31	0.00011	4790.78	0.00021
	7	**18**	23.10	0.04329	556.83	0.00187	1090.56	0.00094	575.4	0.00179
	8	**17**	9.28	0.10775	95.41	0.01161	181.54	0.00581	103.71	0.01036
	9	**16**			23.34	0.05268	42.32	0.02634	28.18	0.04059
	10	**15**			7.91	0.18008	13.46	0.09004	11.83	0.10366
	11	**14**					5.65	0.23808	8.76	0.14759

[*]*Note*: Only ARL_0 values greater than 5 and less than 10 000 are shown.

4.2.7 Performance comparison of the two-sided sign charts

The out-of-control performance of the two-sided sign control charts were compared amongst one another and with that of the two-sided SR charts under the normal, the double exponential and the Cauchy distributions; again we included the Shewhart $\overline{X}$ chart for completeness. The design parameters of the charts were chosen (coupled with randomization) so that the in-control *ARL* values were all equal to 370. As for the one-sided charts various characteristics of the run-length distributions were obtained using simulations and shown in Tables 4.14, 4.15 and 4.16, respectively with a summary of our findings given in Table 4.17.

Table 4.14: In-control and out-of-control properties of the run-length distributions of the two-sided *1-of-1* sign, the *2-of-2* DR sign, the *2-of-2* KL sign, the *2-of-3* sign, the *1-of-1* SR, *2-of-2* DR SR, *2-of-2* KL SR and the *1-of-1* X-bar charts under the Normal distribution

	***1-of-1* sign (*LCL*=0 , *UCL* = 10)**							***1-of-1* SR (*LCL* = -*UCL* = -55)**						
Shift	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**
0.0	370	369.78	19	106	255	507	1106	370	377.3	20	110	264	528	1138
0.2	175.00	173.96	10	51	122	242	523	170.00	169.03	9	49	118	237	506
0.4	56.42	55.87	3	17	39	78	167	51.81	51.31	3	15	36	72	154
0.6	21.57	21.14	2	7	15	30	64	19.15	18.55	1	6	14	26	56
0.8	9.82	9.33	1	3	7	13	29	8.60	8.14	1	3	6	12	25
1.0	5.23	4.71	1	2	4	7	15	4.61	4.09	1	2	3	6	13
1.2	3.24	2.70	1	1	2	4	9	2.85	2.29	1	1	2	4	7
	***2-of-2* sign DR (*LCL*=1 , *UCL*=9)**							***2-of-2* SR DR (*LCL* = -*UCL* = -39)**						
Shift	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**
0.0	370	363.53	20	106	253	501	1086	370	367.32	21	108	257	508	1103
0.2	166.30	164.54	10	49	116	230	496	129.10	127.36	8	38	90	179	385
0.4	43.75	42.3	4	14	31	60	128	27.42	26.04	3	9	19	38	80
0.6	14.68	13.25	2	5	11	20	41	8.77	7.42	2	3	6	12	24
0.8	6.83	5.47	2	3	5	9	18	4.25	2.92	2	2	3	5	10
1.0	4.08	2.71	2	2	3	5	10	2.82	1.40	2	2	2	3	6
1.2	2.97	1.54	2	2	2	4	6	2.29	0.74	2	2	2	2	4
	***2-of-2* sign KL (*LCL*=1, *UCL* = 9)**							***2-of-2* SR KL (*LCL* = -*UCL* = -37)**						
Shift	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**
0.0	370	364.24	21	107	256	512	1099	370	368.73	21	108	256	514	1102
0.2	113.10	111.56	7	34	79	156	337	86.09	85.13	6	26	60	119	257
0.4	28.28	26.81	3	9	20	39	82	18.69	17.48	2	6	13	25	53
0.6	10.42	9.09	2	4	8	14	29	6.68	5.34	2	3	5	9	17
0.8	5.29	3.92	2	2	4	7	13	3.60	2.23	2	2	3	4	8
1.0	3.44	2.06	2	2	3	4	8	2.57	1.11	2	2	2	3	5
1.2	2.65	1.17	2	2	2	3	5	2.18	0.58	2	2	2	2	4
	***2-of-3* sign (*LCL*=1 , *UCL*=9)**							***1-of-1* X-bar (*LCL* = -*UCL* = -0.94858)**						
Shift	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**
0.0	370	363.37	20	106	252	504	1088	370	368.45	19	106	256	510	1105
0.2	108.50	106.69	7	33	76	150	318	110.50	109.73	6	32	77	153	329
0.4	26.72	25.04	3	9	19	36	77	24.04	23.46	2	7	17	33	71
0.6	9.94	8.33	2	4	7	13	27	7.39	6.87	1	2	5	10	21
0.8	5.18	3.58	2	3	4	7	12	3.12	2.58	1	1	2	4	8
1.0	3.43	1.84	2	2	3	4	7	1.77	1.17	1	1	1	2	4
1.2	2.68	1.07	2	2	2	3	5	1.27	0.59	1	1	1	1	2

Table 4.15: In-control and out-of-control properties of the run-length distributions of the two-sided *1-of-1* sign, the *2-of-2* DR sign, the *2-of-2* KL sign, the *2-of-3* sign, the *1-of-1* SR, *2-of-2* DR SR, *2-of-2* KL SR and the *1-of-1* X-bar charts under the double exponential distribution

	***1-of-1* sign (*LCL*=0 , *UCL* = 10)**							***1-of-1* SR (*LCL* = -*UCL* = -55)**						
Shift	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**
0.0	370	369.58	20	107	255	515	1106	370	377.57	21	113	263	528	1132
0.2	91.20	90.31	5	27	63	126	272	88.55	88.69	5	25	61	123	263
0.4	24.44	23.92	2	7	17	34	72	22.9	22.20	2	7	16	32	67.5
0.6	10.04	9.49	1	3	7	14	29	9.19	8.73	1	3	6	13	27
0.8	5.41	4.88	1	2	4	7	15	4.85	4.26	1	2	3	7	13
1.0	3.47	2.92	1	1	3	5	9	3.06	2.48	1	1	2	4	8
1.2	2.52	1.95	1	1	2	3	6	2.29	1.73	1	1	2	3	6
	***2-of-2* sign DR (*LCL*=1 , *UCL*=9)**							***2-of-2* SR DR (*LCL* = -*UCL* = -39)**						
Shift	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**
0.0	370	361.71	20	106	253	503	1086	370	370.37	20	107	257	511	1106
0.2	77.17	75.93	5	23	54	106	229	78.94	77.11	5	24	55	109	233
0.4	16.89	15.57	2	6	12	23	48	15.76	14.52	2	5	11	21	45
0.6	6.93	5.56	2	3	5	9	18	6.17	4.85	2	3	5	8	16
0.8	4.17	2.79	2	2	3	5	10	3.65	2.30	2	2	3	4	8
1.0	3.09	1.67	2	2	2	4	6	2.76	1.33	2	2	2	3	6
1.2	2.59	1.11	2	2	2	3	5	2.36	0.85	2	2	2	2	4
	***2-of-2* sign KL (*LCL*=1, *UCL* = 9)**							***2-of-2* SR KL (*LCL* = -*UCL* = -37)**						
Shift	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**
0.0	370	372.82	20	108	260	512	1114	370	368.05	21	107	257	513	1103
0.2	48.75	47.51	4	15	34	66	143	51.43	49.96	4	16	36	71	151
0.4	11.73	10.26	2	4	9	16	32	11.18	9.82	2	4	8	15	31
0.6	5.45	4.01	2	2	4	7	13.5	4.89	3.56	2	2	4	6	12
0.8	3.55	2.16	2	2	3	4	8	3.13	1.76	2	2	2	4	7
1.0	2.76	1.28	2	2	2	3	5	2.49	1.02	2	2	2	3	5
1.2	2.40	0.86	2	2	2	2	4	2.22	0.63	2	2	2	2	4
	***2-of-3* sign (*LCL*=1 , *UCL*=9)**							***1-of-1* X-bar (*LCL* = -*UCL* = -1.011335)**						
Shift	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**
0	370	363.27	20	106	253	504	1089	370	342.30	23	141	258	504	1058
0.2	46.66	44.73	4	15	33	64	136	159.80	180.02	5.5	46	107	210	416
0.4	11.21	9.55	2	4	8	15	30	42.08	36.77	2.5	13	29	64.5	115
0.6	5.26	3.68	2	3	4	7	13	10.44	8.83	1	3	8	16	24.5
0.8	3.49	1.9	2	2	3	4	7	4.02	3.65	1	2	3	6	9.5
1.0	2.77	1.16	2	2	2	3	5	2.01	1.59	1	1	1	2	5.5
1.2	2.42	0.77	2	2	2	3	4	1.44	0.74	1	1	1	2	3

Table 4.16: In-control and out-of-control properties of the run-length distributions of the two-sided 1-of-1 sign, the *2-of-2* DR sign, the *2-of-2* KL sign, the *2-of-3* sign, the *1-of-1* SR, *2-of-2* DR SR, *2-of-2* KL SR and the *1-of-1* X-bar charts under the Cauchy distribution

	***1-of-1* sign (*LCL*=0 , *UCL* = 10)**							***1-of-1* SR (*LCL* = -*UCL* = -55)**						
Shift	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**
0.0	370	369.19	19	107	255	513	1111	370	381.92	20	113	268	525	1143
0.2	27.06	26.62	2	8	19	37	80	25.38	24.72	2	8	18	35	75
0.4	7.00	6.43	1	2	5	10	20	6.55	6.06	1	2	5	9	18
0.6	3.83	3.31	1	1	3	5	10	3.61	3.08	1	1	3	5	10
0.8	2.76	2.21	1	1	2	4	7	2.59	1.99	1	1	2	3	7
1.0	2.25	1.68	1	1	2	3	6	2.13	1.54	1	1	2	3	5
1.2	1.97	1.38	1	1	1	2	5	1.88	1.30	1	1	1	2	5
	***2-of-2* sign DR (*LCL*=1 , *UCL*=9)**							***2-of-2* SR DR (*LCL* = -*UCL* = -39)**						
Shift	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**
0.0	370	361.28	20	106	252	504	1089	370	368.00	20	107	257	512	1099
0.2	18.92	17.59	2	6	14	26	54	24.99	23.54	3	8	18	34	72
0.4	5.08	3.70	2	2	4	7	12	7.28	5.98	2	3	5	10	19
0.6	3.29	1.87	2	2	2	4	7	4.66	3.34	2	2	4	6	11
0.8	2.71	1.25	2	2	2	3	5	3.77	2.42	2	2	3	5	9
1.0	2.46	0.95	2	2	2	3	4	3.33	1.97	2	2	2	4	7
1.2	2.32	0.76	2	2	2	2	4	3.07	1.69	2	2	2	4	6
	***2-of-2* sign KL (*LCL*=1, *UCL* = 9)**							***2-of-2* SR KL (*LCL* = -*UCL* = -37)**						
Shift	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**
0.0	370	363.78	21	108	257	500	1085	370	369.31	20	107	257	511	1099
0.2	13.04	11.65	2	5	9	18	36	16.61	15.22	2	6	12	22	47
0.4	4.14	2.79	2	2	3	5	10	5.34	4.00	2	2	4	7	13
0.6	2.88	1.45	2	2	2	3	6	3.55	2.20	2	2	3	4	8
0.8	2.46	0.94	2	2	2	3	4	2.96	1.54	2	2	2	4	6
1.0	2.30	0.72	2	2	2	2	4	2.69	1.24	2	2	2	3	5
1.2	2.21	0.58	2	2	2	2	4	2.53	1.04	2	2	2	3	5
	***2-of-3* sign (*LCL*=1 , *UCL*=9)**							***1-of-1* X-bar (*LCL* = -*UCL* = -61.36038)**						
Shift	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**	***ARL***	***SDRL***	**5^{th}**	**Q_1**	***MDRL***	**Q_3**	**95^{th}**
0.0	370	361.35	20	106	253	507	1084	370	368.97	19	106	257	510	1103
0.2	12.34	10.65	2	5	9	16	34	371.00	371.53	20	107	257	515	1109
0.4	4.09	2.51	2	2	3	5	9	369.2	368.89	20	107	257	510	1103
0.6	2.90	1.30	2	2	2	3	6	373.1	372.22	19	107	259	517	1116
0.8	2.50	0.86	2	2	2	3	4	369.8	368.44	20	107	258	513	1103
1.0	2.33	0.66	2	2	2	3	4	370.3	371.86	19	106	256	514	1110
1.2	2.23	0.53	2	2	2	2	3	372.7	371.80	20	107	258	517	1114

Table 4.17: Ranking (from most to least favorable) of the two-sided nonparametric charts under the normal, the double exponential and the Cauchy distributions based on out-of-control *ARL* and run-length percentiles. The ARL_0 = 370

Normal	Double Exponential	Cauchy
2-of-2 KL SR	*2-of-2* KL sign / *2-of-3* sign	*2-of-2* KL sign / *2-of-3* sign
2-of-2 KL sign / *2-of-3* sign	*2-of-2* KL SR	*2-of-2* KL SR
1-of-1 $\overline{X}$	*2-of-2* DR sign	*2-of-2* DR sign
2-of-2 DR SR	*2-of-2* DR SR	*2-of-2* DR SR
2-of-2 DR sign	*1-of-1* SR	*1-of-1* SR
1-of-1 SR	*1-of-1* sign	*1-of-1* sign
1-of-1 sign	*1-of-1* $\overline{X}$	*1-of-1* $\overline{X}$

In general, we observe that:

(i) the *2-of-2* DR sign, the *2-of-2* KL sign and the *2-of-3* sign charts all outperform the original *1-of-1* sign chart under all three the distributions, and

(ii) the *2-of-2* KL sign chart and the *2-of-3* sign chart are best overall; only outperformed by the *2-of-2* KL SR chart in case of the normal distribution. (Note: the *2-of-2* KL charts generally outperform the *2-of-2* DR charts; whether the chart is based on the sign test or the SR test).

More specifically, we note that:

(i) under the normal distribution, the two-sided *2-of-2* KL SR chart performs the best (this was also the case for the one-sided charts), but the *2-of-2* KL sign and the *2-of-3* sign charts are good/close competitors, whereas

(ii) under the double exponential distribution and the Cauchy distribution:

(a) the *2-of-2* KL sign and the *2-of-3* sign charts are the top performers,

(b) the sign charts generally perform better than the SR charts except in case of the *1-of-1* chart (i.e. the *1-of-1* SR chart is better than the *1-of-1* sign chart),

(c) the *2-of-2* KL sign and the *2-of-3* sign charts are both better than the *2-of-2* KL SR chart, and

(d) the *2-of-2* DR sign chart is better than the *2-of-2* DR SR chart.

4.3 Precedence charts for the unknown π^{th} quantile (Case U)

Introduction

Case U is the scenario when the π^{th} percentile of the process distribution is unknown or unspecified; this is unlike Case K and as a consequence the control limits are unknown.

To estimate the control limits a reference sample is obtained; this reference sample is also called the preliminary sample or the calibration sample or the Phase I sample. Once the control limits are estimated, Phase II starts. In Phase II the estimated control limits are used for future monitoring of the process using new incoming samples taken sequentially from the process; this is the prospective monitoring phase.

The new control charts we consider here, in Case U, are based on the median test, which is essentially a modified sign test for two independent samples and is a member of a more general class of nonparametric two-sample tests referred to as precedence tests or precedence statistics (see e.g. Gibbons and Chakraborti, (2003)). We therefore refer to the charts of Case U as precedence charts.

Assumptions

We assume that

(i) the reference sample $(X_1, X_2, ..., X_m)$ is a random sample of size m available from an in-control (IC) distribution with an unknown continuous cumulative distribution function (c.d.f) $F_X(x) = F(x-\theta)$ where θ is the location parameter and F is some continuous c.d.f with median zero,

(ii) each Phase II test sample $(Y_{i1}, Y_{i2}, ..., Y_{in})$ taken at sampling stage (time) $i = 1,2,3,...$ is a random sample (rational subgroup) of size $n > 1$ from an unknown continuous distribution with c.d.f $G_Y(y) = F(y-\theta_i)$ where θ_i is the location parameter of the i^{th} test sample, and

(iii) the Phase II test samples are drawn sequentially and independently of one another and of the reference sample.

Charting statistics and control limits

The control limits, as mentioned before, are estimated from the Phase I reference sample and then used for prospective monitoring of the process. In Phase II, one charting statistic is calculated from each new incoming sample and then compared to the estimated control limits.

The estimated control limits are found by arranging the Phase I observations in ascending order, that is,

$$X_{1:m} < X_{2:m} < \ldots < X_{m:m}$$

where $X_{j:m}$ denotes the j^{th} order statistic of the reference sample of size m, and selecting two order statistics $X_{a:m}$ and $X_{b:m}$ (for a given $1 \le a < b \le m$) so that the estimated control limits for the two-sided precedence charts are given by

$$\hat{LCL} = X_{a:m} \qquad \text{and} \qquad \hat{UCL} = X_{b:m} \tag{4-28}$$

respectively, where a and b are labeled the charting constants; determination of the charting constants will be discussed later. Note that, like the sign charts of Case K, the precedence charts do not have a centerline.

The charting statistic at time $i = 1,2,3,\ldots$ is an order statistic $Y_{j:n}^{i}$ for $1 < j \le n$ from each of the Phase II test samples.

The operation and the signaling rules (i.e. when a process is declared OOC) of the runs-rule enhanced precedence charts is similar to that of the sign charts; however, instead of comparing T_i (the sign statistic) with the known control limits UCL and LCL of (4-7) we now compare $Y_{j:n}^{i}$ with the estimated control limits $\hat{UCL}$ and $\hat{LCL}$ of (4-28) at each sampling stage $i = 1,2,3,\ldots$.

Remark 11

(i) The median is a robust and flexible estimator of location in the sense that it is preferred in situations where large measurement errors are expected and is applicable in more diverse situations (unlike the mean). Thus, although we develop and discuss the theory of the *1-of-1* and the runs-rule enhanced precedence charts so that any order statistic can be used as charting statistic, the median is a popular choice in practice and we therefore focus mainly on the median chart, that is, the case where the charting statistic is taken as the test sample median.

Furthermore, to simplify matters we assume that the sample size $n = 2s+1$ is odd so that the median of the test sample $Y^i_{j:n}$ is uniquely defined with $j = s+1$. Thus, for example, when the subgroup size n is equal to 5, as is fairly common in SPC applications, the charting statistic is the 3rd smallest value in the test sample.

(ii) Only two-sided precedence charts are studied. The required modifications for the one-sided precedence charts are simple and briefly indicated in section 4.3.3.

(iii) The proposed precedence charts do not signal unless the charting statistic $Y^i_{j:n}$ is less than or equal to the estimated lower control limit $X_{a:m}$ or is greater than or equal to the estimated upper control limit $X_{b:m}$. Although this is theoretically negligible as the underling process distributions are assumed to be continuous, in practice, one needs to apply the charts in a correct manner as ties might occur when it is found that the Y order statistic (i.e. charting statistic) is equal to one of the control limits.

(iv) The precedence charts can be applied as soon as the necessary order statistic is available and can be a practical advantage in some applications. We comment more on this point later.

4.3.1 Run-length distributions of the two-sided precedence charts

The run-length distributions and the statistical characteristics of the precedence charts (such as the ARL, $VARL$ etc.) are required to design the charts and reveal important information regarding their performance.

We again use a Markov chain approach to derive the run-length distributions and in some cases draw on the results of the geometric distribution of order k to obtain closed form expressions.

Even though the operation (i.e. the signaling events, when a process is declared OOC etc.) of the runs-rule enhanced precedence charts of Case U are similar to that of the runs-rule enhanced sign charts of Case K, there is a fundamental difference in deriving the run-length distributions of the precedence charts compared to that of the sign charts.

In particular, because the control limits are estimated they are random variables (as indicated by the ^ - notation in (4-28)) and, consequently, the signaling indicators of the (two-sided) precedence charts i.e.

$$\hat{\xi}_i = \begin{cases} 1 & \text{if} \quad Y_{j:n}^i \geq U\hat{C}L \\ 0 & \text{if} \quad L\hat{C}L < Y_{j:n}^i < U\hat{C}L \\ 2 & \text{if} \quad Y_{j:n}^i \leq L\hat{C}L \end{cases} \tag{4-29}$$

and

$$\hat{\xi}_i^{\pm} = I(Y_{j:n}^i \notin (L\hat{C}L, U\hat{C}L)) = \begin{cases} 1 & \text{if} \quad Y_{j:n}^i \notin (L\hat{C}L, U\hat{C}L) \\ 0 & \text{if} \quad Y_{j:n}^i \in (L\hat{C}L, U\hat{C}L) \end{cases} \tag{4-30}$$

for $i = 1,2,3,...$ are dependent tri-variate (or binary) random variables.

The design, analysis and performance of the charts must therefore take account of the additional variability introduced as a result of estimating the control limits; this is the main stumbling block in calculating the run-length distributions, here, in Case U, particularly for the charts that use signaling rules (ii) and (iii) defined in the beginning of section 4.1 on page 258. Like in Chapter 3 we use a two-step approach to derive the run-length distribution which involves the method of conditioning (see e.g. Chakraborti, (2000)).

First we derive the *conditional* run-length distributions i.e. conditioned on the two order statistics (control limits), which lets us focus on specific values of the control limits. The performance of the charts as measured by their *conditional* run-length distributions are therefore different for each user as each user has his/her own control limits based on his/her own Phase I data (sample).

Second we derive the *unconditional* (or marginal) run-length distributions by averaging over the joint distribution of the two order statistics. The *unconditional* run-length distributions reflect the bigger picture or the overall performance of the charts and take into account that the control limits are estimated. The performance of the charts as measured by their *unconditional* run-length distributions are therefore the same for each user.

Signaling probabilities

The key ingredients to the *conditional* run-length distributions are

(i) the one-step transition probabilities $p_{i,k}$ and

(ii) the success probability α.

The one-step transition probabilities are the elements of the transition probability matrix and are required in case one uses the Markov chain approach. The success probability, on the other hand, is a prerequisite if one wishes to use the properties of the geometric distribution of order k as it is a parameter of the distribution (see e.g. expressions (4-16) and (4-17)).

The one-step transition probabilities and the success probability all depend on and are functions of the conditional probability of a signal i.e. the probability for a charting statistic to plot OOC given that (or conditionally on having observed) $X_{a:m} = x_{a:m}$ and $X_{b:m} = x_{b:m}$, which is given by

$$\begin{aligned} p_C^{\pm}(X_{a:m}, X_{b:m}, F, G) &= \Pr(\hat{\xi}_i^{\pm} = 1 \mid \hat{LCL} = x_{a:m}, \hat{UCL} = x_{b:m}) \\ &= 1 - \Pr(X_{a:m} < Y_{j:n}^i < X_{b:m} \mid X_{a:m} = x_{a:m}, X_{b:m} = x_{b:m}) \\ &= 1 - G_j(X_{b:m}) + G_j(X_{a:m}) \end{aligned} \qquad (4\text{-}31)$$

where G_j denotes the c.d.f of the j^{th} order statistic in a sample of size n from a distribution with c.d.f G and the subscript "C" in $p_C^{\pm}(X_{a:m}, X_{b:m}, F, G)$ indicates that (4-31) is a conditional probability.

Using the probability integral transformation (PIT) and the fact that the j^{th} order statistic from a *uniform*(0,1) distribution follows a beta distribution with parameters j and $n - j + 1$ (see e.g. Gibbons and Chakraborti, (2003)) it follows, for example, that

$$\begin{aligned} G_j(X_{a:m}) &= \Pr(Y_{j:n}^i \le X_{a:m} \mid X_{a:m} = x_{a:m}) \\ &= \Pr(G(Y_{j:n}^i) \le G(X_{a:m}) \mid X_{a:m} = x_{a:m}) \\ &= \Pr(U_{j:n}^i \le G(X_{a:m})) = I_{G(X_{a:m})}(j, n - j + 1) \end{aligned}$$

where $U_{j:n}^i$ is the j^{th} order statistic from a *uniform*(0,1) distribution and

$$I_p(u,v) = [\beta(u,v)]^{-1} \int_0^p w^{u-1}(1-w)^{v-1} dw \quad \text{for} \quad u, v > 0$$

is the c.d.f of the $Beta(u,v)$ distribution, also known as the incomplete beta function.

Thus, the conditional probability of a signal in (4-31) can be expressed as

$$p_C^{\pm}(X_{a:m}, X_{b:m}, F, G) = 1 - I_{G(X_{b:m})}(j, n-j+1) + I_{G(X_{a:m})}(j, n-j+1). \quad (4\text{-}32)$$

With the conditional probability of a signal in (4-32) we can without difficulty find the *conditional* run-length distributions of the two-sided *1-of-1*, the *2-of-2* DR, the *2-of-2* KL and the *2-of-3* precedence charts. The *unconditional* run-length distributions, in general, follow straightforwardly from the *conditional* run-length distributions.

Remark 12

(i) We denote, without loss of generality, the two order statistics $(X_{a:m}, X_{b:m})$ by $\mathbf{Z} = (X, Y)$ and their observed values $(x_{a:m}, x_{b:m})$ by $\mathbf{z} = (x, y)$. Thus when writing $\mathbf{Z} = \mathbf{z}$ it means $(X_{a:m}, X_{b:m}) = (x_{a:m}, x_{b:m})$ or $(X, Y) = (x, y)$.

In particular, this notation permits us to write (4-32) as

$$p_C^{\pm}(X, Y, F, G) = 1 - I_{G(Y)}(j, n-j+1) + I_{G(X)}(j, n-j+1). \tag{4-33}$$

(ii) It is instructive to compare the signaling probability of the two-sided sign chart of Case K with that of the two-sided precedence chart of Case U.

Specifically, we note that by substituting $G(X)$ and $G(Y)$ for p, replacing a with $j-1$ and swapping b for $n-j$ in (4-13) we obtain (4-33).

(iii) Because $\{\hat{\xi}_i^{\pm} = 1\}$ if and only if $\{\hat{\xi}_i = 1\} \cup \{\hat{\xi}_i = 2\}$ we can re-express the conditional probability of a signal of the two-sided precedence chart in terms of that of the upper and the lower one-sided charts i.e.

$$p_C^{\pm}(X, Y, F, G) = p_C^{-}(X, F, G) + p_C^{+}(Y, F, G) \tag{4-34}$$

where

$$p_C^{-}(X, F, G) = \Pr(\hat{\xi}_i = 2 \mid X_{a:m} = x_{a:m}) = I_{G(X)}(j, n-j+1) \tag{4-35}$$

and

$$p_C^{+}(Y, F, G) = \Pr(\hat{\xi}_i = 1 \mid X_{b:m} = x_{b:m}) = 1 - I_{G(Y)}(j, n-j+1). \tag{4-36}$$

Expression (4-34) will be particularly useful when deriving the run-length distributions of the two-sided precedence charts via the Markov chain approach.

(iv) For notational simplicity and brevity we denote $p_C^{\pm}(X, Y, F, G)$, $p_C^{-}(X, F, G)$ and $p_C^{+}(Y, F, G)$ simply by $p_C^{\pm}$, p_C^{-} and p_C^{+}, respectively.

4.3.1.1 Distribution of N_{1of1}: Run-length distribution for the *1-of-1* precedence chart

The two-sided *1-of-1* precedence chart was studied in detail by Chakraborti et al. (2004); this chart is called the "basic" precedence chart. The authors derived explicit formulae for both the conditional and the unconditional run-length distributions and their associated statistical characteristics (such as the ARL, $VARL$, FAR etc.) by applying, amongst numerous other techniques, results of the geometric distribution of order $k=1$ coupled with the method of conditioning (expectation by conditioning); doing so they have taken proper account of the dependency between the Phase II signaling events. In the paragraphs that follow, we simply review the most important statistical characteristics of the *1-of-1* precedence chart; for complete details on the derivations of the results, see the original article by Chakraborti et al. (2004).

Conditional run-length distribution

In particular, Chakraborti et al. (2004) showed that given $\mathbf{Z}=\mathbf{z}$ the *conditional* distribution of the run-length N_{1of1} is geometric with parameter (success probability) $p_C^{\pm}=p_C^{\pm}(X,Y,F,G)$. Accordingly, all properties and characteristics of the conditional run-length distribution follow conveniently from the properties of the geometric distribution of order $k=1$.

For example, the conditional p.m.f of N_{1of1} is

$$\Pr(N_{1of1}=t\mid\mathbf{Z}=\mathbf{z})=(1-p_C^{\pm})^{t-1}p_C^{\pm}\quad\text{for}\quad t=1,2,3,...$$

whereas the conditional average run-length ($CARL$) and the conditional variance of the run-length ($CVARL$) are given by

$$CARL_{1of1}=E(N_{1of1}\mid\mathbf{Z}=\mathbf{z})=1/p_C^{\pm}\quad\text{and}\quad CVARL_{1of1}=\text{var}(N_{1of1}\mid\mathbf{Z}=\mathbf{z})=(1-p_C^{\pm})/(p_C^{\pm})^2$$

respectively.

The conditional false alarm rate also follows straightforwardly as it is found by substituting $F=G$ in $p_C^{\pm}$, that is,

$$CFAR_{1of1}=p_C^{\pm}(X,Y,F,F)=1-I_{F(Y)}(j,n-j+1)+I_{F(X)}(j,n-j+1).$$

Unconditional run-length distribution

Most importantly Chakraborti et al. (2004) showed that by averaging over the joint distribution of $\mathbf{Z}$ one obtains the *unconditional* or marginal run-length distribution and its associated characteristics.

The unconditional p.m.f of N_{1of1}, in particular, is given by

$$\Pr(N_{1of1}=t)=E_{\mathbf{Z}}(\Pr(N_{1of1}=t\mid\mathbf{Z}))=E_{\mathbf{Z}}\left((q_C^{\pm})^{t-1}(1-q_C^{\pm})\right)=D^*(t-1)-D^*(t) \quad (4\text{-}37)$$

for $t=1,2,3,...$ and $D^*(0)=1$ where

$$D^*(t)=E_{\mathbf{Z}}\left((q_C^{\pm})^t\right)=\int_0^1\int_0^y\left(\frac{1}{\beta(j,n-j+1)}\sum_{h=0}^{n-j}\frac{(-1)^h}{j+h}\binom{n-j}{h}\left(GF^{-1}(y)^{j+h}-GF^{-1}(x)^{j+h}\right)\right)^t f_{a,b}(x,y)dxdy$$

with $q_C^{\pm}=1-p_C^{\pm}$ and where $f_{a,b}(x,y)$ denotes the joint p.d.f of the a^{th} and the b^{th} order statistics in a reference sample of size m from the *uniform*(0,1) distribution, given by

$$f_{a,b}(x,y)=\frac{m!}{(a-1)!(b-a-1)!(m-b)!}x^{a-1}(y-x)^{b-a-1}(1-y)^{m-b} \qquad 0<x<y<1.$$

Likewise, by writing the conditional ARL as

$$CARL_{1of1}=\sum_{t=0}^{\infty}\Pr(N_{1of1}>t\mid\mathbf{Z}=\mathbf{z})=\sum_{t=0}^{\infty}(q_C^{\pm})^t$$

the unconditional average run-length ($UARL_{1of1}$) follows by averaging over the joint distribution of $\mathbf{Z}$ and then simplifying i.e.

$$\begin{aligned}UARL_{1of1}&=E_{\mathbf{Z}}(CARL_{1of1})=\sum_{t=0}^{\infty}E_{\mathbf{Z}}\left((q_C^{\pm})^t\right)=\sum_{t=0}^{\infty}D^*(t)\\&=\int_0^1\int_0^y\left(1-\frac{1}{\beta(j,n-j+1)}\sum_{h=0}^{n-j}\frac{(-1)^h}{j+h}\binom{n-j}{h}\left(GF^{-1}(y)^{j+h}-GF^{-1}(x)^{j+h}\right)\right)^{-1}f_{a,b}(x,y)dxdy.\end{aligned} \quad (4\text{-}38)$$

The unconditional probability of a signal follows in the same manner and is given by

$$p^{\pm}(F,G) = \Pr(\hat{\xi}_i^{\pm} = 1) = E_{\mathbf{Z}}(\Pr(\hat{\xi}_i^{\pm} = 1 \mid \mathbf{Z})) = E_{\mathbf{Z}}(p_C^{\pm}(x,y,F,G))$$
$$= \int_0^1 \int_0^y (1 - I_{GF^{-1}(y)}(j, n-j+1) + I_{GF^{-1}(x)}(j, n-j+1))\, f_{a,b}(x,y)dxdy. \quad (4\text{-}39)$$

For complete details on the derivation of expressions (4-37), (4-38) and (4-39) see Chakraborti et al. (2004).

Remark 13

Chakraborti et al. (2004) noted that, from equations (4-37), (4-38) and (4-39), "… it is evident that in general the run length distribution depends on the distribution functions F and G through the composite function $\psi = GF^{-1}$. For example, when $F = G$, the process is in control, so $\psi(u) = u$, and the in-control run length distribution follows …".

In particular, the in-control unconditional p.m.f is given by

$$\Pr(N_{1of1,0} = t) = \Pr_{F=G}(N_{1of1} = t) = D(t-1) - D(t) \quad \text{for} \quad t = 1,2,3,... \quad \text{and} \quad D(0) = 1$$

where $N_{1of1,0}$ denotes the in-control run-length random variable and

$$D(t) = \int_0^1 \int_0^y \left(\frac{1}{\beta(j, n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(y^{j+h} - x^{j+h}\right) \right)^t f_{a,b}(x,y)dxdy$$

whereas the unconditional in-control average run-length ($UARL_{1of1,0}$) and the unconditional false alarm rate ($UFAR_{1of1}$) follows from (4-38) and (4-39) and given by

$$UARL_{1of1,0} = \int_0^1 \int_0^y \left(1 - \frac{1}{\beta(j, n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} (y^{j+h} - x^{j+h}) \right)^{-1} f_{a,b}(x,y)dxdy$$

and

$$UFAR_{1of1} = p^{\pm}(F,F) = \int_0^1 \int_0^y \left(1 - I_y(j, n-j+1) + I_x(j, n-j+1)\right) f_{a,b}(x,y)dxdy,$$

respectively.

4.3.1.2 Distribution of N_{2of2}^{DR}: Run-length distribution for the *2-of-2* DR precedence chart

As pointed out in section 4.3.1.1, in Chakraborti et al. (2004) the idea of conditioning on the reference sample order statistics $\mathbf{Z}$ was effectively used to derive the distribution of the run-length N_{1of1} of the two-sided *1-of-1* precedence chart and to study various properties of their chart in a convenient way. Using the same conditioning idea we derive the *conditional* and the *unconditional* run-length distributions of the two-sided *2-of-2* DR chart.

Conditional run-length distribution

Given that $\mathbf{Z}=\mathbf{z}$ the sequence of signaling indicators $\hat{\xi}_1^{\pm},\hat{\xi}_2^{\pm},\hat{\xi}_3^{\pm},...$ in (4-30) are i.i.d. Bernoulli random variables with success probability $p_C^{\pm}=\Pr(I(Y_{j:n}^{i}\notin(L\hat{C}L,U\hat{C}L))=1\,|\,\mathbf{Z}=\mathbf{z})$. Thus, conditionally on the order statistics $\mathbf{Z}$ the run-length N_{2of2}^{DR} of the two-sided *2-of-2* DR chart follows a geometric distribution of order two.

The conditional p.m.f of N_{2of2}^{DR} is therefore given by (4-16) with $\alpha=p_C^{\pm}$ and $k=2$ i.e.

$$\Pr(N_{2of2}^{DR}=t\,|\,\mathbf{Z}=\mathbf{z})=\begin{cases}0 & \text{if} \quad 0\le t<2\\ (p_C^{\pm})^2 & \text{if} \quad t=2\end{cases} \tag{4-40a}$$

and for $t\ge 3$

$$\Pr(N_{2of2}^{DR}=t\,|\,\mathbf{Z}=\mathbf{z})=\sum_{i=1}^{[\frac{t+1}{3}]}(-1)^{i-1}\frac{(p_C^{\pm})^{2i}}{(1-p_C^{\pm})^{1-i}}\left\{\binom{t-2i-1}{i-2}+(1-p_C^{\pm})\binom{t-2i-1}{i-1}\right\} \tag{4-40b}$$

whereas the conditional average run-length (expected value or mean) and the conditional variance of the run-length can be found from (4-17) and given by

$$CARL_{2of2}^{DR}=E(N_{2of2}^{DR}\,|\,\mathbf{Z}=\mathbf{z})=\frac{1+p_C^{\pm}}{(p_C^{\pm})^2} \tag{4-41}$$

and

$$CVARL_{2of2}^{DR}=\mathrm{var}(N_{2of2}^{DR}\,|\,\mathbf{Z}=\mathbf{z})=\frac{1-5(1-p_C^{\pm})(p_C^{\pm})^2-(p_C^{\pm})^5}{(1-p_C^{\pm})^2(p_C^{\pm})^4}, \tag{4-42}$$

respectively (see e.g. Remark 5).

Unconditional run-length distribution

The complexity of the *conditional* distribution in (4-40), particularly for $t \geq 3$, makes a direct application of conditioning to derive a closed form expression for the *unconditional* distribution of N_{2of2}^{DR} unattractive. Instead, we find the *unconditional* distribution of N_{2of2}^{DR} by first conditioning on the total number of successes $S_n = \sum_{i=1}^{n} \hat{\xi}_i^{\pm}$ in the sequence of *n* random variables $\hat{\xi}_1^{\pm}, \hat{\xi}_2^{\pm}, ..., \hat{\xi}_n^{\pm}$. (Note that, here, n is the number of random variables and not the sample size.)

To this end, note that, although $\hat{\xi}_1^{\pm}, \hat{\xi}_2^{\pm}, ..., \hat{\xi}_n^{\pm}$ is a sequence of dependent binary random variables they are exchangeable or symmetrically dependent; this means that any permutation of any subset of these random variables has the same distribution; this can be written as

$$\Pr(\hat{\xi}_{\pi(1)}^{\pm} = 1, ..., \hat{\xi}_{\pi(u)}^{\pm} = 1) = \Pr(\hat{\xi}_1^{\pm} = 1, ..., \hat{\xi}_u^{\pm} = 1) \tag{4-43}$$

for any permutation $\pi(1), ..., \pi(u)$ of $1,2,...,u \leq n$. Using (4-43) we can derive an exact closed form expression for the *unconditional* p.m.f of N_{2of2}^{DR}.

George and Bowman (1995) derived the distribution of the total number of successes S_n in a sequence of n exchangeable binary trials. According to their result

$$\Pr(S_n = s) = \binom{n}{s} \sum_{i=0}^{n-s} (-1)^i \binom{n-s}{i} \lambda_{s+i} \quad \text{for} \quad s = 1,2,...,n \tag{4-44}$$

where

$$\lambda_u = \Pr(\hat{\xi}_1^{\pm} = 1, ..., \hat{\xi}_u^{\pm} = 1) \quad \text{for} \quad u = 1,2,..,n. \tag{4-45}$$

Using (4-44) the unconditional distribution of N_{2of2}^{DR} is given by

$$\Pr(N_{2of2}^{DR} = t) = \begin{cases} 0 & \text{if } 0 \leq t < 2 \\ \lambda_2 & \text{if } t = 2 \end{cases} \tag{4-46a}$$

and for $t \geq 3$

$$\Pr(N_{2of2}^{DR} = t) = \sum_{y=1}^{t-2} \sum_{k=0}^{\min\left(y, \left[\frac{t-y-2}{2}\right]\right)} \sum_{i=0}^{y} (-1)^k (-1)^i \binom{y}{k} \binom{y}{i} \binom{t-2(k+1)-1}{y-1} \lambda_{t-y+i} \tag{4-46b}$$

The proof of (4-46) is straightforward for $t \le 2$ i.e

$$\Pr(N_{2of2}^{DR} = 2) = \Pr(\hat{\xi}_1^{\pm} = 1, \hat{\xi}_2^{\pm} = 1) = \lambda_2$$

where λ_2 is defined in (4-45).

For $t \ge 3,$ we write the unconditional distribution of N_{2of2}^{DR} as

$$\Pr(N_{2of2}^{DR} = t) = \sum_{y=1}^{t-2} \Pr(N_{2of2}^{DR} = t \mid S_t = t - y) \Pr(S_t = t - y) \qquad (4\text{-}47)$$

and then first consider the conditional probability $\Pr(N_{2of2}^{DR} = t \mid S_t = t - y)$.

By de Finetti's theorem a sequence of exchangeable random variables is conditionally i.i.d.. Hence, the conditional distribution of N_{2of2}^{DR} given the number of successes in exchangeable binary random variables is the same as that for a sequence of i.i.d. binary variables; this latter distribution has been worked out in the literature (see e.g. Balakrishnan and Koutras (2002), p 56; note a typo) and is given by

$$\Pr(N_{2of2}^{DR} = t \mid S_t = t - y) = \binom{t}{y}^{-1} \sum_{k=0}^{[\frac{t-y-2}{2}]} (-1)^k \binom{y}{k} \binom{t - 2(k+1) - 1}{y - 1}. \qquad (4\text{-}48)$$

Now, using (4-44) we have

$$\Pr(S_t = t - y) = \binom{t}{y} \sum_{i=0}^{y} (-1)^i \binom{y}{i} \lambda_{t-y+i},$$

so that (4-46) follows by substituting $\Pr(S_t = t - y)$ and (4-48) in (4-47).

Remark 14

(i) Conditionally on the reference sample order statistics that define the Phase II control limits we have that

$$\lambda_u^C = \Pr(\hat{\xi}_1^{\pm} = 1,\ldots,\hat{\xi}_u^{\pm} = 1 \mid \mathbf{Z} = \mathbf{z}) = \prod_{i=1}^{u} \Pr(\hat{\xi}_i^{\pm} = 1 \mid \mathbf{Z} = \mathbf{z}) = (p_C^{\pm})^u$$

so that the unconditional probability λ_u in (4-45) equals

$$\begin{aligned}\lambda_u &= \Pr(\hat{\xi}_1^{\pm} = 1,\ldots,\hat{\xi}_u^{\pm} = 1) = E_{\mathbf{Z}}(\lambda_u^C) = E_{\mathbf{Z}}((p_C^{\pm})^u) \\ &= \int_0^1 \int_0^y \left(1 - I_{GF^{-1}(y)}(j, n-j+1) + I_{GF^{-1}(x)}(j, n-j+1)\right)^u f_{a,b}(x,y)dxdy.\end{aligned} \tag{4-49}$$

(ii) The run-length distribution of the two-sided *2-of-2* DR chart depends on the distribution functions F and G through the composite function $\psi = GF^{-1}$ present in λ_u (see expression (4-49)). Thus, the in-control run-length distribution is obtained by substituting $\lambda_{u,0}$ in (4-46a) and (4-46b) where

$$\lambda_{u,0} = \int_0^1 \int_0^y \left(1 - I_y(j, n-j+1) + I_x(j, n-j+1)\right)^u f_{a,b}(x,y)dxdy \tag{4-50}$$

and is found from (4-49) by substituting $F = G$. It is evident from (4-50) that the in-control run-length distribution of the two-sided *2-of-2* DR (like that of the *1-of-1* chart) is free from either *F* or *G* and that the *2-of-2* DR chart is thus distribution-free.

(iii) The unconditional false alarm rate, the unconditional average run-length etc of the two-sided *2-of-2* DR chart is calculated later in section 4.3.2.

4.3.1.3 Distribution of N_{2of2}^{KL} and N_{2of3}: Run-length distribution for the two-sided *2-of-2* KL and *2-of-3* precedence charts

In sections 4.3.1.1 and 4.3.1.2 we illustrated how to find the *conditional* and *unconditional* distributions of N_{1of1} and N_{2of2}^{DR} via the geometric distribution of order k (= 1 or 2). Here, in section 4.3.1.3, we illustrate how to find the *conditional* and *unconditional* distributions of N_{2of2}^{KL} and N_{2of3} via the Markov chain approach. The conditional and unconditional distributions of N_{1of1} and N_{2of2}^{DR}, via the Markov chain approach, can be found in a similar manner and is not shown here.

The Markov chain approach for finding the *conditional* run-length distributions of the (two-sided) precedence charts in Case U is similar to those of the sign charts in Case K. In particular, the state spaces are identical so that we merely substitute:

(i) $p_C^+(Y,F,G)$ (defined in (4-36)) for $p^+(n,b,\theta)$, and

(ii) $p_C^-(X,F,G)$ (defined in (4-35)) for $p^-(n,a,\theta)$

in any one of the essential transition probability matrices of the two-sided sign charts (i.e. the $\mathbf{Q}_{h\times h}$'s given in (4-21), (4-22), (4-26) etc.) to obtain the *conditional* essential transition probability matrices $\mathbf{Q}_{h\times h}^C$ (say) of the precedence charts. Note that, here, in Case U, the superscript "C" in $\mathbf{Q}_{h\times h}^C$ indicates that we work with a *conditional* essential transition probability matrix i.e. conditioned on the order statistics $\mathbf{Z}$.

Upon substituting $\mathbf{Q}_{h\times h}^C$ into (4-8), (4-9) and (4-10) we obtain the *conditional* p.m.f, the *conditional ARL* (*CARL*) and the *conditional VARL* (*CVARL*), respectively. The *unconditional* run-length distributions and the associated *unconditional* characteristics of the precedence chart is then found by averaging over the distribution of $\mathbf{Z}$.

Conditional distributions of N_{2of2}^{KL} and N_{2of3}

Conditional on $\mathbf{Z}$ the sequence of signaling indicators $\hat{\xi}_1, \hat{\xi}_2, \hat{\xi}_3, \ldots$ in (4-29) are i.i.d. tri-variate random variables with:

(i) $\Pr(\hat{\xi}_i = 2 \mid \mathbf{Z} = \mathbf{z}) = p_C^-$,

(ii) $\Pr(\hat{\xi}_i = 1 \mid \mathbf{Z} = \mathbf{z}) = p_C^+$ and

(iii) $\Pr(\hat{\xi}_i = 0 \mid \mathbf{Z} = \mathbf{z}) = 1 - p_C^- - p_C^+$

respectively.

Thus, *conditional* on $\mathbf{Z}$ the run-length distribution of the two-sided *2-of-2* KL precedence chart is

$$\Pr(N_{2of2}^{KL} = t \mid \mathbf{Z} = \mathbf{z}) = \boldsymbol{\xi}(\mathbf{Q}_{4\times4}^{C})^{t-1}(\mathbf{I} - \mathbf{Q}_{4\times4}^{C})\mathbf{1} \quad \text{for} \quad t = 1,2,3,\ldots \tag{4-51}$$

with the *conditional* average run-length and the *conditional* variance of the run-length given by

$$CARL_{2of2}^{KL} = E(N_{2of2}^{KL} \mid \mathbf{Z} = \mathbf{z}) = \boldsymbol{\xi}(\mathbf{I} - \mathbf{Q}_{4\times4}^{C})^{-1}\mathbf{1} \tag{4-52}$$

and

$$CVARL_{2of2}^{KL} = \operatorname{var}(N_{2of2}^{KL} \mid \mathbf{Z} = \mathbf{z}) = \boldsymbol{\xi}(\mathbf{I} + \mathbf{Q}_{4\times4}^{C})(\mathbf{I} - \mathbf{Q}_{4\times4}^{C})^{-2}\mathbf{1} - (CARL_{2of2}^{KL})^2 \tag{4-53}$$

respectively, where

$$\mathbf{Q}_{4\times4}^{C} = \begin{bmatrix} 0 & 1-p_C^- - p_C^+ & p_C^+ & p_C^- \\ 0 & 1-p_C^- - p_C^+ & p_C^+ & p_C^- \\ 0 & 1-p_C^- - p_C^+ & 0 & p_C^- \\ 0 & 1-p_C^- - p_C^+ & p_C^+ & 0 \end{bmatrix} \tag{4-54}$$

denotes the *conditional* essential transition probability matrix of the *2-of-2* KL chart and follows from (4-26) having substituted the conditional probabilities p_C^- and p_C^+ (defined in (4-35) and (4-36)) for p^- and p^+, respectively.

Likewise, the *conditional* p.m.f, the *conditional* average run-length and the *conditional* variance of the two-sided *2-of-3* precedence chart are

$$\Pr(N_{2of3} = t \mid \mathbf{Z} = \mathbf{z}) = \boldsymbol{\xi}(\mathbf{Q}_{8\times8}^{C})^{t-1}(\mathbf{I} - \mathbf{Q}_{8\times8}^{C})\mathbf{1} \quad \text{for} \quad t = 1,2,3,\ldots, \tag{4-55}$$

$$CARL_{2of3} = E(N_{2of3} \mid \mathbf{Z} = \mathbf{z}) = \boldsymbol{\xi}(\mathbf{I} - \mathbf{Q}_{8\times8}^{C})^{-1}\mathbf{1}, \tag{4-56}$$

and

$$CVARL_{2of3} = \text{var}(N_{2of3} \mid \mathbf{Z} = \mathbf{z}) = \boldsymbol{\xi}(\mathbf{I} + \mathbf{Q}_{8\times8}^{C})(\mathbf{I} - \mathbf{Q}_{8\times8}^{C})^{-2}\mathbf{1} - (CARL_{2of3})^{2} \tag{4-57}$$

respectively, where the *conditional* essential transition probability matrix of the *2-of-3* chart is given by

$$\mathbf{Q}_{8\times8}^{C} = \begin{bmatrix}
0 & 1-p_C^- - p_C^+ & p_C^+ & p_C^- & 0 & 0 & 0 & 0 \\
0 & 1-p_C^- - p_C^+ & 0 & 0 & p_C^+ & 0 & p_C^- & 0 \\
0 & 0 & p_C^+ & p_C^- & 0 & 1-p_C^- - p_C^+ & 0 & 0 \\
0 & 0 & p_C^+ & p_C^- & 0 & 0 & 0 & 1-p_C^- - p_C^+ \\
0 & 0 & 0 & p_C^- & 0 & 1-p_C^- - p_C^+ & 0 & 0 \\
0 & 1-p_C^- - p_C^+ & 0 & 0 & 0 & 0 & p_C^- & 0 \\
0 & 0 & p_C^+ & 0 & 0 & 0 & 0 & 1-p_C^- - p_C^+ \\
0 & 1-p_C^- - p_C^+ & 0 & 0 & p_C^+ & 0 & 0 & 0
\end{bmatrix} \tag{4-58}$$

and follows from (4-27). In particular, expressions (4-55), (4-56) and (4-57) follow from having substituted $\mathbf{Q}_{8\times8}^{C}$ in (4-8), (4-9) and (4-10), respectively.

Remark 15

The *conditional ARL* expressions in (4-52) and (4-56) have been symbolically simplified and closed form expressions are given in Table 4.5; however, here, in Case U, we substitute p_C^+ for p^+ and p_C^- for p^-, respectively. Closed form expressions of the *conditional VARL*'s in (4-53) and (4-57) can be obtained in a similar manner (i.e. simplifying the expressions symbolically).

Unconditional distributions of N_{2of2}^{KL} and N_{2of3}

The unconditional distributions of N_{2of2}^{KL} and N_{2of3} are obtained by averaging the conditional distributions given in (4-51) and (4-55) over the distribution of $\mathbf{Z}$ i.e.

$$\begin{aligned}\Pr(N_{2of2}^{KL}=t) &= \int_0^1\int_0^y \Pr(N_{2of2}^{KL}=t\mid \mathbf{Z})\, f_{a,b}(x,y)dxdy \\ &= \int_0^1\int_0^y \xi(\mathbf{Q}_{4\times4}^{C})^{t-1}(\mathbf{I}-\mathbf{Q}_{4\times4}^{C})\mathbf{1}\, f_{a,b}(x,y)dxdy\end{aligned} \tag{4-59}$$

and

$$\begin{aligned}\Pr(N_{2of3}=t) &= \int_0^1\int_0^y \Pr(N_{2of3}=t\mid \mathbf{Z})\, f_{a,b}(x,y)dxdy \\ &= \int_0^1\int_0^y \xi(\mathbf{Q}_{8\times8}^{C})^{t-1}(\mathbf{I}-\mathbf{Q}_{8\times8}^{C})\mathbf{1}\, f_{a,b}(x,y)dxdy\end{aligned} \tag{4-60}$$

for $t=1,2,3,\ldots.$

Remark 16

The *unconditional* or marginal distributions of N_{2of2}^{KL} and N_{2of3} in (4-59) and (4-60) depend on the distribution functions F and G through the composite function $\psi = GF^{-1}$ present in both

$$p_C^- = I_{GF^{-1}(X)}(j, n-j+1) \quad \text{and} \quad p_C^+ = 1 - I_{GF^{-1}(Y)}(j, n-j+1),$$

which form part of the *conditional* essential transition probability matrices in (4-54) and (4-58).

The in-control unconditional run-length distributions follow by substituting $F = G$ in p_C^- and p_C^+ so that $GF^{-1}(u) = u$.

For example, the in-control marginal (or unconditional) run-length distribution of the two-sided *2-of-2* KL chart is

$$\Pr(N_{2of2,0}^{KL} = t) = \Pr_{F=G}(N_{2of2}^{KL} = t) = \int_0^1 \int_0^y \Pr_{F=G}(N_{2of2}^{KL} = t \mid \mathbf{Z})\, f_{a,b}(x,y)\,dxdy$$
$$= \int_0^1 \int_0^y \boldsymbol{\xi}(\mathbf{Q}_{4\times4,0}^C)^{t-1}(\mathbf{I} - \mathbf{Q}_{4\times4,0}^C)\mathbf{1}\, f_{a,b}(x,y)\,dxdy$$

for $t = 1,2,3,\ldots$ with

$$\mathbf{Q}_{4\times4,0}^C = \begin{bmatrix} 0 & I_Y(j,n-j+1) - I_X(j,n-j+1) & 1 - I_Y(j,n-j+1) & I_X(j,n-j+1) \\ 0 & I_Y(j,n-j+1) - I_X(j,n-j+1) & 1 - I_Y(j,n-j+1) & I_X(j,n-j+1) \\ 0 & I_Y(j,n-j+1) - I_X(j,n-j+1) & 0 & I_X(j,n-j+1) \\ 0 & I_Y(j,n-j+1) - I_X(j,n-j+1) & 1 - I_Y(j,n-j+1) & 0 \end{bmatrix}.$$

Once again, the in-control run-length distribution is seen to be free from both F and G, and thus the *2-of-2* KL chart is distribution-free; the same being true for *2-of-3* precedence chart.

4.3.2 Unconditional *ARL*, *VARL* and *FAR* calculations

In order to design and study the performance of a Phase II control chart one typically examines the average and the variance of the *unconditional* run-length distribution ($UARL$ & $UVARL$) along with the *unconditional* false alarm rate ($UFAR$).

For the proposed nonparametric runs-rule enhanced precedence charts, expressions for the average and the variance of the *unconditional* run-length distribution can be obtained exactly and most conveniently derived by using the characteristics of the *conditional* run-length distributions coupled with conditional expectation.

To this end, note that,

$$E(N) = E_{\mathbf{Z}}(E(N \mid \mathbf{Z})) \qquad \text{and} \qquad \text{var}(N) = E_{\mathbf{Z}}(\text{var}(N \mid \mathbf{Z})) + \text{var}_{\mathbf{Z}}(E(N \mid \mathbf{Z})) \qquad (4\text{-}61)$$

where $E(N)$ and $\text{var}(N)$ denote the *unconditional* characteristics whilst $E(N \mid \mathbf{Z})$ and $\text{var}(N \mid \mathbf{Z})$ denote the *conditional* characteristics (i.e. conditioned on $\mathbf{Z}$). The *unconditional* false alarm rate ($UFAR$) can be obtained in a similar manner and is shown below.

4.3.2.1 Unconditional *ARL*, *VARL* and *FAR* of the *2-of-2* DR chart

Unconditional *ARL*

The *unconditional ARL* for the *2-of-2* DR chart is computed by averaging expression (4-41) over the joint distribution of the order statistics.

Thus

$$UARL_{2of2}^{DR} = E_{\mathbf{Z}}(E(N_{2of2}^{DR} \mid \mathbf{Z})) = E_{\mathbf{Z}}(CARL_{2of2}^{DR})$$

$$= \int_0^1 \int_0^y \left(\frac{1 + p_C^{\pm}}{(p_C^{\pm})^2} \right) f_{a,b}(x,y)dxdy \tag{4-62}$$

$$= \int_0^1 \int_0^y \left(\frac{2 - I_{GF^{-1}(y)}(j, n-j+1) + I_{GF^{-1}(x)}(j, n-j+1)}{(1 - I_{GF^{-1}(y)}(j, n-j+1) + I_{GF^{-1}(x)}(j, n-j+1))^2} \right) f_{a,b}(x,y)dxdy.$$

The in-control unconditional average run-length ($UARL_{2of2,0}^{DR}$) is obtained by substituting $F = G$ in (4-62) and given by

$$UARL_{2of2,0}^{DR} = \int_0^1 \int_0^y \left(\frac{2 - I_y(j, n-j+1) + I_x(j, n-j+1)}{(1 - I_y(j, n-j+1) + I_x(j, n-j+1))^2} \right) f_{a,b}(x,y)dxdy. \tag{4-63}$$

Unconditional *VARL*

The *unconditional* variance of the *2-of-2* DR chart is obtained by noting that, in general, the unconditional variance in (4-61) can be re-written as

$$\begin{aligned} \text{var}(N) &= E_{\mathbf{Z}}(\text{var}(N \mid \mathbf{Z})) + \text{var}_{\mathbf{Z}}(E(N \mid \mathbf{Z})) \\ &= E_{\mathbf{Z}}(\text{var}(N \mid \mathbf{Z})) + \{E_{\mathbf{Z}}[(E(N \mid \mathbf{Z}))^2] - [E_{\mathbf{Z}}(E(N \mid \mathbf{Z}))]^2\}. \end{aligned} \tag{4-64}$$

For the *2-of-2* DR chart, in particular, we have that $E(N_{2of2}^{DR} \mid \mathbf{Z})$ and $\text{var}(N_{2of2}^{DR} \mid \mathbf{Z})$ are given by (4-41) and (4-42), respectively so that the unconditional variance of the run-length of the *2-of-2* DR chart is given by

$$UVARL_{2of2}^{DR} = \text{var}(N_{2of2}^{DR}) = E_{\mathbf{z}}\left(\frac{1-5(1-p_C^{\pm})(p_C^{\pm})^2-(p_C^{\pm})^5}{(1-p_C^{\pm})^2(p_C^{\pm})^4}\right)+E_{\mathbf{z}}\left(\left(\frac{1+p_C^{\pm}}{(p_C^{\pm})^2}\right)^2\right)-\left(E_{\mathbf{z}}\left(\frac{1+p_C^{\pm}}{(p_C^{\pm})^2}\right)\right)^2. \quad (4\text{-}65)$$

The in-control variance of the unconditional run-length distribution of the *2-of-2* DR chart is found by substituting $1-I_Y(j,n-j+1)+I_X(j,n-j+1)$ for $p_C^{\pm}$ in expression (4-65), where $p_C^{\pm}$ is defined in (4-33).

Unconditional *FAR*

The *conditional* false alarm rate of the *2-of-2* DR chart follows from Table 4.5 by substituting $p_C^{+} = p_C^{+}(Y,F,G)$ (with $F=G$) for p_0^{+} and $p_C^{-} = p_C^{-}(X,F,G)$ (with $F=G$) for p_0^{-} and is given by

$$\begin{aligned} CFAR_{2of2}^{DR} &= (p_C^{+}(Y,F,F))^2+(p_C^{+}(X,F,F))^2+2p_C^{+}(Y,F,F)p_C^{+}(X,F,F) \\ &= (1-I_Y(j,n-j+1))^2+(I_X(j,n-j+1))^2+2(1-I_Y(j,n-j+1))(I_X(j,n-j+1)). \end{aligned}$$

By averaging over the joint distribution of the order statistics we obtain the *unconditional* false alarm rate of the *2-of-2* DR chart i.e.

$$\begin{aligned} UFAR_{2of2}^{DR} = E_{\mathbf{z}}(CFAR_{2of2}^{DR}) &= \int_0^1 (1-I_y(j,n-j+1))^2 f_b(y)dy+\int_0^1 (I_x(j,n-j+1))^2 f_a(x)dx \\ &\quad +2\int_0^1\int_0^y (1-I_y(j,n-j+1))(I_x(j,n-j+1)) f_{a,b}(x,y)dxdy \end{aligned} \quad (4\text{-}66)$$

where $f_a(x)$ and $f_b(x)$ denotes the marginal p.d.f's of the a^{th} and the b^{th} order statistics in a random sample (the reference or Phase I sample) of size m from a *uniform*(0,1) distribution, which are known to be a $Beta(a,m-a+1)$ distribution and a $Beta(b,m-b+1)$ distribution, respectively.

4.3.2.2 Unconditional *ARL*, *VARL* and *FAR* of the *2-of-2* KL chart

Unconditional *ARL*

The *conditional ARL* (or the conditional expected value) of the *2-of-2* KL chart follows from (4-52), with a symbolically simplified version given in Table 4.5, i.e.

$$CARL_{2of2}^{KL} = E(N_{2of2}^{KL} \mid \mathbf{Z}) = \boldsymbol{\xi}(\mathbf{I} - \mathbf{Q}_{4\times4}^{C})^{-1}\mathbf{1} = \left(\frac{(p_C^+)^2}{(p_C^+ + 1)} + \frac{(p_C^-)^2}{(p_C^- + 1)}\right)^{-1}$$

so that by averaging over $\mathbf{Z}$, the *unconditional ARL* of the *2-of-2* KL chart is found to be

$$\begin{aligned} UARL_{2of2}^{KL} &= E_{\mathbf{Z}}(E(N_{2of2}^{KL} \mid \mathbf{Z})) = E_{\mathbf{Z}}(CARL_{2of2}^{KL}) = \int_0^1\int_0^y \left(\frac{(p_C^+)^2}{(p_C^+ + 1)} + \frac{(p_C^-)^2}{(p_C^- + 1)}\right)^{-1} f_{a,b}(x,y)dxdy \\ &= \int_0^1\int_0^y \left(\frac{(1 - I_{GF^{-1}(y)}(j, n-j+1))^2}{(2 - I_{GF^{-1}(y)}(j, n-j+1))} + \frac{(I_{GF^{-1}(x)}(j, n-j+1))^2}{(I_{GF^{-1}(x)}(j, n-j+1)+1)}\right)^{-1} f_{a,b}(x,y)dxdy. \end{aligned} \tag{4-67}$$

The in-control unconditional average run-length is again obtained by substituting $F = G$ in (4-67)

$$UARL_{2of2,0}^{KL} = \int_0^1\int_0^y \left(\frac{(1 - I_y(j, n-j+1))^2}{(2 - I_y(j, n-j+1))} + \frac{(I_x(j, n-j+1))^2}{(I_x(j, n-j+1)+1)}\right)^{-1} f_{a,b}(x,y)dxdy \tag{4-68}$$

which is distribution-free.

Unconditional *VARL*

Substituting $\text{var}(N_{2of2}^{KL} \mid \mathbf{Z})$ (given in (4-53)) and $E(N_{2of2}^{KL} \mid \mathbf{Z})$ (given in (4-52)) in (4-64) we find that the *unconditional* variance of N_{2of2}^{KL} is given by

$$\begin{aligned} UVARL_{2of2}^{KL} = \text{var}(N_{2of2}^{KL}) = E_{\mathbf{Z}}\left(\boldsymbol{\xi}(\mathbf{I} + \mathbf{Q}_{4\times4}^{C})(\mathbf{I} - \mathbf{Q}_{4\times4}^{C})^{-2}\mathbf{1} - (CARL_{2of2}^{KL})^2\right) \\ + E_{\mathbf{Z}}\left(\left(\boldsymbol{\xi}(\mathbf{I} - \mathbf{Q}_{4\times4}^{C})^{-1}\mathbf{1}\right)^2\right) - \left(E_{\mathbf{Z}}\left(\boldsymbol{\xi}(\mathbf{I} - \mathbf{Q}_{4\times4}^{C})^{-1}\mathbf{1}\right)\right)^2. \end{aligned} \tag{4-69}$$

Unconditional *FAR*

The *conditional* false alarm rate of the *2-of-2* KL chart follows from Table 4.5 by substituting $p_C^+ = p_C^+(Y,F,G)$ (with $F=G$) for p_0^+ and $p_C^- = p_C^-(X,F,G)$ (with $F=G$) for p_0^- and is given by

$$\begin{aligned} CFAR_{2of2}^{KL} &= (p_C^+(Y,F,F))^2 + (p_C^+(X,F,F))^2 \\ &= (1-I_Y(j,n-j+1))^2 + (I_X(j,n-j+1))^2. \end{aligned} \tag{4-70}$$

By averaging over the joint distribution of the order statistics the *unconditional* false alarm rate is obtain as

$$UFAR_{2of2}^{KL} = E_{\mathbf{Z}}(CFAR_{2of2}^{KL}) = \int_0^1 (1-I_y(j,n-j+1))^2 f_b(y)dy + \int_0^1 (I_x(j,n-j+1))^2 f_a(x)dx \tag{4-71}$$

which is again distribution-free.

4.3.2.3 Unconditional *ARL*, *VARL* and *FAR* of the *2-of-3* chart

Unconditional *ARL* and *VARL*

The *unconditional ARL* and unconditional *VARL* of the *2-of-3* chart are obtained in the same manner as that of the *2-of-2* KL chart; that is, we use the *conditional* counterparts derived via the Markov chain approach and find that

$$UARL_{2of3} = E_Z(CARL_{2of3}) = \int_0^1 \int_0^y \boldsymbol{\xi}(\mathbf{I} - \mathbf{Q}_{8\times8}^C)\mathbf{1}\, f_{a,b}(x,y)dxdy$$

and

$$\begin{aligned} UVARL_{2of3} &= E_{\mathbf{Z}}(\text{var}(N_{2of3} \mid \mathbf{Z})) + \{E_{\mathbf{Z}}[(E(N_{2of3} \mid \mathbf{Z}))^2] - [E_{\mathbf{Z}}(E(N_{2of3} \mid \mathbf{Z}))]^2\} \\ &= E_{\mathbf{Z}}\left(\boldsymbol{\xi}(\mathbf{I} + \mathbf{Q}_{8\times8}^C)(\mathbf{I} - \mathbf{Q}_{8\times8}^C)^{-2}\mathbf{1} - (CARL_{2of3})^2\right) + E_{\mathbf{Z}}\left(\left(\boldsymbol{\xi}(\mathbf{I} - \mathbf{Q}_{8\times8}^C)^{-1}\mathbf{1}\right)^2\right) - \left(E_{\mathbf{Z}}\left(\boldsymbol{\xi}(\mathbf{I} - \mathbf{Q}_{8\times8}^C)^{-1}\mathbf{1}\right)\right)^2, \end{aligned}$$

respectively.

Unconditional *FAR*

The *conditional FAR* of the *2-of-3* chart is found (like that of the *2-of-2* DR chart and the *2-of-2* KL chart) from Table 4.5 by substituting the conditional probabilities p_C^- and p_C^+ (defined in (4-35) and (4-36) with $F = G$) for p^- and p^+, respectively and is given by

$$\begin{aligned} UFAR_{2of3} = {} & 2(1 - I_Y(j, n-j+1))^2(1 - I_Y(j, n-j+1) + I_X(j, n-j+1)) \\ & + 2(I_X(j, n-j+1))^2(1 - I_Y(j, n-j+1) + I_X(j, n-j+1)). \end{aligned}$$

The *unconditional* FAR is thus given by

$$\begin{aligned} UFAR_{2of3} = {} & 2\int_0^1 \int_0^y (1 - I_y(j, n-j+1))^2(1 - I_y(j, n-j+1) + I_x(j, n-j+1))\, f_{a,b}(x,y)dxdy \\ & + 2\int_0^1 \int_0^y (I_x(j, n-j+1))^2(1 - I_y(j, n-j+1) + I_x(j, n-j+1))\, f_{a,b}(x,y)dxdy. \end{aligned}$$

4.3.3 Run-length distributions of the one-sided precedence charts

If detecting higher (lower) values is of interest, that is, whether the parameter or percentile of interest has shifted to the right (left), we can use a one-sided upper (lower) control chart with an estimated upper (lower) control limit $U\hat{C}L = X_{b:m}$ ($L\hat{C}L = X_{a:m}$) only.

The operation of the one-sided upper and lower runs-rules enhanced precedence charts of Case U is similar to that of the one-sided upper and lower runs-rules enhanced sign charts of Case K. For example, the *2-of-2* one-sided upper (lower) precedence chart signals on the first occurrence of a run of length two of the charting statistic $Y_{j:n}^{i}$ on or above (below) the estimated upper (lower) control limit.

The derivation of the run-length distributions of the one-sided runs-rules enhanced precedence charts parallels that of the two-sided precedence charts. In particular, we let

$$\hat{\xi}_i^{+} = I(Y_{j:n}^{i} \geq U\hat{C}L) = \begin{cases} 1 & \text{if} \quad Y_{j:n}^{i} \geq U\hat{C}L \\ 0 & \text{if} \quad Y_{j:n}^{i} < U\hat{C}L \end{cases}$$

and

$$\hat{\xi}_i^{-} = I(Y_{j:n}^{i} \leq L\hat{C}L) = \begin{cases} 1 & \text{if} \quad Y_{j:n}^{i} \leq L\hat{C}L \\ 0 & \text{if} \quad Y_{j:n}^{i} > L\hat{C}L \end{cases}$$

for $i = 1,2,3,...$ denote the indicator functions for the one-sided precedence charts corresponding to the events $\{Y_{j:n}^{i} \geq U\hat{C}L\}$ and $\{Y_{j:n}^{i} \leq L\hat{C}L\}$, respectively. Then, we can again use a two-step approach to derive the run-length distribution. In other words, we first derive the *conditional* run-length distribution i.e. conditioned on the particular order statistic (control limit) and then, second, we derive the *unconditional* or marginal run-length distribution by "averaging over" the distribution of the order statistic that constitutes the Phase II control limit.

In particular, given $X_{b:m} = x_{b:m}$ the sequence of signaling indicators $\hat{\xi}_1^{+}, \hat{\xi}_2^{+}, \hat{\xi}_3^{+}, ...$ are i.i.d. Bernoulli random variables with success probability

$$p_C^{+} = \Pr(\hat{\xi}_i^{\pm} = 1 \mid X_{b:m} = x_{b:m}) = \Pr(\hat{\xi}_i^{+} = 1 \mid X_{b:m} = x_{b:m}) = p_C^{+}(Y, F, G)$$

so that the *conditional* distribution of the run-length variable N^{+}_{2of2} of the *2-of-2* upper one-sided precedence chart, for example, is geometric of order $k=2$ with parameter (success probability) $p^{+}_{C}=p^{+}(y,F,G)$. Consequently, all the properties and the characteristics of the *conditional* run-length distribution follow conveniently from the properties of the geometric distribution of order $k=2$ by substituting $\alpha=p^{+}_{C}$ and $k=2$ in expressions (4-16) and (4-17).

Alternatively, we can use a Markov chain approach; doing so we find that

$$\Pr(N^{+}_{2of2}=t \mid X_{b:m}=x_{b:m})=\boldsymbol{\xi}(\mathbf{Q}^{C}_{3\times3})^{t-1}(\mathbf{I}-\mathbf{Q}^{C}_{3\times3})\mathbf{1} \qquad \text{for} \qquad t=1,2,3,...$$

with the *conditional* average run-length and the *conditional* variance of the run-length given by

$$CARL^{+}_{2of2}=E(N^{+}_{2of2} \mid X_{b:m}=x_{b:m})=\boldsymbol{\xi}(\mathbf{I}-\mathbf{Q}^{C}_{3\times3})^{-1}\mathbf{1}$$

and

$$CVARL^{+}_{2of2}=\text{var}(N^{+}_{2of2} \mid X_{b:m}=x_{b:m})=\boldsymbol{\xi}(\mathbf{I}+\mathbf{Q}^{C}_{3\times3})(\mathbf{I}-\mathbf{Q}^{C}_{3\times3})^{-2}\mathbf{1}-(CARL^{+}_{2of2})^{2}$$

respectively, where

$$\mathbf{Q}^{C}_{3\times3}=\begin{bmatrix}0 & 1-p^{+}_{C} & p^{+}_{C}\\ 0 & 1-p^{+}_{C} & p^{+}_{C}\\ 0 & 1-p^{+}_{C} & 0\end{bmatrix}$$

denotes the *conditional* essential transition probability matrix of the *2-of-2* upper one-sided precedence chart and follows from (4-15) having substituted p^{+}_{C} for p^{+}.

The *unconditional* p.m.f of N^{+}_{2of2}, for example, is obtained by averaging the *conditional* run-length distribution over the distribution of $X_{b:m}$ i.e.

$$\Pr(N^{+}_{2of2}=t)=\int_{0}^{1}\Pr(N^{+}_{2of2}=t \mid Y=y)\,f_{b}(y)dy=\int_{0}^{1}\boldsymbol{\xi}(\mathbf{Q}^{C}_{3\times3})^{t-1}(\mathbf{I}-\mathbf{Q}^{C}_{3\times3})\mathbf{1}\,f_{b}(y)dy.$$

To obtain a closed form expression of the *unconditional* p.m.f of N^{+}_{2of2} requires the same steps as carried out in case of the *2-of-2* DR precedence chart of section 4.3.1.2 and therefore not shown here.

4.3.4 Design and implementation of the two-sided precedence charts

In order to implement the proposed precedence charts in practice we need the upper and the lower control limits. This means that we need to find the indices (charting constants) a and b that specify the reference sample order statistics, which constitute the lower and the upper control limit, respectively.

Determination of charting constants

In Phase II applications one typically determines the charting constants a and b so that a specified in-control unconditional average run-length (say, $UARL_0^*$ equal to 370 or 500) is obtained. This means that we have to solve

$$UARL_{1of1,0}^{*} = \int_0^1 \int_0^y \left(1 - \frac{1}{\beta(j,n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(y^{j+h} - x^{j+h}\right)\right)^{-1} f_{a,b}(x,y)dxdy \qquad (4\text{-}72)$$

for the *1-of-1* chart,

$$UARL_{2of2,0}^{DR*} = \int_0^1 \int_0^y \left(\frac{2 - I_y(j,n-j+1) + I_x(j,n-j+1)}{(1 - I_y(j,n-j+1) + I_x(j,n-j+1))^2}\right) f_{a,b}(x,y)dxdy \qquad (4\text{-}73)$$

for the *2-of-2* DR chart,

$$UARL_{2of2,0}^{KL*} = \int_0^1 \int_0^y \left(\frac{(1 - I_y(j,n-j+1))^2}{(2 - I_y(j,n-j+1))} + \frac{(I_x(j,n-j+1))^2}{(I_x(j,n-j+1)+1)}\right)^{-1} f_{a,b}(x,y)dxdy \qquad (4\text{-}74)$$

for the *2-of-2* KL chart and

$$UARL_{2of3,0}^{*} = \int_0^1 \int_0^y \left(\xi(\mathbf{I} - \mathbf{Q}_{8\times8,0}^{C})^{-1}\mathbf{1}\right) f_{a,b}(x,y)dxdy \qquad (4\text{-}75)$$

for the *2-of-3* chart where $\mathbf{Q}_{8\times8,0}^{C}$ follows from (4-58) by substituting $1 - I_Y(j,n-j+1)$ for p_C^+ and $I_Y(j,n-j+1)$ for p_C^-.

4.3.4.1 Charting constants of the *1-of-1* chart

Chakraborti et al. (2004) provided values for the charting constants a and b for the two-sided *1-of-1* precedence chart for a number of different choices (combinations) of the size m of the Phase I reference sample, the size n of Phase II samples and j (the selected order statistic) so that the in-control unconditional average run-length (i.e. $UARL_{1of1,0}$) is close to 370 , 500 and 1000, respectively.

4.3.4.2 Charting constants of the *2-of-2* DR, *2-of-2* KL and *2-of-3* charts

Tables 4.18, 4.19 and 4.20 display various choices (combinations) of the charting constants a and b for the two-sided *2-of-2* DR, the two-sided *2-of-2* KL and the two-sided *2-of-3* charts, for a given or specified in-control unconditional ARL in the neighborhood of 300 and 500, when reference samples of size $m = 50$, 100, 200 and 500 are used to estimate the control limits in Phase I and these limits are used to monitor the location (center) of a process using the medians of Phase II (test) samples of size $n = 5$, 7, or 9, respectively. Thus, j equals 3, 4, and 5, respectively in the tables.

Note that for each combination of values of *n*, *j* and *m* the tables display (in each cell) the $UARL_0$, the $UFAR$ and (a,b) values, where the $UARL_0$ values are in the neighborhood of 300 to 500.

Since the Phase II (test) sample median is used as the charting statistic and the Phase II sample size n is odd, it seems reasonable to use symmetric control limits, and thus we take $b = m - a + 1$, so that only a needs to be determined. However, this needs not be the case when the chart constants are to be determined for a charting statistic other than the median that might be of interest.

In addition, note that, in general, it is rare to achieve an $UARL_0$ (or an $UFAR$) exactly as specified (i.e. 300 or 500) with the nonparametric charts because the in-control distribution of the run-length distribution is discrete. However, as can be seen, one can get reasonably close to the values typically used in practice.

For example, from Table 4.18 for $m = 500$, $n = 5$ and $j = 3$, one set of constants for the *2-of-2* DR chart are given by $a = 72$ and $b = 500 - 72 + 1 = 429$ so that $\hat{LCL} = X_{72:500}$ and $\hat{UCL} = X_{429:500}$. In this case the achieved (or attained) $UARL_0$ and the attained unconditional FAR of the chart are 496.90 and 0.0025, respectively. Moreover, these are the exact values and remain the same for all continuous distributions. If instead we took $a = 71$ and $b = 500 - 71 + 1 = 430$, so that $\hat{LCL} = X_{71:500}$

and $\hat{UCL} = X_{430:500}$, the achieved $UARL_0$ increases to 536.72 and the attained FAR decreases to 0.0023. For a more moderate reference sample size, such as $m = 50$, Table 4.18 shows that it is possible to obtain $UARL_0$ values such as 275.30 or 605.44; the latter which may be deemed reasonably large in practice. Obviously as *m* and/or *n* increase, the available choices for the $UARL_0$ values also increase.

Similar behavior is observed in the case of the two-sided *2-of-2* KL and the *2-of-3* charts shown in Tables 4.19 and 4.20, respectively. For instance, in case of the *2-of-2* KL chart, when $m = 500$ and one uses $a = 80$ and $b = 500 - 80 + 1 = 421$, so that $\hat{LCL} = X_{80:500}$ and $\hat{UCL} = X_{421:500}$, the ARL_0 of the *2-of-2* KL chart (when $n = 5$ and $j = 3$) is 524.39, whereas the FAR is 0.0023. However, if instead one chooses to use $a = 81$ and $b = 500 - 81 + 1 = 420$, so that $\hat{LCL} = X_{81:500}$ and $\hat{UCL} = X_{420:500}$, the $UARL_0$ decreases to 490.21, whereas the $UFAR$ slightly increases to 0.0024. Although for $m = 500$, $n = 5$ and $j = 3$ a specified $UARL_0$ such as 500 cannot be obtained exactly, by increasing the size of the reference sample m and/or the test sample size *n*, the range of possible $UARL_0$ and $UFAR$ values that can be attained increases.

All equations (i.e. (4-73), (4-74) and (4-75)) are solved using the software package Mathcad®14.0.

Table 4.18: Unconditional in-control average run-length ($UARL_0$), unconditional false alarm rate ($UFAR$) and chart constants (a,b) [1] for the *2-of-2* DR nonparametric chart for m = 50,100,200,500 and $(n, j) = (5, 3), (7, 4), (9, 5)$

n=5 , j=3				n=7 , j=4				n=9 , j=5			
m=50	100	200	500	m=50	100	200	500	m=50	100	200	500
605.44	548.99	537.62	536.72	597.80	509.54	597.72	526.08	976.53	739.47	558.51	528.95
0.0072	0.0040	0.0029	0.0023	0.0090	0.0048	0.0027	0.0024	0.0084	0.0040	0.0031	0.0024
(8,43)	(15,86)	(29,172)	(71,430)	(10,41)	(19,82)	(36,165)	(90,411)	(11,40)	(21,80)	(42,159)	(104,397)
275.30	373.31	443.56	496.90	264.91	345.93	490.44	487.01	383.92	481.18	456.18	488.41
0.0121	0.0055	0.0034	0.0025	0.0150	0.0065	0.0033	0.0026	0.0144	0.0056	0.0037	0.0026
(9,42)	(16,85)	(30,171)	(72,429)	(11,40)	(20,81)	(37,164)	(91,410)	(12,39)	(22,79)	(43,158)	(105,396)
	261.69	368.80	460.60		241.21	405.20	451.33	172.47	322.26	375.04	451.43
	0.0074	0.0040	0.0026		0.0088	0.0039	0.0028	0.0236	0.0077	0.0044	0.0028
	(17,84)	(31,170)	(73,428)		(21,80)	(38,163)	(92,409)	(13,38)	(23,78)	(44,157)	(106,395)
		308.82	427.48			336.97	418.70		221.57	310.28	417.68
		0.0047	0.0028			0.0046	0.0030		0.0104	0.0053	0.0030
		(32,169)	(74,427)			(39,162)	(93,408)		(24,77)	(45,156)	(107,394)
		260.37	397.20			281.98	388.83			258.24	386.83
		0.0056	0.0031			0.0054	0.0032			0.0062	0.0033
		(33,168)	(75,426)			(40,161)	(94,407)			(46,155)	(108,393)
			369.50				361.45				358.60
			0.0033				0.0034				0.0035
			(76,425)				(95,406)				(109,392)
			344.12				336.33				332.75
			0.0035				0.0037				0.0038
			(77,424)				(96,405)				(110,391)
			320.83				313.25				309.06
			0.0037				0.0039				0.0041
			(78,423)				(97,404)				(111,390)
			299.44				292.03				287.31
			0.0040				0.0042				0.0044
			(79,422)				(98,403)				(112,389)

[1]The three rows of each cell shows the achieved (attained) $UARL_0$, the $UFAR$ and the charting constants (a , b), respectively

Table 4.19: Unconditional in-control average run-length ($UARL_0$), unconditional false alarm rate ($UFAR$) and chart constants (a,b) [1] for the *2-of-2* KL nonparametric chart for m = 50,100,200,500 and $(n, j) = (5, 3), (7, 4), (9, 5)$

n=5 , j=3				n=7 , j=4				n=9 , j=5			
m=50	100	200	500	m=50	100	200	500	m=50	100	200	500
1010.37	650.75	559.01	524.39	985.39	594.56	504.01	506.61	1591.68	547.12	548.41	530.19
0.0048	0.0033	0.0026	0.0023	0.0063	0.0041	0.0031	0.0024	0.0062	0.0049	0.0031	0.0023
(8,43)	(16,85)	(32,169)	(80,421)	(10,41)	(20,81)	(40,161)	(99,402)	(11,40)	(23,78)	(45,156)	(112,389)
460.89	456.52	471.18	490.21	437.32	414.67	424.10	472.95	626.67	376.11	456.29	493.12
0.0079	0.0044	0.0031	0.0024	0.0102	0.0054	0.0036	0.0026	0.0103	0.0066	0.0036	0.0025
(9,42)	(17,84)	(33,168)	(81,420)	(11,40)	(21,80)	(41,160)	(100,401)	(12,39)	(24,77)	(46,155)	(113,388)
237.00	328.69	399.60	458.70	217.33	296.08	358.81	441.90	281.29	264.69	381.78	459.05
0.0123	0.0057	0.0036	0.0026	0.0160	0.0070	0.0042	0.0027	0.0165	0.0086	0.0042	0.0027
(10,41)	(18,83)	(34,167)	(82,419)	(12,39)	(22,79)	(42,159)	(101,400)	(13,38)	(25,76)	(47,154)	(114,387)
	242.15	340.87	429.62			305.16	413.24			321.15	427.69
	0.0074	0.0041	0.0027			0.0048	0.0029			0.0049	0.0029
	(19,82)	(35,166)	(83,418)			(43,158)	(102,399)			(48,153)	(115,386)
		292.37	402.76			260.82	386.77			271.54	398.81
		0.0047	0.0029			0.0056	0.0031			0.0057	0.0031
		(36,165)	(84,417)			(44,157)	(103,398)			(49,152)	(116,385)
			377.91				362.28				372.18
			0.0031				0.0033				0.0033
			(85,416)				(104,397)				(117,384)
			354.91				339.62				347.61
			0.0033				0.0035				0.0035
			(86,415)				(105,396)				(118,383)
			333.60				318.62				324.92
			0.0035				0.0037				0.0038
			(87,414)				(106,395)				(119,382)
			313.83				299.16				303.95
			0.0037				0.0040				0.0040
			(88,413)				(107,394)				(120,381)
			295.48								284.55
			0.0039								0.0043
			(89,412)								(121,380)

[1]The three rows of each cell shows the achieved (attained) $UARL_0$, the $UFAR$ and the charting constants (a , b), respectively

Table 4.20: Unconditional in-control average run-length ($UARL_0$), unconditional false alarm rate ($UFAR$) and chart constants (a,b) [1] for the *2-of-3* nonparametric chart for m = 50,100,200,500 and $(n, j) = (5, 3), (7, 4), (9, 5)$

n=5 , j=3				n=7 , j=4				n=9 , j=5			
m=50	100	200	500	m=50	100	200	500	m=50	100	200	500
1336.27	755.74	623.35	672.76	1313.99	690.80	685.53	653.58	2423.24	1025.16	786.51	712.63
0.0049	0.0033	0.0025	0.0018	0.0065	0.0040	0.0025	0.0019	0.0062	0.0035	0.0024	0.0018
(7,44)	(14,87)	(28,73)	(68,433)	(9,42)	(18,83)	(35,166)	(87,414)	(10,41)	(20,81)	(40,161)	(100,401)
527.33	502.46	513.03	621.56	513.55	460.68	561.53	604.19	819.75	653.26	636.94	656.49
0.0084	0.0045	0.0030	0.0020	0.0107	0.0055	0.0030	0.0021	0.0105	0.0049	0.0029	0.0020
(8,43)	(15,86)	(29,172)	(69,432)	(10,41)	(19,82)	(36,165)	(88,413)	(11,40)	(21,80)	(41,160)	(101,400)
246.51	346.18	425.78	575.05	233.72	316.82	463.41	559.19	329.94	430.22	519.65	605.46
0.0134	0.0060	0.0035	0.0021	0.0169	0.0073	0.0035	0.0022	0.0169	0.0066	0.0034	0.0021
(9,42)	(16,85)	(30,171)	(70,431)	(11,40)	(20,81)	(37,164)	(89,412)	(12,39)	(22,79)	(42,159)	(102,399)
130.77	246.05	356.16	532.74	120.09	224.02	385.16	518.13	152.81	291.98	426.99	559.02
0.0201	0.0078	0.0042	0.0023	0.0252	0.0095	0.0041	0.0024	0.0257	0.0087	0.0041	0.0023
(10,41)	(17,84)	(31,170)	(71,430)	(12,39)	(21,80)	(38,163)	(90,411)	(13,38)	(23,78)	(43,158)	(103,398)
	179.74	300.11	494.18		162.45	322.29	480.62	79.60	203.71	353.24	516.70
	0.0101	0.0048	0.0024		0.0122	0.0048	0.0026	0.0374	0.0114	0.0048	0.0025
	(18,83)	(32,169)	(72,429)		(22,79)	(39,162)	(91,410)	(14,37)	(24,77)	(44,157)	(104,397)
		254.64	459.00			271.43	446.32		145.81	294.15	478.10
		0.0056	0.0026			0.0056	0.0028		0.0147	0.0056	0.0027
		(33,168)	(73,428)			(40,161)	(92,409)		(25,76)	(45,156)	(105,396)
		217.47	426.85			230.00	414.91			246.49	442.85
		0.0065	0.0028			0.0065	0.0030			0.0065	0.0029
		(34,167)	(74,427)			(41,160)	(93,408)			(46,155)	(106,395)
			397.43				386.11			207.81	410.63
			0.0030				0.0032			0.0076	0.0031
			(75,426)				(94,407)			(47,154)	(107,394)
			370.48				359.69				381.14
			0.0031				0.0034				0.0033
			(76,425)				(95,406)				(108,393)
			345.75				335.41				354.12
			0.0034				0.0030				0.0035
			(77,424)				(96,405)				(109,392)
			323.04				313.09				329.35
			0.0037				0.0039				0.0038
			(78,423)				(97,404)				(110,391)
			302.15				292.53				306.61
			0.0039				0.0042				0.0041
			(79,422)				(98,403)				(111,390)
			282.91				273.59				285.71
			0.0042				0.0044				0.0044
			(80,421)				(99,402)				(112,389)
			265.18				256.12				266.50
			0.0045				0.0047				0.0047
			(81,420)				(100,401)				(113,388)

[1]The three rows of each cell shows the achieved (attained) $UARL_0$, the $UFAR$ and the charting constants (a , b), respectively

Example 3

In order to illustrate the runs-rule enhanced nonparametric precedence charts we use the data given in Table 5.1 on p. 213 and Table 5.2 on p. 219 of Montgomery (2001).

The goal of this study was to establish statistical control of the inside diameter of the piston rings for an automotive engine manufactured in a forging process. Twenty-five retrospective or Phase I samples, each of size five, were collected when the process was thought to be in-control. As shown in Example 5.1 on p. 213 of Montgomery (2001), the traditional Shewhart $\bar{X}$ and R charts provide no indication of an out-of-control condition, so these data are considered to be Phase I reference data and these "trial" limits were adopted for use in on-line process control.

In order to implement the nonparametric control charts the charting constants are needed. Possible symmetric control limits ($b = m - a + 1$) for the four charts are shown in Table 4.21, for $m = 125$, $n = 5$ and $j = 3$, along with the corresponding $UARL_0$ and $UFAR$ values.

Table 4.21: Unconditional in-control average run-length ($UARL_0$), unconditional false alarm rate ($UFAR$) and chart constants (a, b) for the two-sided *1-of-1*, *2-of-2* DR, *2-of-2* KL and *2-of-3* precedence charts when $m = 125$, $n = 5$ and $j = 3$

1-of-1				*2-of-2* DR				*2-of-2* KL				*2-of-3*			
a	***b***	$UARL_0$	*UFAR*	***a***	***b***	$UARL_0$	*UFAR*	***a***	***b***	$UARL_0$	*UFAR*	***a***	***b***	$UARL_0$	*UFAR*
5	121	1315.98	0.0019	17	109	898.74	0.0023	18	108	1125.44	0.0018	17	109	822.40	0.0026
6	120	695.09	0.0029	18	108	638.60	0.0031	19	107	819.47	0.0024	18	108	590.03	0.0034
7	119	413.80	0.0044	19	107	464.38	0.0040	20	106	608.81	0.0030	19	107	433.39	0.0043
8	118	267.40	0.0062	20	106	344.73	0.0052	21	105	460.54	0.0038	20	106	325.09	0.0055
9	117	183.47	0.0084	21	105	260.69	0.0066	22	104	354.09	0.0048	21	105	248.51	0.0069
				22	104	200.46	0.0084	23	103	276.28	0.0059	22	104	193.27	0.0086

Using Table 4.21, for an $UARL_0$ of 500, one can take $a = 7$ so that $b = 119$, and therefore the control limits for the *1-of-1* precedence chart are the 7th and the 119th ordered values of the reference sample. Thus $\hat{LCL} = X_{7:125} = 73.984$ and $\hat{UCL} = X_{119:125} = 74.017$, which yield an in-control unconditional ARL of 413.80 and an unconditional FAR of 0.0044.

A plot of the medians for the *1-of-1* chart is shown in Figure 4.12 for all forty samples, the first twenty five of which are from Phase I. It is seen that the 37th median is outside the control limits and so the *1-of-1* precedence chart signals on the 12th (i.e. 37th – 25th) sample in the prospective phase.

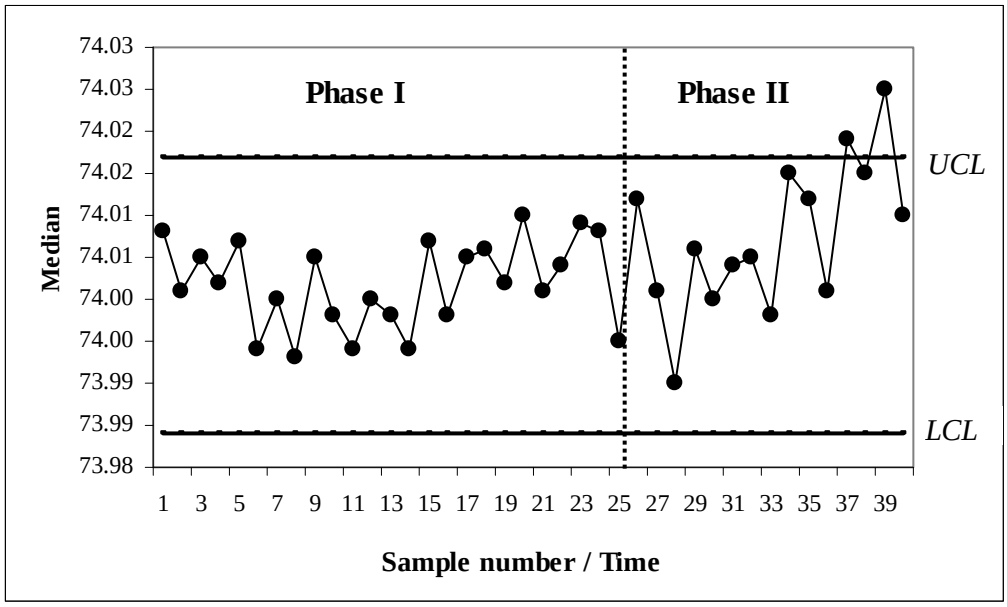


Figure 4.12: The *1-of-1* Phase II Precedence chart for the Montgomery (2001) piston-ring data

For the *2-of-2* DR chart, we take $a = 19$ so that $b = 125 - 19 + 1 = 107$ and the resulting limits, $\hat{LCL} = X_{19:125} = 73.990$ and $\hat{UCL} = X_{107:125} = 74.012$, yield an $UARL_0$ and $UFAR$ of 464.38 and 0.0040, respectively. Note, however, that if one chooses $a = 20$ so that $b = 106$, the control limits become $\hat{LCL} = X_{20:125}$ and $\hat{UCL} = X_{106:125}$ and the corresponding $UARL_0$ decreases to 344.73, whereas the $UFAR$ slightly increases to 0.0052. The *2-of-2* DR chart is shown in Figure 4.13.

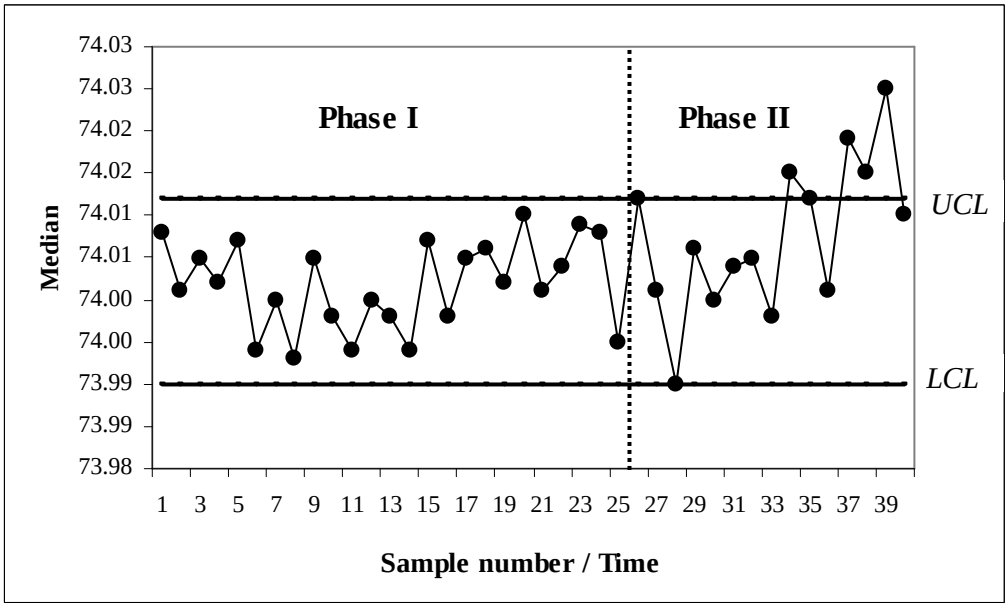


Figure 4.13: The *2-of-2* Phase II DR chart for the Montgomery (2001) piston-ring data

For the *2-of-2* KL chart we take $a = 21$ so that $b = 125 - 21 + 1 = 105$ and thus $L\hat{C}L = X_{21:125} = 73.992$ and $U\hat{C}L = X_{105:125} = 74.011$; this yields an $UARL_0$ of 460.54 and an *UFAR* of 0.0038, respectively. The *2-of-2* KL chart is shown in Figure 4.14.

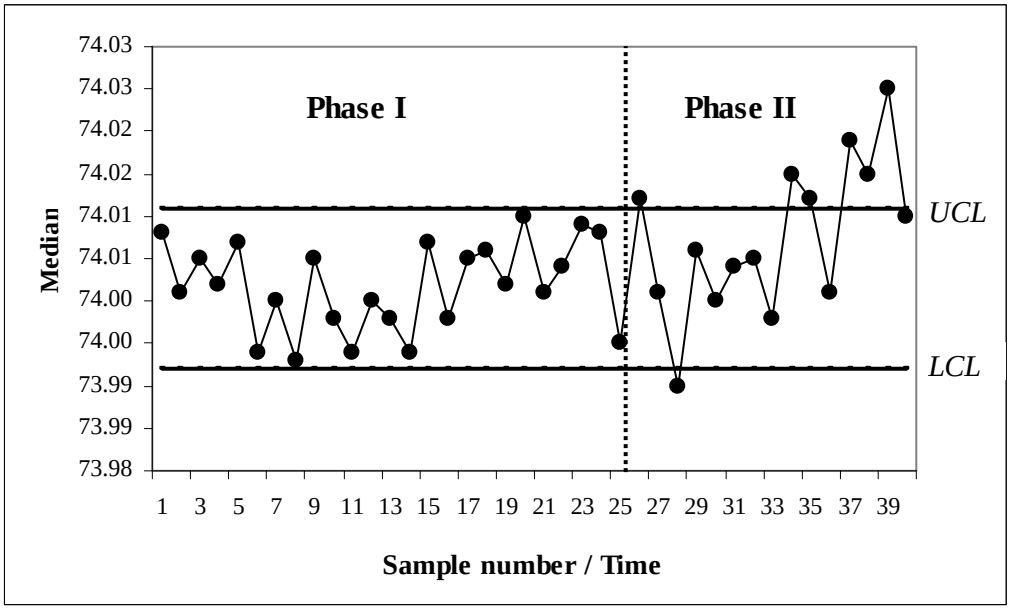


Figure 4.14: The *2-of-2* Phase II KL chart for the Montgomery (2001) piston ring data

Finally, for the *2-of-3* chart we take $a = 19$ so that $b = 125 - 19 + 1 = 107$ and thus $L\hat{C}L = X_{19:125} = 73.990$ and $U\hat{C}L = X_{107:125} = 74.012$ which yields an $UARL_0$ of 433.39 and an *UFAR* of 0.0043, respectively. The *2-of-3* chart is identical to the *2-of-2* DR chart shown in Figure 4.14 and is thus omitted; this is so because the control limits (in this example) of the *2-of-3* chart are exactly the same as that of the *2-of-2* DR chart.

The *2-of-2* DR charts signals on the 3rd sample whereas both the *2-of-2* KL and the *2-of-3* charts signal on the 10th sample in the prospective phase. Note, however, that the achieved *UFAR* values for the four charts are much larger (63%, 48%, 41% and 59%, respectively) than the nominal *FAR* of 0.0027.

4.3.5 Performance comparison of the two-sided precedence charts

The performance of Phase II control charts is typically compared by first designing each control chart to (roughly) have the same in-control *unconditional* average run-length ($UARL_0$) and then examining their out-of-control *unconditional* average run-length ($UARL_1$) values at some out-of-control value(s) of the parameter of interest. The control chart with the shorter (or smaller) out-of-control average run-length is usually preferred. Since the proposed run rules enhanced Phase II charts are nonparametric Shewhart-type charts applicable in Case U, their main competitor is the basic *1-of-1* precedence control chart of Chakraborti et al. (2004).

To study robustness, three different underlying process distributions i.e. the normal distribution, the *t*-distribution and the gamma distribution, were used in a simulation study with 100 000 repetitions for each distribution investigated. Because the shape of the *t*-distribution is very similar to that of the normal distribution (it is symmetric, but with more probability in the tails) it was used to study the effect of heavier tails. The gamma distribution was used to study the effect of skewness (see e.g. Figure 4.15). In order for the results of the three distributions to be comparable, the t and gamma distributions were scaled so that they also had a mean of zero and a variance of one. Thus, the $n(0,1)$, the $\frac{1}{\sqrt{2}}t(4)$ and the $Gamma(1,1)-1$ distributions were used.

The parametric Shewhart $\bar{X}$ chart was included in the comparison for the normal distribution but not for the *t* and the gamma distribution since the $\bar{X}$ chart is well-known to be non-robust under non-normality (see e.g. Chakraborti et al. (2004)).

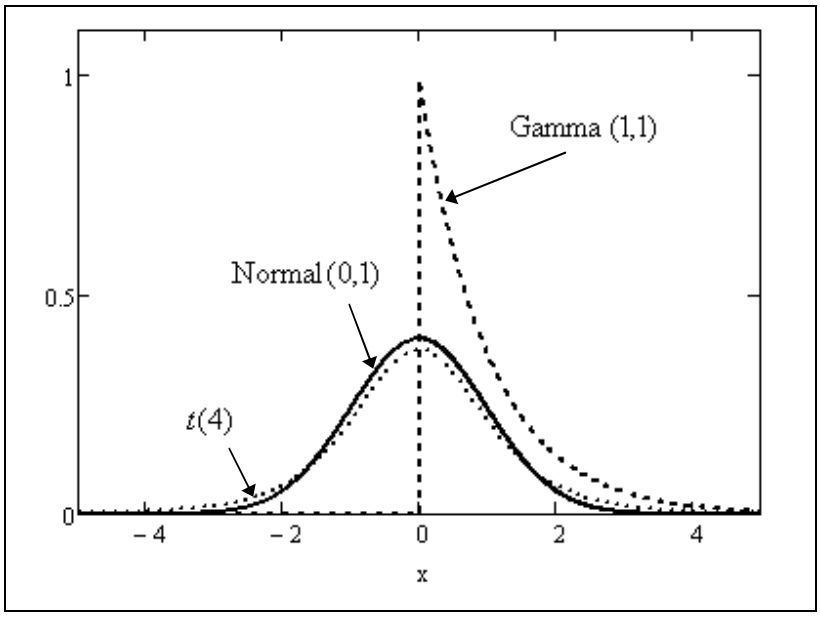


Figure 4.15: Probability distributions used for the performance comparison of the two-sided precedence control charts

Tables 4.22, 4.23 and 4.24 display the performance comparison results when a reference sample of size $m = 500$ is used to estimate the control limits to monitor location in Phase II (future or test) samples of size $n = 5$ and using $Y_{3:5}$, the median, as the charting statistic. The charts were designed so that an $UARL_0$ value close to 500 was achieved.

Instead of using randomization to get an $UARL_0$ exactly equal to 500, two combinations of chart constants are used for each nonparametric chart for which the $UARL_0$ was the nearest below and the nearest above the target $UARL_0^*$ of 500. The tables show the *unconditional* average run-length ($UARL_0$) along with the *unconditional* standard deviation of the run-length ($SDRL_0$). The shift refers to a shift in the mean of the distribution.

From Table 4.22 it is seen that even under the normal distribution, the nonparametric charts can be quite efficient i.e. good at detecting shifts. The *2-of-2* KL and the *2-of-3* charts are both almost as efficient as the $\overline{X}$ chart, with shorter ARL's but a slightly higher $SDRL$'s when the process is OOC, especially for small shifts.

When the distribution is $t(4)$, that is symmetric yet with heavier tails than the normal, Table 4.23 shows that the *2-of-2* DR, the *2-of-2* KL and the *2-of-3* schemes perform better than the basic precedence *1-of-1* chart in detecting small shifts, with the *2-of-2* KL chart being the best and is closely followed by the *2-of-3* chart. Thus, the three new nonparametric Shewhart-type charts with signaling rules provide better alternatives than the basic precedence *1-of-1* chart and the $\overline{X}$ chart, especially for smaller shifts. The same observation applies in the case of a right-skewed distribution such as the $Gamma(1,1)$ as shown in Table 4.24 but with the *2-of-2* KL chart doing the best. So the runs-type signaling rules enhance the nonparametric chart's sensitivity to a location shift.

Moreover, the gain in efficiency (relative to the *1-of-1* chart) can be substantial; for example, for the $t(4)$ distribution for a shift of 0.5, the OOC ARL values of the *1-of-1*, *2-of-2* DR, *2-of-3* and the *2-of-2* KL charts are 117.63, 40.98, 26.64 and 26.28, respectively when the corresponding ARL_0 values are very comparable, 520.27, 536.72, 532.74 and 524.39, respectively. Note that in Table 4.24 for the $Gamma(1,1)$ distribution the basic precedence chart display somewhat of a strange behavior in that both the ARL and $SDRL$ values first increase from their corresponding values for the in-control case for a shift of 0.25; thereafter the ARL and $SDRL$ values decrease for increasing shifts as it might be expected. We have not been able to explain this phenomenon. A repeat of the simulations produced similar results.

Table 4.22: Unconditional *ARL* and *SDRL* values for the *normal* distribution for the *2-of-2* DR, *2-of-2* KL, *2-of-3*, Basic (*1-of-1*) Precedence chart and the Shewhart X-bar chart when $m = 500$, $n = 5$, $j = 3$

	2-of-2 DR				*2-of-2* KL				*2-of-3*			
Shift	***ARL***	***SDRL***	***ARL***	***SDRL***	***ARL***	***SDRL***	***ARL***	***SDRL***	***ARL***	***SDRL***	***ARL***	***SDRL***
0.00	496.90	573.05	536.72	621.20	490.21	554.18	524.39	594.55	494.18	569.01	532.74	615.81
0.25	233.82	278.56	250.23	299.59	170.07	203.00	180.06	215.88	161.67	196.04	171.95	209.89
0.50	58.22	66.10	61.33	69.99	39.37	43.17	41.11	45.28	36.62	39.34	38.26	41.33
0.75	17.55	17.85	18.23	18.64	12.99	12.60	13.39	13.06	12.75	11.06	13.15	11.58
1.00	7.36	6.41	7.56	6.63	5.99	4.90	6.12	5.04	6.54	4.23	6.64	4.36
1.25	4.12	2.90	4.19	2.97	3.62	2.34	3.67	2.39	4.43	1.98	4.47	2.03
1.50	2.88	1.52	2.91	1.55	2.67	1.26	2.69	1.29	3.60	1.04	3.62	1.06
1.75	2.36	0.86	2.37	0.87	2.27	0.72	2.28	0.74	3.24	0.58	3.25	0.58
2.00	2.13	0.49	2.14	0.50	2.10	0.41	2.10	0.42	3.09	0.32	3.09	0.33
2.25	2.04	0.28	2.05	0.28	2.03	0.23	2.03	0.23	3.03	0.18	3.03	0.18
2.50	2.01	0.14	2.01	0.15	2.01	0.12	2.01	0.12	3.01	0.09	3.01	0.09
2.75	2.00	0.07	2.00	0.07	2.00	0.06	2.00	0.06	3.00	0.05	3.00	0.05
3.00	2.00	0.03	2.00	0.03	2.00	0.02	2.00	0.02	3.00	0.02	3.00	0.02
	(a=72,b=429)		**(a=71,b=430)**		**(a=81,b=420)**		**(a=80,b=421)**		**(a=72,b=429)**		**(a=71,b=430)**	

	1-of-1				X-bar	
Shift	***ARL***	***SDRL***	***ARL***	***SDRL***	***ARL***	***SDRL***
0.00	460.22	538.61	520.27	613.67	500.00	571.14
0.25	233.27	290.26	261.60	329.17	184.12	216.66
0.50	70.42	85.43	77.73	95.38	43.38	48.51
0.75	23.74	27.01	25.79	29.64	13.12	13.71
1.00	9.58	10.11	10.26	10.93	5.19	4.93
1.25	4.63	4.43	4.88	4.72	2.63	2.15
1.50	2.66	2.21	2.76	2.34	1.67	1.08
1.75	1.78	1.22	1.83	1.28	1.26	0.58
2.00	1.36	0.72	1.39	0.75	1.09	0.32
2.25	1.16	0.44	1.17	0.45	1.03	0.17
2.50	1.06	0.26	1.07	0.27	1.01	0.08
2.75	1.02	0.15	1.02	0.16	1.00	0.03
3.00	1.01	0.08	1.01	0.09	1.00	0.01
	(a=25,b=476)		**(a=24,b=477)**		**3.084500892**	

Table 4.23: Unconditional *ARL* and *SDRL* values for the *2-of-2* DR, *2-of-2* KL, *2-of-3* and the Basic (*1-of-1*) Precedence chart for the *t*(4) distribution when $m = 500$, $n = 5$, $j = 3$

	***2-of-2* DR**				***2-of-2* KL**			
Shift	***ARL***	***SDRL***	***ARL***	***SDRL***	***ARL***	***SDRL***	***ARL***	***SDRL***
0.00	496.90	573.05	536.72	621.20	490.21	554.18	524.39	594.55
0.25	200.92	248.00	215.95	268.18	138.19	170.25	146.90	182.03
0.50	38.68	45.31	40.98	48.41	25.09	27.66	26.28	29.19
0.75	10.01	9.77	10.41	10.26	7.43	6.66	7.65	6.92
1.00	4.26	3.11	4.35	3.22	3.61	2.36	3.67	2.43
1.25	2.72	1.33	2.75	1.37	2.52	1.08	2.54	1.10
1.50	2.23	0.67	2.24	0.68	2.17	0.55	2.17	0.56
1.75	2.07	0.36	2.08	0.36	2.05	0.30	2.05	0.30
2.00	2.02	0.19	2.02	0.20	2.02	0.16	2.02	0.16
2.25	2.01	0.11	2.01	0.11	2.00	0.09	2.00	0.09
2.50	2.00	0.06	2.00	0.06	2.00	0.05	2.00	0.05
2.75	2.00	0.03	2.00	0.03	2.00	0.03	2.00	0.03
3.00	2.00	0.02	2.00	0.02	2.00	0.02	2.00	0.02
	(*a*=72,*b*=429)		**(*a*=71,*b*=430)**		**(*a*=81,*b*=420)**		**(*a*=80,*b*=421)**	
	2-of-3				***1-of-1***			
Shift	***ARL***	***SDRL***	***ARL***	***SDRL***	***ARL***	***SDRL***	***ARL***	***SDRL***
0.00	494.18	569.01	532.74	615.81	460.22	538.61	520.27	613.67
0.25	138.65	175.16	147.99	189.23	288.43	370.47	328.18	426.13
0.50	25.36	27.21	26.64	28.92	102.82	143.88	117.63	167.75
0.75	8.19	6.28	8.43	6.57	32.84	45.71	37.43	53.44
1.00	4.53	2.14	4.59	2.20	11.19	14.51	12.58	16.84
1.25	3.49	0.92	3.51	0.94	4.47	5.01	4.91	5.71
1.50	3.16	0.53	3.16	0.46	2.25	1.97	2.40	2.20
1.75	3.05	0.35	3.05	0.38	1.46	0.90	1.51	0.98
2.00	3.02	0.18	3.01	0.13	1.16	0.46	1.18	0.49
2.25	3.01	0.60	3.00	0.07	1.05	0.24	1.06	0.26
2.50	3.00	0.04	3.00	0.05	1.02	0.13	1.02	0.14
2.75	3.00	0.02	3.00	0.02	1.01	0.07	1.01	0.08
3.00	3.00	0.01	3.00	0.01	1.00	0.04	1.00	0.04
	(*a*=72,*b*=429)		**(*a*=71,*b*=430)**		**(*a*=25,*b*=476)**		**(*a*=24,*b*=477)**	

Table 4.24: Unconditional *ARL* and *SDRL* values for the *2-of-2* DR, *2-of-2* KL *2-of-3*and the Basic (*1-of-1*) Precedence chart for the *gamma*(1,1) distribution when $m = 500$, $n = 5$, $j = 3$

	2-of-2 DR				*2-of-2* KL			
Shift	***ARL***	***SDRL***	***ARL***	***SDRL***	***ARL***	***SDRL***	***ARL***	***SDRL***
0.00	496.90	573.05	536.72	621.20	490.21	554.18	524.39	594.55
0.25	233.82	815.92	250.23	887.98	310.12	405.49	331.64	436.02
0.50	58.22	216.59	61.33	234.84	88.52	111.41	94.24	119.33
0.75	17.55	61.43	18.23	66.27	28.03	33.05	29.65	35.23
1.00	7.36	18.96	7.56	20.33	10.26	10.74	10.75	11.39
1.25	4.12	6.42	4.19	6.84	4.55	3.79	4.72	4.00
1.50	2.88	2.30	2.91	2.45	2.61	1.34	2.66	1.42
1.75	2.36	0.74	2.37	0.80	2.05	0.34	2.06	0.37
2.00	2.13	0.13	2.14	0.15	2.00	0.03	2.00	0.04
2.25	2.04	0.00	2.05	0.00	2.00	0.00	2.00	0.00
2.50	2.01	0.00	2.01	0.00	2.00	0.00	2.00	0.00
2.75	2.00	0.00	2.00	0.00	2.00	0.00	2.00	0.00
3.00	2.00	0.00	2.00	0.00	2.00	0.00	2.00	0.00
	(*a*=72,*b*=429)		**(*a*=71,*b*=430)**		**(*a*=81,*b*=420)**		**(*a*=80,*b*=421)**	
	2-of-3				***1-of-1***			
Shift	***ARL***	***SDRL***	***ARL***	***SDRL***	***ARL***	***SDRL***	***ARL***	***SDRL***
0.00	494.18	569.01	532.74	615.81	460.22	538.61	520.27	613.67
0.25	314.31	425.10	336.97	459.01	527.27	730.48	600.16	844.59
0.50	90.55	115.92	98.04	126.21	255.49	351.96	290.46	406.50
0.75	30.22	35.15	31.84	37.14	124.76	170.53	141.61	196.68
1.00	11.98	11.68	12.57	12.35	61.56	83.20	69.72	95.80
1.25	6.01	4.18	6.23	4.48	30.80	40.94	34.78	47.05
1.50	3.88	1.58	3.96	1.67	15.70	20.35	17.67	23.33
1.75	3.15	0.51	3.17	0.54	8.22	10.23	9.20	11.69
2.00	3.01	0.08	3.01	0.09	4.47	5.19	4.96	5.92
2.25	3.00	0.01	3.00	0.01	2.58	2.64	2.83	3.02
2.50	3.00	0.00	3.00	0.00	1.63	1.32	1.75	1.52
2.75	3.00	0.00	3.00	0.00	1.19	0.61	1.24	0.72
3.00	3.00	0.00	3.00	0.00	1.03	0.23	1.05	0.29
	(*a*=72,*b*=429)		**(*a*=71,*b*=430)**		**(*a*=25,*b*=476)**		**(*a*=24,*b*=477)**	

Because the run-length distribution is highly right-skewed, exclusive use of the *ARL* (and the *SDRL*) to characterize chart performance has been criticized in the literature and some researchers have strongly suggested an examination of the percentiles too (see e.g. Radson and Boyd, (2005) and Chakraborti, (2007)).

To this end, the three quartiles (Q_1, Q_2 and Q_3) along with the 5^{th} and the 95^{th} percentiles are shown in Tables 4.25, 4.26, 4.27 and 4.28, for the *2-of-2* DR, the *2-of-2* KL, the *2-of-3* and the *1-of-1* precedence chart, for the normal, $t(4)$ and $Gamma(1,1)$ distributions, respectively for $n = 5$ and $j = 3$. Note that (i) these values are all *unconditional* i.e. being averaged over the joint distribution of the order statistics $X_{a:m}$ and $X_{b:m}$, and (ii) these values were obtained via simulations (200 000 repetitions) using SAS®9.1 since exact calculations, via enumeration of the c.d.f, took too long for the upper percentiles. The SAS®-programs used in the simulations are provided in Appendix 4A.

A comparison of the quartiles lead to the same general observation that the newly proposed nonparametric charts are more efficient than the basic precedence chart, with the *2-of-2* KL and the *2-of-3* charts having a slight edge.

For example, in the in-control case and with the *t*-distribution, for the *2-of-2* DR chart (with $a = 72$ & $b = 429$) the three quartiles are 127, 313 and 658, respectively, which are very close to those for the *2-of-2* KL chart (with $a = 81$ & $b = 420$) and the *2-of-3* chart (with $a = 72$ & $b = 429$): 126, 312 & 650 and 127, 312 & 653, respectively. By contrast, for the *1-of-1* precedence chart (with $a = 25$ & $b = 476$) the three quartiles are 116, 287 and 603, respectively, which are all smaller. Since we want the in-control percentiles to be larger, the new charts are better. On the other hand, in the out-of-control case, for a shift of 0.50 in the mean, the quartiles for the *2-of-2* KL chart (with $a = 81$ & $b = 420$) and *2-of-3* chart (with $a = 72$ & $b = 429$) are all shorter: 7, 16 & 33 and 8, 17 & 32, respectively, compared to both the *2-of-2* DR chart (with $a = 72$ & $b = 429$): 11, 24 and 50 and the *1-of-1* precedence chart (with $a = 25$ & $b = 476$): 23, 57 and 127. This shows that the *2-of-2* KL and the *2-of-3* charts are superior.

Table 4.25: The three quartiles (Q_1, Q_2 and Q_3) and the 5th & 95th percentiles of the run-length distribution of the *2-of-2* DR chart; charting constants (a=72 , b=429)

	normal					*t*(4)					*gamma*(1,1)				
Shift	**5th**	**Q_1**	**Q_2**	**Q_3**	**95th**	**5th**	**Q_1**	**Q_2**	**Q_3**	**95th**	**5th**	**Q_1**	**Q_2**	**Q_3**	**95th**
0.00	23	128	314	657	1587	23	127	313	658	1588	23	127	313	653	1586
0.25	11	58	144	306	758	10	49	121	260	662	23	129	329	734	2002
0.50	4	16	37	76	183	3	11	24	50	122	8	37	93	205	547
0.75	2	6	12	23	52	2	4	7	13	29	3	12	29	61	158
1.00	2	3	5	10	20	2	2	3	5	10	2	5	10	21	51
1.25	2	2	3	5	10	2	2	2	3	5	2	2	4	8	18
1.50	2	2	2	3	6	2	2	2	2	4	2	2	2	4	8
1.75	2	2	2	2	4	2	2	2	2	2	2	2	2	2	4
2.00	2	2	2	2	3	2	2	2	2	2	2	2	2	2	2
2.25	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
2.50	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
2.75	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
3.00	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2

Table 4.26: The three quartiles (Q_1, Q_2 and Q_3) and the 5th & 95th percentiles of the run-length distribution of the *2-of-2* KL chart; charting constants (a=81 , b=420)

	normal					*t*(4)					*gamma*(1,1)				
Shift	**5th**	**Q_1**	**Q_2**	**Q_3**	**95th**	**5th**	**Q_1**	**Q_2**	**Q_3**	**95th**	**5th**	**Q_1**	**Q_2**	**Q_3**	**95th**
0.00	23	128	313	652	1554	24	126	312	650	1549	24	128	314	655	1552
0.25	9	43	105	223	549	7	34	84	179	449	14	72	181	394	1037
0.50	3	11	26	52	122	2	7	16	33	77	5	22	53	114	292
0.75	2	4	9	17	38	2	3	5	10	20	2	8	17	36	88
1.00	2	2	4	8	16	2	2	3	4	8	2	4	7	13	30
1.25	2	2	3	4	8	2	2	2	3	5	2	2	3	6	12
1.50	2	2	2	3	5	2	2	2	2	4	2	2	2	3	5
1.75	2	2	2	2	4	2	2	2	2	2	2	2	2	2	2
2.00	2	2	2	2	3	2	2	2	2	2	2	2	2	2	2
2.25	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
2.50	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
2.75	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
3.00	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2

Table 4.27: The three quartiles (Q_1, Q_2 and Q_3) and the 5th & 95th percentiles of the run-length distribution of the *2-of-3* chart; charting constants (a=72 , b=429)

	normal					*t*(4)					*gamma*(1,1)				
Shift	**5th**	**Q_1**	**Q_2**	**Q_3**	**95th**	**5th**	**Q_1**	**Q_2**	**Q_3**	**95th**	**5th**	**Q_1**	**Q_2**	**Q_3**	**95th**
0.00	25	127	312	653	1576	24	127	312	653	1580	25	128	313	655	1582
0.25	10	41	99	208	524	8	34	82	176	457	15	72	179	392	1051
0.50	4	11	24	48	111	4	8	17	32	76	6	23	54	114	295
0.75	3	5	9	16	34	3	4	6	10	20	4	9	19	38	94
1.00	3	4	5	8	15	3	3	4	5	9	3	5	8	15	33
1.25	3	3	4	5	8	3	3	3	4	5	3	3	4	7	14
1.50	3	3	3	4	6	3	3	3	3	4	3	3	3	4	7
1.75	3	3	3	3	4	3	3	3	3	3	3	3	3	3	4
2.00	3	3	3	3	4	3	3	3	3	3	3	3	3	3	3
2.25	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
2.50	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
2.75	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
3.00	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3

Table 4.28: The three quartiles (Q_1, Q_2 and Q_3) and the 5th & 95th of the run-length distribution of the Basic (*1-of-1*) precedence chart; charting constants (a=25 , b=476)

	normal					*t*(4)					*gamma*(1,1)				
Shift	**5th**	**Q_1**	**Q_2**	**Q_3**	**95th**	**5th**	**Q_1**	**Q_2**	**Q_3**	**95th**	**5th**	**Q_1**	**Q_2**	**Q_3**	**95th**
0.00	21	116	288	608	1489	21	116	287	603	1479	21	115	287	606	1477
0.25	10	56	140	300	765	12	66	168	368	956	20	115	294	655	1786
0.50	3	18	43	92	228	4	23	57	127	350	10	57	144	319	864
0.75	2	6	15	31	75	2	8	19	41	110	5	28	71	157	420
1.00	1	3	6	13	29	1	3	7	14	36	3	14	35	77	206
1.25	1	2	3	6	13	1	1	3	6	13	2	7	18	39	103
1.50	1	1	2	3	7	1	1	2	3	6	1	4	9	20	51
1.75	1	1	1	2	4	1	1	1	2	3	1	2	5	10	26
2.00	1	1	1	2	3	1	1	1	1	2	1	1	3	5	14
2.25	1	1	1	1	2	1	1	1	1	2	1	1	2	3	7
2.50	1	1	1	1	2	1	1	1	1	1	1	1	1	2	4
2.75	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2
3.00	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

4.4 Concluding remarks: Summary and recommendations

The new class of runs-rule enhanced nonparametric sign charts of Case K and the distribution-free precedence charts of Case U can be useful for the quality practitioner in that they

(i) enhance the in-control and the out-of-control performance of the *1-of-1* sign chart of Amin et al. (1995) and the basic precedence chart of Chakraborti et al. (2004), respectively, and

(ii) outperform the classical and well-known Shewhart $\overline{X}$ chart (especially for heavy-tailed or skewed distributions).

In particular, the charts based on the *k-of-k* and the *k-of-w* signaling rules facilitate larger ARL_0 and smaller FAR values which allow practitioners greater flexibility while designing charts to best suit their needs.

The key advantage and main benefits of the nonparametric charts are:

(i) their in-control run-length distributions (and all associated performance characteristics such as the ARL_0 and FAR, for example) are the same for all continuous distributions, and

(ii) one does not have to assume symmetry of the underlying distribution (unlike the SR charts). Thus, practitioners need not worry about what the underlying distribution is (and the serious consequences/ramifications/costs if it is not normal, for example) as far as implementing and understanding the charts' properties are concerned.

The sign charts have an added advantage as they can be applied in situations where the data are just dichotomous.

A further practical advantage of the precedence charts is their potential to save time and resources in situations where the data are naturally collected in an ordered fashion, as is common in "life-testing" type situations, where one observes the "time to failure" of some item and it is costly and time consuming to wait for all units to fail. Because the control limits and the charting statistic of the precedence charts are based on order statistics, they can be applied as soon as the required order statistics are observed, whereas the Shewhart or CUSUM or EWMA $\overline{X}$ charts can not be applied since one needs the full dataset to calculate the average.

Also, the precedence charts can be adapted to and applied in the case of ordinal data. The charting statistic can be chosen to be any order statistic of the Phase II sample suitable in a specific application. The median, used in this chapter, of course enjoys the robustness property and is therefore less affected by the presence of outliers (very small or large observations) than the $\overline{X}$ chart, for example.

Finally, the implementation and application of the sign and precedence charts are easy using the tables with the charting constants (and attained ARL_0 and FAR values) and it is recommend that they be used more frequently in practice.

4.5 Appendix 4A: SAS® programs

4.5.1 SAS® programs to simulate the run-length distributions of the upper one-sided X-bar, sign and SR charts in Case K

4.5.1.1 The *1-of-1* X-bar, sign and SR charts

```
*1-of-1 upper one-sided X-bar chart;

proc iml;
ARL = 370;
sim = 100000;
n = 10;
UCL = probit(1-1/ARL)/sqrt(n);
simrl = j(sim,13,.);
do d = 1 to 2.2 by 0.2;
do j = 1 to sim;
ct = 0;
do k = 1 to 10000000 while ( ^((ct>=UCL))  );
x = j(n,1,.);
call randgen(x,'NORMAL',d-1,1);
ct = sum(x)/n;
rl = k;
end;
simrl[j,d*5+1]=rl;
end;
end;
create RL1of1_Xbar from simrl[colname={delta000
delta020 delta040 delta060 delta080 delta100 delta120}];
append from simrl;
quit;
proc univariate data=RL1of1_Xbar;
var delta000 delta020 delta040 delta060 delta080 delta100 delta120;
run;
```

```
*1-of-1 upper one-sided sign chart;

proc iml;
ARL = 370;
sim = 100000;
a = 0;
n = 10;
UCL = n-a;
med = j(n,1,0);
simrl = j(sim,13,.);
q = (1/ARL - 1 + probbnml(0.5,n,n-a-1) ) / (probbnml(0.5,n,n-a-1) - probbnml(0.5,n,n-
a-2)) ;
do d = 1 to 2.2 by 0.2;
do j = 1 to sim;
ct = 0;
random = 0;
do k = 1 to 10000000 while ( ^((ct>=UCL)|((ct=UCL-1)&(random<=q)))  );
x = j(n,1,.);
call randgen(x,'NORMAL',d-1,1);
vec = x > med;
ct = sum(vec);
random = ranuni(0);
rl = k;
end;
simrl[j,d*5+1]=rl;
end;
end;
*print simrl;
create RL1of1_sign from simrl[colname={delta000
delta020 delta040 delta060 delta080 delta100 delta120}];
append from simrl;
quit;
proc univariate data=RL1of1_sign;
var delta000 delta020 delta040 delta060 delta080 delta100 delta120;
run;
```

```
*1-of-1 upper one-sided SR chart;

proc iml;
ARL = 370;
sim = 100000;
n = 10;
UCL = 53;
UCL1 = 51;
cdfUCL = 0.002;
pmfUCL1 = 0.0009;
med = j(n,1,0);
simrl = j(sim,13,.);
q = (1/ARL - cdfUCL) / ( pmfUCL1 ) ;
do d = 1 to 2.2 by 0.2;
do j = 1 to sim;
random = 0;
ct = 0;
do k = 1 to 10000000 while ( ^(   (ct>=UCL)  |  ((ct=UCL1)&(random<=q))   )  );
x = j(n,1,.);
call randgen(x,'NORMAL',d-1,1);
vec   = x > med;
wplus = (vec`)*rank(abs(x));
ct = 2*wplus - n*(n+1)/2;
random = ranuni(0);
rl = k;
end;
simrl[j,d*5+1]=rl;
end;
end;
create RL1of1_SR from simrl[colname={delta000
delta020 delta040 delta060 delta080 delta100 delta120}];
append from simrl;
quit;
proc univariate data=RL1of1_SR;
var delta000 delta020 delta040 delta060 delta080 delta100 delta120;
run;
```

4.5.1.2 The *2-of-2* sign and SR charts

```
*2-of-2 upper one-sided sign chart;

proc iml;
ARL = 370;
sim = 100000;
a = 1;
n = 10;
UCL = n-a;
med = j(n,1,0);
simrl = j(sim,13,.);
q = ( (sqrt(4*ARL+1)+1)/(2*ARL) - 1 + probbnml(0.5,n,n-a-1) ) /
    (probbnml(0.5,n,n-a-1) - probbnml(0.5,n,n-a-2)) ;
do d = 1 to 2.2 by 0.2;
do j = 1 to sim;
x = j(n,1,.);
ct1 = 0;
ct = 0;
random1 = 0;
random = 0;
do k = 1 to 10000000 while (
^( (  (ct1>=UCL)&(ct>=UCL)  )                        |
   (  ((ct1=UCL-1)&(random1<=q))&(ct>=ucl)  )        |
   (  (ct1>=UCL)&((ct=UCL-1)&(random<=q))  )         |
   (  ((ct1=UCL-1)&(random1<=q))&((ct=UCL-1)&(random<=q)))  )  );
ct1 = ct;
call randgen(x,'NORMAL',d-1,1);
vec = x > med;
ct = sum(vec);
random = ranuni(0);
random1 = ranuni(0);
rl = k;
end;
simrl[j,d*5+1]=rl;
end;
end;
create RL2of2_sign from simrl[colname={delta000
delta020 delta040 delta060 delta080 delta100 delta120}];
append from simrl;
quit;
proc univariate data=RL2of2_sign;
var delta000 delta020 delta040 delta060 delta080 delta100 delta120;
run;
```

```
*2-of-2 upper one-sided SR chart;

proc iml;
ARL = 370;
sim = 100000;
n = 10;
UCL = 33;
UCL1 = UCL - 2;
cdfUCL = 0.0527;
pmfUCL1 = 0.0127;
med = j(n,1,0);
simrl = j(sim,13,.);
q = (  (sqrt(4*ARL+1)+1)/(2*ARL) -  cdfUCL ) / ( pmfUCL1 ) ;
do d = 1 to 2.2 by 0.2;
do j = 1 to sim;
x = j(n,1,.);
ct1 = 0;
ct = 0;
random1 = 0;
random = 0;
do k = 1 to 10000000 while (
^( (  (ct1>=UCL)&(ct>=UCL)  )                        |
 (  ((ct1=UCL1)&(random1<=q))&(ct>=ucl)  )     |
 (  (ct1>=UCL)&((ct=UCL1)&(random<=q))  )      |
 (  ((ct1=UCL1)&(random1<=q))&((ct=UCL1)&(random<=q)))  )  );
ct1 = ct;
call randgen(x,'NORMAL',d-1,1);
vec   = x > med;
wplus = (vec`)*rank(abs(x));
ct = 2*wplus - n*(n+1)/2;
random = ranuni(0);
random1 = ranuni(0);
rl = k;
end;
simrl[j,d*5+1]=rl;
end;
end;
create RL2of2_SR from simrl[colname={delta000
delta020 delta040 delta060 delta080 delta100 delta120}];
append from simrl;
quit;
proc univariate data=RL2of2_SR;
var delta000 delta020 delta040 delta060 delta080 delta100 delta120;
run;
```

4.5.1.3 The *2-of-3* sign chart

```
*2-of-3 upper-sided sign chart;

proc iml;
ARL = 370;
sim = 100000;
a = 1;
n = 10;
UCL = n-a;
med = j(n,1,0);
simrl = j(sim,13,.);
q = 0.632202808;
do d = 1 to 2.2 by 0.2;
do j = 1 to sim;
x = j(n,1,.);
ct2 = 0;
ct1 = 0;
ct = 0;
random2 = 0;
random1 = 0;
random = 0;
do k = 1 to 10000000 while (
^(
 ( (ct2>=UCL)                 & (ct1<UCL) &  (ct>=UCL)               )     |
 ( ((ct2=UCL-1)&(random2<=q)) & (ct1<UCL) &  (ct>=UCL)               )     |
 ( (ct2>=UCL)                 & (ct1<UCL) & ((ct=UCL-1)&(random<=q)) )     |
 ( ((ct2=UCL-1)&(random2<=q)) & (ct1<UCL) & ((ct=UCL-1)&(random<=q)) )     |
 ( (ct1>=UCL)                 & (ct2<UCL) &  (ct>=UCL)               )     |
 ( ((ct1=UCL-1)&(random1<=q)) & (ct2<UCL) &  (ct>=UCL)               )     |
 ( (ct1>=UCL)                 & (ct2<UCL) & ((ct=UCL-1)&(random<=q)) )     |
 ( ((ct1=UCL-1)&(random1<=q)) & (ct2<UCL) & ((ct=UCL-1)&(random<=q)) )      )   );
ct2 = ct1;
ct1 = ct;
call randgen(x,'NORMAL',d-1,1);
vec = x > med;
ct = sum(vec);
random = ranuni(0);
random1 = ranuni(0);
random2 = ranuni(0);
rl = k;
end;
simrl[j,d*5+1]=rl;
end;
end;
ARL = sum(simrl)/sim;
create RL2of3_sign from simrl[colname={delta000
delta020 delta040 delta060 delta080 delta100 delta120}];
append from simrl;
quit;
proc univariate data=RL2of3_sign;
var delta000 delta020 delta040 delta060 delta080 delta100 delta120;
run;
```

4.5.2 SAS® programs to simulate the run-length distributions of the two-sided precedence charts in Case U

4.5.2.1 The *1-of-1* precedence chart

```
proc iml;
m = 500;
n = 5;
j = (n+1)/2;
sim = 100000;
a = 25;
b = 476;
rlvec = j(sim,13,.);
do delta = 1 to 3 by 0.25;
shift = 4*delta-3;
do k = 1 to sim;
xref = j(m,1,0);
call randgen(xref,'NORMAL'); yref = xref;
call sort(yref, {1});
lcl = yref[a,1];
ucl = yref[b,1];
count = 1;
signal = 0;
above = j(2,1,0);
below = j(2,1,1);
do while (signal = 0);
xfut = j(n,1,0);
call randgen(xfut,'NORMAL',delta-1,1); yfut = xfut;
call sort(yfut, {1});
plotstat = yfut[j,1];
cl=j(2,1,0);
cl[1,1]=ucl;
cl[2,1]=lcl;
plotstatvec=j(2,1,plotstat);
check = plotstatvec <= cl;
if check = above then signal = 1;
else if check = below then signal = 1;
else count = count + 1;
count1 = count;
rlvec[k,shift] = count1;
end;
end;
end;
create RL1of1_Precedence_normal from rlvec[colname={delta000 delta025 delta050
delta075 delta100 delta125 delta150 delta175 delta200 delta225 delta250 delta275
delta300}];
append from rlvec;
quit;
proc univariate data=RL1of1_Precedence_normal;
var delta000 delta025 delta050 delta075 delta100 delta125 delta150 delta175
    delta200 delta225 delta250 delta275 delta300;
run;
```

4.5.2.2 The *2-of-2* DR and the *2-of-2* KL precedence charts

```
proc iml;
m = 500; n = 5; j = (n+1)/2;
sim = 100000;
a = 81; b = 420;
rlvec = j(sim,13,.);
do delta = 1 to 3 by 0.25;
shift = 4*delta-3;
do k = 1 to sim;
xref = j(m,1,0);
call randgen(xref,'NORMAL'); yref = xref;
call sort(yref, {1});
lcl = yref[a,1];
ucl = yref[b,1];
count = 1;
signal = 0;
dummy = j(2,1,0);
check = {1,0};
above = j(2,2,0);
below = j(2,2,1);
abovebelow = {0 1 , 0 1};
belowabove = {1 0 , 1 0};
matrix = j(2,2,0);
do while (signal = 0);
dummy = check;
xfut = j(n,1,0);
call randgen(xfut,'NORMAL',delta-1,1); yfut = xfut;
call sort(yfut, {1});
plotstat = yfut[j,1];
cl=j(2,1,0);
cl[1,1]=ucl;
cl[2,1]=lcl;
plotstatvec=j(2,1,plotstat);
check = plotstatvec <= cl;
matrix = dummy||check;
if matrix = above then signal = 1; * DR and KL;
else if matrix = below then signal = 1; * DR and KL;
else if matrix = abovebelow then signal = 1; * DR only;
else if matrix = belowabove then signal = 1; * DR only;
else count = count + 1;
count1 = count;
rlvec[k,shift] = count1;
end;
end;
end;
create RL2of2_Precedence_normal from rlvec[colname={delta000 delta025 delta050
delta075 delta100 delta125 delta150 delta175 delta200 delta225 delta250 delta275
delta300}];
append from rlvec;
quit;
proc univariate data= RL2of2_Precedence_normal;
var delta000 delta025 delta050 delta075 delta100 delta125 delta150 delta175
    delta200 delta225 delta250 delta275 delta300;
run;
```

4.5.2.3 The *2-of-3* precedence chart

```
proc iml;
m = 500; n = 5; j = (n+1)/2;
sim = 100000;
a = 72; b = 429;
rlvec = j(sim,13,.);
do delta = 1 to 3 by 0.25;
shift = 4*delta-3;
do k = 1 to sim;
xref = j(m,1,0);
call randgen(xref,'NORMAL'); yref = xref;
call sort(yref, {1});
lcl = yref[a,1];
ucl = yref[b,1];
count = 2;
signal = 0;
dummy1 = {1,0}; dummy2 = j(2,1,.); check = {1,0};
between_above_above = {1 0 0,
                       0 0 0};
between_below_below = {1 1 1,
                        0 1 1};
below_between_below = {1 1 1,
                       1 0 1};
above_between_above = {0 1 0,
                       0 0 0};
matrix = j(2,3,.);
do while (signal = 0);
dummy2 = dummy1;
dummy1 = check;
xfut = j(n,1,0);
call randgen(xfut,'NORMAL',delta-1,1); yfut = xfut;
call sort(yfut, {1});
plotstat = yfut[j,1];
cl=j(2,1,0);
cl[1,1]=ucl;
cl[2,1]=lcl;
plotstatvec=j(2,1,plotstat);
check = plotstatvec <= cl;
matrix = dummy2||dummy1||check;
if matrix = between_above_above then signal = 1;
else if matrix = between_below_below then signal = 1;
else if matrix = below_between_below then signal = 1;
else if matrix = above_between_above then signal = 1;
else count = count + 1;
count1 = count;
rlvec[k,shift] = count1;
end;
end;
end;
create RL2of3_Precedence_normal from rlvec[colname={delta000 delta025 delta050
delta075 delta100 delta125 delta150 delta175 delta200 delta225 delta250 delta275
delta300}];
append from rlvec;
quit;
proc univariate data= RL2of3_Precedence_normal;
var delta000 delta025 delta050 delta075 delta100 delta125 delta150 delta175
    delta200 delta225 delta250 delta275 delta300;
run;
```

4.5 Appendix 4A: SAS® programs

4.5.1 SAS® programs to simulate the run-length distributions of the upper one-sided X-bar, sign and SR charts in Case K

4.5.1.1 The *1-of-1* X-bar, sign and SR charts

```
*1-of-1 upper one-sided X-bar chart;

proc iml;
ARL = 370;
sim = 100000;
n = 10;
UCL = probit(1-1/ARL)/sqrt(n);
simrl = j(sim,13,.);
do d = 1 to 2.2 by 0.2;
do j = 1 to sim;
ct = 0;
do k = 1 to 10000000 while ( ^((ct>=UCL))  );
x = j(n,1,.);
call randgen(x,'NORMAL',d-1,1);
ct = sum(x)/n;
rl = k;
end;
simrl[j,d*5+1]=rl;
end;
end;
create RL1of1_Xbar from simrl[colname={delta000
delta020 delta040 delta060 delta080 delta100 delta120}];
append from simrl;
quit;
proc univariate data=RL1of1_Xbar;
var delta000 delta020 delta040 delta060 delta080 delta100 delta120;
run;
```

```
*1-of-1 upper one-sided sign chart;

proc iml;
ARL = 370;
sim = 100000;
a = 0;
n = 10;
UCL = n-a;
med = j(n,1,0);
simrl = j(sim,13,.);
q = (1/ARL - 1 + probbnml(0.5,n,n-a-1) ) / (probbnml(0.5,n,n-a-1) - probbnml(0.5,n,n-
a-2)) ;
do d = 1 to 2.2 by 0.2;
do j = 1 to sim;
ct = 0;
random = 0;
do k = 1 to 10000000 while ( ^((ct>=UCL)|((ct=UCL-1)&(random<=q)))  );
x = j(n,1,.);
call randgen(x,'NORMAL',d-1,1);
vec = x > med;
ct = sum(vec);
random = ranuni(0);
rl = k;
end;
simrl[j,d*5+1]=rl;
end;
end;
*print simrl;
create RL1of1_sign from simrl[colname={delta000
delta020 delta040 delta060 delta080 delta100 delta120}];
append from simrl;
quit;
proc univariate data=RL1of1_sign;
var delta000 delta020 delta040 delta060 delta080 delta100 delta120;
run;
```

```
*1-of-1 upper one-sided SR chart;

proc iml;
ARL = 370;
sim = 100000;
n = 10;
UCL = 53;
UCL1 = 51;
cdfUCL = 0.002;
pmfUCL1 = 0.0009;
med = j(n,1,0);
simrl = j(sim,13,.);
q = (1/ARL - cdfUCL) / ( pmfUCL1 ) ;
do d = 1 to 2.2 by 0.2;
do j = 1 to sim;
random = 0;
ct = 0;
do k = 1 to 10000000 while ( ^(   (ct>=UCL)  |  ((ct=UCL1)&(random<=q))   )  );
x = j(n,1,.);
call randgen(x,'NORMAL',d-1,1);
vec   = x > med;
wplus = (vec`)*rank(abs(x));
ct = 2*wplus - n*(n+1)/2;
random = ranuni(0);
rl = k;
end;
simrl[j,d*5+1]=rl;
end;
end;
create RL1of1_SR from simrl[colname={delta000
delta020 delta040 delta060 delta080 delta100 delta120}];
append from simrl;
quit;
proc univariate data=RL1of1_SR;
var delta000 delta020 delta040 delta060 delta080 delta100 delta120;
run;
```

4.5.1.2 The *2-of-2* sign and SR charts

```
*2-of-2 upper one-sided sign chart;

proc iml;
ARL = 370;
sim = 100000;
a = 1;
n = 10;
UCL = n-a;
med = j(n,1,0);
simrl = j(sim,13,.);
q = ( (sqrt(4*ARL+1)+1)/(2*ARL) - 1 + probbnml(0.5,n,n-a-1) ) /
    (probbnml(0.5,n,n-a-1) - probbnml(0.5,n,n-a-2)) ;
do d = 1 to 2.2 by 0.2;
do j = 1 to sim;
x = j(n,1,.);
ct1 = 0;
ct = 0;
random1 = 0;
random = 0;
do k = 1 to 10000000 while (
^( (  (ct1>=UCL)&(ct>=UCL)  )                        |
   (  ((ct1=UCL-1)&(random1<=q))&(ct>=ucl)  )        |
   (  (ct1>=UCL)&((ct=UCL-1)&(random<=q))  )         |
   (  ((ct1=UCL-1)&(random1<=q))&((ct=UCL-1)&(random<=q)))  )  );
ct1 = ct;
call randgen(x,'NORMAL',d-1,1);
vec = x > med;
ct = sum(vec);
random = ranuni(0);
random1 = ranuni(0);
rl = k;
end;
simrl[j,d*5+1]=rl;
end;
end;
create RL2of2_sign from simrl[colname={delta000
delta020 delta040 delta060 delta080 delta100 delta120}];
append from simrl;
quit;
proc univariate data=RL2of2_sign;
var delta000 delta020 delta040 delta060 delta080 delta100 delta120;
run;
```

```
*2-of-2 upper one-sided SR chart;

proc iml;
ARL = 370;
sim = 100000;
n = 10;
UCL = 33;
UCL1 = UCL - 2;
cdfUCL = 0.0527;
pmfUCL1 = 0.0127;
med = j(n,1,0);
simrl = j(sim,13,.);
q = (  (sqrt(4*ARL+1)+1)/(2*ARL) -  cdfUCL ) / ( pmfUCL1 ) ;
do d = 1 to 2.2 by 0.2;
do j = 1 to sim;
x = j(n,1,.);
ct1 = 0;
ct = 0;
random1 = 0;
random = 0;
do k = 1 to 10000000 while (
^( (  (ct1>=UCL)&(ct>=UCL)  )                        |
 (  ((ct1=UCL1)&(random1<=q))&(ct>=ucl)  )     |
 (  (ct1>=UCL)&((ct=UCL1)&(random<=q))  )      |
 (  ((ct1=UCL1)&(random1<=q))&((ct=UCL1)&(random<=q)))  )  );
ct1 = ct;
call randgen(x,'NORMAL',d-1,1);
vec   = x > med;
wplus = (vec`)*rank(abs(x));
ct = 2*wplus - n*(n+1)/2;
random = ranuni(0);
random1 = ranuni(0);
rl = k;
end;
simrl[j,d*5+1]=rl;
end;
end;
create RL2of2_SR from simrl[colname={delta000
delta020 delta040 delta060 delta080 delta100 delta120}];
append from simrl;
quit;
proc univariate data=RL2of2_SR;
var delta000 delta020 delta040 delta060 delta080 delta100 delta120;
run;
```

4.5.1.3 The *2-of-3* sign chart

```
*2-of-3 upper-sided sign chart;

proc iml;
ARL = 370;
sim = 100000;
a = 1;
n = 10;
UCL = n-a;
med = j(n,1,0);
simrl = j(sim,13,.);
q = 0.632202808;
do d = 1 to 2.2 by 0.2;
do j = 1 to sim;
x = j(n,1,.);
ct2 = 0;
ct1 = 0;
ct = 0;
random2 = 0;
random1 = 0;
random = 0;
do k = 1 to 10000000 while (
^(
 ( (ct2>=UCL)                 & (ct1<UCL) &  (ct>=UCL)               )     |
 ( ((ct2=UCL-1)&(random2<=q)) & (ct1<UCL) &  (ct>=UCL)               )     |
 ( (ct2>=UCL)                 & (ct1<UCL) & ((ct=UCL-1)&(random<=q)) )     |
 ( ((ct2=UCL-1)&(random2<=q)) & (ct1<UCL) & ((ct=UCL-1)&(random<=q)) )     |
 ( (ct1>=UCL)                 & (ct2<UCL) &  (ct>=UCL)               )     |
 ( ((ct1=UCL-1)&(random1<=q)) & (ct2<UCL) &  (ct>=UCL)               )     |
 ( (ct1>=UCL)                 & (ct2<UCL) & ((ct=UCL-1)&(random<=q)) )     |
 ( ((ct1=UCL-1)&(random1<=q)) & (ct2<UCL) & ((ct=UCL-1)&(random<=q)) )      )   );
ct2 = ct1;
ct1 = ct;
call randgen(x,'NORMAL',d-1,1);
vec = x > med;
ct = sum(vec);
random = ranuni(0);
random1 = ranuni(0);
random2 = ranuni(0);
rl = k;
end;
simrl[j,d*5+1]=rl;
end;
end;
ARL = sum(simrl)/sim;
create RL2of3_sign from simrl[colname={delta000
delta020 delta040 delta060 delta080 delta100 delta120}];
append from simrl;
quit;
proc univariate data=RL2of3_sign;
var delta000 delta020 delta040 delta060 delta080 delta100 delta120;
run;
```

4.5.2 SAS® programs to simulate the run-length distributions of the two-sided precedence charts in Case U

4.5.2.1 The *1-of-1* precedence chart

```
proc iml;
m = 500;
n = 5;
j = (n+1)/2;
sim = 100000;
a = 25;
b = 476;
rlvec = j(sim,13,.);
do delta = 1 to 3 by 0.25;
shift = 4*delta-3;
do k = 1 to sim;
xref = j(m,1,0);
call randgen(xref,'NORMAL'); yref = xref;
call sort(yref, {1});
lcl = yref[a,1];
ucl = yref[b,1];
count = 1;
signal = 0;
above = j(2,1,0);
below = j(2,1,1);
do while (signal = 0);
xfut = j(n,1,0);
call randgen(xfut,'NORMAL',delta-1,1); yfut = xfut;
call sort(yfut, {1});
plotstat = yfut[j,1];
cl=j(2,1,0);
cl[1,1]=ucl;
cl[2,1]=lcl;
plotstatvec=j(2,1,plotstat);
check = plotstatvec <= cl;
if check = above then signal = 1;
else if check = below then signal = 1;
else count = count + 1;
count1 = count;
rlvec[k,shift] = count1;
end;
end;
end;
create RL1of1_Precedence_normal from rlvec[colname={delta000 delta025 delta050
delta075 delta100 delta125 delta150 delta175 delta200 delta225 delta250 delta275
delta300}];
append from rlvec;
quit;
proc univariate data=RL1of1_Precedence_normal;
var delta000 delta025 delta050 delta075 delta100 delta125 delta150 delta175
    delta200 delta225 delta250 delta275 delta300;
run;
```

4.5.2.2 The *2-of-2* DR and the *2-of-2* KL precedence charts

```
proc iml;
m = 500; n = 5; j = (n+1)/2;
sim = 100000;
a = 81; b = 420;
rlvec = j(sim,13,.);
do delta = 1 to 3 by 0.25;
shift = 4*delta-3;
do k = 1 to sim;
xref = j(m,1,0);
call randgen(xref,'NORMAL'); yref = xref;
call sort(yref, {1});
lcl = yref[a,1];
ucl = yref[b,1];
count = 1;
signal = 0;
dummy = j(2,1,0);
check = {1,0};
above = j(2,2,0);
below = j(2,2,1);
abovebelow = {0 1 , 0 1};
belowabove = {1 0 , 1 0};
matrix = j(2,2,0);
do while (signal = 0);
dummy = check;
xfut = j(n,1,0);
call randgen(xfut,'NORMAL',delta-1,1); yfut = xfut;
call sort(yfut, {1});
plotstat = yfut[j,1];
cl=j(2,1,0);
cl[1,1]=ucl;
cl[2,1]=lcl;
plotstatvec=j(2,1,plotstat);
check = plotstatvec <= cl;
matrix = dummy||check;
if matrix = above then signal = 1; * DR and KL;
else if matrix = below then signal = 1; * DR and KL;
else if matrix = abovebelow then signal = 1; * DR only;
else if matrix = belowabove then signal = 1; * DR only;
else count = count + 1;
count1 = count;
rlvec[k,shift] = count1;
end;
end;
end;
create RL2of2_Precedence_normal from rlvec[colname={delta000 delta025 delta050
delta075 delta100 delta125 delta150 delta175 delta200 delta225 delta250 delta275
delta300}];
append from rlvec;
quit;
proc univariate data= RL2of2_Precedence_normal;
var delta000 delta025 delta050 delta075 delta100 delta125 delta150 delta175
    delta200 delta225 delta250 delta275 delta300;
run;
```

4.5.2.3 The *2-of-3* precedence chart

```
proc iml;
m = 500; n = 5; j = (n+1)/2;
sim = 100000;
a = 72; b = 429;
rlvec = j(sim,13,.);
do delta = 1 to 3 by 0.25;
shift = 4*delta-3;
do k = 1 to sim;
xref = j(m,1,0);
call randgen(xref,'NORMAL'); yref = xref;
call sort(yref, {1});
lcl = yref[a,1];
ucl = yref[b,1];
count = 2;
signal = 0;
dummy1 = {1,0}; dummy2 = j(2,1,.); check = {1,0};
between_above_above = {1 0 0,
                       0 0 0};
between_below_below = {1 1 1,
                        0 1 1};
below_between_below = {1 1 1,
                       1 0 1};
above_between_above = {0 1 0,
                       0 0 0};
matrix = j(2,3,.);
do while (signal = 0);
dummy2 = dummy1;
dummy1 = check;
xfut = j(n,1,0);
call randgen(xfut,'NORMAL',delta-1,1); yfut = xfut;
call sort(yfut, {1});
plotstat = yfut[j,1];
cl=j(2,1,0);
cl[1,1]=ucl;
cl[2,1]=lcl;
plotstatvec=j(2,1,plotstat);
check = plotstatvec <= cl;
matrix = dummy2||dummy1||check;
if matrix = between_above_above then signal = 1;
else if matrix = between_below_below then signal = 1;
else if matrix = below_between_below then signal = 1;
else if matrix = above_between_above then signal = 1;
else count = count + 1;
count1 = count;
rlvec[k,shift] = count1;
end;
end;
end;
create RL2of3_Precedence_normal from rlvec[colname={delta000 delta025 delta050
delta075 delta100 delta125 delta150 delta175 delta200 delta225 delta250 delta275
delta300}];
append from rlvec;
quit;
proc univariate data= RL2of3_Precedence_normal;
var delta000 delta025 delta050 delta075 delta100 delta125 delta150 delta175
    delta200 delta225 delta250 delta275 delta300;
run;
```

Chapter 5

Concluding remarks: Summary and recommendations for future research

To finish-off this thesis, we give here a brief summary of the research conducted in the thesis and offer concluding remarks concerning unanswered questions and/or future research opportunities.

In this thesis, in general, we focused on a variety of aspects related to the basic (yet powerful) statistical tool often used in quality improvement efforts within the realm of statistical quality control, that is, the Shewhart-type of control chart. First, we looked at Shewhart-type Phase I variables charts for the variance, the standard deviation and the range; this was followed by an overview of the literature on Shewhart-type Phase I variables charts for the location and the spread of a process. Second, we studied the Shewhart-type Phase II *p*-chart and the Shewhart-type Phase II *c*-chart in Case U (when the parameters are unknown) and assessed the influence when the parameters are estimated from a Phase I sample on the performance of these charts; both these charts are attributes charts and are widely used in practice. Lastly, we developed a new class of nonparametric Shewhart-type Phase I and Phase II control charts, for monitoring or controlling a certain quantile of the underlying probability distribution of a process, based on runs-type signaling rules using the well-known sign test and the two-sample median test statistic as plotting statistics. In the next few paragraphs we point out the highlights of the research carried out in this thesis and state some research ideas to be pursued in the near future. We also list the research outputs related to this thesis; this includes a list of technical reports and peer-reviewed articles that were published in international journals, contributions to local and international conferences where the author of this thesis presented papers and some draft articles that were submitted and are currently under review.

Variables control charts

Assuming that the underlying process distribution follows a normal distribution with an unknown mean and an unknown variance, in Chapter 2 we studied the design of the well-known Shewhart-type

S^2, S and R charts for Phase I applications based on the availability of m independent rational subgroups each of size $n > 1$. We showed that, because multiple plotting statistics are simultaneously compared to the same set of estimated control limits, the signaling events (i.e. the event when a plotting statistic plots outside the control limits) are mutually dependent. We further argued (with reference to the article of Champ and Jones, (2004)) that the correct design criterion of Shewhart-type Phase I charts is the false alarm probability (FAP), which is the probability of at least one false alarm, and not the false alarm rate (FAR), which deals with only one plotting statistic at a time and is defined as the probability of a signal at any particular sampling stage. Accordingly, we found the appropriate charting constants for a variety of (m, n)-combinations for each of the three charts (using intensive computer simulation experiments) so that the FAP of each chart does not exceed 0.01, 0.05 and 0.10, respectively.

The literature overview, in Chapter 2, regarding univariate parametric Phase I Shewhart-type charts for the location and the spread of a process not only presented the current state of the art of constructing these charts, but also brought several important points under our attention:

(i) There is a lack of proper guidance to the practitioner on the correct statistical design and implementation of Phase I charts. In a search of the standard statistical process control textbooks on the market, none to very little material was found, including the standard book of Montgomery (2005), who discusses the topic without the necessary statistical theory.

(ii) Some of the authors that studied the Phase I problem (especially when the process parameters are estimated) ignore the dependency between the Phase I plotting statistics and incorrectly used the FAR (which only deals with a single plotting statistic at a time) to design the charts as apposed to the FAP (which takes into account that multiple charting statistics are simultaneously compared to the estimated control limits). This would certainly deteriorate the performance of these charts. Our methodology provides the correct control limits for the applications studied.

(iii) There seems to be no consensus on exactly how one should compare the performance of competing Phase I charts. This boils down to the question of how to formulate and define an out-of-control situation in Phase I. One current proposal is to adopt the scenario that one of the Phase I samples is out-of-control and that the remaining $m-1$ samples are in-control and then (via computer simulation) compare the empirical

probability that at least one point plots outside the estimated control limits. The chart with the highest empirical probability of detecting the out-of-control sample is then declared the winner; this can be investigated further.

(iv) There is a genuine need to develop a Phase I control chart for the case when $n = 1$, that is, for individuals data. Admittedly there are some articles available in the literature that address the problem (see e.g. Nelson (1982), Roes, Does and Schurink (1993), Rigdon, Cruthis and Champ (1994) and Bryce, Gaudard and Joiner (1997)) but the problem has not yet been solved satisfactorily. The main stumbling block appears to be finding a suitable point estimator for the variance or the standard deviation and deriving the exact joint distribution of the standardized plotting statistics. Since individuals data is so common nowadays in many industries, this problem is important and will be studied using methods similar to the ones in this thesis.

(v) The design of Phase I control charts for correlated data needs to be looked at. A good starting point is the articles by Boyles (2000) and Maragah and Woodall (1992).

(vi) Except for the study by Borror and Champ (2001), there is apparently no other published work regarding the design of Phase I Shewhart-type attributes charts. This is an important aspect because the study of the Phase II *p*-chart and the Phase II *c*-chart is based on the availability of an in-control reference sample, which is usually obtained at the end of a successful Phase I study.

(vii) It would definitely be helpful and beneficial to the quality practitioner if a unified approach to the design and implementation of Phase I variables and attributes charts is available; this is a topic currently under investigation by the author of this thesis and his supervisors.

Attributes control charts

The Phase II Shewhart-type *p*-chart and *c*-chart were studied in detail in Chapter 3. The aim was to determine the effect of estimating the unknown process parameters from a Phase I reference sample on the performance of the charts in their Phase II application. The methodology that we used was based on the two-step procedure which was introduced in the statistical process control arena by Chakraborti (2000). The procedure entails that we first condition on a particular observed value of the point

estimator from Phase I (in order to obtain the conditional Phase II run-length distribution and the associated characteristics of the conditional run-length distribution) and then calculate the unconditional Phase II run-length distribution and the associated properties of the unconditional run-length distribution by averaging over all the values of the point estimator. We numerically investigated the various properties of the conditional and the unconditional run-length distributions, for the in-control and the out-of-control scenarios, and compared the results to the benchmark values of Case K (i.e. when the parameters are known). It was found that the widely-followed guidelines regarding the number of Phase I rational subgroups, m, and the sample size, n, is not adequate to control the average run-length and/or the false alarm rate at acceptable levels. The cause of the discrepancy between the attained ARL and the attained FAR values and the industry standards of 370 and 0.0027 (respectively) is twofold. The discrepancy is partly due to the fact that the underlying process distributions are discrete and to some extent it is caused by the fact that the standard formula, i.e. mean $\pm$ 3 $\times$ standard deviation, for calculating the control limits, is not 100% correct; this is so because the normal approximation to the binomial and the Poisson distributions is not very good for all values of the parameters p and c (especially p close to 0 or 1 and c close to 0).

The question of how we can correctly design the Phase II Shewhart-type *p*-chart and *c*-chart remains, in some way, unanswered. As pointed out in an earlier section, the formulae for the characteristics of the unconditional run-length distribution can be helpful in this regard and there are two possible routes to follow:

(i) The usual approach is to specify a certain attribute of the unconditional Phase II run-length distribution (such as the unconditional average run-length, which is common in routine applications (see e.g. Chakraborti, (2006))) and then solve for the charting constant(s).

Even though this approach is viable, it would only be successful insofar it is possible to accurately specify the unknown parameters p and c. The reason for this drawback is the fact that the unconditional properties of the charts are unconditional only with respect to the point estimators and not with respect to the unknown parameters.

(ii) A second approach one can pursue is to also uncondition on the properties of the charts with respect to the parameters p and c. This approach, which is closely linked to a

Bayes approach, entails that we treat p and c as random variables and that we choose appropriate (prior) distributions to model the uncertainty in the parameters.

As suggested earlier, the standard beta distribution (with support on the interval (0,1)) and the gamma distribution (with the positive real numbers as support) would work. However, the dilemma in this approach is that we still require expert knowledge and guidance when choosing the parameters of the beta and the gamma distributions.

Currently, the topic of finding and comparing suitable charting constants for the application of the Shewhart-type Phase II *p*-chart and *c*-chart is underway by the author of this thesis and co-workers.

Nonparametric Shewhart-type control charts with runs-type signaling rules

Lastly, in Chapter 4 we designed new nonparametric control charts based on runs-type signaling rules using the well-known sign test statistic and the two-sample median test statistic as plotting statistics. The sign test was used in the design of the charts when the percentile under investigation of the underlying process distribution was known (or specified) whereas the two-sample median test was used to construct the charts when the percentile was unknown. The main advantages of the nonparametric charts are:

(i) The fact that the underlying distribution needs not be specified (as we only require continuity of the distribution);

(ii) The precise numerical measurements need not be available (because we only count the number of observations greater or smaller than a specified value or simply rank the observations within each sample). Neither the counting nor the ranking procedure requires exact measurements;

(iii) The sign charts have the added advantage that they can be applied in scenarios where the data are just dichotomous (e.g. yes/no); and

(iv) The precedence charts give us the flexibility to apply the chart in situations where the data is naturally collected in an ordered fashion (e.g. time to failure).

We derived the run-length distributions of this new class of distribution-free control charts using a Markov chain approach and, where possible, we also used the results related to the geometric

distribution of order k. Where necessary we again used the two-step conditioning and unconditioning idea by Chakraborti (2000) to obtain the Phase II run-length distributions.

Having derived the run-length distributions and the associated characteristics of the new charts, extensive tables were provided with the suitable charting constants for each chart which should help the practitioner in the setting up of the charts. A numerical example was also given to illustrate the implementation and operation of the charts. However, having pointed out the benefits of the new runs rules enhanced charts, there are two important aspects concerning nonparametric control charts (in general) that are worth mentioning:

(i) There is a major shortcoming regarding the application of the nonparametric charts in industry because there is a lack of a proper understanding (and perhaps an appreciation) of the topic nonparametrics and consequently the important role these charts can play in practice.

The main reason for this limitation seems to be that distribution-free (nonparametric) methods are typically only touched on in undergraduate statistics courses in most programs and are not necessarily taught at a post-graduate level and, in most cases, not even taught to the engineers and/or the operator personnel who have to deal with the monitoring of the processes. What is more, is the fact that none of the available (standard) textbooks on statistical process control covers the topic of nonparametric control charting procedures in any detail.

(ii) It would be a great improvement and definitely to the advantage of the quality practitioner if software developers were to include the nonparametric control charts that are already available, as standard options or procedures in their statistical computer packages. Currently, these nonparametric control chart procedures are not available for practitioners and they simply resort to the standard parametric control chart methodologies.

Research outputs

A number of research outputs related to and based on this thesis have seen the light. Below we provide a list with the details of the technical reports and the peer-reviewed articles that were published, the articles that were accepted for publication, the local and the international conferences where papers were presented and the draft articles that were submitted and currently under review.

Published articles

(i) Chakraborti, S., Human, S. W. (2006). "Parameter estimation and performance of the *p*-chart for attributes data". *IEEE Transactions on Reliability*, 55(3):559-566;

(ii) Chakraborti, S., Human, S. W. (2008). "Properties and performance of the *c*-chart for attributes data". *Journal of Applied Statistics*, 35(1):89-100;

(iii) Chakraborti, S., Human, S.W., Graham, M.A. (2008). "Phase I statistical process control charts: An overview and some results". *Quality Engineering*, 21(1):52-62; and

(iv) Chakraborti, S., Eryilmaz, S., Human, S. W. (2009). "A Phase II nonparametric control chart based on precedence statistics with runs-type signaling rules". *Computational Statistics and Data Analysis*, 53(1):1054-1065.

Articles accepted for publication

(i) Human, S. W., Chakraborti, S., Smit, C. F. "Nonparametric Shewhart-type sign control charts based on runs". *Communications in Statistics – Theory and Methods.*

Articles under review

(ii) Human, S. W., Chakraborti, S., Smit, C. F. "Control charts for variation in Phase I applications", Submitted to *Computational Statistics and Data Analysis*.

Technical reports

(i) Human, S. W., Chakraborti, S., Smit, C. F. (2009). "Shewhart-type S^2, S and R control charts for Phase I applications". Technical Report 09/01, Department of Statistics, University of Pretoria, ISBN: 978-1-86854-735-7.

(ii) Human, S. W., Chakraborti, S., Smit, C. F. (2009). "Nonparametric Shewhart-type control charts with runs-type signaling rules". Technical Report 09/02, Department of Statistics, University of Pretoria, ISBN: 978-1-86854-738-8.

International conference

(i) The 7th World Congress in Probability and Statistics in Singapore jointly sponsored by the Bernoulli Society and the Institute of Mathematical Statistics (2008) where the results related to the nonparametric control charts of Chapter 4 was presented.

Local conferences

(i) The annual conference of the South African Statistical Association (SASA) hosted by the Department of Statistics of the Rhodes University in Grahamstown (2005) where the results related to the Phase II *p*-chart of Chapter 3 was presented;

(ii) The annual conference of the South African Statistical Association (SASA) hosted by the Department Statistics and Actuarial Science of the University of Stellenbosch (2006) where the results related to the Phase II *c*-chart of Chapter 3 was presented;

(iii) The annual conference of the South African Statistical Association (SASA) hosted by the Department of Statistics and Actuarial Science of the University of Witwatersrand (2007) where the results related to the Phase I S^2, S and R control charts of Chapter 2 was presented; and

(iv) The annual conference of the South African Statistical Association (SASA) hosted by the Department of Statistics of the University of Pretoria (2008) where the results related to the nonparametric control charts of Chapter 4 was presented.

The end.

References

Acosta-Mejia, C. A. (1999). "Improved p charts to monitor process quality". *IIE Transactions*, 31(6):509-516.

Albers, W., Kallenberg, W. C. M. (2004). "Empirical non-parametric control charts: Estimation effects and corrections". *Journal of Applied Statistics*, 31(3):345-360.

Albers, W., Kallenberg, W. C. M., Nurdiati, S. (2006). "Data driven choice of control charts". *Journal of Statistical Planning and Inference*, 136(3):909-941.

Amin, R. W., Reynolds, M. R. Jr., Bakir, S. (1995). "Nonparametric quality control charts based on the sign statistic". *Communication in Statistics - Theory and Methods*, 24(6):1597-1623.

Bain, L. J., Engelhardt, M. (1992). *Introduction to probability and mathematical statistics*, 2nd edition, Duxbury Press.

Bakir, S. T. (2004). "A distribution-free Shewhart quality control chart based on signed-ranks". *Quality Engineering*, 16(4):613-623.

Balakrishnan, N., Read, C. B., Vidakovic, B. (2006). *Encyclopedia of Statistical Sciences*, 2nd edition. John Wiley.

Balakrishnan, N., Koutras, M. V. (2002). *Runs and scans with applications*, John Wiley.

Borror, C. M., Champ, C. W. (2001). "Phase I control charts for independent Bernoulli data". *Quality and Reliability Engineering International*, 17(5):391-396.

Boyles, R. A. (2000). "Phase I analysis for autocorrelated processes". *Journal of Quality Technology*, 32(4):395-409.

Bryce, G. R., Gaudard, M. A., Joiner, B. L. (1997). "Estimating the standard deviation for individuals control charts". *Quality Engineering*, 10(2):331-341.

Casella, G., Berger, R. L. (2002). *Statistical inference*, 2nd edition. Duxbury.

Chakraborti, S. (2000). "Run length, average run length and false alarm rate of Shewhart X-bar chart: Exact derivations by conditioning". *Communications in Statistics - Simulation and Computation*, 29(1):61-81.

Chakraborti, S. (2006). "Parameter estimation and design considerations in prospective applications of the $\overline{X}$ chart". *Journal of Applied Statistics*, 33(4):439-459.

Chakraborti, S. (2007). "Run length distribution and percentiles: The Shewhart $\overline{X}$ chart with unknown parameters". *Quality Engineering*, 19(2):119-127.

Chakraborti, S., Eryilmaz, S. A. (2007). "A nonparametric Shewhart-type signed-rank control chart based on runs". *Communications in Statistics- Simulation and Computation*, 36(2):335-356.

Chakraborti, S., Eryilmaz, S., Human S. W. (2009). "A phase II nonparametric control chart based on precedence statistics with runs-type signaling rules". *Computational Statistics and Data Analysis* 53(4):1054-1065.

Chakraborti, S., Graham, M. A. (2007). "Nonparametric control charts". *Encyclopedia of Quality and Reliability*, Volume 1:415-429, John Wiley.

Chakraborti, S., Human, S. (2006). "Parameter estimation and performance of the *p*-chart for attributes data". *IEEE Transactions on Reliability*, 55(3):559-566.

Chakraborti, S., Human, S. (2008). "Properties and performance of the *c*-chart for attributes data". *Journal of Applied Statistics*, 35(1):89-100.

Chakraborti, S., Human, S., Graham, M. A. (2009). "Phase I statistical process control charts: An overview and some results". *Quality Engineering*, 21(1):52-62.

Chakraborti, S., Van der Laan, P., Bakir, S. T. (2001). "Nonparametric control charts: An overview and some results". *Journal of Quality Technology*, 33(3):304-315.

Chakraborti, S., Van der Laan, P., Van de Wiel, M. A. (2004). "A class of distribution-free control charts". *Applied Statistics*, 53(3):443-462.

Champ, C. W., Chou, S. P. (2003). "Comparison of standard and individual limits Phase I Shewhart $\overline{X}$, R, and S charts". *Quality and Reliability Engineering International*, 19(2):161-170.

Champ, C. W., Jones, L. A. (2004). "Designing Phase I $\overline{X}$ charts with small sample sizes". *Quality and Reliability Engineering International*, 20(5):497-510.

Chou, S. P., Champ, C. W. (1995). "A comparison of two Phase I control charts". *Proceedings of the Quality and Productivity Section of the American Statistical Association*, 31-35.

Crowder, S. W. (1987). "A simple method for studying run-length distributions of exponentially weighted moving average charts". *Technometrics*, 29(4):401-407.

Crowder, S. W. (1989). "Design of exponentially weighted moving average schemes". *Journal of Quality Technology*, 21(3):155-162.

Deming, W. E. (1986): *Out of the crisis*, Cambridge.

Derman, C., Ross, S. M. (1997). *Statistical aspects of quality control*, Academic Press.

Eisenhart, C., Hastay, M. W., Wallis, W. A. (1947). *Selected techniques of statistical analysis for scientific and industrial research and production and management engineering*, 1st edition, McGraw-Hill.

Farnum, N. R. (1994) *Modern statistical quality control and improvement*, Duxbury

Fu, J. C., Lou, W. Y. W. (2003). Distribution theory of runs and patterns and its applications: A finite Markov chain imbedding technique, World Scientific Publishing.

George, E. O., Bowman, D. (1995). "A full likelihood procedure for analyzing exchangeable binary data". *Biometrics*, 51(2):512-523.

Gibbons, J. D., Chakraborti, S. (2003). *Nonparametric statistical inference*, 4th edition, Marcel Dekker.

Gupta, A. K., Nadarajah, S. (2004). *Handbook of beta distribution and its applications*, Marcel Dekker.

Hawkins, D. M. (1977). "Testing a sequence of observations for a shift in location". *Journal of the American Statistical Association*, 72(357):180-186.

Hawkins, D. M., Olwell, D. H. (1998). *Cumulative sum charts and charting for quality improvement*, Springer.

Hawkins, D. M., Qiu, P., Kang, C. W. (2003). "The changepoint model for statistical process control". *Journal of Quality Technology*, 35(4):355-366.

Hawkins, D. M., Zamba, K. D. (2005). "Statistical process control for shifts in mean or variance using a change-point formulation". *Technometrics*, 47(2):164-173.

Hillier, F.S. (1969). "$\overline{X}$ and R chart control limits based on a small number of subgroups". *Journal of Quality Technology*, 1(1):17-26.

Hogg, R. V., McKean, J. W., Craig, A. T. (2005). *Introduction to mathematical statistics*, 6th edition, Prentice Hall.

Human, S. W., Chakraborti, S., Smit, C. F. (2009). "Shewhart-type S^2, S and R control charts for Phase I applications". Technical Report 09/01, Department of Statistics, University of Pretoria, ISBN: 978-1-86854-735-7.

Human, S. W., Chakraborti, S., Smit, C. F. (2009). "Nonparametric Shewhart-type control charts with runs-type signaling rules". Technical Report 09/02, Department of Statistics, University of Pretoria, ISBN: 978-1-86854-738-8

Human, S. W., Chakraborti, S., Smit, C. F. (Accepted). "Nonparametric Shewhart-type sign control charts based on runs". *Communications in Statistics – Theory and Methods*.

Human, S. W., Chakraborti, S., Smit, C. F. (Submitted). "Design of S^2, S and R control charts for Phase I application". *Computational Statistics and Data Analysis*.

Human, S. W., Graham, M. A. (2007). "Average run lengths and operating characteristic curves". *Encyclopedia of Quality and Reliability*, Volume 1:159-168, John Wiley.

Jensen, W. A., Jones-Farmer, L. A., Champ, C. W., Woodall, W. H. (2006). "Effects of parameter estimation on control chart properties: A literature review". *Journal of Quality Technology*, 38(4):349-364.

Johnson, N. L., Kemp, A. W., Kotz, S. (2005). *Univariate discrete distributions*, 3rd edition, John Wiley.

Johnson, N. L., Kotz, S., Balakrishnan, N. (1994). *Continuous univariate distributions: Volume 1*, 2nd edition, John Wiley.

Johnson, N. L., Kotz, S., Balakrishnan, N. (1995). *Continuous univariate distributions: Volume 2*, 2nd edition, John Wiley.

Jones, L. A., Champ, C. W. (2002). "Phase I control charts for times between events". *Quality and Reliability Engineering International*, 18(6):479-488.

Jones, L. A., Champ, C. W., Rigdon, S. E. (2004). "The run length distribution of the CUSUM with estimated parameters". *Journal of Quality Technology*, 36(1):95-108.

King, E. P. (1954). "Probability limits for the average chart when process standards are unspecified". *Industrial Quality Control*, 10(6):62-64.

Klein, M. (2000). "Two alternatives to the Shewhart $\bar{X}$ control chart". *Journal of Quality Technology*, 32(4):427-431.

Koning, A. J. (2006). "Model-based control charts in phase 1 statistical process control". *Statistica Neerlandica*, 60(3):327-338.

Kotz, S., Balakrishnan, N., Johnson, N. L. (2000). *Continuous multivariate distributions, Volume 1: Models and Applications*, 2nd edition, John Wiley.

Krishnan, M. (1967). "The noncentral bivariate chi distribution". *SIAM Review*, 9(4):708-714.

Lucas, J. M., Saccucci, M. S. (1990). "Exponentially weighted moving average control schemes: Properties and enhancements". *Technometrics*, 32(1):1-12.

Maragah, H. D., Woodall, W. H. (1992). "The effect of autocorrelation on retrospective X chart". *Journal of Statistical Computation and Simulation*, 40(1):29-42.

Mason, R. L., Young, J. C. (2002). *Multivariate statistical process control with industrial applications*, ASA-SIAM Series on Statistics and Applied Probability.

Mathcad®14.0. Parametric Technology Corporation.

Mathematica®. Wolfram Research.

Montgomery, D. C. (2001). *Introduction to statistical quality control*, 4th edition, John Wiley.

Montgomery, D. C. (2005). *Introduction to statistical quality control*, 5th edition, John Wiley.

Nedumaran, G., Pignatiello, J. J. (2005). "On constructing retrospective $\bar{X}$ control chart limits". *Quality and Reliability Engineering International*, 21(1):81-89.

Nelson, L. S. (1982). "Control charts for individual measurements", *Journal of Quality Technology*, 14(3):172-173.

Nelson, P. R. (1982). "Multivariate normal and *t* distributions with $\rho_{jk} = \alpha_j \alpha_k$". *Communications in Statistics- Simulation and Computation*, 11(2):239-248.

Nelson, L. S. (1984). "The Shewhart control chart-tests for special causes.". *Journal of Quality Technology*, 16(4):237-239.

Nelson, P. R., Wludyka, P. S. Copeland, K. A. F. (2005): *The analysis of means: A graphical method for comparing means, rates and proportions*, ASA-SIAM Series on Statistics and Applied Probability.

Patnaik, P. B. (1950). "The use of mean range as an estimator of variance in statistical tests". *Biometrika*, 37(1):78-87.

Quesenberry, C. P. (1991). "SPC *Q* charts for start-up processes and short or long runs". *Journal of Quality Technology*, 23(3):213-224.

Page, E. S. (1954). "Continuous inspection schemes". *Biometrika*, 41(1):100-115.

Page, E. S. (1961). "Cumulative sum charts". Technometrics, 3(1):1-9.

Radson, D., Boyd, A. H. (2005). "Graphical representation of run length distributions". *Quality Engineering*, 17(2):301-308.

Rigdon, S. E., Cruthis, E. N., Champ, C. W. (1994). "Design strategies for individuals and moving range control charts". *Journal of Quality Technology*, 26(4):274-287.

Roberts, S. W. (1959). "Control chart tests based on geometric moving averages". *Technometrics*, 1(3):239-250.

Roes, K. T. C., Does, R. J. J. M., Schurink, Y. (1993). "Shewhart-type control charts for individual observations", *Journal of Quality Technology*, 25(3):188-198.

Ruggeri, F., Kennett, R. S., Faltin, F. W. (2007). *Encyclopedia of statistics in quality and reliability*, John Wiley.

Ryan, T. P. (1989). *Statistical methods for quality improvement*, 1st edition, John Wiley.

Ryan, T. P. (2000). *Statistical methods for quality improvement*, 2nd edition, John Wiley.

SAS®9.1 SAS Institute, Inc.

Scientific WorkPlace® Mackichan Software, Inc

Shewhart, W. A. *Economic control of quality of manufactured product*, 50th Anniversary, Commemorative Issue, Quality Press.

Shmueli, G. Cohen, A. (2003). “Run-length distribution or control charts with runs and scans rules”. *Communications in Statistics – Theory and Methods*, 32(2):475-495.

Sullivan, J. H., Woodall, W. H. (1996). “A control chart for preliminary analysis of individual observations”. *Journal of Quality Technology*, 28(3):265-278.

Woodall, W. H. (1997). “Control charts based on attribute data: Bibliography and review”. *Journal of Quality Technology*, 29(2):172-183.

Woodall, W. H. (2000). “Controversies and contradictions in statistical process control” *Journal of Quality Technology*, 32(4):341-350.

Yang, C. H., Hillier, F. S. (1970). “Mean and variance control chart limits based on a small number of subgroups”. *Journal of Quality Technology*, 2(1):9-16.

Zhang, C. W., Xie, M., Goh, T. N. (2006). “Design of exponential control charts using a sequential sampling scheme”. *IEEE Transactions*, 38(12):1105-1116.