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.

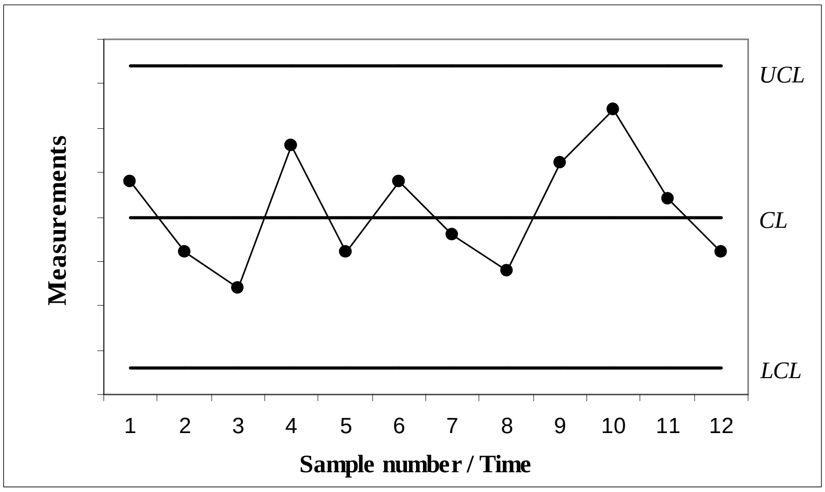

Figure 1.1: A Shewhart-type of control chart

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.

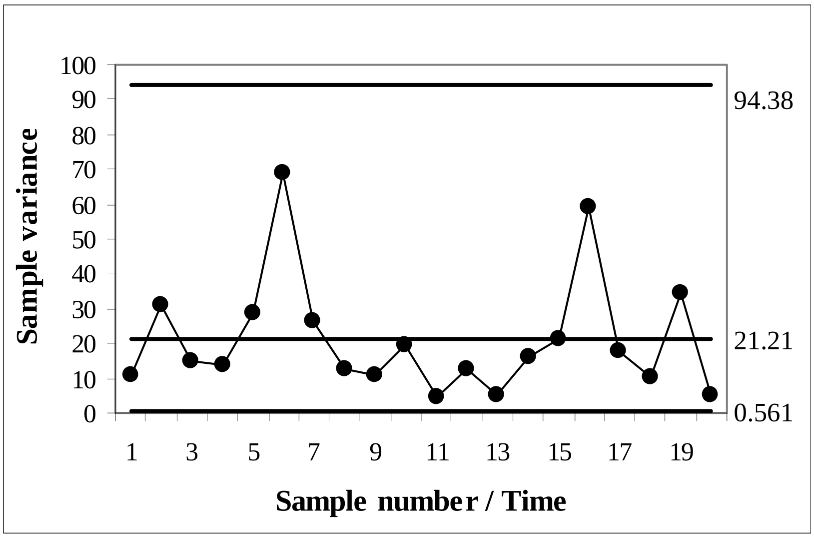

Figure 1.2: The Shewhart-type Phase I S^2 control chart for the data in Table 1.1

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.

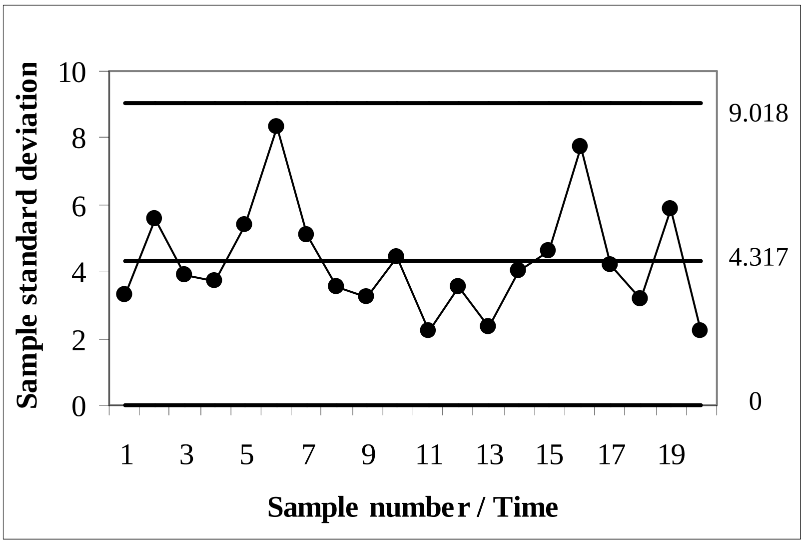

Figure 1.3: The Shewhart-type Phase I S control chart for the data in Table 1.1

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).

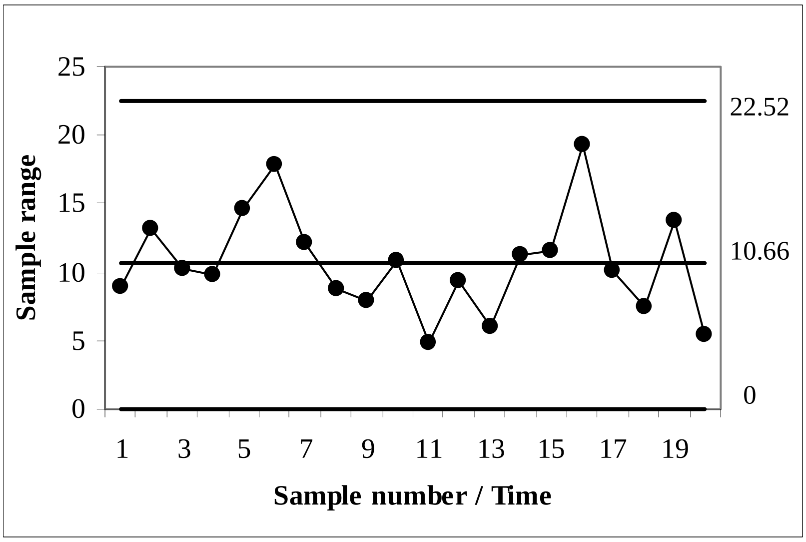

Figure 1.4: The Shewhart-type Phase I S control chart for the data in Table 1.1

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.

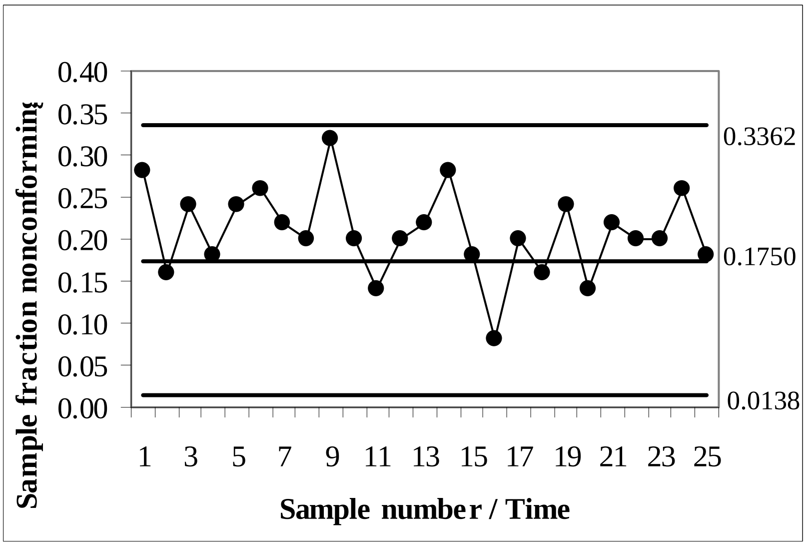

Figure 1.5: The Shewhart-type Phase II p-chart for the data in Table 1.2 with $\bar{p} = 0.1750$

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 9$^{\text{th}}$ 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.

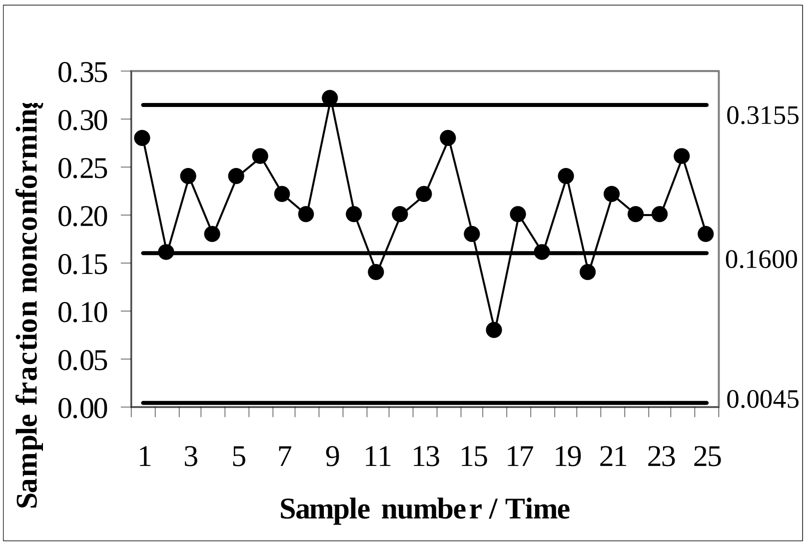

Figure 1.6: The Shewhart-type Phase II p-chart for the data in Table 1.2 with $\bar{p} = 0.1600$

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 $L\hat{C}L = X_{a:m}$ and $U\hat{C}L = X_{b:m}$, 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 $\hat{\underline{\boldsymbol{\theta}}} = (\hat{\theta}_1, \hat{\theta}_2,...,\hat{\theta}_k)$ using the mn individual observations combined (pooled). Using statistical distribution theory and some given performance criterion, an estimated lower and upper control limit, denoted

$$L\hat{C}L = g_1(\hat{\underline{\boldsymbol{\theta}}}) \qquad \text{and} \qquad U\hat{C}L = g_2(\hat{\underline{\boldsymbol{\theta}}})$$

are then obtained, where g_1 and g_2 are two specified functions of $\hat{\underline{\boldsymbol{\theta}}}$ such that $L\hat{C}L < U\hat{C}L$.

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 $\overline{X}_i$ (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

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

to estimate μ and use the pooled variance

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

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

(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 $\underline{\hat{\boldsymbol{\theta}}} = (\bar{\bar{X}},\bar{V})$, which would be used to estimate the unknown parameters $\underline{\boldsymbol{\theta}} = (\mu,\sigma^2)$. Thus, writing the estimated control limits as functions of $\underline{\hat{\boldsymbol{\theta}}}$ i.e.

$$L\hat{C}L = g_1(\bar{\bar{X}},\bar{V}) \qquad \text{and} \qquad U\hat{C}L = g_2(\bar{\bar{X}},\bar{V})$$

it is clear that the events

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

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

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

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

$$\begin{aligned} FAP &= \Pr(\text{At least one false alarm from the } m \text{ samples}) \\ &= 1 - \Pr(\text{No false alarm from the } m \text{ samples}) \\ &= 1 - \Pr(\bigcap_{i=1}^{m} \{E_i\} \mid IC) \qquad\qquad (2\text{-}1) \\ &= 1 - \Pr(\bigcap_{i=1}^{m} \{L\hat{C}L < C_i < U\hat{C}L\} \mid IC) \\ &= 1 - \int_l^u \int_l^u \cdots \int_l^u f_{C_1, C_2, \ldots, C_m}(c_1, c_2, \ldots, c_m) dc_1 dc_2 \ldots dc_m \end{aligned}$$

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

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

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

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

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

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

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

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

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

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

and

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

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

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

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

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

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

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

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 $0 \leq w \leq 1$.

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

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

If it is desired that the $FAP \leq FAP_0$, 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.

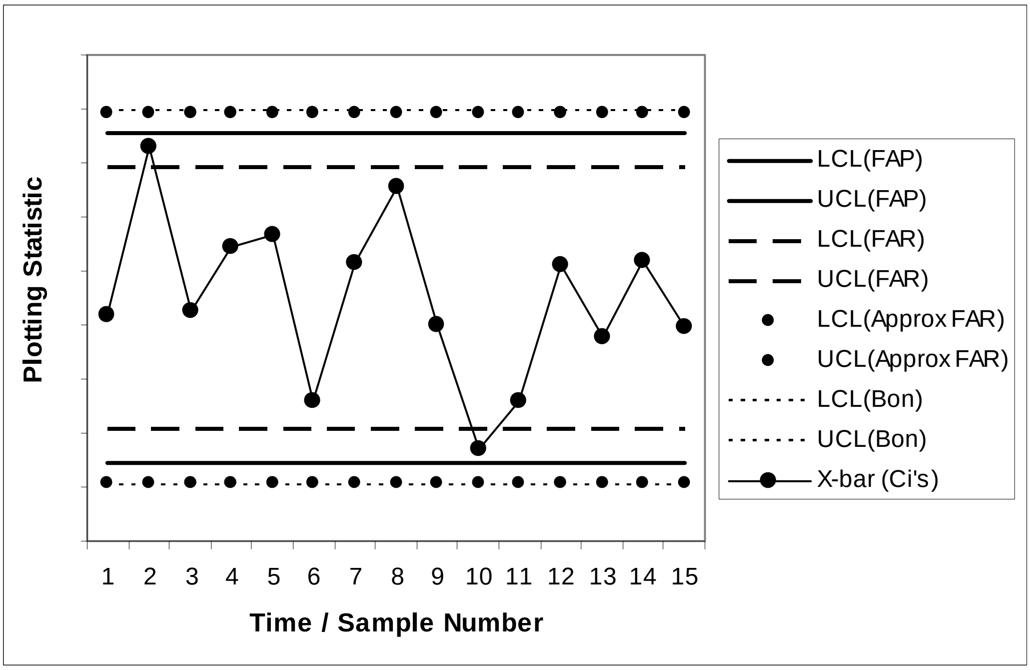

Figure 2.1: Shewhart-type Phase I $\overline{X}$ control charts in Case U

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.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

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 $100(1-\xi)$ 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

$$\begin{aligned}
FAP &= \Pr(\text{At least one false alarm from the } m \text{ samples}) \\
&= 1-\Pr(\text{No false alarm from the } m \text{ samples}) \\
&= 1-\Pr\left(\bigcap_{i=1}^{m}\{L\hat{C}L < S_i^2 < U\hat{C}L\} \mid IC\right) \\
&= 1-\Pr\left(\bigcap_{i=1}^{m}\left\{\frac{\overline{V}\,\chi^2_{1-\alpha_L,n-1}}{n-1} < S_i^2 < \frac{\overline{V}\,\chi^2_{\alpha_U,n-1}}{n-1}\right\} \mid IC\right) \\
&= 1-\Pr\left(\bigcap_{i=1}^{m}\left\{\frac{\chi^2_{1-\alpha_L,n-1}}{n-1} < \frac{S_i^2}{\overline{V}} < \frac{\chi^2_{\alpha_U,n-1}}{n-1}\right\} \mid IC\right) \\
&= 1-\Pr\left(\bigcap_{i=1}^{m}\left\{\frac{\chi^2_{1-\alpha_L,n-1}}{m(n-1)} < \frac{\dfrac{(n-1)S_i^2}{\sigma^2}}{\dfrac{m(n-1)\overline{V}}{\sigma^2}} < \frac{\chi^2_{\alpha_U,n-1}}{m(n-1)}\right\} \mid IC\right) \\
&= 1-\Pr\left(\bigcap_{i=1}^{m}\{a < Y_i < b\} \mid IC\right) \\
&= 1-\int_a^b\int_a^b \ldots \int_a^b f(y_1, y_2, \ldots, y_m)\,dy_1 dy_2 \ldots dy_m
\end{aligned} \qquad (2\text{-}19)$$

where

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

are called the charting constants, and

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

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:

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

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

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

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

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

(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

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

or, equivalently, such that

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

where

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

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 $f(y_1, y_2, ..., y_m)$ 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

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

which implies that

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

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

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

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 $Y_i = \dfrac{X_i}{SUM_1}$ 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 $\dfrac{FAP_0}{2}$ 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 $\dfrac{FAP_0}{2}$.

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 $\overline{v}$ 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 $FAP \le FAP_0 = 0.05$.

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

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

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

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 0.9088	0.0156 0.8413	0.0301 0.7877	0.0433 0.7483	0.0568 0.7173	0.0694 0.6896	0.0790 0.6682	0.0889 0.6498
4	0.0020 0.8151	0.0089 0.7303	0.0179 0.6722	0.0273 0.6303	0.0361 0.5964	0.0444 0.5712	0.0524 0.5514	0.0580 0.5335
5	0.0012 0.7329	0.0056 0.6433	0.0123 0.5854	0.0191 0.5440	0.0257 0.5116	0.0319 0.4867	0.0380 0.4673	0.0431 0.4507
6	0.0008 0.6673	0.0041 0.5765	0.0089 0.5183	0.0144 0.4789	0.0197 0.4492	0.0245 0.4253	0.0292 0.4063	0.0337 0.3929
7	0.0005 0.6067	0.0031 0.5203	0.0069 0.4662	0.0115 0.4271	0.0156 0.4008	0.0198 0.3789	0.0239 0.3617	0.0272 0.3482
8	0.0004 0.5625	0.0024 0.4740	0.0055 0.4216	0.0094 0.3876	0.0129 0.3612	0.0162 0.3415	0.0197 0.3249	0.0230 0.3116
9	0.0003 0.5210	0.0019 0.4389	0.0046 0.3882	0.0078 0.3546	0.0109 0.3301	0.0139 0.3119	0.0167 0.2960	0.0196 0.2839
10	0.0002 0.4866	0.0016 0.4072	0.0039 0.3599	0.0066 0.3282	0.0093 0.3034	0.0121 0.2863	0.0146 0.2724	0.0171 0.2597
15	0.0001 0.3676	0.0008 0.3026	0.0020 0.2629	0.0036 0.2374	0.0053 0.2190	0.0070 0.2054	0.0086 0.1943	0.0100 0.1857
20	0.0000 0.2972	0.0005 0.2405	0.0013 0.2085	0.0023 0.1872	0.0035 0.1720	0.0047 0.1609	0.0059 0.1521	0.0070 0.1446
25	0.0000 0.2492	0.0003 0.2016	0.0009 0.1734	0.0017 0.1554	0.0026 0.1430	0.0035 0.1328	0.0044 0.1256	0.0052 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 $FAP \le FAP_0$, 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

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

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

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

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

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

It follows from (2-33) that

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

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

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

Thus, a and b can be approximated 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 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 0.2265	0.0006 0.1834	0.0014 0.1592	0.0024 0.1433	0.0034 0.1320	0.0044 0.1235	0.0054 0.1168	0.0063 0.1114
30	0.0001 0.1965	0.0004 0.1583	0.0010 0.1369	0.0018 0.1229	0.0026 0.1130	0.0035 0.1056	0.0042 0.0998	0.0050 0.0950
50	0.0000 0.1306	0.0002 0.1038	0.0005 0.0890	0.0009 0.0794	0.0013 0.0727	0.0018 0.0676	0.0022 0.0637	0.0026 0.0605
100	0.0000 0.0734	0.0001 0.0575	0.0002 0.0489	0.0003 0.0433	0.0005 0.0394	0.0007 0.0365	0.0009 0.0343	0.0011 0.0325
300	0.0000 0.0285	0.0000 0.0220	0.0000 0.0184	0.0001 0.0162	0.0001 0.0147	0.0002 0.0135	0.0002 0.0126	0.0003 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

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

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

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

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:

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

where

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

with

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

and

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

where

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

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

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

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

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

and then calculating the values of k_L and k_U as

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

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 $V_i = \dfrac{\sqrt{X_i}}{SUM_2}$ 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

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

and

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

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

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

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:

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

where

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

and

$$U_i = \frac{R_i}{\sum_{j=1}^{m} R_j}. \qquad (2\text{-}49)$$

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

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

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 2.5799	1.9853 2.6011	2.0686 2.5898	2.1222 2.5591	2.1340 2.5560	2.1564 2.5255	2.1782 2.5127	2.1890 2.5265
4	1.8112 3.0321	2.0487 2.9575	2.1569 2.8946	2.2172 2.8711	2.2463 2.8618	2.2845 2.8664	2.2760 2.8450	2.2863 2.8501
5	1.8246 3.2993	2.0719 3.1739	2.2143 3.1538	2.2695 3.1107	2.3112 3.0805	2.3366 3.0484	2.3488 3.0619	2.3771 3.0799
6	1.8368 3.5111	2.0997 3.3599	2.2318 3.3006	2.3194 3.2727	2.3599 3.2454	2.3991 3.2296	2.4120 3.2383	2.4323 3.2356
7	1.8411 3.6426	2.1138 3.4942	2.2681 3.4457	2.3398 3.4167	2.3985 3.3980	2.4318 3.3525	2.4407 3.3640	2.4697 3.3344
8	1.8455 3.7124	2.1226 3.6070	2.2808 3.5041	2.3738 3.4735	2.4229 3.4798	2.4665 3.4636	2.4789 3.4258	2.5026 3.4356
9	1.8516 3.8091	2.1418 3.6639	2.3045 3.5978	2.3965 3.5685	2.4398 3.5464	2.4814 3.5150	2.5146 3.5394	2.5203 3.5449
10	1.8570 3.8817	2.1432 3.7647	2.3098 3.6748	2.4055 3.6187	2.4735 3.6356	2.5102 3.6247	2.5326 3.6169	2.5528 3.6264
15	1.8665 4.1505	2.1783 3.9741	2.3610 3.9184	2.4638 3.9056	2.5352 3.8580	2.5762 3.8504	2.6171 3.8834	2.6397 3.8311
20	1.8722 4.2783	2.1947 4.1461	2.3852 4.0705	2.5101 4.0580	2.5839 4.0381	2.6248 4.0344	2.6612 4.0139	2.6879 4.0165
25	1.8779 4.3803	2.2052 4.2291	2.4095 4.1593	2.5250 4.1088	2.6131 4.1225	2.6647 4.0882	2.6925 4.1260	2.7420 4.0937
30	1.8837 4.4708	2.2134 4.3216	2.4095 4.2455	2.5489 4.2014	2.6326 4.2427	2.6908 4.1837	2.7384 4.2087	2.7613 4.2250
50	1.8875 4.6996	2.2462 4.5158	2.4633 4.4824	2.5848 4.4973	2.6943 4.4472	2.7602 4.4441	2.7936 4.4293	2.8386 4.4606
100	1.9065 4.9189	2.2696 4.7731	2.5037 4.7112	2.6595 4.7214	2.7592 4.7068	2.8470 4.7219	2.9038 4.7050	2.9351 4.7116
300	1.9065 5.3002	2.3398 5.1709	2.6114 5.0612	2.7193 5.0800	2.8566 5.0315	2.9512 5.0690	3.0141 5.0358	3.0510 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 2.1401	1.7523 2.1070	1.7900 2.0891	1.8210 2.0607	1.8175 2.0564	1.8273 2.0516	1.8353 2.0485	1.8286 2.0411
4	1.6945 2.4594	1.8428 2.4025	1.8974 2.3594	1.9292 2.3475	1.9490 2.3320	1.9610 2.3318	1.9555 2.3113	1.9634 2.3187
5	1.7273 2.6863	1.8917 2.5901	1.9653 2.5683	1.9991 2.5385	2.0239 2.5190	2.0380 2.5033	2.0400 2.5032	2.0584 2.5084
6	1.7487 2.8316	1.9298 2.7478	2.0089 2.7061	2.0577 2.7043	2.0833 2.6514	2.1012 2.6505	2.1099 2.6473	2.1218 2.6516
7	1.7611 2.9593	1.9565 2.8718	2.0571 2.8351	2.0992 2.8143	2.1327 2.7959	2.1498 2.7668	2.1551 2.7697	2.1697 2.7694
8	1.7693 3.0505	1.9748 2.9575	2.0783 2.9140	2.1348 2.8854	2.1606 2.8696	2.1859 2.8664	2.2054 2.8494	2.2091 2.8548
9	1.7813 3.1416	1.9965 3.0405	2.1082 2.9920	2.1653 2.9715	2.1885 2.9533	2.2095 2.9307	2.2301 2.9307	2.2353 2.9401
10	1.7902 3.2049	2.0075 3.1096	2.1187 3.0502	2.1784 3.0360	2.2268 3.0156	2.2464 3.0137	2.2606 3.0141	2.2708 3.0123
15	1.8150 3.4670	2.0590 3.3424	2.1874 3.2804	2.2621 3.2870	2.3112 3.2591	2.3314 3.2567	2.3525 3.2549	2.3674 3.2402
20	1.8303 3.6072	2.0918 3.5003	2.2291 3.4459	2.3129 3.4364	2.3632 3.4214	2.3956 3.4095	2.4186 3.3964	2.4330 3.3985
25	1.8350 3.7320	2.1058 3.6033	2.2614 3.5671	2.3457 3.5261	2.4021 3.5220	2.4390 3.5154	2.4627 3.5103	2.4910 3.5047
30	1.8436 3.8131	2.1222 3.7039	2.2802 3.6478	2.3696 3.6187	2.4281 3.6194	2.4720 3.5900	2.4958 3.5912	2.5180 3.5994
50	1.8589 4.0514	2.1643 3.9074	2.3287 3.8767	2.4354 3.8697	2.5157 3.8466	2.5519 3.8365	2.5914 3.8412	2.6068 3.8427
100	1.8875 4.3278	2.1994 4.1882	2.3960 4.1459	2.5101 4.1237	2.5969 4.1225	2.6734 4.1316	2.6833 4.1168	2.7420 4.1323
300	1.9065 4.7282	2.2696 4.6094	2.5306 4.5766	2.6296 4.5421	2.7592 4.5445	2.8470 4.5483	2.9038 4.4844	2.9351 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	1.6105	1.6325	1.6426	1.6461	1.6513	1.6522	1.6456
	1.8970	1.8620	1.8460	1.8267	1.8149	1.8110	1.8092	1.7989
4	1.6091	1.7136	1.7499	1.7690	1.7828	1.7887	1.7805	1.7873
	2.1826	2.1339	2.0966	2.0882	2.0788	2.0762	2.0599	2.0654
5	1.6549	1.7747	1.8266	1.8542	1.8649	1.8766	1.8765	1.8846
	2.3851	2.3129	2.2856	2.2710	2.2528	2.2394	2.2477	2.2399
6	1.6812	1.8232	1.8780	1.9160	1.9314	1.9492	1.9489	1.9596
	2.5319	2.4586	2.4267	2.4246	2.3865	2.3859	2.3826	2.3828
7	1.6997	1.8582	1.9327	1.9633	1.9873	2.0016	2.0033	2.0102
	2.6470	2.5885	2.5411	2.5277	2.5050	2.4970	2.4892	2.4991
8	1.7159	1.8812	1.9620	1.9985	2.0256	2.0415	2.0525	2.0546
	2.7378	2.6673	2.6189	2.5985	2.5917	2.5915	2.5760	2.5767
9	1.7298	1.9018	1.9919	2.0308	2.0541	2.0721	2.0779	2.0893
	2.8327	2.7520	2.7061	2.6837	2.6728	2.6494	2.6561	2.6551
10	1.7407	1.9210	2.0083	2.0589	2.0872	2.1005	2.1136	2.1241
	2.8922	2.8124	2.7648	2.7492	2.7365	2.7290	2.7274	2.7188
15	1.7778	1.9818	2.0864	2.1500	2.1846	2.2012	2.2201	2.2342
	3.1582	3.0476	2.9936	2.9957	2.9718	2.9650	2.9682	2.9563
20	1.7960	2.0216	2.1376	2.2053	2.2463	2.2707	2.2937	2.3017
	3.3097	3.2008	3.1606	3.1496	3.1357	3.1178	3.1097	3.1282
25	1.8064	2.0473	2.1739	2.2486	2.2885	2.3175	2.3341	2.3558
	3.4270	3.3166	3.2777	3.2422	3.2380	3.2376	3.2438	3.2248
30	1.8207	2.0660	2.1995	2.2710	2.3307	2.3575	2.3745	2.3906
	3.5156	3.4090	3.3571	3.3408	3.3370	3.3192	3.3155	3.3213
50	1.8398	2.1175	2.2614	2.3607	2.4183	2.4477	2.4811	2.4910
	3.7559	3.6383	3.5940	3.5858	3.5707	3.5588	3.5471	3.5723
100	1.8684	2.1760	2.3422	2.4503	2.5320	2.5692	2.6098	2.6261
	4.0418	3.9308	3.8767	3.8548	3.8629	3.8539	3.8595	3.8620
300	1.9065	2.2696	2.4498	2.6296	2.6618	2.7428	2.7936	2.8192
	4.4422	4.3286	4.2536	4.2731	4.2524	4.2358	4.2639	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

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

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 = *0.001025th percentile* = 0.0009
$\hat{b}$	0.3599	0.1734	b = *0.998975th 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

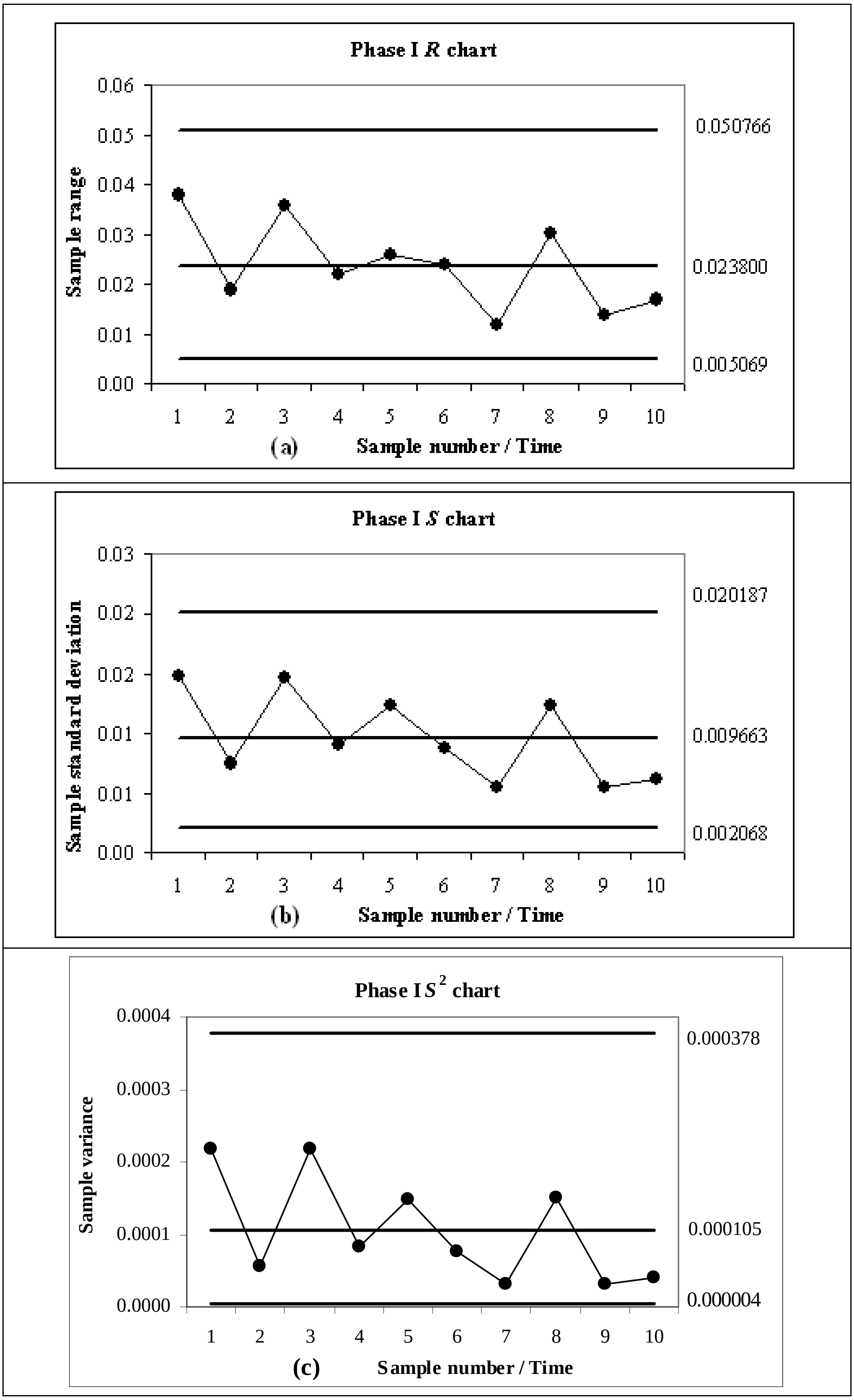

Figure 2.2: Shewhart-type Phase I R, S and S^2 charts for $m = 10$

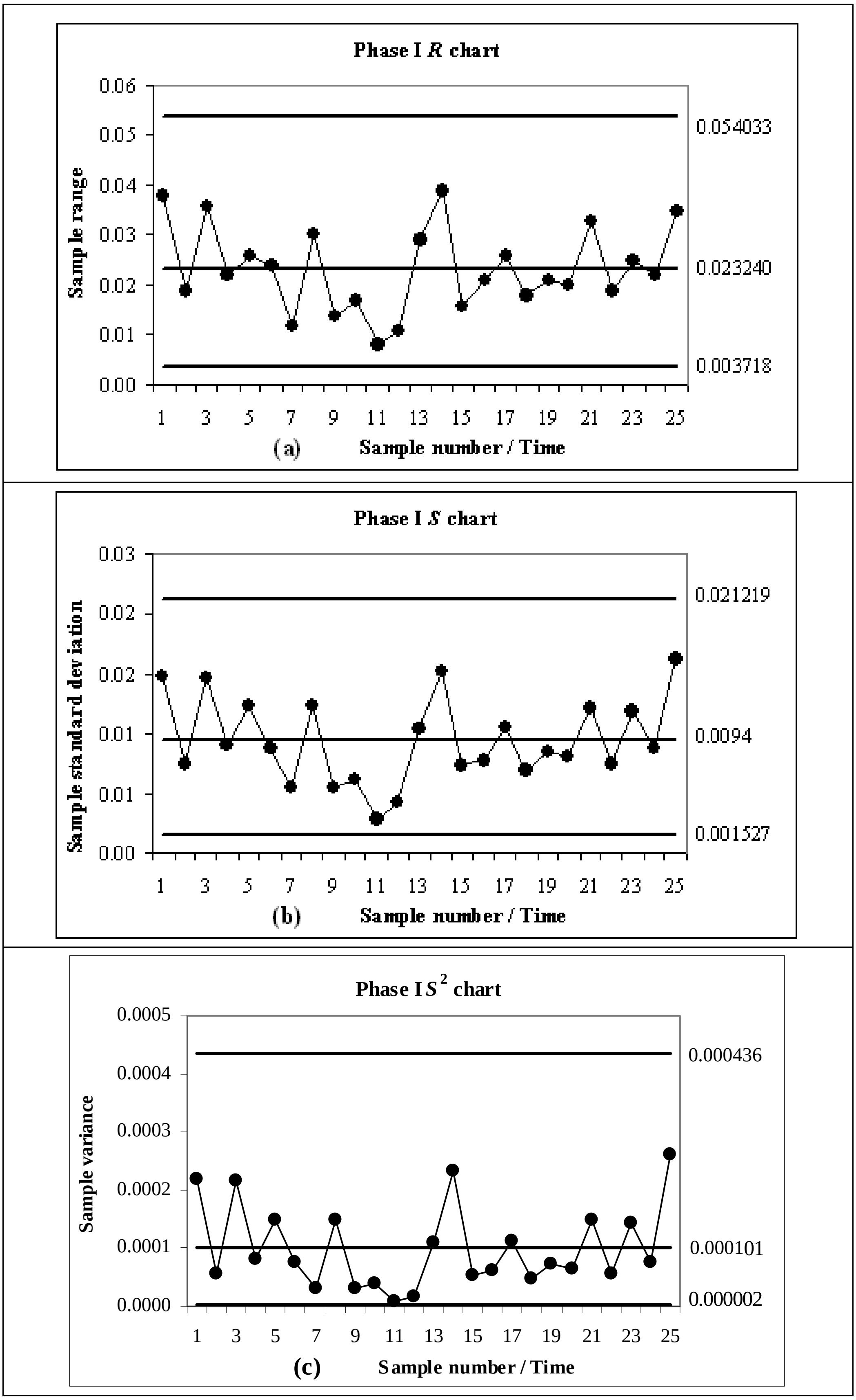

Figure 2.3: Shewhart-type Phase I R, S and S^2 charts for $m = 25$

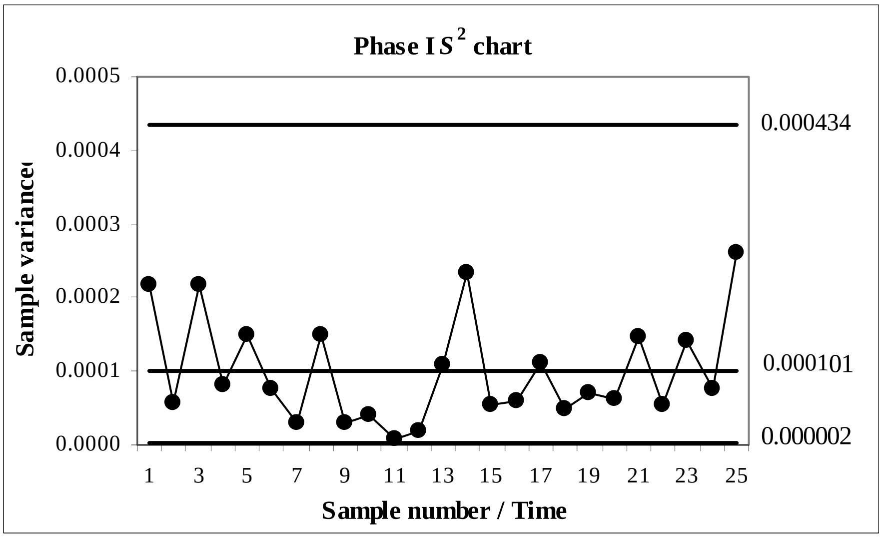

Figure 2.4: Shewhart-type Phase I S^2 assuming independence of the charting statistics

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

$$
\begin{aligned}
FAP_{\overline{X}} &= 1 - \Pr(|\overline{X}_1 - \overline{\overline{X}}| \le k_m \sigma / \sqrt{n}, ..., |\overline{X}_m - \overline{\overline{X}}| \le k_m \sigma / \sqrt{n} \mid IC) \\
&= 1 - \Pr(\bigcap_{i=1}^{m} \{|\overline{X}_i - \overline{\overline{X}}| \le k_m \sigma / \sqrt{n}\} \mid IC) \\
&= 1 - \Pr\left(\frac{\max |\overline{X}_i - \overline{\overline{X}}|}{\sigma / \sqrt{n}} \le k_m \text{ for } i = 1,2,...m \mid IC \right)
\end{aligned}
$$

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

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

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

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

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

(which is an exact result) and approximates the distribution of $\dfrac{v\overline{R}^2}{c^2\sigma^2}$ 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 $i = 1,2,\ldots,m$ 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

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

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 $\overline{X}_i$ and its range R_i fall inside the control limits for the $\overline{X}$ 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 $\overline{\overline{X}}$, the mean range $\overline{R}$ 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 $\overline{\overline{X}}$ and $\overline{R}$. 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.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

(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 $\overline{X}$ chart using the average (pooled) sample variance $\overline{V} = \frac{1}{m}\sum_{i=1}^{m} S_i^2$ (instead of the mean range $\overline{R}$) 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 $\overline{V}$.

Phase I $\overline{X}$ chart

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

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

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

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

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

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

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

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

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

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

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

and that the random variable

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

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

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

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

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

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

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

$$L\hat{C}L_{S^2} = B_7^{**}\overline{V} \qquad \text{and} \qquad U\hat{C}L_{S^2} = B_8^{**}\overline{V}$$

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.

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

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

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

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 $\overline{V}$) 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 $i=1,2,\ldots,m$ 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

$$L\hat{C}L = \sqrt{B_7^{**}\overline{V}} \qquad \text{and} \qquad U\hat{C}L = \sqrt{B_8^{**}\overline{V}}\,.$$

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

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

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

where

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

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

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

was chosen using Boole's inequality such that, if the process is in-control, the probability that at least one sample mean $\overline{X}_i$ 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

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

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

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

and

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

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

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

and

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

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

Based on these results, they conclude that the standard limits $\overline{X}$ chart slightly outperforms the corresponding individual limits $\overline{X}$ 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

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

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 $\overline{X}$ 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 $\overline{X}$ 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 $\overline{X}$ 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

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

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

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

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

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 $\overline{X}$ chart is shown to be

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

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 $\overline{X}$ 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 $(1-\alpha+\tau)$ 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 $(1-\alpha+\tau)$ 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

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

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

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

where X_i for $i = 1,2,\ldots,m$ 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;
```