Chapter 3

Attributes control charts: Case K and Case U

3.0 Chapter overview

Introduction

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

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

Motivation

The *p*-chart and *c*-chart are particularly useful in the service industries and in non-manufacturing quality improvements efforts since many of the quality characteristics found in these environments are in actual fact attributes. SPC with attributes data therefore constitutes an important area of research and applications (see e.g. Woodall (1997) for a review).

The classical application of the *p*-chart and the *c*-chart requires that the parameters of the distributions are known. In many situations the true fraction nonconforming, p, and the true average number of nonconformities in an inspection unit, c, are unknown or unspecified and need to be estimated from a reference sample or historical (past) data. While there are empirical rules and guidelines for setting up the charts, little is known about their run-length distributions when the fact that the parameters are estimated is taken into account. Understanding the effect of estimating the parameters on the in-control (IC) and the out-of-control (OOC) performance of the charts are therefore of interest from a practical and a theoretical point of view.

In this chapter we derive and evaluate expressions for the run-length distributions of the Shewhart-type *p*-chart and the Shewhart-type *c*-chart when the parameters are estimated. An exact approach based on the binomial and the Poisson distributions is used since in many applications the values of p and c are such that the normal approximation to the binomial and the Poisson distributions is quite poor, especially in the tails. The results are used to discuss the appropriateness of the widely followed empirical rules for choosing the size of the Phase I sample used to estimate the unknown parameters; this includes both the number of reference samples (or inspection units) m and the sample size n. Note that, in our developments, we assume that the size of each subgroup or the size of each inspection unit stays constant over time.

Methodology

We examine the effect of estimating p and c on the performance of the *p*-chart and the *c*-chart via their run-length distributions and associated characteristics such as the average run-length (ARL), the false alarm rate (FAR) and the probability of a "no-signal". Exact expressions are derived for the Phase II run-length distributions and the related Phase II characteristics using expectation by conditioning (see e.g. Chakraborti, (2000)). We first obtain the characteristics of the run-length distributions conditioned on point estimates from Phase I and then find the unconditional characteristics by averaging over the distributions of the point estimators. This two-step analysis provides valuable insight into the specific as well as the overall effects of parameter estimation on the performance of the charts in Phase II.

The conditional characteristics let us focus on specific values of the estimators and look at the performance of the charts in more detail for the particular value(s) at hand. The unconditional characteristics characterize the overall performance of the charts i.e. averaged over all possible values of the estimators.

In practice we will obviously have only a single realization for each of the point estimators and the characteristics of the conditional run-length distribution therefore provide important information specific only to our particular situation; but, since each user will have his own values for each of the point estimators the conditional run-length performance will be different from user to user. The unconditional run-length, on the other hand, lets us look at the bigger picture, averaged over all possible values of the point estimators, and is therefore the same for all users.

Layout of Chapter 3

This chapter consists of two main sections and an appendix. The first section is labeled "The *p*-chart and the *c*-chart for standards known (Case K)" and the second section is called "The *p*-chart and the *c*-chart for standards unknown (Case U)". In the first section we study the charts when the parameters are known. The second section focuses on the situation when the parameters are unknown and forms the heart of Chapter 3. In both sections we study the *p*-chart and the *c*-chart in unison; this points out the similarity and the differences between the charts and helps one to understand the theory and/or methodology better.

Appendix 3A gives an example of each chart and contains a discussion on the characteristics of the *p*-chart and the *c*-chart in Case K. To the author's knowledge none of the standard textbooks and/or articles currently available in the literature give a detailed discussion of the Case K *p*-chart's and the Case K *c*-chart's characteristics.

3.1 The *p*-chart and the *c*-chart for standards known (Case K)

Introduction

Case K is the scenario where known values for the parameters are available. This will happen in high volume manufacturing processes where ample reliable information is available so that it is possible to specify values for the parameters.

Studying Case K not only sets the stage for the situation when the parameters are unknown (Case U), but the characteristics and the performance of the charts in Case K are also important. In particular, it helps us understand the operation and the performance of the charts in the simplest of cases (when the parameters are known) and provides us with benchmark values which we can use to determine the effect of estimating the parameters on the operation and the performance of the charts in Case U (when the parameters are unknown).

The *p*-chart is used when we monitor the fraction of nonconforming items in a sample of size $n \geq 1$ and is based on the binomial distribution. The *c*-chart is based on the Poisson distribution and used when we focus on monitoring the number of nonconformities in an inspection unit, where the inspection unit may consist of one or more than one physical unit.

Assumptions

We derive and study the characteristics of the charts in Case K assuming that: (i) the sample size and the size of an inspection unit (whichever is applicable) stay constant over time, (ii) the nonconforming items occur independently i.e. the occurrence of a nonconforming item at a particular point in time does not affect the probability of a nonconforming item in the time periods that immediately follow, and (iii) the probability of observing a nonconformity in an inspection unit is small, yet the number of possible nonconformities in an inspection unit is infinite.

To this end, let $X_i \sim iidBin(n, p)$ for $i = 1,2,...$ denote the number of nonconforming items in a sample of size $n \geq 1$ with true fraction nonconforming $0 < p < 1$; the sample fraction nonconforming is then defined as $p_i = X_i / n$. Similarly, let $Y_i \sim iidPoi(c)$, $c > 0$ for $i = 1,2,...$ denote the number of nonconformities in an inspection unit where c denotes the true average number of nonconformities in an inspection unit.

Charting statistics

The charting statistics of the *p*-chart is the sample fraction nonconforming $p_i = X_i / n$ for $i = 1,2,...$; the charting statistics of the *c*-chart is the number of nonconformities Y_i for $i = 1,2,...$, in an inspection unit.

Control limits

For known values of the true fraction nonconforming and the true average number of nonconformities in an inspection unit, denoted by p_0 and c_0 respectively, the upper control limits (UCL's), the centerlines (CL's), and the lower control limits (LCL's) of the traditional *p*-chart and the traditional *c*-chart are

$$UCL_p = p_0 + 3\sqrt{p_0(1-p_0)/n} \qquad CL_p = p_0 \qquad LCL_p = p_0 - 3\sqrt{p_0(1-p_0)/n} \tag{3-1}$$

and

$$UCL_c = c_0 + 3\sqrt{c_0} \qquad CL_c = c_0 \qquad LCL_c = c_0 - 3\sqrt{c_0} \tag{3-2}$$

respectively (see e.g. Montgomery, (2005) p. 268 and p. 289).

The control limits in (3-1) and (3-2) are k-sigma limits (where $k = 3$) and based on the tacit assumption that both the binomial distribution and the Poisson distribution are well approximated by the normal distribution.

The subscripts "*p*" and "*c*" in (3-1) and (3-2) are used to distinguish the control limits of the two charts; where no confusion is possible the subscripts are dropped.

Implementation

The actual operation of the charts consist of: (i) taking independent samples and independent inspection units at equally spaced successive time intervals, (ii) computing the charting statistics, and then (iii) plotting the charting statistics (one at a time) reflected on the vertical axis of the control charts versus the sample number and the inspection unit number $i = 1,2,...$ reflected on the horizontal axis.

The control limits are also displayed on the charts so that every time a new charting statistic is plotted it is in actual fact compared to the control limits. The aim is to detect when (or if) the true process parameters p and c change (moves away) from their known or specified or target values p_0 and c_0, respectively.

Signaling and non-signaling events

The event when a charting statistic (point) plots outside the control limits, which is called a signaling event and denoted by A_i for $i = 1,2,...$, is interpreted as evidence that the parameter is no longer equal to its specified value. The charting procedure therefore stops, a signal (alarm) is given, and we declare the process out-of-control (OOC) i.e. we say that $p \neq p_0$ or state that $c \neq c_0$. Investigation and corrective action is typically required to find and eliminate the possible assignable cause(s) and/or source(s) of variability responsible for the behavior.

The complimentary event is when a plotted point lies between (within) the control limits and labeled a non-signaling event or a "no-signal". In case of a no-signal the charting procedure continues, no user intervention is necessary, and we consider the process to be in-control (IC) i.e. we say that $p = p_0$ or declare that $c = c_0$. We denote the non-signaling event by

$$A_i^C : \{LCL < Q_i < UCL\}$$

where $Q_i = p_i$ or Y_i for $i = 1,2,...$ and LCL and UCL denote the control limits in either (3-1) or (3-2).

Note that, in a hypothesis-testing framework, concluding that the process is out-of-control when the process is actually in-control is called a type I error; similarly, concluding that the process is in-control when it is really out-of-control is a called a type II error.

3.1.1 Probability of a no-signal

Introduction

The probability of a no-signal refers to the probability of a non-signaling event and is denoted by

$$\beta = \Pr(A_i^C) \quad \text{for} \quad i = 1,2,\ldots .$$

The probability of a no-signal is important because: (i) it is the key for the derivation of the run-length distribution, and (ii) plays a central role when we assess the performance of a control chart. Once we have the probability of a no-signal, the run-length distribution is completely known.

Probability of a no-signal: *p*-chart

The probability of a no-signal on the *p*-chart is the probability of the event

$$\{LCL_p < p_i < UCL_p\} \quad \text{for} \quad i = 1,2,\ldots . \tag{3-3}$$

Since p is known and equal to p_0 the control limits LCL_p and UCL_p are known values (constants) which makes $p_i = X_i / n$ the only random quantity in (3-3).

The cumulative distribution function of the sample fraction nonconforming p_i is known and given by

$$\Pr(p_i \le a) = \Pr(X_i / n \le a) = \Pr(X_i \le na) = \sum_{j=0}^{[na]} \Pr(X_i = j) = \sum_{j=0}^{[na]} \binom{n}{j} p^j (1-p)^{n-j}$$

for $0 \le a \le 1$, $0 < p < 1$ and where $[na]$ denotes the largest integer not exceeding na. Because the distribution of p_i is defined in terms of that of $X_i \sim Bin(n, p)$ we re-express the non-signaling event in (3-3) as

$$\{nLCL_p < X_i < nUCL_p\}$$

and use the properties of the distribution of X_i to derive the probability of a no-signal.

Thus, at the i^{th} observation the non-signaling probability for the *p*-chart is a function of and depends on p, p_0 and n, and is derived as follows

$$\begin{aligned}\beta(p, p_0, n) &= \Pr(LCL_p < p_i < UCL_p)\\ &= \Pr(nLCL_p < X_i < nUCL_p)\\ &= \Pr(X_i < nUCL_p) - \Pr(X_i \le nLCL_p)\end{aligned}$$

$$= \begin{cases} H(b; p, n) & \text{if } nLCL_p < 0 \\ H(b; p, n) - H(a; p, n) & \text{if } nLCL_p \ge 0 \end{cases} \tag{3-4}$$

$$= \begin{cases} 1 - I_p(b+1, n-b) & \text{if } nLCL_p < 0 \\ I_p(a+1, n-a) - I_p(b+1, n-b) & \text{if } nLCL_p \ge 0 \end{cases}$$

$$= 1 - I_p(b+1, n-b) - 1_{\{nLCL_p : nLCL_p \ge 0\}}(nLCL_p)(1 - I_p(a+1, n-a))$$

for $0 < p, p_0 < 1$, where UCL_p and LCL_p are defined in (3-1) and both are functions of n and p_0,

$$H(b; p, n) = \Pr(X_i \le b) = \sum_{j=0}^{b} \binom{n}{j} p^j (1-p)^{n-j}$$

denotes the cumulative distribution function (c.d.f) of the $Bin(n, p)$ distribution,

$$I_t(u, v) = (\beta(u, v))^{-1} B(t; u, v) \quad \text{for} \quad 0 < t < 1 \qquad \text{and} \qquad B(t; u, v) = \int_0^t s^{u-1}(1-s)^{v-1} ds \quad \text{for} \quad u, v > 0$$

denotes the c.d.f of the $Beta(u, v)$ distribution (also known as the incomplete beta function) with $\beta(u, v) = B(1; u, v)$,

$$1_{\{x : x \ge 0\}}(x) = \begin{cases} 1 & \text{if } x \ge 0 \\ 0 & \text{if } x < 0 \end{cases},$$

and where

$$a = [nLCL_p] \quad \& \quad b = \begin{cases} \min\{nUCL_p - 1, n\} & \text{if } nUCL_p \text{ is an integer} \\ \min\{[nUCL_p], n\} & \text{if } nUCL_p \text{ is not an integer} \end{cases} \tag{3-5}$$

and $[x]$ denotes the largest integer not exceeding x.

Remark 1

(i) Making use of the c.d.f of the beta distribution and the indicator function $1_{\{x: x \geq 0\}}(x)$ helps us write the probability of a no-signal in a more compact way (see e.g. the last line of (3-4)).

(ii) The relationship between the c.d.f of the binomial distribution and the c.d.f of the type I or standard beta distribution is evident from (3-4) and given by

$$H(b; n, p) = 1 - I_p(b+1, n-b) = I_{1-p}(n-b, b+1).$$

(iii) The charting constants a and b in (3-5) are suitably modified to take account of the fact that the $Bin(n, p)$ distribution assigns nonzero probabilities only to integers from 0 to n.

(iv) To cover both the in-control and the out-of-control scenarios we do not assume that the specified value for the fraction nonconforming p_0 in (3-4) is necessarily equal to the true fraction nonconforming p.

Probability of a no-signal: *c*-chart

The probability of a no-signal on the *c*-chart is the probability that the event

$$\{LCL_c < Y_i < UCL_c\} \quad \text{for} \quad i = 1,2,... \tag{3-6}$$

occurs. Since c is specified and equal to c_0 the control limits LCL_c and UCL_c are constants. As a result Y_i is the only random variable in (3-6). Because the distribution of Y_i is known (assumed) to be Poisson with parameter (in general) c, we derive the probability of a no-signal on the *c*-chart (directly) in terms of the distribution of Y_i.

The probability of a no-signal on the *c*-chart is a function of and depends on c and c_0, and is derived as follows

$$\begin{aligned}\beta(c,c_0) &= \Pr(LCL_c < Y_i < UCL_c) \\ &= \Pr(Y_i < UCL_c) - \Pr(Y_i \le LCL_c) \\ &= G(f;c) - G(d;c) \\ &= \Gamma_{f+1}(c) - \Gamma_{d+1}(c)\end{aligned} \tag{3-7}$$

for $c, c_0 > 0$, where UCL_c and LCL_c are defined in (3-2) and both are functions of c_0,

$$G(f;c) = \Pr(Y_i \le f) = \sum_{j=0}^{f} \frac{e^{-c}c^j}{j!}$$

denotes the c.d.f of the $Poi(c)$ distribution,

$$\Gamma_t(u) = (\Gamma(t))^{-1}\Gamma(t;u) \qquad \text{where} \qquad \Gamma(t;u) = \int_t^{\infty} s^{u-1}e^{-s}ds \quad \text{for} \quad t,u > 0$$

denotes the upper incomplete gamma function,

$$\Gamma(t) = (t-1)!$$

for positive integer values of t, and where

$$d = \max\{0,[LCL_c]\} \quad \& \quad f = \begin{cases} UCL_c - 1 & \text{if } UCL_c \text{ is an integer} \\ [UCL_c] & \text{if } UCL_c \text{ is not an integer.} \end{cases} \tag{3-8}$$

Remark 2

(i) The relationship between the c.d.f of the Poisson distribution and the lower incomplete gamma function is evident from (3-7) and given by $G(f;c) = \Gamma_{f+1}(c)$.

(ii) The constants d and f in (3-8) incorporate the fact that the $Poi(c)$ distribution only assigns nonzero probabilities to non-negative integers.

(iii) We do not assume that c in (3-7) is necessarily equal to c_0; this enables us to study both the in-control and the out-of-control properties of the *c*-chart.

3.1.2 Operating characteristic and the OC-curve

The Operating Characteristic (OC) or the β-risk is the probability that a chart does not signal on the first sample or the first inspection unit following a sustained (permanent) step shift in the parameter and thus failing to detect the shift. For the *p*-chart the OC is the probability of a no-signal $\beta(p, p_0, n)$ with $p \neq p_0$ and for the *c*-chart the OC is the probability $\beta(c, c_0)$ with $c \neq c_0$.

A graphical display (plot) of the OC as a function of $0 < p < 1$ (in case of the *p*-chart), or as a function of $c > 0$ (in case of the *c*-chart), is called the operating characteristic curve or simply the OC-curve. The OC-curve lets us see a chart's ability to detect a shift in the process parameter and therefore describes the performance of the chart.

3.1.3 False alarm rate

As an alternative to the OC-curve we can graph the probability of a signal as a function of p for values of $0 < p < 1$ or as a function of c for values of $c > 0$. The probability of a signal is $1-\beta$ i.e. one minus the probability of a no-signal, and is in some situations intuitively easier understood than the OC.

For the *p*-chart the probability of a signal is $1-\beta(p, p_0, n)$ where $\beta(p, p_0, n)$ is defined in (3-4) and for the *c*-chart the probability of a signal is $1-\beta(c, c_0)$ where $\beta(c, c_0)$ is given in (3-7).

When we substitute p with p_0 in $1-\beta(p,p_0,n)$ and replace c with c_0 in $1-\beta(c,c_0)$ we obtain the false alarm rate (FAR) of the charts, that is,

$$FAR(p_0,p_0,n)=1-\beta(p_0,p_0,n) \qquad \text{and} \qquad FAR(c_0,c_0)=1-\beta(c_0,c_0).$$

The false alarm rate is the probability of a signal when the process is in-control (i.e. no shift occurred) and often used a measure of a control chart's in-control performance.

The OC-curve and the probability of a signal as functions of p or c i.e. given a shift in the process, focus on the probability of a single event and involves only one charting statistic. A more popular and perhaps more useful method to evaluate and examine the performance of a control chart is its run-length distribution.

3.1.4 Run-length distribution

The number of rational subgroups to be collected or the number of charting statistics to be plotted on a control chart before the first or next signal, is called the run-length of a chart. The discrete random variable defining the run-length is called the run-length random variable and denoted by N. The distribution of N is called the run-length distribution.

Characteristics of the run-length distribution give us more insight into the performance of a chart. The characteristics of the run-length distribution most often looked at are, for example, its moments (such as the expected value and the standard deviation) as well as the percentiles or the quartiles (see e.g. Shmueli and Cohen, (2003)).

If no shift occurred (i.e. $p = p_0$ or $c = c_0$) the distribution of N is called the in-control run-length distribution. In contrast, if the process did encounter a shift (i.e. $p \neq p_0$ or $c \neq c_0$) the distribution of N is labeled the out-of-control run-length distribution. To distinguish between the in-control and the out-of-control situations the notations N_0 and N_1 are used; this notation is also used for the characteristics of the run-length distribution.

Assuming that the rational subgroups are independent and that the probability of a signal is the same for all samples (inspection units) the run-length distribution is given by

$$\Pr(N=j)=\beta^{j-1}(1-\beta) \qquad j=1,2,... \tag{3-9}$$

where β denotes the probability of a no-signal defined in (3-4) or (3-7).

The distribution in (3-9) is recognized as the geometric distribution (of order 1) with probability of "success" $1-\beta$ so that we write, symbolically, $N \sim Geo(1-\beta)$. The success probability is the probability of a signal and, as mentioned before, completely characterizes the geometric (run-length) distribution.

Various statistical characteristics of the run-length distribution provide insight into how a control chart functions and performs. Typically we want the chart to signal quickly once a change takes place and not signal too often when the process is actually in-control, which is when no shift or no change has occurred. We are interested in the typical value as well as the spread or the variation in the run-length distribution.

3.1.5 Average run-length

A popular measure of the central tendency of a distribution is the expected value (mean) or the average. Accordingly, the average has been the most popular index or measure of a control chart's performance and is called the average run-length (*ARL*). The *ARL* is defined as the expected number of rational subgroups that must be collected before the chart signals.

When the process is in-control the expected number of charting statistics that must be plotted before the control chart signals erroneously is called the in-control average run-length and denoted by ARL_0. The out-of-control average run-length is denoted by ARL_1 and is the expected number of charting statistics to be plotted before a chart signals after the process has gone out-of-control. Obviously, for an efficient control chart the in-control average run-length should be large and the out-of-control average run-length should be small.

From the properties of the geometric distribution the *ARL* is the expected value of N so that

$$ARL = E(N) = 1/(1-\beta). \qquad (3\text{-}10)$$

Therefore, when the signaling events are independent and have the same probability the *ARL* of the chart is simply the reciprocal of the probability of a signal $1-\beta$. If the process is in-control, the in-control *ARL* is equal to the reciprocal of the *FAR*, that is, $ARL_0 = 1/FAR$. It is this simple relationship between the average run-length and the probability of a signal, or the in-control average run-length and the false alarm rate, that accounts for the popularity of the (in-control) average run-length and the probability of a signal (false alarm rate) as measures of a control chart's performance.

3.1.6 Standard deviation and percentiles of the run-length

Other characteristics of the run-length distribution are also of interest. For example, in addition to the mean we should also look at the standard deviation of the run-length distribution to get an idea about the variation or spread.

Using results for the geometric distribution, the standard deviation of the run-length, denoted by *SDRL*, is given by

$$SDRL = \text{stdev}(N) = \sqrt{\beta}/(1-\beta). \qquad (3\text{-}11)$$

Since the geometric distribution is skewed to the right the mean and the standard deviation become questionable measures of central tendency and spread so that additional descriptive measures are useful. For example, the percentiles, such as the median and the quartiles (which are more robust or outlier resistant), can provide valuable information about the location as well as the variation in the run-length distribution.

Because the run-length distribution is discrete, the $100q^{\text{th}}$ percentile ($0 < q < 1$) is defined as the smallest integer *j* such that the cumulative probability is at least q, that is, $\Pr(N \leq j) \geq q$. The median run-length (denoted by *MDRL*) is the 50^{th} percentile so that $q = 0.5$, whereas the first quartile (Q_1) is the 25^{th} percentile so that $q = 0.25$.

3.1.7 In-control and out-of-control run-length distributions

The characteristics of the in-control run-length distributions are essential in the design and implementation of a control chart. Furthermore, for out-of-control performance comparisons we need the out-of-control run-length distributions and/or characteristics. For example, the in-control average run-lengths of the charts are typically fixed at an acceptably high level so that the number of false alarms or the false alarm rate is reasonably small. The chart with the smallest or the lowest out-of-control average run-length for a certain change (or shift of a specified size) in the process parameter is then selected to be the winner (i.e. the best performing chart). Alternatively, we can fix the false alarm rate of the charts at an acceptably small value and then select that chart with the highest probability of a signal (given a specified shift in the parameter) as the winner.

Note that, the average run-length and the probability of a signal are two equivalent performance measures in that they both lead to the same decision and follows from the relationship between the average run-length and the probability of a signal given in (3-10).

The run-length distributions and some related characteristics of the run-length distributions of the *p*-chart and the *c*-chart, which all conveniently follow from the properties of the geometric distribution of order 1, are summarized in Table 3.1 and Table 3.2, respectively.

The characteristics of the *p*-chart and the *c*-chart are seen to be all functions of and depend entirely on the probability of a no-signal, that is, $\beta(p, p_0, n)$ or $\beta(c, c_0)$; once we have expressions and/or numerical values for the two probabilities $\beta(p, p_0, n)$ and $\beta(c, c_0)$ the run-length distributions are completely known.

The in-control run-length distributions and the in-control characteristics of the run-length distributions are obtained when $p = p_0$ and $c = c_0$. The out-of-control run-length distributions and the out-of-control characteristics are found by setting $p \neq p_0$ and $c \neq c_0$, respectively.

An in-depth analysis and discussion of the in-control run-length distributions of the *p*-chart and the *c*-chart in Case K (and their related in-control properties) are given in Appendix 3A. From time to time we will refer to the results therein; especially when we study and look at the effects of parameter estimation on the performance of the charts in Case U.

Table 3.1: The probability mass function (p.m.f), the cumulative distribution function (c.d.f), the false alarm rate (*FAR*), the average run-length (*ARL*), the standard deviation of the run-length (*SDRL*) and the quantile function (qf) of the run-length distribution of the *p*-chart in Case K

p.m.f	$\Pr(N_p = j; p, p_0, n) = \beta(p, p_0, n)^{j-1}(1-\beta(p, p_0, n)) \qquad j = 1,2,\ldots$	(3-12)
c.d.f	$\Pr(N_p \le j; p, p_0, n) = 1-(\beta(p, p_0, n))^j \qquad j = 1,2,\ldots$	(3-13)
FAR	$FAR(p_0, n) = 1-\beta(p_0, p_0, n)$	(3-14)
ARL	$ARL(p, p_0, n) = E(N_p) = 1/(1-\beta(p, p_0, n))$	(3-15)
SDRL	$SDRL(p, p_0, n) = \text{stdev}(N_p) = \sqrt{\beta(p, p_0, n)}/(1-\beta(p, p_0, n))$	(3-16)
qf	$Q_{N_p}(q; p, p_0, n) = \inf\{\text{int } x : \Pr(N_p \le x; p, p_0, n) \ge q\} \quad 0 < q < 1$	(3-17)

Table 3.2: The probability mass function (p.m.f), the cumulative distribution function (c.d.f), the false alarm rate (*FAR*), the average run-length (*ARL*), the standard deviation of the run-length (*SDRL*) and the quantile function (qf) of the run-length distribution of the *c*-chart in Case K

p.m.f	$\Pr(N_c = j; c, c_0) = \beta(c, c_0)^{j-1}(1-\beta(c, c_0)) \qquad j = 1,2,\ldots$	(3-18)
c.d.f	$\Pr(N_c \le j; c, c_0) = 1-(\beta(c, c_0))^j \qquad j = 1,2,\ldots$	(3-19)
FAR	$FAR(c_0) = 1-\beta(c_0, c_0)$	(3-20)
ARL	$ARL(c, c_0) = E(N_c) = 1/(1-\beta(c, c_0))$	(3-21)
SDRL	$SDRL(c, c_0) = \text{stdev}(N_c) = \sqrt{\beta(c, c_0)}/(1-\beta(c, c_0))$	(3-22)
qf	$Q_{N_c}(q; c, c_0) = \inf\{\text{int } x : \Pr(N_c \le x; c, c_0) \ge q\} \quad 0 < q < 1$	(3-23)

3.2 The *p*-chart and the *c*-chart for standards unknown (Case U)

Introduction

Case U is the scenario when the parameters p and c are unknown. Case U occurs more often in practice than Case K particularly when not much historical knowledge or expert opinion is available. In the service industries, non-manufacturing environments and job-shop environments, which all involve low-volume of "production", it often happens that there is a scarcity of historical data.

Setting up a control chart in Case U consists of two phases: Phase I and Phase II. The former is the so-called retrospective phase whereas the latter is labeled the prospective or the monitoring phase (see e.g. Woodall, (2000)). In Phase I the parameters and the control limits are estimated from an in-control reference sample or calibration sample. In Phase II, new incoming subgroups are collected independently from the Phase I reference sample. The charting statistic for each Phase II subgroup is then calculated and individually compared to the estimated Phase II control limits until the first point plots outside the limits. The goal is to detect when (or if) the process parameters change.

We study and analyze the performance of the *p*-chart and *c*-chart following a Phase I analysis. In other words, we focus on the run-length distributions and the associated characteristics of the run-length distributions of the *p*-chart and the *c*-chart in Phase II.

3.2.1 Phase I of the Phase II *p*-chart and *c*-chart

The charting procedures to ensure that the Phase I data is representative of the in-control state of the process were discussed in Chapter 2. Here we consider the matter only in very general terms and assume that such in-control Phase I data is available; this implies that each sample and each inspection unit in the reference sample has identical (unknown) parameters.

Phase I data and assumptions

The Phase I data is the in-control reference sample or the historical (past) data that is used to estimate the unknown parameters. In case of the *p*-chart the Phase I data consists of m mutually independent samples each of size $n \geq 1$. The Phase I data for the *c*-chart consists of m mutually independent inspection units.

To this end, let $X_i \sim iidBin(n, p)$ for $i = 1,2,...,m$ denote the number of nonconforming items in the i^{th} reference sample of size $n \geq 1$ with unknown true fraction nonconforming $0 < p < 1$. The sample fraction nonconforming of each preliminary sample is $p_i = X_i / n$ for $i = 1,2,...,m$. Similarly, let $Y_i \sim iidPoi(c)$, $c > 0$ for $i = 1,2,...,m$ denote the number of nonconformities in the i^{th} reference inspection unit where c denotes the unknown true average number of nonconformities in an inspection unit.

Phase I point estimators for *p* and *c*

The average of the m Phase I sample fractions nonconforming $p_1, p_2,..., p_m$ and the average of the numbers of nonconformities in each Phase I inspection unit $Y_1, Y_2,..., Y_m$, are used to estimate p and c, respectively. In other words, we estimate p by

$$\bar{p} = \frac{1}{m}\sum_{i=1}^{m} p_i = \frac{1}{mn}\sum_{i=1}^{m} X_i = \frac{U}{mn} \qquad (3\text{-}24)$$

and c by

$$\bar{c} = \frac{1}{m}\sum_{i=1}^{m} Y_i = \frac{V}{m} \qquad (3\text{-}25)$$

where the random variable

$$U = \sum_{i=1}^{m} X_i \sim Bin(mn, p)$$

denotes the total number of nonconforming items in the entire set of mn reference observations and the random variable

$$V = \sum_{i=1}^{m} Y_i \sim Poi(mc)$$

denotes the total number of nonconformities in the entire set of m reference inspection units.

Remark 3

(i) It can be verified that the point estimators $\bar{p}$ and $\bar{c}$ in (3-24) and (3-25) are: (a) the maximum likelihood estimators (MLE's), and (b) the minimum variance unbiased estimators (MVUE's), of p and c, respectively (see e.g. Johnson, Kemp and Kotz, (2005) p. 126 and p. 174).

In particular, note that, the expected value and the variance of $\bar{p}$ are

$$E(\bar{p}) = \frac{E(U)}{mn} = \frac{mnp}{mn} = p,$$

and

$$\text{var}(\bar{p}) = \frac{\text{var}(U)}{(mn)^2} = \frac{mnp(1-p)}{(mn)^2} = \frac{p(1-p)}{mn},$$

respectively, whereas the expected value and the variance of $\bar{c}$ are

$$E(\bar{c}) = \frac{E(V)}{m} = \frac{mc}{m} = c,$$

and

$$\text{var}(\bar{c}) = \frac{\text{var}(V)}{m^2} = \frac{mc}{m^2} = \frac{c}{m},$$

respectively.

(ii) It is essential to note that the distribution of U depends on the unknown parameter p and the distribution of V depends on the unknown parameter c so that it is technically correct to write

$$U \mid p \sim Bin(mn, p) \qquad \text{and} \qquad V \mid c \sim Poi(mc)\,.$$

This observation will become vital when we study the unconditional run-length distributions and the characteristics of the unconditional run-length distribution in later sections.

3.2.2 Phase II *p*-chart and *c*-chart

A Phase II chart refers to the operation and implementation of a chart following a Phase I analysis in which any unknown parameters were estimated from the Phase I reference sample.

Phase II estimated control limits

It is standard practice to replace p_0 with $\bar{p}$ in (3-1) and substitute $\bar{c}$ for c_0 in (3-2) when the parameters p and/or c are unknown (see e.g. Ryan, (2000) p. 155 and p. 169 and, Montgomery, (2005) p. 269 and p. 290). The estimated upper control limits ($U\hat{C}L$'s), the estimated centerlines ($\hat{C}L$'s), and the estimated lower control limits ($L\hat{C}L$'s) of the *p*-chart and the *c*-chart are therefore given by

$$U\hat{C}L_p = \bar{p} + 3\sqrt{\bar{p}(1-\bar{p})/n} \qquad \hat{C}L_p = \bar{p} \qquad L\hat{C}L_p = \bar{p} - 3\sqrt{\bar{p}(1-\bar{p})/n} \tag{3-26}$$

and

$$U\hat{C}L_c = \bar{c} + 3\sqrt{\bar{c}} \qquad \hat{C}L_c = \bar{c} \qquad L\hat{C}L_c = \bar{c} - 3\sqrt{\bar{c}} \tag{3-27}$$

respectively.

By the invariance property of MLE's the estimated control limits in (3-26) and (3-27) are the MLE's of the control limits of (3-1) and (3-2) in Case K (see e.g. Theorem 7.2.10 in Casella and Berger, (2002) p. 320). However, unlike in Case K, the Phase II estimated control limits are functions of and depend on the point estimators (variables) $\bar{p}$ or $\bar{c}$ and are random variables. We therefore need to account for the variability in the estimated control limits while determining and understanding the chart's properties.

Phase II charting statistics

Let $p_i = X_i / n$ for $i = m+1, m+2,...$ denote the Phase II charting statistics for the *p*-chart where $X_i \sim iidBin(n, p_1)$ denote the number of nonconforming items in the i^{th} Phase II sample of size $n \geq 1$ with fraction nonconforming $0 < p_1 < 1$. Similarly, let $Y_i \sim iidPoi(c_1)$, $c_1 > 0$ for $i = m+1, m+2,...$ denote the number of nonconformities in the i^{th} Phase II inspection unit where c_1 denotes the average number of nonconformities in an inspection unit in Phase II. These Y_i's are the Phase II charting statistics of the *c*-chart.

Remark 4

(i) The *p*-chart

It is important to note that the application of the *p*-chart in Case U depends on three parameters: the unknown true fraction nonconforming p, the point estimate $\overline{p}$ and p_1.

In Phase II we denote p with p_1 so that p_1 denotes the probability of an item being nonconforming in the prospective monitoring phase and p denotes the probability of an item being nonconforming in the retrospective phase. To maintain greater generality and to cover both the in-control (IC) and the out-of-control (OOC) cases, we do not assume that p_1 is necessarily equal to p. We therefore write $p_1 = p$ for the IC scenario and $p_1 \neq p$ for the OOC case.

Also, in Phase I we estimate p by $\overline{p}$, which (due to sampling variability) is not necessarily equal to p; we write this as $p = \overline{p}$ and $p \neq \overline{p}$. When $p = \overline{p}$ we say that p is estimated without error.

This is a key observation. Because we use $\overline{p}$ to calculate the estimated control limits, in Phase II we are actually comparing p_1 against $\overline{p}$ and not against p; this leads to the following four unique scenarios:

(i) $p_1 = p = \overline{p}$: the process is IC in Phase II and p is estimated without error,

(ii) $p_1 \neq p = \overline{p}$: the process is OOC in Phase II and p is estimated without error,

(iii) $p_1 = p \neq \overline{p}$: the process is IC in Phase II and p is *not* estimated without error, and

(iv) $p_1 \neq p \neq \overline{p}$: the process is OOC in Phase II and p is *not* estimated without error.

To simplify matters we assume, without loss of generality, that the process operates IC in Phase II and $\overline{p}$ is not necessarily equal to p; this is scenario (iii) listed above.

(ii) The *c*-chart

For the *c*-chart in Case U we have a similar situation as that for the *p*-chart i.e. the application of the *c*-chart in Case U depends on three parameters: the true (but unknown) average number of nonconformities in an inspection unit c, the point estimate $\bar{c}$ and c_1.

In Phase II we denote c with c_1 so that c_1 denotes the average number of nonconformities in an inspection unit in the prospective monitoring phase and c denotes the average number of nonconformities in an inspection unit in the retrospective phase. To maintain greater generality and to cover both the in-control (IC) and the out-of-control (OOC) cases, we do not assume that c_1 is necessarily equal to c, which we write as $c_1 = c$ for the IC scenario and $c_1 \neq c$ for the OOC case.

In Phase I however we estimate c by $\bar{c}$, which (due to sampling variability) is not necessarily equal to c and we write this as $c = \bar{c}$ and $c \neq \bar{c}$. When $c = \bar{c}$ we say that c is estimated without error.

Now, because we use $\bar{c}$ to calculate the estimated control limits, in Phase II we are actually comparing c_1 against $\bar{c}$ and not c; this leads to the following four unique scenarios for the Phase II *c*-chart:

(i) $c_1 = c = \bar{c}$: the process is IC in Phase II and c is estimated without error,

(ii) $c_1 \neq c = \bar{c}$: the process is OOC in Phase II and c is estimated without error,

(iii) $c_1 = c \neq \bar{c}$: the process is IC in Phase II and c is *not* estimated without error, and

(iv) $c_1 \neq c \neq \bar{c}$: the process is OOC in Phase II and c is *not* estimated without error.

To simplify matters we assume, without loss of generality, that the process operates IC in Phase II and we assume that $\bar{c}$ is not necessarily equal to c; this is scenario (iii) listed above.

Phase II implementation and operation

The actual operation of the *p*-chart and the *c*-chart in Phase II consists of: (i) taking independent samples and independent inspection units (independent from the Phase I data), (ii) calculating the Phase II sample fractions nonconforming $p_i = X_i / n$ and the numbers of nonconformities in each Phase II inspection unit Y_i for $i = m+1, m+2, \ldots$, and then (iii) comparing these charting statistics (one at a time) to the estimated control limits in (3-26) and (3-27), respectively.

The moment that the first charting statistic plots on or outside the estimated limits a signal is given and the charting procedure stops. The process is then declared out-of-control and we say (in practice) that $p_1 \neq \bar{p}$ (in case of the *p*-chart) or state that $c_1 \neq \bar{c}$ (in case of the *c*-chart).

By comparing the Phase II charting statistics with the estimated control limits, the Phase II characteristics of the charts are (unlike in case K) affected by the variation in the point estimates $\bar{p} = U / mn$ and $\bar{c} = V / m$ where $U \mid p \sim Bin(mn, p)$ and $V \mid c \sim Poi(mc)$ are random variables but the values of m and n can be controlled or decided upon by the user.

The variation in the estimated control limits has significant implications on the properties of the charts. Most importantly the Phase II run-length distributions are no longer geometric since the Phase II signaling events are no longer independent. Intuitively, since estimating the limits introduces extra uncertainty it is expected that the run-length distributions in Case U will be more skewed to the right than the geometric. The additional variation must therefore be accounted for while determining and understanding the chart's properties. We give a systematic examination and detailed derivations of the Phase II run-length distributions of the *p*-chart and *c*-chart in what follows.

Phase II signaling event and Phase II non-signaling event

The event that occurs when a Phase II charting statistic plots outside the estimated control limits is called a Phase II signaling event and denoted by B_i for $i = m+1, m+2, \ldots$. In case of a Phase II signaling event, an alarm or signal is given and we declare the process out-of-control, that is, we say that $p_1 \neq \bar{p}$ or state that $c_1 \neq \bar{c}$. This means, for instance, that in practice we conclude that the probability p_1 of an item being nonconforming in Phase II is not equal to the estimated value $\bar{p}$.

The Phase II non-signaling event is the complementary event of the Phase II signaling event and occurs when a Phase II charting statistic plots within or between the estimated control limits. We denote the Phase II non-signaling event by

$$B_i^C : \{L\hat{C}L < Q_i < U\hat{C}L\}$$

where $Q_i = p_i$ or Y_i for $i = m+1, m+2, \ldots$ and $L\hat{C}L$ and $U\hat{C}L$ are the control limits in either (3-26) or (3-27), respectively.

In case of a Phase II non-signaling event no signal is given and we consider the process in-control, that is, we say that $p_1 = \bar{p}$ or state that $c_1 = \bar{c}$.

Dependency of the Phase II non-signaling events

If the Phase II signaling events were independent, the sequence of trials comparing each Phase II charting statistic Q_i with the estimated limits $U\hat{C}L$ and $L\hat{C}L$, would be a sequence of independent Bernoulli trials. The run-length between occurrences of the signaling event would therefore be a geometric random variable with probability of success equal to $\Pr(B_i)$. Moreover, the average run-length would be $ARL = 1/\Pr(B_i)$.

However, the signaling events B_i and B_j (or, equivalently, the non-signaling events B_i^C and B_j^C) are *not* mutually independent for $i \neq j = m+1, m+2, \ldots$ and the distribution of the run-length between the occurrences of the event B_i is as a result *not* geometric. In particular, because each Phase II p_i (or Y_i) for $i = m+1, m+2, \ldots$ is compared to the same set of estimated control limits, which are random variables, the signaling events are dependent.

To derive exact closed form expressions for the Phase II run-length distributions we use a two-step approach called the "method of conditioning" (see e.g. Chakraborti, (2000)). First we condition on the observed values of the random variables U and V to obtain the *conditional* Phase II run-length distribution and then use the *conditional* Phase II run-length distributions to obtain the marginal or *unconditional* Phase II run-length distributions.

To this end, note that, *given* (or *conditional on or having observed*) particular estimates of $\bar{p}$ and $\bar{c}$ (say $\bar{p}_{\text{obs}}$ and $\bar{c}_{\text{obs}}$), the Phase II non-signaling events *are mutually independent* each with the *same* probability so that the *conditional* Phase II run-length distributions *are* geometric. For instance, for a given or observed value of $\bar{p}$ (say $\bar{p}_{\text{obs}}$), the estimated Phase II control limits of the *p*-chart are constant i.e. they are *not* random variables, so that the conditional Phase II non-signaling events of the *p*-chart

$$\{\bar{p}-3\sqrt{\bar{p}(1-\bar{p})/n} < p_i < \bar{p}+3\sqrt{\bar{p}(1-\bar{p})/n} \mid \bar{p}=\bar{p}_{\text{obs}}\} \quad \text{for} \quad i=m+1, m+2, \ldots$$

are mutually independent each with the *same* probability given by

$$1-\hat{\beta}_p = 1-\Pr(\bar{p}-3\sqrt{\bar{p}(1-\bar{p})/n} < p_i < \bar{p}+3\sqrt{\bar{p}(1-\bar{p})/n} \mid \bar{p}=\bar{p}_{\text{obs}}). \tag{3-28}$$

The same is true for the *c*-chart. That is, for an observed value of $\bar{c}$ (say $\bar{c}_{\text{obs}}$) the events

$$\{\bar{c}-3\sqrt{\bar{c}} < Y_i < \bar{c}+3\sqrt{\bar{c}} \mid \bar{c}=\bar{c}_{\text{obs}}\} \quad \text{for} \quad i=m+1, m+2, \ldots$$

are mutually independent each with the *same* probability given by

$$1-\hat{\beta}_c = 1-\Pr(\bar{c}-3\sqrt{\bar{c}} < Y_i < \bar{c}+3\sqrt{\bar{c}} \mid \bar{c}=\bar{c}_{\text{obs}}). \tag{3-29}$$

The parameters of the *conditional* Phase II (geometric) run-length distributions are the conditional probabilities $1-\hat{\beta}_p$ and $1-\hat{\beta}_c$ so that, symbolically, we write

$$(N \mid \bar{p} = \bar{p}_{\text{obs}}) \sim Geo(1-\hat{\beta}_p) \qquad \text{and} \qquad (N \mid \bar{c} = \bar{c}_{\text{obs}}) \sim Geo(1-\hat{\beta}_c).$$

Thus, once the Phase I reference samples are gathered and the control limits are estimated, the Phase II run-length of a particular chart will follow some *conditional* distribution which will depend on the realization of the random variable $U = u$ or $V = v$, or, alternatively, on the observed values $\bar{p} = \bar{p}_{\text{obs}}$ or $\bar{c} = \bar{c}_{\text{obs}}$.

Note that the distributions of $U \mid p \sim Bin(mn, p)$ and $V \mid c \sim Poi(mc)$, or, equivalently, the distributions of $\bar{p}$ and $\bar{c}$, depend on the values of the unknown parameters p or c (see e.g. Remark 3(ii) as well as expressions (3-24) and (3-25), respectively). It is therefore better to write the conditional run-length distributions as

$$(N \mid \bar{p} = \bar{p}_{\text{obs}}, p) \sim Geo(1-\hat{\beta}_p) \qquad \text{and} \qquad (N \mid \bar{c} = \bar{c}_{\text{obs}}, c) \sim Geo(1-\hat{\beta}_c).$$

Moreover the conditional Phase II run-length distribution therefore provides only hypothetical information about the performance of a control chart with an estimated parameter. We can, for example, only assume some hypothetical value for p or c and then suppose that this estimate of p or c is the 25^{th} or the 75^{th} percentile of the sampling distributions of $\bar{p}$ or $\bar{c}$ so that the run-length distribution, conditioned on such a value, gives some insight into how a chart with this estimate performs in practice. This gives the user an idea of just how poorly or how well a chart will perform in a hypothetical case with an estimated parameter.

To overcome this abovementioned dilemma, the *marginal* or the *unconditional* run-length distribution can give a practitioner insight into a chart's general performance. The marginal distribution incorporates the additional variability which is introduced to the run-length through estimation of p or c by averaging over all possible values of the random variable U or V (while, of course, assuming a particular value for p or c). With the unconditional run-length distribution the practitioner therefore sees the overall effect of estimation on the run-length distribution *before* any data is collected.

3.2.3 Conditional Phase II run-length distributions and characteristics

The conditional run-length distributions and the associated conditional characteristics focus on the performance of the charts given $\overline{p} = p_{\text{obs}}$ and $\overline{c} = c_{\text{obs}}$.

Conditional probability of a no-signal

The probability of a no-signal in Phase II conditional on the point estimate $\overline{p} = p_{\text{obs}}$ or $\overline{c} = c_{\text{obs}}$ is called the conditional probability of a no-signal. This probability, which we previously denoted by $\hat{\beta}_p$ or $\hat{\beta}_c$, is in general denoted by

$$\hat{\beta} = \Pr(B_i^C \mid \hat{\theta}) \quad \text{for} \quad i = m+1, m+2, \ldots$$

where $\hat{\theta} = (\overline{p}, p)$ in case of the *p*-chart and $\hat{\theta} = (\overline{c}, c)$ in case of the *c*-chart.

The conditional probability of a no-signal, like in Case K (see e.g. Tables 3.1 and 3.2), completely characterizes the conditional Phase II run-length distribution and is thus the key to derive and examine the conditional Phase II run-length distributions of Case U. We derive exact expressions for $\hat{\beta}$ for both charts in what follows.

Conditional probability of a no-signal: *p*-chart

This probability is derived by conditioning on an observed value u of the random variable U or, equivalently, conditioning on an observed value $\bar{p}_{obs}$ of the point estimator $\bar{p} = U / mn$ (see e.g. (3-28)).

In doing so, the Phase II charting statistic $p_i = X_i / n$ for $i = m+1, m+2,...$ is the only random variable in (3-28). The cumulative distribution function of p_i for $i = m+1, m+2,...$, as mentioned earlier, is completely known and given by

$$\Pr(p_i \le a) = \Pr(X_i / n \le a) = \Pr(X_i \le na) = \sum_{j=0}^{[na]} \Pr(X_i = j) = \sum_{j=0}^{[na]} \binom{n}{j} p_1^j (1-p_1)^{n-j} \text{ for } 0 \le a \le 1 \text{ and}$$

p_1 denotes the true fraction nonconforming in Phase II (see Remark 4).

We therefore derive the conditional probability of a no-signal by first re-expressing the Phase II conditional non-signaling event in terms of X_i. This is done by making use of the relationship $X_i = np_i$. We then use the properties of X_i to derive an explicit and exact expression for the conditional probability of a no-signal.

For the p-chart the conditional probability of a no-signal in Phase II is

$$\hat{\beta}(p_1, m, n \mid \bar{p} = \bar{p}_{\text{obs}}, p)$$
$$= \Pr(L\hat{C}L_p < p_i < U\hat{C}L_p \mid \bar{p} = \bar{p}_{\text{obs}}, p)$$
$$= \Pr(X_i < nU\hat{C}L_p | \bar{p} = \bar{p}_{\text{obs}}, p) - \Pr(X_i \le nL\hat{C}L_p | \bar{p} = \bar{p}_{\text{obs}}, p)$$

$$= \Pr(X_i < n\{\bar{p} + 3\sqrt{\bar{p}(1-\bar{p})/n}\} | \bar{p} = \bar{p}_{\text{obs}}, p) - \Pr(X_i \le n\{\bar{p} - 3\sqrt{\bar{p}(1-\bar{p})/n}\} | \bar{p} = \bar{p}_{\text{obs}}, p)$$

$$= \Pr(X_i < n\{\frac{U}{mn} + 3\sqrt{\frac{U}{mn}(1-\frac{U}{mn})/n}\} \mid U = u, p) - \Pr(X_i \le n\{\frac{U}{mn} - 3\sqrt{\frac{U}{mn}(1-\frac{U}{mn})/n}\} | U = u, p)$$

$$= \Pr(X_i < m^{-1}(U + 3\sqrt{mU - n^{-1}U^2}) \mid U = u, p) - \Pr(X_i \le m^{-1}(U - 3\sqrt{mU - n^{-1}U^2}) \mid U = u, p) \quad \text{(3-30)}$$

$$= \begin{cases} 0 & \text{if } U = 0 \text{ or } U = mn \\ \bar{H}(\hat{b}, p_1, n) & \text{if } nL\hat{C}L_p < 0 \\ \bar{H}(\hat{b}, p_1, n) - \bar{H}(\hat{a}, p_1, n) & \text{if } nL\hat{C}L_p \ge 0 \end{cases}$$

$$= \begin{cases} 0 & \text{if } U = 0 \text{ or } U = mn \\ 1 - I_{p_1}(\hat{b}+1, n-\hat{b}) & \text{if } nL\hat{C}L_p < 0 \\ I_{p_1}(\hat{a}+1, n-\hat{a}) - I_{p_1}(\hat{b}+1, n-\hat{b}) & \text{if } nL\hat{C}L_p \ge 0 \end{cases}$$

$$= \begin{cases} 0 & \text{if } U = 0 \text{ or } U = mn \\ 1 - I_{p_1}(\hat{b}+1, n-\hat{b}) - 1_{\{nL\hat{C}L_p : nL\hat{C}L_p \ge 0\}}(nL\hat{C}L_p)(1 - I_{p_1}(\hat{a}+1, n-\hat{a})) & \text{if } U = 1,2,\dots,mn-1 \end{cases}$$

for $0 < p, \bar{p}, p_1 < 1$, where

$$\bar{H}(\hat{b}, p_1, n) = \sum_{j=0}^{\hat{b}} \binom{n}{j} p_1^j (1-p_1)^{n-j}$$

denotes the c.d.f of the $Bin(n, p_1)$ distribution and

$$\hat{a} = \hat{a}(m, n \mid U, p) = [nL\hat{C}L_p] \quad \text{(3-31a)}$$

and

$$\hat{b} = \hat{b}(m, n \mid U, p) = \begin{cases} \min\{nU\hat{C}L_p - 1, n\} & \text{if } nU\hat{C}L \text{ is an integer} \\ \min\{[nU\hat{C}L_p], n\} & \text{if } nU\hat{C}L \text{ is not an integer.} \end{cases} \quad \text{(3-31b)}$$

Remark 5

(i) The conditional probability of a no-signal for the *p*-chart is a function of and depends on

a. the fraction nonconforming in Phase II p_1,
b. the number of reference samples m,
c. the sample size n,
d. the point estimator $\overline{p}$ or, equivalently, the random variable U, and
e. the unknown true fraction nonconforming p; indirectly via the random variable $U \mid p \sim Bin(mn, p)$.

As noted earlier in Remark 4(i), p_1 is not necessarily equal to p and because of sampling variability $\overline{p}$ is typically different from p.

(ii) When none of the Phase I reference sample observations are nonconforming, that is, when $U = 0$ or $\overline{p} = 0$, it makes sense not to continue to Phase II but examine the situation in more detail. Similar logic applies to the other extreme, that is when all the observations are nonconforming so that $U = mn$ or $\overline{p} = 1$.

Based on this intuitive reasoning the conditional probability of a no-signal $\hat{\beta}(p_1, m, n \mid \overline{p}, p)$ is defined to be zero in both of these boundary situations. It then follows that the conditional probability of a signal $1 - \hat{\beta}(p_1, m, n \mid \overline{p}, p)$ is one. Effectively the control chart signals, in these cases, when p_i for $i = m+1, m+2, \ldots$ plots on or beyond either of the two estimated control limits or is equal to either 0 or n; this, in actual fact, implies that the *p*-chart signals on the first Phase II sample.

Conditional probability of a no-signal: *c*-chart

By conditioning on an observed value v of the random variable V or, equivalently, conditioning on an observed value $\bar{c}_{\text{obs}}$ of the point estimator $\bar{c} = V/m$, the Phase II charting statistic Y_i for $i = m+1, m+2,...$ is the only random quantity (variable) in (3-29).

Because the distribution of Y_i is known (assumed) to be Poisson with parameter c_1, we use the properties of this distribution to derive an explicit and exact expression for the conditional probability of a no-signal for the *c*-chart.

The conditional probability of a no-signal in Phase II is

$$\begin{aligned}
&\hat{\beta}(c_1, m \,|\, \bar{c} = \bar{c}_{\text{obs}}, c)\\
&= \Pr(L\hat{C}L_c < Y_i < U\hat{C}L_c \,|\, \bar{c} = \bar{c}_{\text{obs}}, c)\\
&= \Pr(Y_i < U\hat{C}L_c \,|\, \bar{c} = \bar{c}_{\text{obs}}, c) - \Pr(Y_i \le L\hat{C}L_c \,|\, \bar{c} = \bar{c}_{\text{obs}}, c)\\
&= \Pr(Y_i < \bar{c} + 3\sqrt{\bar{c}} \,|\, \bar{c} = \bar{c}_{\text{obs}}, c) - \Pr(Y_i \le \bar{c} - 3\sqrt{\bar{c}} \,|\, \bar{c} = \bar{c}_{\text{obs}}, c)\\
&= \Pr\left(Y_i < \frac{V}{m} + 3\sqrt{\frac{V}{m}} \,\Big|\, V = v, c\right) - \Pr\left(Y_i \le \frac{V}{m} - 3\sqrt{\frac{V}{m}} \,\Big|\, V = v, c\right) \qquad (3\text{-}32)\\
&= \begin{cases} 0 & \text{if} \quad V = 0 \\ \bar{G}(\hat{f}; c_1) - \bar{G}(\hat{d}; c_1) & \text{if} \quad V = 1,2,3,... \end{cases}\\
&= \begin{cases} 0 & \text{if} \quad V = 0 \\ \Gamma_{\hat{f}+1}(c_1) - \Gamma_{\hat{d}+1}(c_1) & \text{if} \quad V = 1,2,3,... \end{cases}
\end{aligned}$$

for $c, \bar{c}, c_1 > 0$, where

$$\bar{G}(\hat{f}; c_1) = \sum_{j=0}^{\hat{f}} \frac{e^{-c_1} c_1^j}{j!}$$

denotes the c.d.f of the $Poi(c_1)$ distribution and

$$\hat{d} = \hat{d}(m \,|\, V, c) = \max\{0, [L\hat{C}L_c]\} \qquad (3\text{-}33a)$$

and

$$\hat{f} = \hat{f}(m \,|\, V, c) = \begin{cases} U\hat{C}L_c - 1 & \text{if } U\hat{C}L_c \text{ is an integer} \\ [U\hat{C}L_c] & \text{if } U\hat{C}L_c \text{ is not an integer.} \end{cases} \qquad (3\text{-}33b)$$

Remark 6

(i) The probability of a no-signal for the *c*-chart is a function of and depends on

a. the average number of nonconformities in an inspection unit in Phase II c_1,

b. the number of reference inspection units m from Phase I,

c. the point estimator $\bar{c}$ or, equivalently, the random variable V, and

d. the unknown true average number of nonconformities in an inspection unit c; indirectly via the random variable $V \mid c \sim Poi(mc)$.

Again, note that, c_1 is not necessarily equal to c, and since $\bar{c}$ is subject to sampling variation it is typically different from c.

(ii) When we observe no nonconformities in the Phase I reference sample i.e. when $V = 0$ or $\bar{c} = 0$, it is essential to pause and examine the situation in more detail. Thus, for $V = 0$ the conditional probability of a no-signal in Phase II is defined to be zero so that the conditional probability of a signal in Phase II is one.

Summary of the conditional run-length distributions and the related conditional characteristics

Given observed values u and v of the random variables U and V, the conditional run-length distributions of the charts are geometric with the probability of success equal to the conditional probability of a signal i.e.

$$1-\hat{\beta}(p_1, m, n \mid U = u, p) \quad \text{and} \quad 1-\hat{\beta}(c_1, m \mid V = v, c)$$

respectively.

This is so, because for given or fixed values of $U = u$ and $V = v$ the control limits can be calculated exactly and the analyses continue as if the parameters p and c are known. This is similar to the standards known case (Case K) where the run-length distribution was seen to be geometric. All the characteristics of the conditional run-length distributions therefore follow from the well-known properties of the geometric distribution. In particular, the conditional run-length distributions and the associated conditional characteristics for the *p*-chart and the *c*-chart are summarized in Table 3.3 and Table 3.4, respectively.

The conditional run-length distribution and the conditional characteristics of the run-length distributions all depend on either the observed value of the random variable U or that of V; these observed values cannot be controlled by the user and is a direct result of estimating p and c. Thus, as the values of U and V change (randomly), the conditional run-length distributions and the conditional characteristics of the run-length distributions will also change randomly. This implies, for example, that the conditional characteristics are random variables which all have their own probability distributions so that one can present a quantity such as the expected conditional *SDRL* i.e. $E_U(CSDRL(p_1, m, n \mid U, p)$ or $E_V(CSDRL(c_1, m \mid V, c)$. Although this is technically correct it is not the best approach; a better approach would be to calculate the unconditional standard deviation i.e.

$$USDRL = \sqrt{E_U(\text{var}(p_1, m, n \mid U, p)) + \text{var}_U(E(p_1, m, n \mid U, p))}$$

or

$$USDRL = \sqrt{E_V(\text{var}(c_1, m \mid V, c)) + \text{var}_V(E(c_1, m \mid V, c))}$$

which is computed from the marginal run-length distribution and incorporates both the expected conditional *SDRL* and the variation in the expected conditional *ARL*. We discuss this in more detail later when we examine the conditional and unconditional properties of the charts.

Table 3.3: The conditional probability mass function (c.p.m.f), the conditional cumulative distribution function (c.c.d.f), the conditional false alarm rate (*CFAR*), the conditional average run-length (*CARL*) and the conditional standard deviation of the run-length (*CSDRL*) of the *p*-chart in Phase II of Case U

c.p.m.f	$\Pr(N_p = j; p_1, m, n \mid U, p) = [\hat{\beta}(p_1, m, n \mid U, p)]^{j-1}[1 - \hat{\beta}(p_1, m, n \mid U, p)] \quad j = 1,2,....$	(3-34)
c.c.d.f	$\Pr(N_p \le j; p_1, m, n \mid U, p) = 1 - [\hat{\beta}(p_1, m, n \mid U, p)]^{j} \quad j = 1,2,....$	(3-35)
CFAR	$CFAR(p_1, m, n \mid U, p = p_1) = 1 - \hat{\beta}(p_1, m, n \mid U, p = p_1)$	(3-36)
CARL	$CARL(p_1, m, n \mid U, p) = 1/[1 - \hat{\beta}(p_1, m, n \mid U, p)]$	(3-37)
CSDRL	$CSDRL(p_1, m, n \mid U, p) = \sqrt{\hat{\beta}(p_1, m, n \mid U, p)}\,/[1 - \hat{\beta}(p_1, m, n \mid U, p)]$	(3-38)
cqf	$Q_{N_p}(q; p_1, m, n \mid U, p) = \inf\{\text{int}\, x : \Pr(N_p \le j; p_1, m, n \mid U, p) \ge q\} \quad 0 < q < 1$	(3-39)

Table 3.4: The conditional probability mass function (c.p.m.f), the conditional cumulative distribution function (c.c.d.f), the conditional false alarm rate (*CFAR*), the conditional average run-length (*CARL*) and the conditional standard deviation of the run-length (*CSDRL*) of the *c*-chart in Phase II of Case U

c.p.m.f	$\Pr(N_c = j; c_1, m \mid V, c) = [\hat{\beta}(c_1, m \mid V, c)]^{j-1}[1 - \hat{\beta}(c_1, m \mid V, c)] \quad j = 1,2,....$	(3-40)
c.c.d.f	$\Pr(N_c \le j; c_1, m \mid V, c) = 1 - [\hat{\beta}(c_1, m \mid V, c)]^{j} \quad j = 1,2,....$	(3-41)
CFAR	$CFAR(c_1, m \mid V, c = c_1) = 1 - \hat{\beta}(c_1, m \mid V, c = c_1)$	(3-42)
CARL	$CARL(c_1, m \mid V, c) = 1/[1 - \hat{\beta}(c_1, m \mid V, c)]$	(3-43)
CSDRL	$CSDRL(c_1, m \mid V, c) = \sqrt{\hat{\beta}(c_1, m \mid V, c)}\,/[1 - \hat{\beta}(c_1, m \mid V, c)]$	(3-44)
cqf	$Q_{N_c}(q; c_1, m \mid V) = \inf\{\text{int}\, x : \Pr(N_c \le j; c_1, m \mid V, c) \ge q\} \quad 0 < q < 1$	(3-45)

It is important to note that the conditional run-length distributions and the associated characteristics of the conditional run-length distributions do not only depend on the random variables U and V; they also indirectly depend on the unknown parameters p and c.

The dependency on U and V follows from the fact that we estimate p using $\bar{p} = U/mn$ and we estimate c using $\bar{c} = V/m$. The indirect dependency on p and c follows from the fact that the distribution of U (which is binomial with parameters mn and p) and the distribution of V (which is Poisson with parameter mc) depend on the unknown parameters p and c. To evaluate any of the conditional characteristics we need the observed values of U and V but we also need to assume values for p and c.

The aforementioned point is demonstrated in the following two examples which illustrate the operation and the implementation of the Phase II *p*-chart and the Phase II *c*-chart when we are given a particular Phase I sample.

Example 1: A Phase II *p*-chart

Consider Example 6.1 on p. 289 of Montgomery (2001) concerning a frozen orange juice concentrate that is packed in 6-oz cardboard cans. A machine is used to make the cans and the goal is to set up a control chart to improve i.e. decrease, the fraction of nonconforming cans produced by the machine. Since no specific value of the fraction nonconforming p is given the scenario is an example of Case U, that is, when the standard is unknown. The chart is therefore implemented in two stages.

Phase I

To establish the control chart $m = 30$ reference samples were taken each with $n = 50$ cans, selected in half hour intervals over a three-shift period in which the machine was in continuous operation. Once the Phase I control chart was established samples 15 and 23 were found to be out-of-control and eliminated after further investigation. Revised control limits were calculated using the remaining $m = 28$ samples. Based on the revised control limits sample 21 was found out-of-control, but since further investigations regarding sample 21 did not produce any reasonable or logical assignable cause it was not discarded. This is the retrospective phase (or Phase I) of the analysis. The final 28 samples were used to estimate the control limits and then monitor the process in Phase II.

Phase II (conditional)

Although the random variable U could theoretically take on any integer value from 0 to $mn = 28 \times 50 = 1400$, for the given set of reference data it was found that $U = 301$; this was the total number of nonconforming cans after discarding samples 15 and 23. It follows from (3-24) that the point estimate of p is $\bar{p} = 301/1400 = 0.215$.

The estimated control limits and centerline corresponding to $U = 301$ are found from (3-26) to be

$$U\hat{C}L_p = 0.215 + 3\sqrt{0.215(0.785)/50} = 0.3893 \text{ and } L\hat{C}L_p = 0.215 - 3\sqrt{0.215(0.785)/50} = 0.0407.$$

We find the constants $\hat{a}$ and $\hat{b}$ using (3-31) to be

$$\hat{b}(m = 28, n = 50 \mid U = 301, p) = 19 \quad \text{and} \quad \hat{a}(m = 28, n = 50 \mid U = 301, p) = 2.$$

Because U is unequal to 0 or mn it follows from (3-30) that the conditional probability of a no-signal in Phase II is

$$\begin{aligned}\hat{\beta}(p_1, m = 28, n = 50 \mid \bar{p} = 0.215, p) &= 1 - I_{p_1}(19, 50 - 19 - 1) - (1 - I_{p_1}(2, 50 - 2 - 1)) \\ &= I_{p_1}(2,47) - I_{p_1}(19,30)\end{aligned}$$

for $0 < p, p_1 < 1$.

Assuming, without loss of generality, that the process is in-control at a fraction nonconforming of 0.2, that is, $p_1 = p = 0.2$, the conditional false alarm rate (*CFAR*) is equal to

$$1 - \hat{\beta}(p_1 = 0.2, 28, 50 \mid \bar{p} = 0.215, p = 0.2) = 1 - I_{0.2}(2,47) + I_{0.2}(19,30) = 0.002218.$$

The in-control conditional average run-length therefore equals

$$CARL_0 = 1/0.002218 = 450.89$$

and is found using (3-37).

Compared to the Case K *FAR* and *ARL* of 0.0027 and 369.84 (see e.g. Tables A3.4 and A3.5 of Appendix 3A) we see that our *p*-chart (here, in Case U, with $\bar{p} = 0.215$ and assuming that $p_1 = p = 0.2$) would signal less often, if the process is in-control, than what it would if p had in fact been known to be equal to 0.2.

However, note that, since each user has his/her own unique reference sample, the point estimate $\bar{p}$ will differ from one user to the next so that the performance of each user's chart will also vary. To this end, the unconditional characteristics are useful as they do not depend on any specific observed value of the point estimate. This, however, is looked at later when we continue Example 1 after having derived expressions for the unconditional characteristics of the *p*-chart's Phase II run-length distribution. ■

Example 2: A Phase II *c*-chart

Consider Example 6.3 on p. 310 in Montgomery (2001) about the quality control of manufactured printed circuit boards. Since c is not specified it had to be estimated. The chart was therefore implemented in two phases.

Phase I

A total of 26 successive inspection units each consisting of 100 individual items of product were obtained to estimate the unknown true average number of nonconformities in an inspection unit *c*. It was found that units number 6 and 20 were out-of-control and therefore eliminated. The revised control limits were calculated using the remaining $m = 24$ inspection units with the number of nonconformities in an inspection unit shown in Table 6.7 on p. 311 of Montgomery (2001). The revised control limits were used for monitoring the process in Phase II.

Phase II (conditional)

Theoretically the variable V, the total number of nonconformities in the 24 inspection units, could take on any positive integer value including zero i.e. $V \in \{0,1,2,...\}$. For the given Phase I data it is found that $V = 472$. Using (3-25) the average number of nonconformities in an inspection unit c is estimated as $\bar{c} = 472/24 = 19.67$ so that the estimated 3-sigma control limits are found from (3-27) to be

$$U\hat{C}L_c = 32.97 \quad \text{and} \quad L\hat{C}L_c = 6.36\,.$$

These estimated limits yield

$$\hat{d}(m = 24 \,|\, V = 472, c) = 6 \quad \text{and} \quad \hat{f}(m = 24 \,|\, V = 472, c) = 32\,.$$

Because V is unequal to zero it follows from (3-32) that the probability of a no-signal is

$$\hat{\beta}(c_1, 24 \,|\, \bar{c} = 19.67, c) = \Gamma_{33}(c_1) - \Gamma_7(c_1) \quad \text{for} \quad c, c_1 > 0\,.$$

For the given (observed) value of $V = 472$ one can investigate the chart's performance using the conditional properties. Assuming, without loss of generality, that the process operates in-control at an average of twenty nonconformities in an inspection unit, that is, $c_1 = c = 20$ is the true in-control average number of nonconformities in an inspection unit, the conditional false alarm rate i.e. the false alarm rate given $V = 472$, is found to be equal to

$$CFAR = 1 - \Gamma_{33}(20) - \Gamma_7(20) = 0.004983.$$

The *CFAR* is approximately 72% larger than the value of 0.0029 one would have obtained in Case K for $c_0 = 20$ and is 85% higher than the nominal value 0.0027 (see e.g. Table A3.12 in Appendix 3A); this is true even though the estimated average number of nonconformities in an inspection unit ($\bar{c} = 19.67$) is within $|(19.67 - 20)/\sqrt{20}| = 0.07$ standard deviation units of the true average number of nonconformities in an inspection unit ($c = 20$). However, note that, like the *p*-chart of Example 1, each user typically has his/her own distinct Phase I data so that the performance of the *c*-chart in Case U will be different for each user. ■

To get an overall picture of a *p*-chart's or a *c*-chart's performance one needs to look at the unconditional properties of the chart; this is looked at later. First we look at the conditional run-length distribution and the related conditional characteristics of the *p*-chart and *c*-chart.

The characteristics of the conditional run-length distribution depend on and are functions of the random variables U or V; as a result, these characteristics are random variables themselves and vary as U or V changes.

To understand the effect of U or V on the characteristics of the conditional run-length distribution, it is instructive to study the conditional characteristics of the charts as functions of U and V as they show precisely how the conditional characteristics of each chart vary as the point estimates $\bar{p}$ and $\bar{c}$ fluctuate.

First we look at the conditional characteristics of the *p*-chart and then at those of the *c*-chart.

3.2.3.1 Conditional characteristics of the *p*-chart

Once we observed a value u of the random variable U we can calculate the conditional probability of a signal. The Phase II conditional run-length distribution is then completely known (see e.g. Table 3.3).

Tables 3.5 and 3.6 illustrate the exact steps to calculate the conditional probability of a no-signal, the conditional probability of a signal or the conditional false alarm rate (*CFAR*), the conditional average run-length (*CARL*) and the conditional standard deviation of the run-length (*CSDRL*) for the *p*-chart. These are all conditional Phase II properties as they all depend on an observed value from Phase I.

For illustration purposes we assume a total of $T = mn = 20$ individual Phase I observations is used to estimate p using $\bar{p} = U / mn$ as point estimate and that $p_1 = p = 0.5$. The latter assumption implies that the process operated at a fraction nonconforming of $p = 0.5$ during Phase I and that in Phase II the process continues to operate at this same level so that $p_1 = 0.5$; this is the same as saying that the process is in-control in Phase II. However, note that, because of sampling variation the observed value of $\bar{p}$ may of course not be equal to p (see e.g. Remark 4(i)).

The calculations of Table 3.5 are based on the assumption that $m = 4$ independent Phase I reference samples each of size $n = 5$ are used whereas the computations of Table 3.6 are based on $m = 1$ with $n = 20$.

In particular, column 1 lists all the values of U (the total number of possible nonconforming items in the entire Phase I reference sample) that can possibly be attained. This ranges from a minimum of zero to a maximum of twenty. Column 2 converts the observed value u of U into a point estimate of the unknown true fraction of nonconforming items, that is, we calculate $\bar{p} = u / 20 = \bar{p}_{\text{obs}}$ which estimates p. Because each row entry in each of the succeeding columns (i.e. columns 3 to 12) is computed by conditioning on a row entry from column 1 (or, equivalently, from column 2) we start calculating the conditional properties in columns 1 and/or 2 and sequentially proceed to the right-hand side of the tables. Thus, given a value u or $\bar{p}_{\text{obs}}$ the lower and the upper control limits are estimated in columns 3 and 4 using (3-26). These estimated limits are then used to compute the two constants $\hat{a}$ and $\hat{b}$ defined in (3-31), which are shown in columns 5 and 6, respectively. Finally, columns 7 through 10 list the probability of a no-signal, the *FAR*, the in-control *ARL* and the in-control *SDRL* given the observed value u from column 1, respectively. These properties are labeled $\Pr(\text{No Signal} \mid U, p)$, *CFAR*, $CARL_0$ and $CSDRL_0$, and calculated using (3-30) and the expressions in Table 3.3. Columns 11 and 12 show the values of the probability mass function (p.m.f) and the cumulative distribution function (c.d.f) of the random variable $U \mid p = 0.5 \sim Bin(20,0.5)$, that is,

$$\Pr(U = u \mid p = 0.5) = \binom{20}{u} 0.5^{20} \quad \text{and} \quad \Pr(U \leq u \mid p = 0.5) = \sum_{j=0}^{u} \binom{20}{j} 0.5^{20} \quad \text{for} \quad u = 0,1,2,...,20.$$

Both these probability functions are useful when interpreting the characteristics of the conditional run-length distribution. The former shows the exact probability of obtaining a particular value u of U whereas the latter can be used to find the percentiles of the distribution of U.

T = 20 with m = 4 and n = 5

Consider Table 3.5 which uses a total of $T = 20$ individual in-control Phase I reference observations from $m = 4$ independent samples each of size $n = 5$.

There are two unique scenarios. The first takes place when $U = 0$ (the minimum value possible) and the second occurs when $U = 4 \times 5 = 20$ (the maximum value). In both these cases the probability of a no-signal is zero by definition and the chart signals once the first Phase II sample is observed. As a result the conditional in-control average run-length is $CARL_0 = 1$. In the former situation the estimated control limits are $L\hat{C}L_p = U\hat{C}L_p = 0$ and in the latter the limits are $L\hat{C}L_p = U\hat{C}L_p = 1$. In both these situations the constants $\hat{a}$ and $\hat{b}$ need not be calculated; this is indicated by NA (read as "not applicable") in columns 5 and 6, respectively (see e.g. (3-30) and Remark 5(ii)).

The probability that none or all of the Phase I reference observations are nonconforming is of course rather small. The probabilities of these two events are $P(U = 0 \mid 0.5) = P(U = 20 \mid 0.5) = 0.5^{20}$ which are zero when rounded to four decimal places (see e.g. column 11). For all other values of $U \neq 0$ and $U \neq mn = 20$, that is, when $U \in \{1,2,...,19\}$, we proceed with the calculation of the conditional characteristics as follows.

Table 3.5: Conditional probability of a no-signal, the conditional false alarm rate (*CFAR*), the in-control conditional average run-length ($CARL_0$) and the in-control conditional standard deviation of the run-length ($CSDRL_0$) of the *p*-chart in Case U for $m = 4$ and $n = 5$, assuming that $p_1 = p = 0.5$

(1) u	(2) $\bar{p}_{obs}$	(3) $L\hat{C}L_p$	(4) $U\hat{C}L_p$	(5) $\hat{a}$	(6) $\hat{b}$	(7) Pr(No Signal | *U*, *p*)	(8) *CFAR*	(9) $CARL_0$	(10) $CSDRL_0$	(11) Pr(*U*=*u*| *p*)	(12) Pr(*U*<=*u*| *p*)
0	0.00	0.00	0.00	NA	NA	0.0000	1.0000	1.00	0.00	0.0000	0.0000
1	0.05	-0.24	0.34	NA	1	0.1875	0.8125	1.23	0.53	0.0000	0.0000
2	0.10	-0.30	0.50	NA	2	0.5000	0.5000	2.00	1.41	0.0002	0.0002
3	0.15	-0.33	0.63	NA	3	0.8125	0.1875	5.33	4.81	0.0011	0.0013
4	0.20	-0.34	0.74	NA	3	0.8125	0.1875	5.33	4.81	0.0046	0.0059
5	0.25	-0.33	0.83	NA	4	0.9688	0.0313	32.00	31.50	0.0148	0.0207
6	0.30	-0.31	0.91	NA	4	0.9688	0.0313	32.00	31.50	0.0370	0.0577
7	0.35	-0.29	0.99	NA	4	0.9688	0.0313	32.00	31.50	0.0739	0.1316
8	0.40	-0.26	1.06	NA	5	1.0000	0.0000	∞	∞	0.1201	0.2517
9	0.45	-0.22	1.12	NA	5	1.0000	0.0000	∞	∞	0.1602	0.4119
10	0.50	-0.17	1.17	NA	5	1.0000	0.0000	∞	∞	0.1762	0.5881
11	0.55	-0.12	1.22	NA	5	1.0000	0.0000	∞	∞	0.1602	0.7483
12	0.60	-0.06	1.26	NA	5	1.0000	0.0000	∞	∞	0.1201	0.8684
13	0.65	0.01	1.29	0	5	0.9688	0.0313	32.00	31.50	0.0739	0.9423
14	0.70	0.09	1.31	0	5	0.9688	0.0313	32.00	31.50	0.0370	0.9793
15	0.75	0.17	1.33	0	5	0.9688	0.0313	32.00	31.50	0.0148	0.9941
16	0.80	0.26	1.34	1	5	0.8125	0.1875	5.33	4.81	0.0046	0.9987
17	0.85	0.37	1.33	1	5	0.8125	0.1875	5.33	4.81	0.0011	0.9998
18	0.90	0.50	1.30	2	5	0.5000	0.5000	2.00	1.41	0.0002	1.0000
19	0.95	0.66	1.24	3	5	0.1875	0.8125	1.23	0.53	0.0000	1.0000
20	1.00	1.00	1.00	NA	NA	0.0000	1.0000	1.00	0.00	0.0000	1.0000

Suppose, for instance, that we observe seven nonconforming items out of the possible twenty in the entire Phase I reference sample. Our chance to find exactly seven nonconforming items is approximately 0.0739 (which is relatively high, see e.g. column 11); the probability to find less than seven nonconforming items is $P(U < 7 \mid 0.5) \approx 0.0577$ (see e.g. column 12).

A value of $U = 7$ gives a point estimate for p of $\bar{p} = 7/20 = 0.35$ so that (3-26) yields an estimated upper control limit and an estimated lower control limit of

$$U\hat{C}L_p = 0.35 + 3\sqrt{0.35(0.65)/5} = 0.99 \quad \text{and} \quad L\hat{C}L_p = 0.35 - 3\sqrt{0.35(0.65)/5} = -0.29$$

respectively .

Because $nL\hat{C}L_p = (5)(-0.29) = -1.45$ is less than zero the chart has no lower control limit. We therefore do not calculate a value for $\hat{a}$ in this case. The constant $\hat{b}$, on the other hand, is found to be

$$\hat{b} = \min\{[nU\hat{C}L_p],5\} = \min\{[(5)(0.99)],5\} = \min\{[4.95],5\} = 4.$$

Finally, after substituting $\hat{b}$ in (3-30) we calculate the conditional probability of a no-signal and then also the *CFAR*, the $CARL_0$ and the $CSDRL_0$ using expressions (3-36), (3-37) and (3-38) in Table 3.3.

The conditional probability of a no-signal is

$$\hat{\beta}(p_1 = 0.5, m = 4, n = 5 \mid U = 7, p = 0.5) = \hat{\beta}(p_1 = 0.5, m = 4, n = 5 \mid \bar{p} = 0.35, p = 0.5) = 0.9688,$$

so that the conditional false alarm rate is

$$CFAR(p_1 = 0.5,4,5 \mid U = 7, p = 0.5) = 1 - 0.9688 = 0.0313.$$

The Phase II p-chart then has an in-control conditional *ARL* of

$$CARL_0(p_1 = 0.5,4,5 \mid U = 7, p = 0.5) = 1/0.0313 = 32.00$$

and an in-control conditional *SDRL* of

$$SDRL_0(p_1 = 0.5,4,5 \mid U = 7, p = 0.5) = \sqrt{0.9688}/0.0313 = 31.50.$$

If the process remains to operate at $p_1 = 0.5$ (i.e. the process stays in-control) we expect that the chart would, on average, give a false alarm or erroneous signal on every 32nd sample. This is more often that what we would nominally expect from a 3-sigma Shewhart-type control chart, which typically has an in-control *ARL* of 370.4. We also see that the conditional false alarm rate (*CFAR*), particularly for $U = 7$, is much higher than the nominally expected 0.0027 even though the point estimate $\bar{p} = 0.35$ is $\left|\sqrt{5}(0.35 - 0.50)/\sqrt{0.35(1 - 0.35)}\right| = 0.70$ standard deviation units from the supposedly known value of $p = 0.5$.

For values of U from 8 to 12 the *CFAR* is equal to zero and as a result the moments of the run-length distribution, such as the $CARL_0$ and the $CSDRL_0$, are all undefined; this implies that, in practice, the conditional Phase II chart will not signal and that the $CARL_0$ and the $CSDRL_0$ are both infinite. Although we typically want a high in-control ARL, an *ARL* of infinity is not practical. Thus, $m = 4$ subgroups each of size $n = 5$ is not adequate to control the false alarm rate (*FAR*) at a small yet practically desirable level, and at the same time ensure that a high in-control *ARL* is achieved. This suggests that one needs more reference data and that n needs to be larger relative to m in order to achieve any reasonable probability of a false alarm with attributes data.

T = 20 with m = 1 and n = 20

To study the effect of choosing a larger value of n relative to m suppose that a total of $T = 20$ in-control Phase I reference observations are available but in one sample of twenty observations, that is, $m = 1$ and $n = 20$. Calculations for this situation are shown in Table 3.6.

We observe that the conditional probability of a no-signal i.e.

$$\Pr(\text{No Signal} \mid U, p) = \hat{\beta}(p_1 = 0.5, m = 1, n = 20 \mid U = u, p = 0.5)$$

is non-zero for all values of $U = 0,1,...,20$. As a result none of the *CFAR*'s values are zero and therefore all the moments (such as the in-control *ARL*, the in-control *SDRL* etc.) of the conditional run-length distribution are defined and finite. This suggests the need for a very careful choice of the number of reference samples m and the size n of each of the samples before a *p*-chart with an unknown value of p is implemented in practice.

Table 3.6: Conditional probability of a no-signal, the conditional false alarm rate (*CFAR*), the in-control conditional average run-length (*CARL*$_0$) and the in-control conditional standard deviation of the run-length (*CSDRL*$_0$) of the *p*-chart in Case U for $m = 1$ and $n = 20$, assuming that $p_1 = p = 0.5$

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
u	$\bar{p}_{obs}$	$\hat{LCL}_p$	$\hat{UCL}_p$	$\hat{a}$	$\hat{b}$	Pr(No Signal | *U, p*)	*CFAR*	*CARL*$_0$	*CSDRL*$_0$	Pr(*U*=*u*| *p*)	Pr(*U*<=*u*| *p*)
0	0.00	0.00	0.00	NA	NA	0.0000	1.0000	1.00	0.00	0.0000	0.0000
1	0.05	-0.10	0.20	NA	3	0.0013	0.9987	1.00	0.04	0.0000	0.0000
2	0.10	-0.10	0.30	NA	6	0.0577	0.9423	1.06	0.25	0.0002	0.0002
3	0.15	-0.09	0.39	NA	7	0.1316	0.8684	1.15	0.42	0.0011	0.0013
4	0.20	-0.07	0.47	NA	9	0.4119	0.5881	1.70	1.09	0.0046	0.0059
5	0.25	-0.04	0.54	NA	10	0.5881	0.4119	2.43	1.86	0.0148	0.0207
6	0.30	-0.01	0.61	NA	12	0.8684	0.1316	7.60	7.08	0.0370	0.0577
7	0.35	0.03	0.67	0	13	0.9423	0.0577	17.34	16.84	0.0739	0.1316
8	0.40	0.07	0.73	1	14	0.9793	0.0207	48.27	47.77	0.1201	0.2517
9	0.45	0.12	0.78	2	15	0.9939	0.0061	163.66	163.16	0.1602	0.4119
10	0.50	0.16	0.84	3	16	0.9974	0.0026	388.07	387.57	0.1762	0.5881
11	0.55	0.22	0.88	4	17	0.9939	0.0061	163.66	163.16	0.1602	0.7483
12	0.60	0.27	0.93	5	18	0.9793	0.0207	48.27	47.77	0.1201	0.8684
13	0.65	0.33	0.97	6	19	0.9423	0.0577	17.34	16.84	0.0739	0.9423
14	0.70	0.39	1.01	7	20	0.8684	0.1316	7.60	7.08	0.0370	0.9793
15	0.75	0.46	1.04	9	20	0.5881	0.4119	2.43	1.86	0.0148	0.9941
16	0.80	0.53	1.07	10	20	0.4119	0.5881	1.70	1.09	0.0046	0.9987
17	0.85	0.61	1.09	12	20	0.1316	0.8684	1.15	0.42	0.0011	0.9998
18	0.90	0.70	1.10	13	20	0.0577	0.9423	1.06	0.25	0.0002	1.0000
19	0.95	0.80	1.10	16	20	0.0013	0.9987	1.00	0.04	0.0000	1.0000
20	1.00	1.00	1.00	NA	NA	0.0000	1.0000	1.00	0.00	0.0000	1.0000

The conditional false alarm rate

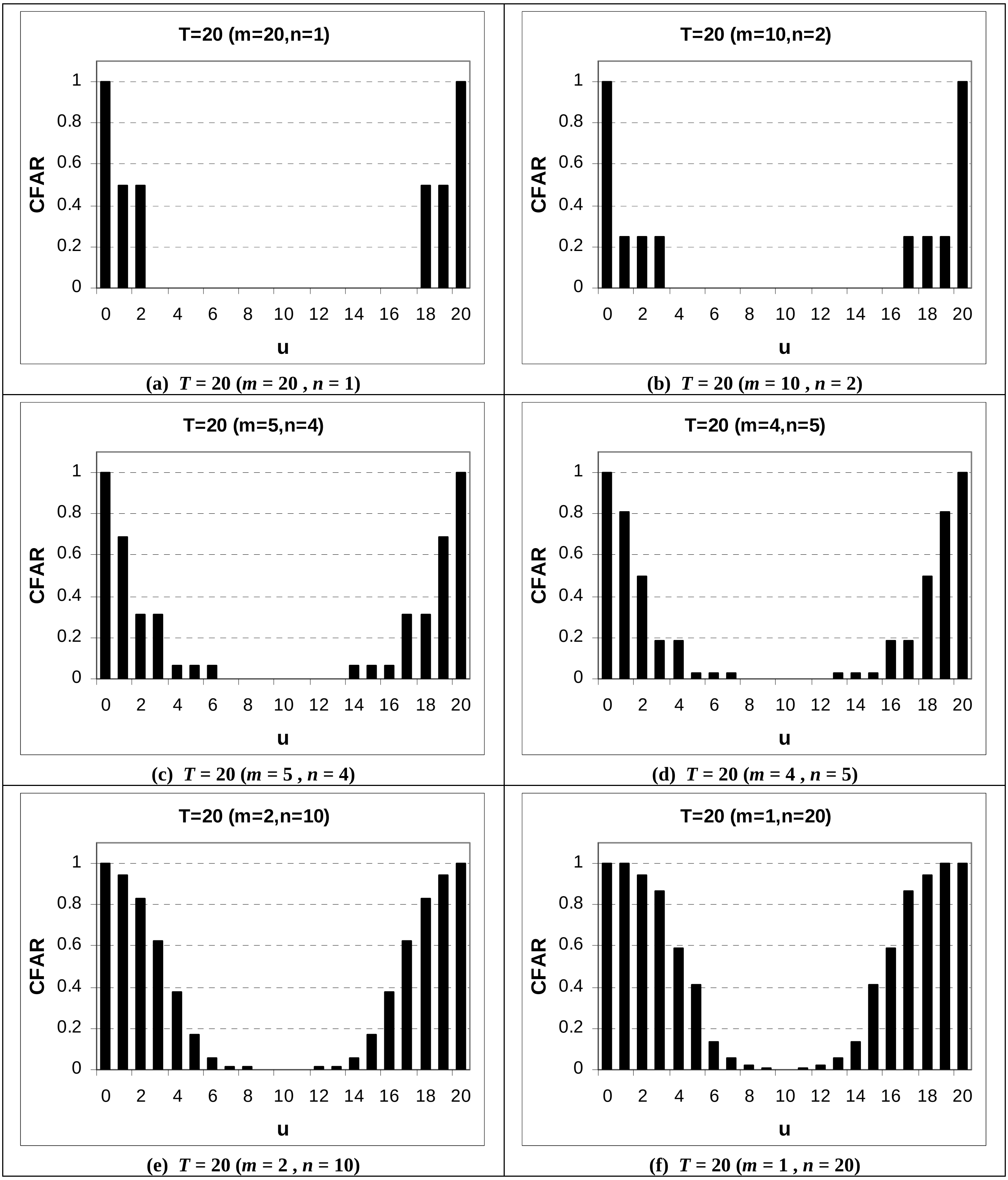
Panels (a) to (f) of Figures 3.1 and 3.2 display the conditional false alarm rate (*CFAR*)

$$1-\hat{\beta}(p_1 = 0.5, m, n \mid U = u, p = 0.5) \quad \text{as a function of} \quad u = 0,1,\ldots,mn$$

for various combinations of m and n when a total of $T = 20$ and a total of $T = 50$ individual Phase I reference observations are used to estimate p. For illustration purposes we assume that $p_1 = p = 0.50$.

The impact of the actual number of nonconforming items u in the entire Phase I reference sample is easily noticed. The distribution of the *CFAR* is seen to be U-shaped and symmetric at the point $mn/2$; this is the mean value of *U*. For values of *U* near the two tails the *CFAR* can be very high, sometimes close to 1 or 100%, which obviously means many false alarms. Of course, this only happens at the rather extreme values of *U* that occur with very small probabilities (see e.g. columns 11 and 12 in Tables 3.5 and 3.6). However, even when U is not as extreme there can be a significantly high probability of a false alarm and it is seen that only when U takes on a value in the neighbourhood of its mean, will the *CFAR* be reasonably small. A potential problem is that for some combinations of *m* and *n* values, especially with smaller values of *n* relative to m, some of the *CFAR* values equal 0, which (as mentioned before) leads to an in-control average run-length that is undefined.

Note that, panels (d) and (f) of Figure 3.1 are in fact displaying the *CFAR*'s of column 8 in Tables 3.5 and 3.6, respectively.



(a) $T = 20$ ($m = 20$, $n = 1$)

(b) $T = 20$ ($m = 10$, $n = 2$)

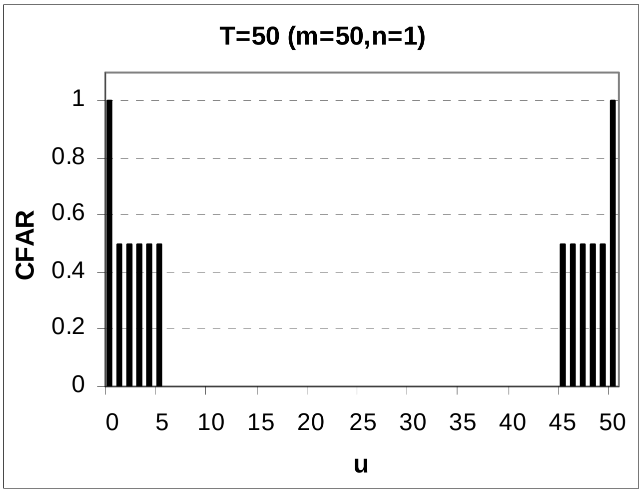
(c) $T = 20$ ($m = 5$, $n = 4$)

(d) $T = 20$ ($m = 4$, $n = 5$)

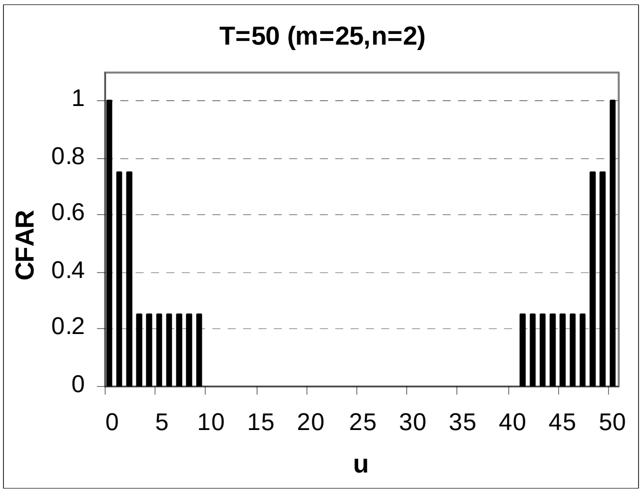
(e) $T = 20$ ($m = 2$, $n = 10$)

(f) $T = 20$ ($m = 1$, $n = 20$)

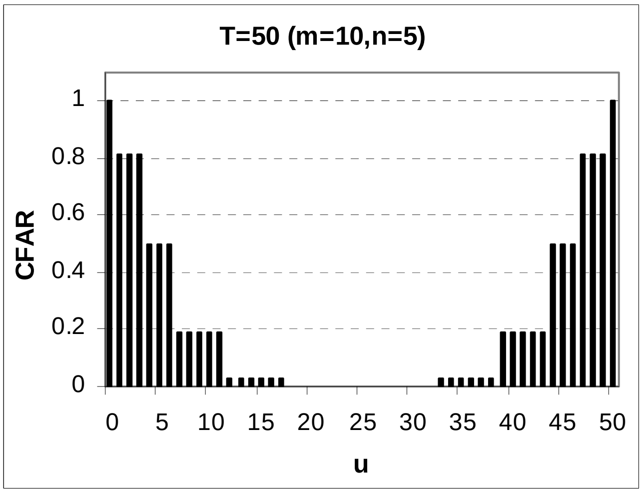
Figure 3.1: The conditional false alarm rate (*CFAR*) as a function of $u = 0,1,\ldots,20$ for various combinations of m and n such that $T = mn = 20$



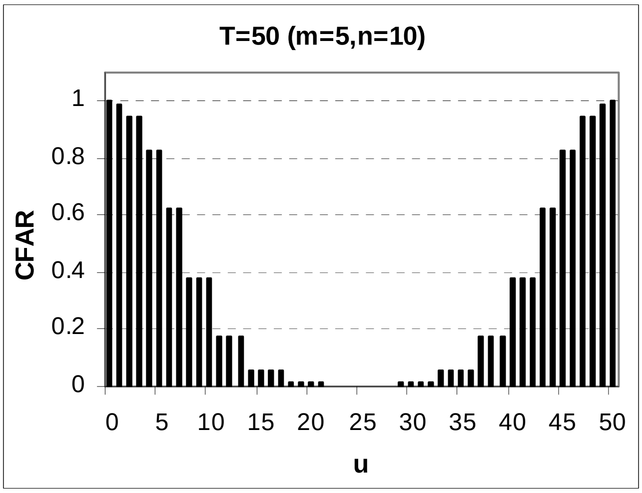
(a) $T = 50$ ($m = 50$, $n = 1$)



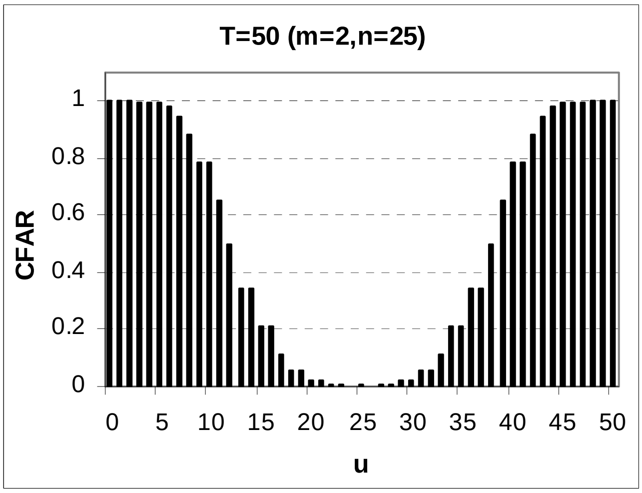
(b) $T = 50$ ($m = 25$, $n = 2$)



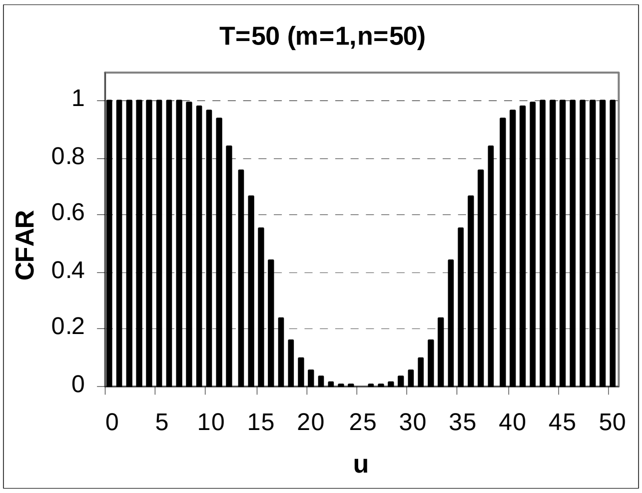
(c) $T = 50$ ($m = 10$, $n = 5$)



(d) $T = 50$ ($m = 5$, $n = 10$)



(e) $T = 50$ ($m = 2$, $n = 25$)



(f) $T = 50$ ($m = 1$, $n = 50$)

Figure 3.2: The conditional false alarm rate (*CFAR*) as a function of $u = 0,1,...,50$ for various combinations of m and n such that $T = mn = 50$

The conditional probability of a no-signal

The distribution of *1-CFAR*, which is the conditional probability of a no-signal when the process is in-control, is shown in panels (a) to (d) of Figure 3.3 for $T = 20$, 50, 100 and 200 when $m = 1$ and $n = T$ i.e. for large n relative to m.

It is seen that the distribution of *1-CFAR* is bell-shaped and symmetric; these two characteristics follow from that of *CFAR* shown in Figure 3.1 and 3.2.

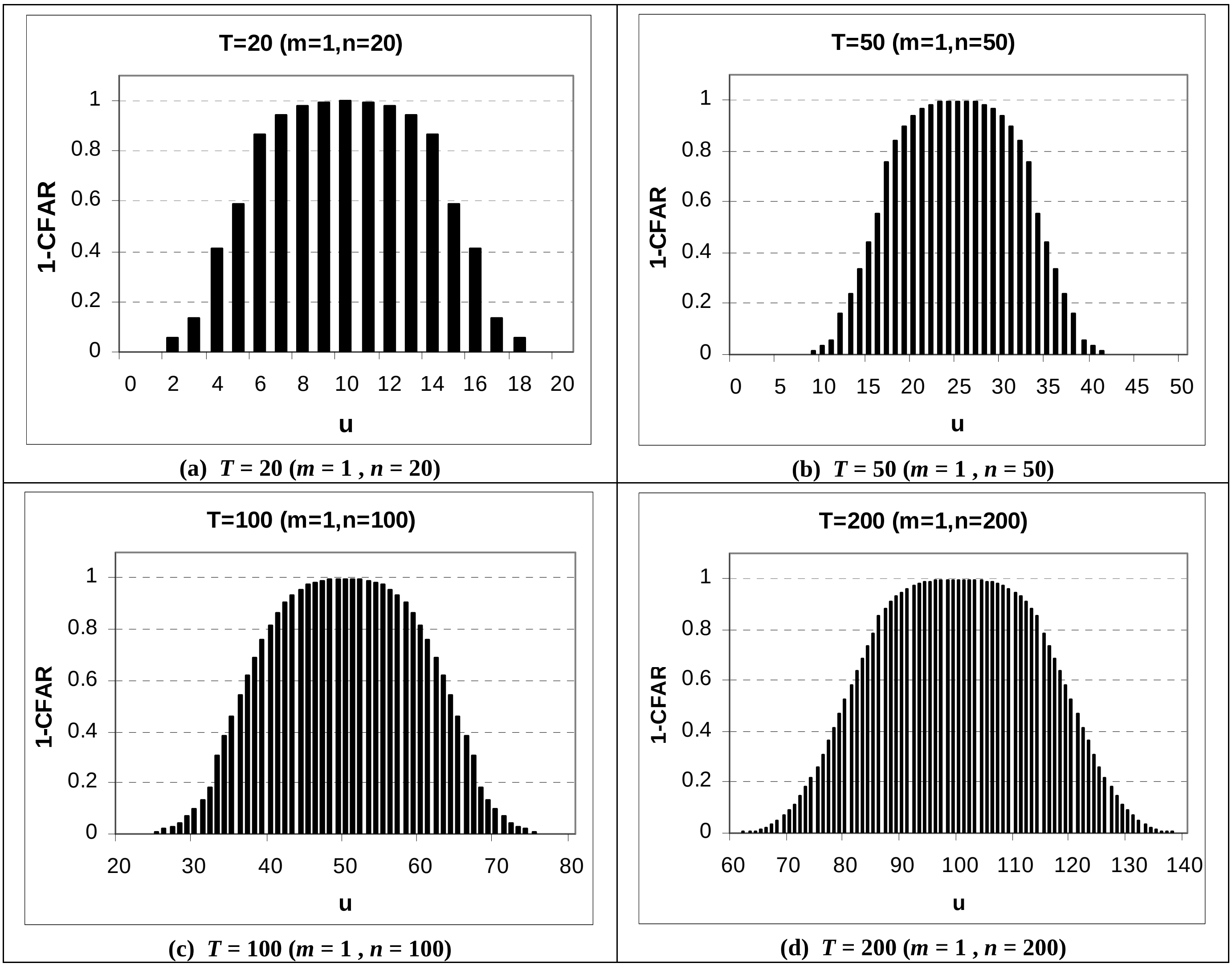


Figure 3.3: The conditional probability of a no-signal when the process is in-control (*1-CFAR*) as a function of $u = 0,1,\ldots,T$ for $m = 1$ and $n = T$ = 20 ,50, 100 and 200

The out-of-control conditional performance of the *p*-chart

The in-control performance of the Phase II *p*-chart (in theory) refers to the characteristics of the chart in the situation where the process operates at the same level in Phase II as what it did in Phase I; this is the scenario when $p_1 = p$. However, because p is unknown and estimated by $\bar{p}$, the observed value $\bar{p}_{\text{obs}}$ plays the role of p so that the conditional in-control performance (in practice) refers to the situation when $p_1 = \bar{p}_{\text{obs}}$ (see e.g. the earlier section labelled "Phase II implementation and operation"). The out-of-control performance (in practice) then refers to the characteristics of the *p*-chart when $p_1 \neq \bar{p}_{\text{obs}}$.

Taking into consideration the aforementioned, we can study the out-of-control performance of the Phase II *p*-chart by making use of the results from the previous section. In particular, by conditioning on a specific observed value $\bar{p}_{\text{obs}}$, the run-length distribution is affected in the same way it would be if the unknown true fraction nonconforming was to change from p (in Phase I) to p_1 (in Phase II). In other words, the out-of-control performance of the Phase II p-chart (i.e. when p has incurred either a downward or an upward shift to p_1 so that $p_1 \neq p$) is equivalent to the performance of the conditional *p*-chart when $p \neq \bar{p}_{\text{obs}}$ i.e. if p was either overestimated or underestimated (see e.g. Jones, Champ and Rigdon, (2004)); this correspondence allows us to examine the out-of-control performance of the *p*-chart by using the conditional statistical characteristics.

To this end, consider, for example, Table 3.7 which lists the false alarm rate (*CFAR*), the average run-length ($CARL_0$) and the standard deviation of the run-length ($CSDRL_0$) of the conditional run-length distribution for different combinations of m and n, provided that $T = mn = 20$ and $p_1 = p = 0.5$. In each case the run-length distribution is conditioned on an estimate of p through a particular realization u of the random variable U or, equivalently, on a specific realization $\bar{p}_{\text{obs}}$.

The values on which we condition are, for illustration proposes only, $U = 7$ (i.e. $\bar{p} = 7/20 = 0.35$), $U = 8$ (i.e. $\bar{p} = 8/20 = 0.40$) and $U = 10$ (i.e. $\bar{p} = 10/20 = 0.50$). These values correspond to the 10$^{\text{th}}$, the 25$^{\text{th}}$ and the 50$^{\text{th}}$ percentiles of the probability distribution of $U \sim Bin(mn = 20, p = 0.5)$, respectively; note that, because the $Bin(20{,}0.5)$ distribution is symmetric, conditioning on $U = 7$ and $U = 8$ are like conditioning on $U = 20 - 7 = 13$ (i.e. $\bar{p} = 13/20 = 0.65$) and $U = 20 - 8 = 12$ (i.e. $\bar{p} = 12/20 = 0.60$), which are the 90$^{\text{th}}$ and the 75$^{\text{th}}$ percentiles of the probability distribution of U, respectively.

In particular, by assuming that $p_1 = p = 0.5$ and then conditioning on $U = 7$ or $U = 8$ (i.e. $\bar{p} = 0.35$ or $\bar{p} = 0.40$) the performance of the Phase II *p*-chart are comparable to that of a process that has sustained a permanent step shift from 0.35 to 0.5 or encountered a lasting step shift from 0.4 to 0.5 i.e. an increase of either 43% or 25%, respectively. Similarly, if we assume that $p_1 = p = 0.5$ and then condition on $U = 13$ or $U = 12$ (i.e. $\bar{p} = 0.65$ or $\bar{p} = 0.60$) the performance of the Phase II *p*-chart is like that of a process that has sustained a permanent step shift from 0.65 to 0.5 (a decrease of 23%) or incurred a step shift from 0.6 to 0.5 (a decrease of 17%).

When $(m,n) = (1,20)$ and we condition on a value of $U = 8$ (or 12), which is the 25th (or the 75th) percentile of the distribution of $U \mid p = 0.5 \sim Bin(20,0.5)$, the *CFAR* is 0.0207 and the $CARL_0$ is 48.27. The *CFAR* is approximately $(0.0207/0.0160 - 1)\times 100\% \approx 29\%$ higher than the probability of a signal of $1-\beta(p = 0.4 \text{ or } 0.6, p_0 = 0.5, n = 20) = 0.0160$ of Case K whereas the $CARL_0$ is roughly $(48.27/62.5 - 1)\times 100\% \approx 23\%$ lower than the out-of-control (OOC) *ARL* of Case K following a sustained shift from 0.4 or 0.6 to 0.5, which is equal to $ARL(p = 0.4 \text{ or } 0.6, p_0 = 0.5, n = 20) = 62.5$ (see e.g. Tables A3.4 and A3.5 in Appendix 3A).

This means that when $m = 1$ and $n = 20$, and p is either underestimated or overestimated by 25% (i.e. the process fraction nonconforming has endured either a 25% decrease or increase and is out-of-control), the *p*-chart of Case U would be better at detecting such a shift than the *p*-chart of Case K. However, note that, this superior performance is a side-effect of estimating p.

The same is true for other combinations of (m,n). For example, if our Phase I reference data consisted of $m = 2$ samples each of size $n = 10$ and we then condition on $U = 8$ (or 12), the conditional *FAR* is $CFAR = 0.0107$ and the in-control conditional *ARL* is $CARL_0 = 93.09$. These values are approximately 73% higher and 43% lower than the probability of a signal and the out-of-control *ARL* of 0.0062 and 162.6 if p had been known.

In contrast, it is noteworthy to see what happens if we condition on $U = 10$ (i.e. the 50th percentile of the distribution of U), which implies that our estimate of p is spot on, that is, the point estimate $\bar{p} = 0.5$ on which we condition is equal to p, so that we are in actual fact dealing with the in-control (IC) performance of the *p*-chart in Case U.

In this case, the *CFAR* and the in-control conditional *ARL* for both the scenarios $(m,n) = (1,20)$ and $(m,n) = (2,10)$, are exactly equal to the in-control performance of the *p*-chart in Case K with

$FAR = 0.0026$ & $ARL_0 = 388.07$ and $FAR = 0.0020$ & $ARL_0 = 512.00$, respectively (see e.g. Tables A3.4 and A3.5 in Appendix 3A). Furthermore, note that, as mentioned before, for some combinations of (m, n), especially when $m >> n$, it happens that for certain values of U the $CFAR$ equals zero which causes the $CARL_0$ and $CSDRL_0$ to be undefined, which is undesirable.

To summarize, when $T = 20$ and p is either underestimated or overestimated (i.e. the process is OOC), the Case U *p*-chart would do better than the Case K chart at detecting a shift, and only if our estimate $\bar{p}$ of p is on target (i.e. the process is IC) would the performance of the Case U and Case K charts be similar.

Table 3.7: The false alarm rate (*CFAR*), the average run-length (*CARL*) and the standard deviation of the run-length (*CSDRL*) of the conditional run-length distribution for different combinations of m and n, provided that $T = mn = 20$ and $p_1 = p = 0.5$

T = 20		**U = 7 or 13 (OOC) ($\bar{p} = 0.35$ or 0.65)**			**U = 8 or 12 (OOC) ($\bar{p} = 0.4$ or 0.6)**			**U = 10 (IC) ($\bar{p} = 0.5$)**		
		10th or 90th Percentile			**25th or 75th Percentile**			**50th Percentile**		
m	***n***	***CFAR***	***CARL*₀**	***CSDRL*₀**	***CFAR***	***CARL*₀**	***CSDRL*₀**	***CFAR***	***CARL*₀**	***CSDRL*₀**
1	**20**	0.0577	17.34	16.84	0.0207	48.27	47.77	0.0026	388.07	387.57
2	**10**	0.0107	93.09	92.59	0.0107	93.09	92.59	0.0020	512.00	511.50
4	**5**	0.0313	32.00	31.50	0.0	∞	∞	0.0	∞	∞
5,4 20,1	**10,2**	0.0	∞	∞	0.0	∞	∞	0.0	∞	∞

Calculations similar to those in Table 3.7 are shown in Tables 3.8, 3.9, 3.10, and 3.11 for a larger range of values for T; we specifically look at $T = 10$,15, 25, 30, 50, 75, 100, 200, 250, 300, 500, 750, 1000 and 1500.

For each value of T we look at all possible combinations of m and n such that $T = mn$ where both m and n are integers. We again condition on the 10th (or the 90th), the 25th (or the 75th), and the 50th percentiles of $U \mid p = 0.5 \sim Bin(T = mn, 0.5)$ so that the interpretation of these conditional characteristics is similar to those for $T = 20$ of Table 3.7. The values of the percentiles of U and the corresponding values of $\bar{p}$ are clearly indicated.

The characteristics that are highlighted in grey indicate those (m, n) combinations for which the Case U *p*-chart performs worse than the Case K *p*-chart; for all the other (m, n) combinations the Case U *p*-chart performs better or just as well as the Case K *p*-chart.

The conditional characteristics of Tables 3.8, 3.9, 3.10, and 3.11 are of great help to the practitioner as he/she gets an idea of the ramifications when (or if) p is underestimated or overestimated for his/her particular combination of m and n values at hand (even before any data is collected); this is similar to investigating the power of a test.

Table 3.8: The false alarm rate (*CFAR*), the average run-length (*CARL*) and the standard deviation of the run-length (*CSDRL*) of the conditional run-length distribution for different combinations of m and n, provided that $T = 10$, 15, 25 and 30 and $p_1 = p = 0.5$

		10th or 90th Percentile			25th or 75th Percentile			50th Percentile		
m	***n***	***CFAR***	***$CARL_0$***	***$CSDRL_0$***	***CFAR***	***$CARL_0$***	***$CSDRL_0$***	***CFAR***	***$CARL_0$***	***$CSDRL_0$***
***T* = 10**		**U = 3 or 7 (OOC) ($\bar{p} = 0.3$ or 0.7)**			**U = 4 or 6 (OOC) ($\bar{p} = 0.4$ or 0.6)**			**U = 5 (IC) ($\bar{p} = 0.5$)**		
1	**10**	0.0547	18.29	17.78	0.0107	93.09	92.59	0.0020	512.00	511.50
2	**5**	0.0313	32.00	31.50	0.0	∞	∞	0.0	∞	∞
5,2	**10,1**	0.0	∞	∞	0.0	∞	∞	0.0	∞	∞
***T* = 15**		**U = 5 or 10 (OOC) ($\bar{p} = 0.3\dot{3}$ or $0.6\dot{6}$)**			**U = 6 or 9 (OOC) ($\bar{p} = 0.4$ or 0.6)**			**U = 7 (IC) ($\bar{p} = 0.4\dot{6}$)**		
1	**15**	0.0592	16.88	16.37	0.0176	56.79	56.29	0.0042	239.18	238.68
3	**5**	0.0	32.00	31.50	0.0	∞	∞	0.0	∞	∞
5,3	**15,1**	0.0	∞	∞	0.0	∞	∞	0.0	∞	∞
***T* = 25**		**U = 9 or 16 (OOC) ($\bar{p} = 0.36$ or 0.64)**			**U = 11 or 14 (OOC) ($\bar{p} = 0.44$ or 0.56)**			**U = 12 (IC) ($\bar{p} = 0.48$)**		
1	**25**	0.0539	18.56	18.05	0.0074	135.23	134.73	0.0025	400.98	400.48
5,5	**25,1**	0.0	∞	∞	0.0	∞	∞	0.0	∞	∞
***T* = 30**		**U = 11 or 19 (OOC) ($\bar{p} = 0.3\dot{6}$ or $0.6\dot{3}$)**			**U = 13 or 17 (OOC) ($\bar{p} = 0.4\dot{3}$ or $0.5\dot{6}$)**			**U = 15 (IC) ($\bar{p} = 0.5$)**		
1	**30**	0.1002	9.98	9.46	0.0081	123.58	123.08	0.0014	698.86	698.36
2	**15**	0.0176	56.89	56.39	0.0037	268.59	268.09	0.0010	1024.00	1023.50
3	**10**	0.0107	93.09	92.59	0.0010	1024.00	1023.50	0.0020	512.00	511.50
5	**6**	0.0	64.00	63.50	0.0	∞	∞	0.0	∞	∞
6,5 15,2	**10,3 30,1**	0.0	∞	∞	0.0	∞	∞	0.0	∞	∞

Table 3.9: The false alarm rate (*CFAR*), the average run-length (*CARL*) and the standard deviation of the run-length (*CSDRL*) of the conditional run-length distribution for different combinations of m and n, provided that $T = 50$, 75, 100, 200 and 250 and $p_1 = p = 0.5$

		10th or 90th Percentile			25th or 75th Percentile			50th Percentile		
m	***n***	***CFAR***	$CARL_0$	$CSDRL_0$	***CFAR***	$CARL_0$	$CSDRL_0$	***CFAR***	$CARL_0$	$CSDRL_0$
***T* = 50**		***U* = 20 or 30 (OOC)** ($\bar{p} = 0.4$ or 0.6)			***U* = 23 or 27 (OOC)** ($\bar{p} = 0.46$ or 0.54)			***U* = 25 (IC)** ($\bar{p} = 0.5$)		
1	**50**	0.0595	16.82	16.31	0.0078	127.77	127.27	0.0026	384.29	383.79
2	**25**	0.0217	46.18	45.68	0.0078	128.67	128.17	0.0041	245.26	244.76
5	**10**	0.0107	93.09	92.59	0.0010	1024.00	1023.50	0.0020	512.00	511.50
10,5 **50,1**	**25,2**	0.0	∞	∞	0.0	∞	∞	0.0	∞	∞
***T* = 75**		***U* = 32 or 43 (OOC)** ($\bar{p} = 0.42\dot{6}$ or $0.57\dot{3}$)			***U* = 35 or 40 (OOC)** ($\bar{p} = 0.4\dot{6}$ or $0.5\dot{3}$)			***U* = 37 (IC)** ($\bar{p} = 0.49\dot{3}$)		
1	**75**	0.0527	18.98	18.47	0.0104	96.39	95.89	0.0038	260.67	260.17
3	**25**	0.0074	135.23	134.73	0.0025	400.98	400.48	0.0025	400.98	400.48
5	**15**	0.0037	268.59	268.09	0.0042	239.18	238.68	0.0010	1024.00	1023.50
15,5 **75,1**	**25,3**	0.0	∞	∞	0.0	∞	∞	0.0	∞	∞
***T* = 100**		***U* = 44 or 56 (OOC)** ($\bar{p} = 0.44$ or 0.56)			***U* = 47 or 53 (OOC)** ($\bar{p} = 0.47$ or 0.53)			***U* = 50 (IC)** ($\bar{p} = 0.5$)		
1	**100**	0.0443	22.56	22.05	0.0107	93.51	93.01	0.0035	284.28	283.78
2	**50**	0.0165	60.74	60.23	0.0035	289.59	289.09	0.0026	384.29	383.79
4	**25**	0.0074	135.23	134.73	0.0025	400.98	400.48	0.0041	245.26	244.76
5	**20**	0.0061	163.66	163.16	0.0015	671.30	670.80	0.0026	388.07	387.57
10	**10**	0.0010	1024.00	1023.50	0.0010	1024.00	1023.50	0.0020	512.00	511.50
20,5 **50,2**	**25,4** **100,1**	0.0	∞	∞	0.0	∞	∞	0.0	∞	∞
***T* = 200**		***U* = 91 or 109 (OOC)** ($\bar{p} = 0.455$ or 0.545)			***U* = 95 or 105 (OOC)** ($\bar{p} = 0.475$ or 0.525)			***U* = 100 (IC)** ($\bar{p} = 0.5$)		
1	**200**	0.0384	26.02	25.52	0.0098	102.24	101.74	0.0023	438.70	438.20
2	**100**	0.0176	56.69	56.19	0.0062	160.75	160.25	0.0035	284.28	283.78
4	**50**	0.0078	127.77	127.27	0.0038	265.37	264.87	0.0026	384.29	383.79
5	**40**	0.0084	119.25	118.75	0.0036	281.45	280.95	0.0022	450.16	449.66
8	**25**	0.0074	135.23	134.73	0.0025	400.98	400.48	0.0041	245.26	244.76
10	**20**	0.0061	163.66	163.16	0.0015	671.30	670.80	0.0026	388.07	387.57
20	**10**	0.0010	1024.00	1023.50	0.0020	512.00	511.50	0.0020	512.00	511.50
25	**8**	0.0039	256.00	255.50	0.0	∞	∞	0.0	∞	∞
40,5 **100,2**	**50,4** **200,1**	0.0	∞	∞	0.0	∞	∞	0.0	∞	∞
***T* = 250**		***U* = 115 or 135 (OOC)** ($\bar{p} = 0.46$ or 0.54)			***U* = 120 or 130 (OOC)** ($\bar{p} = 0.48$ or 0.52)			***U* = 125 (IC)** ($\bar{p} = 0.5$)		
1	**250**	0.0438	22.85	22.35	0.0097	103.13	102.63	0.0029	347.38	346.88
2	**125**	0.0157	63.53	63.03	0.0063	159.02	158.52	0.0022	449.14	448.64
5	**50**	0.0078	127.77	127.27	0.0038	265.37	264.87	0.0026	384.29	383.79
10	**25**	0.0078	128.67	128.17	0.0025	400.98	400.48	0.0041	245.26	244.76
25	**10**	0.0010	1024.00	1023.50	0.0020	512.00	511.50	0.0020	512.00	511.50
50,5 **250,1**	**125,2**	0.0	∞	∞	0.0	∞	∞	0.0	∞	∞

Table 3.10: The false alarm rate (*CFAR*), the average run-length (*CARL*) and the standard deviation of the run-length (*CSDRL*) of the conditional run-length distribution for different combinations of m and n, provided that $T = 300$, 500 and 750 and $p_1 = p = 0.5$

		10th or 90th Percentile			25th or 75th Percentile			50th Percentile		
m	***n***	***CFAR***	$CARL_0$	$CSDRL_0$	***CFAR***	$CARL_0$	$CSDRL_0$	***CFAR***	$CARL_0$	$CSDRL_0$
T = 300		**U = 139 or 161 (OOC) ($\bar{p}$ = 0.463 or 0.536)**			**U = 144 or 156 (OOC) ($\bar{p}$ = 0.48 or 0.52)**			**U = 150 (IC) ($\bar{p}$ = 0.5)**		
1	**300**	0.0470	21.29	20.79	0.0122	81.82	81.32	0.0032	315.53	315.03
2	**150**	0.0205	48.81	48.30	0.0058	173.44	172.94	0.0024	415.71	415.21
3	**100**	0.0106	94.51	94.01	0.0065	154.96	154.46	0.0035	284.28	283.78
4	**75**	0.0102	97.68	97.18	0.0058	171.50	171.00	0.0024	409.13	408.63
5	**60**	0.0069	144.06	143.56	0.0036	274.60	274.10	0.0027	374.47	373.97
6	**50**	0.0078	127.77	127.27	0.0038	265.37	264.87	0.0026	384.29	383.79
10	**30**	0.0028	360.50	360.00	0.0033	300.58	300.08	0.0014	698.86	698.36
12	**25**	0.0025	400.98	400.48	0.0025	400.98	400.48	0.0041	245.26	244.76
15	**20**	0.0061	163.66	163.16	0.0015	671.30	670.80	0.0026	388.07	387.57
20	**15**	0.0042	239.18	238.68	0.0010	1024.00	1023.50	0.0010	1024.00	1023.50
25	**12**	0.0034	292.57	292.07	0.0034	292.57	292.07	0.0005	2048.00	2047.50
30	**10**	0.0010	1024.00	1023.50	0.0020	512.00	511.50	0.0020	512.00	511.50
50,6 75,4 150,2	**60,5 100,3 300,1**	0.0	∞	∞	0.0	∞	∞	0.0	∞	∞
T = 500		**U = 236 or 264 (OOC) ($\bar{p}$ = 0.472 or 0.528)**			**U = 242 or 258 (OOC) ($\bar{p}$ = 0.484 or 0.516)**			**U = 250 (IC) ($\bar{p}$ = 0.5)**		
1	**500**	0.0405	24.68	24.17	0.0113	88.24	87.74	0.0027	370.81	370.31
2	**250**	0.0184	54.39	53.88	0.0070	143.25	142.75	0.0029	347.38	346.88
4	**125**	0.0100	100.06	99.56	0.0038	260.91	260.41	0.0022	449.14	448.64
5	**100**	0.0062	160.75	160.25	0.0038	266.28	265.78	0.0035	284.28	283.78
10	**50**	0.0038	265.37	264.87	0.0038	265.37	264.87	0.0026	384.29	383.79
20	**25**	0.0025	400.98	400.48	0.0025	400.98	400.48	0.0041	245.26	244.76
25	**20**	0.0015	671.30	670.80	0.0015	671.30	670.80	0.0026	388.07	387.57
50	**10**	0.0010	1024.00	1023.50	0.0020	512.00	511.50	0.0020	512.00	511.50
100,5 250,2	**125,4 500,1**	0.0	∞	∞	0.0	∞	∞	0.0	∞	∞
T = 750		**U = 357 or 393 (OOC) ($\bar{p}$ = 0.476 or 0.524)**			**U = 366 or 384 (OOC) ($\bar{p}$ = 0.488 or 0.512)**			**U = 375 (IC) ($\bar{p}$ = 0.5)**		
1	**750**	0.0430	23.24	22.73	0.0089	112.45	111.95	0.0024	413.68	413.18
2	**375**	0.0194	51.54	51.03	0.0051	196.82	196.32	0.0027	370.96	370.46
3	**250**	0.0134	74.53	74.02	0.0051	197.15	196.65	0.0029	347.38	346.88
5	**150**	0.0090	111.07	110.56	0.0038	262.77	262.27	0.0024	415.71	415.21
6	**125**	0.0061	163.01	162.51	0.0041	242.72	242.22	0.0022	449.14	448.64
10	**75**	0.0055	181.29	180.79	0.0032	317.07	316.57	0.0024	409.13	408.63
15	**50**	0.0038	265.37	264.87	0.0018	565.23	564.73	0.0026	384.29	383.79
25	**30**	0.0033	300.58	300.08	0.0033	300.58	300.08	0.0014	698.86	698.36
30	**25**	0.0025	400.98	400.48	0.0025	400.98	400.48	0.0041	245.26	244.76
50	**15**	0.0042	239.18	238.68	0.0010	1024.00	1023.50	0.0010	1024.00	1023.50
75	**10**	0.0020	512.00	511.50	0.0020	512.00	511.50	0.0020	512.00	511.50
125,6 250,3 750,1	**150,5 375,2**	0.0	∞	∞	0.0	∞	∞	0.0	∞	∞

Table 3.11: The false alarm rate (*FAR*), the average run-length (*ARL*) and the standard deviation of the run-length (*SDRL*) of the conditional run-length distribution for different combinations of m and n, provided that $T = 1000$ and 1500 and $p_1 = p = 0.5$

		10^{th} or 90^{th} Percentile			25^{th} or 75^{th} Percentile			50^{th} Percentile		
m	***n***	***CFAR***	***CARL***	***CSDRL***	***CFAR***	***CARL***	***CSDRL***	***CFAR***	***CARL***	***CSDRL***
***T* = 1000**		***U* = 480 or 520 (OOC)** ($\bar{p} = 0.48$ or 0.52)			***U* = 489 or 511 (OOC)** ($\bar{p} = 0.489$ or 0.511)			***U* = 500 (IC)** ($\bar{p} = 0.5$)		
1	**1000**	0.0410	24.40	23.90	0.0106	94.61	94.11	0.0026	378.00	377.50
2	**500**	0.0178	56.25	55.75	0.0056	179.73	179.23	0.0027	370.81	370.31
4	**250**	0.0097	103.13	102.63	0.0051	197.15	196.65	0.0029	347.38	346.88
5	**200**	0.0067	149.04	148.53	0.0033	306.27	305.77	0.0023	438.70	438.20
8	**125**	0.0063	159.02	158.52	0.0041	242.72	242.22	0.0022	449.14	448.64
10	**100**	0.0065	154.96	154.46	0.0038	266.28	265.78	0.0035	284.28	283.78
20	**50**	0.0038	265.37	264.87	0.0018	565.23	564.73	0.0026	384.29	383.79
25	**40**	0.0036	281.45	280.95	0.0022	450.16	449.66	0.0022	450.16	449.66
40	**25**	0.0025	400.98	400.48	0.0025	400.98	400.48	0.0041	245.26	244.76
50	**20**	0.0015	671.30	670.80	0.0026	388.07	387.57	0.0026	388.07	387.57
100	**10**	0.0020	512.00	511.50	0.0020	512.00	511.50	0.0020	512.00	511.50
125,8 **250,4** **1000,1**	**200,5** **500,2**	0.0	∞	∞	0.0	∞	∞	0.0	∞	∞
***T* = 1500**		***U* = 725 or 775 (OOC)** ($\bar{p} = 0.48\dot{3}$ or $0.51\dot{6}$)			***U* = 737 or 763 (OOC)** ($\bar{p} = 0.491\dot{3}$ or $0.508\dot{6}$)			***U* = 750 (IC)** ($\bar{p} = 0.5$)		
1	**1500**	0.0418	23.92	23.41	0.0095	105.36	104.85	0.0025	398.62	398.12
2	**750**	0.0187	53.44	52.94	0.0061	163.72	163.22	0.0024	413.68	413.18
3	**500**	0.0113	88.24	87.74	0.0045	221.21	220.71	0.0027	370.81	370.31
4	**375**	0.0089	112.99	112.48	0.0040	247.16	246.66	0.0027	370.96	370.46
5	**300**	0.0091	109.95	109.45	0.0038	265.71	265.21	0.0032	315.53	315.03
6	**250**	0.0070	143.25	142.75	0.0038	261.95	261.45	0.0029	347.38	346.88
10	**150**	0.0059	168.33	167.83	0.0027	364.75	364.25	0.0024	415.71	415.21
12	**125**	0.0038	260.91	260.41	0.0026	383.35	382.84	0.0022	449.14	448.64
15	**100**	0.0038	266.28	265.78	0.0027	376.82	376.32	0.0035	284.28	283.78
20	**75**	0.0032	317.07	316.57	0.0032	317.07	316.57	0.0024	409.13	408.63
25	**60**	0.0036	274.60	274.10	0.0019	535.30	534.80	0.0027	374.47	373.97
30	**50**	0.0038	265.37	264.87	0.0018	565.23	564.73	0.0026	384.29	383.79
50	**30**	0.0033	300.58	300.08	0.0033	300.58	300.08	0.0014	698.86	698.36
60	**25**	0.0025	400.98	400.48	0.0025	400.98	400.48	0.0041	245.26	244.76
75	**20**	0.0015	671.30	670.80	0.0026	388.07	387.57	0.0026	388.07	387.57
100	**15**	0.0010	1024.00	1023.50	0.0010	1024.00	1023.50	0.0010	1024.00	1023.50
125	**12**	0.0034	292.57	292.07	0.0005	2048.00	2047.50	0.0005	2048.00	2047.50
150	**10**	0.0020	512.00	511.50	0.0020	512.00	511.50	0.0020	512.00	511.50
250,6 **375,4** **750,2**	**300,5** **500,3** **1500,1**	0.0	∞	∞	0.0	∞	∞	0.0	∞	∞

3.2.3.2 Conditional characteristics of the *c*-chart

Like the *p*-chart, once we observe a value v of the random variable V we can calculate the conditional probability of a no-signal of the *c*-chart so that the Phase II conditional run-length distribution and its associated conditional characteristics are completely known (see e.g. Table 3.4). To this end, Table 3.12 illustrates the steps to calculate the conditional probability of a no-signal, the conditional false alarm rate (*CFAR*), the conditional average run-length (*CARL*) and the conditional standard deviation of the run-length (*CSDRL*) of the *c*-chart.

For illustration purposes we assume that $c_1 = c = 20$; this implies that the process operated at a level of twenty nonconformities (on average) in an inspection unit during Phase I and that in Phase II the process continues to operate at this same level. In addition, we assume that $m = 100$ Phase I inspection units are available to estimate c using $\bar{c} = V/m = \bar{c}_{\text{obs}}$, which (because of sampling variation) may or may not be equal to c.

In particular, column 1 lists some values of $V = 0(200)6000$, which (in theory) can be any integer greater than or equal to zero. Column 2 converts the observed value v of V of column 1 into a point estimate of c by calculating $\bar{c}_{\text{obs}} = v/100$. Because each row entry in each of the succeeding columns (i.e. columns 3 to 10) is computed by conditioning on a row entry from column 1 or column 2, we start calculating the conditional properties in column 1 or 2 and sequentially proceed to the right-hand side of the table. So, given a value v or $\bar{c}_{\text{obs}}$ the lower and the upper control limits are estimated in columns 3 and 4 using (3-27) and then used to compute the two constants $\hat{d}$ and $\hat{f}$ defined in (3-33), which are shown in columns 5 and 6, respectively. Finally, columns 7 through 10 list the probability of a no-signal, the *FAR*, the *ARL* and the *SDRL* conditioned on the observed value v from column 1, respectively. These properties are labeled $\Pr(\text{No Signal} \mid V, c)$, *CFAR*, $CARL_0$ and $CSDRL_0$, and calculated using (3-32) and the expressions in Table 3.4.

An examination of Table 3.12 reveals one special scenario i.e. when $V = 0$ (the minimum possible value). In this particular case the estimated control limits are $L\hat{C}L_c = U\hat{C}L_c = 0$ so that the constants $\hat{d}$ and $\hat{f}$ need not be calculated (see e.g. expression (3-32) and Remark 6(ii)); as a result, the probability of a no-signal is defined to be zero so that the *c*-chart signals with probability one once the first Phase II inspection unit is sampled i.e. both the conditional *FAR* and the conditional *ARL* are one (as shown in columns 8 and 9, respectively). For values of $V \neq 0$ we proceed as follows.

Table 3.12: Conditional Probability of a no-signal, the conditional false alarm rate (*CFAR*), the conditional average run-length (*CARL*) and the conditional standard deviation of the run-length (*CSDRL*) of the *c*-chart in Case U for $m = 100$ and assuming that $c_1 = c = 20$

(1) v	(2) $\bar{c}_{obs}$	(3) $L\hat{C}L_c$	(4) $U\hat{C}L_c$	(5) $\hat{d}$	(6) $\hat{f}$	(7) Pr(No Signal | *V*, c)	(8) *CFAR*	(9) $CARL_0$	(10) $CSDRL_0$
0	0	0.00	0.00	NA	NA	0.0000	1.0000	1.00	0.00
200	2	-2.24	6.24	0	6	0.0003	0.9997	1.00	0.02
400	4	-2.00	10.00	0	9	0.0050	0.9950	1.01	0.07
600	6	-1.35	13.35	0	13	0.0661	0.9339	1.07	0.28
800	8	-0.49	16.49	0	16	0.2211	0.7789	1.28	0.60
1000	10	0.51	19.49	0	19	0.4703	0.5297	1.89	1.29
1200	12	1.61	22.39	1	22	0.7206	0.2794	3.58	3.04
1400	14	2.78	25.22	2	25	0.8878	0.1122	8.91	8.40
1600	16	4.00	28.00	4	27	0.9475	0.0525	19.05	18.54
1800	18	5.27	30.73	5	30	0.9865	0.0135	73.82	73.32
2000	20	6.58	33.42	6	33	0.9971	0.0029	339.72	339.22
2200	22	7.93	36.07	7	36	0.9988	0.0012	832.30	831.80
2400	24	9.30	38.70	9	38	0.9949	0.0051	195.92	195.42
2600	26	10.70	41.30	10	41	0.9892	0.0108	92.39	91.89
2800	28	12.13	43.87	12	43	0.9610	0.0390	25.63	25.13
3000	30	13.57	46.43	13	46	0.9339	0.0661	15.12	14.61
3200	32	15.03	48.97	15	48	0.8435	0.1565	6.39	5.87
3400	34	16.51	51.49	16	51	0.7789	0.2211	4.52	3.99
3600	36	18.00	54.00	18	53	0.6186	0.3814	2.62	2.06
3800	38	19.51	56.49	19	56	0.5297	0.4703	2.13	1.55
4000	40	21.03	58.97	21	58	0.3563	0.6437	1.55	0.93
4200	42	22.56	61.44	22	61	0.2794	0.7206	1.39	0.73
4400	44	24.10	63.90	24	63	0.1568	0.8432	1.19	0.47
4600	46	25.65	66.35	25	66	0.1122	0.8878	1.13	0.38
4800	48	27.22	68.78	27	68	0.0525	0.9475	1.06	0.24
5000	50	28.79	71.21	28	71	0.0343	0.9657	1.04	0.19
5200	52	30.37	73.63	30	73	0.0135	0.9865	1.01	0.12
5400	54	31.95	76.05	31	76	0.0081	0.9919	1.01	0.09
5600	56	33.55	78.45	33	78	0.0027	0.9973	1.00	0.05
5800	58	35.15	80.85	35	80	0.0008	0.9992	1.00	0.03
6000	60	36.76	83.24	36	83	0.0004	0.9996	1.00	0.02
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Suppose, for example, that we observe two thousand four hundred nonconformities in the entire Phase I reference sample. The value of $V = 2400$ gives an observed value of the point estimate for c of $\bar{c}_{obs} = 2400/100 = 24$ so that (3-27) yields an estimated upper control limit and an estimated lower control of

$$U\hat{C}L_c = 24 + 3\sqrt{24} = 38.70 \quad \text{and} \quad L\hat{C}L_c = 24 - 3\sqrt{24} = 9.30$$

respectively.

The constants $\hat{d}$ and $\hat{f}$ are thus found to be

$$\hat{d} = \max\{0,[L\hat{C}L_c]\} = \max\{0,[9.3]\} = 9 \qquad \text{and} \qquad \hat{f} = [U\hat{C}L_c - 1] = [38.70] = 38$$

so that upon substituting $\hat{d}$ and $\hat{f}$ in (3-32) we calculate the conditional probability of a no-signal and then also the *CFAR*, the $CARL_0$ and the $CSDRL_0$ using expressions (3-42), (3-43) and (3-44) in Table 3.4.

The conditional probability of a no-signal, in particular, is

$$\hat{\beta}(c_1 = 20, m = 100 \mid V = 2400, c = 20) = \hat{\beta}(c_1 = 20, m = 100 \mid \bar{c} = 24, c = 20) = 0.9949$$

so that the conditional false alarm rate is

$$CFAR(c_1 = 20,100 \mid V = 2400, c = 20) = 1 - 0.9949 = 0.0051.$$

The Phase II c-chart then has a conditional in-control *ARL* of

$$CARL_0(c_1 = 20,100 \mid V = 2400, c = 20) = 1/0.0051 = 195.92$$

and a conditional in-control *SDRL* of

$$SDRL_0(c_1 = 20,100 \mid V = 2400, c = 20) = \sqrt{0.9949}/0.0051 = 195.42.$$

The conditional false alarm rate

Figure 3.4 displays the conditional false alarm rate (*CFAR*), that is, $1-\hat{\beta}(c_1 = 20, m \mid V = v, c = 20)$ as a function of $v = 0,1,2,...$ when m = 50 or 75 or 100 individual Phase I inspection units are used to estimate c; the curve labeled $m = 100$ corresponds to the $CFAR$'s of column 8 in Table 3.12.

The impact of the actual observed number of nonconformities v in the entire Phase I reference sample is easily noticed. The distribution function of the *CFAR* is seen to be slightly negatively U-shaped. For values of V near the two tails (i.e. the extreme left and right) the *CFAR* can be very high, sometimes close to 1 or 100%, which obviously means many false alarms.

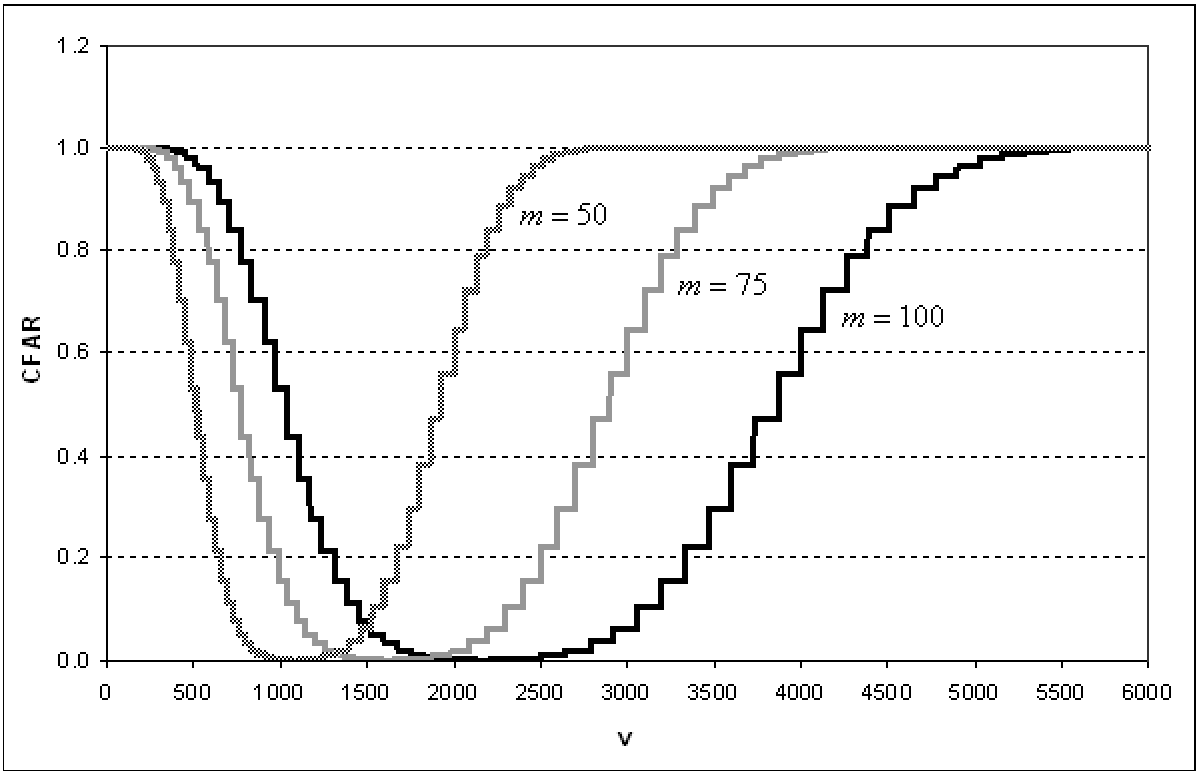


Figure 3.4: The conditional false alarm rate (*CFAR*) as a function of $v = 0,1,2,...$ for m = 50,75 and 100

However, even when V is not near the two tails there can be a significantly high probability of a false alarm; this is more easily seen from Figure 3.5, which (for illustration purposes) displays values of $1-\hat{\beta}(c_1 = 20, m = 100 \mid V = v, c = 20)$ for values of v between 1800 and 2600 only.

It is seen that only when V takes on a value in the neighbourhood of its mean i.e. $E(V \mid c = 20) = mc = 100 \times 20 = 2000$ (or, equivalently, when $\bar{c}$ is close to the true average number of nonconformities, which is 20 in this case) will the *CFAR* be reasonably small and close to its Case K value of 0.0029 (see e.g. Table A3.12 in Appendix 3A).

However even though the *CFAR* may be small, it is (for most values of v) still far from the typical or nominal expected value of 0.0027 of a Shewhart X-bar chart with 3-sigma limits. Thus, the performance of the c-chart, as measured by the false alarm rate, is considerably degraded and unfavourably affected by a poor point estimate $\bar{c}$.

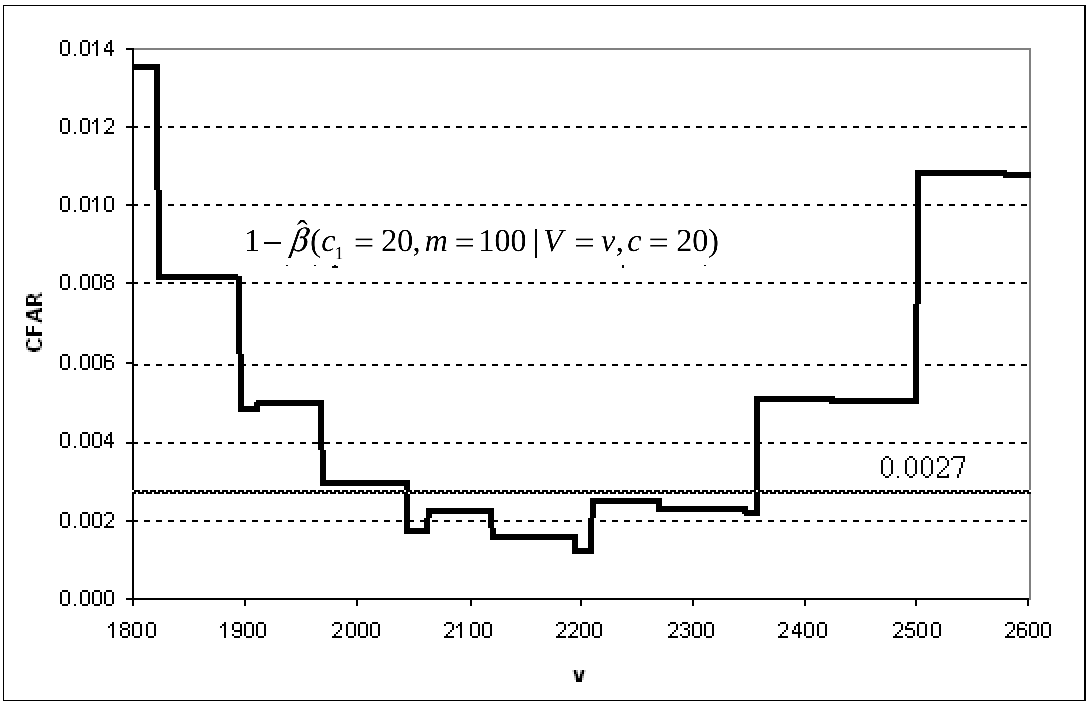


Figure 3.5: The conditional false alarm rate (*CFAR*) as a function of $v = 1800,...,2600$ when m = 100 in relation to the nominal *FAR* of 0.0027

The out-of-control conditional performance of the *c*-chart

The in-control performance of the Phase II *c*-chart (in theory) refers to the characteristics of the chart in the situation where the process operates at the same level in Phase II than what it did in Phase I; this is the scenario when $c_1 = c$. But, because c is unknown and estimated by $\bar{c}$, the observed value $\bar{c}_{\text{obs}}$ plays the role of c so that the conditional in-control performance (in practice) refers to the situation when $c_1 = \bar{c}_{\text{obs}}$ (see e.g. the earlier section labelled "Phase II implementation and operation"). The out-of-control performance (in practice) then refers to the characteristics of the *c*-chart when $c_1 \neq \bar{c}_{\text{obs}}$.

In view of the abovementioned, we can study the out-of-control performance of the Phase II *c*-chart by making use of the results from the previous section concerning the conditional characteristics of the Phase II *c*-chart. In particular, note that, by conditioning on a specific observed value $\bar{c}_{\text{obs}}$ the run-length distribution of the Phase II *c*-chart is affected in the same way it would be if the unknown true average number of nonconformities in an inspection was to change from c (in Phase I) to c_1 (in Phase II). In other words, the out-of-control performance of the Phase II *c*-chart (i.e. when c has incurred either a downward or an upward shift to c_1 so that $c_1 \neq c$) is equivalent to the performance of the *c*-chart when $c \neq \bar{c}_{\text{obs}}$ i.e. if c was either overestimated or underestimated (see e.g. Jones, Champ and Rigdon, (2004)); this correspondence allows us to examine the out-of-control performance of the *c*-chart by using the conditional statistical characteristics of the Phase II *c*-chart.

Consider, for example, Tables 3.13, 3.14 and 3.15 which list the false alarm rate (*CFAR*), the average run-length (*CARL*) and the standard deviation of the run-length (*CSDRL*) of the conditional run-length distribution assuming that $c = 5, 10, 15, 20$ and 30 with $m = 10, 15, 20, 25, 50, 75, 100, 150$ and 200. For each combination of (m, c)-values the run-length distribution is conditioned on (for illustration purposes only) the 10$^{\text{th}}$, 25$^{\text{th}}$, 50$^{\text{th}}$, 75$^{\text{th}}$ and 90$^{\text{th}}$ percentiles of the distribution of $V \mid c \sim Poi(mc)$.

In particular, suppose that $(m, c) = (20, 20)$ and we observed $V = 400$ so that our estimate $\bar{c} = 20$ is spot on. In this case, Table 3.14 shows that the conditional false alarm rate is

$$CFAR(c_1 = 20, m = 20 \mid V = 400, c = 20) = 0.0029,$$

the conditional average run-length is

$$CARL_0(c_1 = 20, m = 20 \mid V = 400, c = 20) = 1/0.0029 = 339.72$$

and the conditional standard deviation of the run-length is

$$CSDRL_0(c_1 = 20, m = 20 \mid V = 400, c = 20) = \sqrt{1-0.0029}/0.0029 = 339.22.$$

These conditional characteristics i.e. conditioned on the 50th percentile of $V \mid c = 20 \sim Poi(400)$, are identical to the in-control characteristics of the Case K *c*-chart, that is, $1-\beta(c = 20, c_0 = 20) = 0.0029$, $ARL(c = 20, c_0 = 20) = 339.72$ and $SDRL(c = 20, c_0 = 20) = 339.22$, which can be found from Table A3.12 in Appendix 3A. To illustrate the out-of-control (OOC) performance of the Case U *c*-chart we should condition on a percentile of $V \mid c \sim Poi(mc)$ other than the 50th percentile. To this end, consider again, for example, the situation when $(m,c) = (20,20)$, but conditioning on the 25th percentile of $V \mid c = 20 \sim Poi(mc = 400)$, that is, $V = 386$ or $\bar{c} = 386/20 = 19.30$; this implies that c is underestimated by approximately $(20/19.3-1)100\% \approx 4\%$ or, equivalently, that the average number of nonconformities in an inspection unit has increased by 4%.

Table 3.14 shows that the $CFAR = 0.0050$, the $CARL_0 = 200.70$ and the $CSDRL_0 = 200.20$. Compared to the probability of a signal of $1-\beta(c = 19.30, c_0 = 20) = 0.0020$, the OOC average run-length of $ARL(c = 19.30, c_0 = 20) = 507.85$ and the OOC standard deviation of the run-length of $SDRL(c = 19.30, c_0 = 20) = 507.35$ of Case K (which are found by evaluating expressions (3-7), (3-21) and (3-22), respectively) we observe that the Case U *c*-chart would detect an increase from 19.30 to 20 quicker than the *c*-chart of Case K. However, this is (as mentioned earlier in case of the *p*-chart) a side-effect of estimating c and not due to improved performance.

On the other hand, when $(m,c) = (20,20)$, and we condition on the 90th percentile of $V \mid c = 20 \sim Poi(mc = 400)$, that is, $V = 426$ or $\bar{c} = 426/20 = 21.30$, which implies that c is overestimated by $(21.3/20-1)100\% = 6.5\%$ (or, equivalently, that the average number of nonconformities in an inspection unit has decreased by 6.5%), Table 3.14 shows that the $CFAR = 0.0016$, the $CARL_0 = 632.01$ and the $CSDRL_0 = 631.51$ which implies that the Case U *c*-chart performs worse than the Case K *c*-chart with probability of a signal of $1-\beta(c = 21.3, c_0 = 20) = 0.0068$, an out-of-control ARL of $ARL(c = 21.30, c_0 = 20) = 146.15$ and an out-of-control $SDRL$ of $SDRL(c = 21.30, c_0 = 20) = 145.65$.

Note that, when conditioning on a particular percentile of V, the OOC performance of the Case U *c*-chart is the same for two or more (m,c) combinations and thus the overlap of certain of the cells as seen in Tables 3.13, 3.14 and 3.15. For example, the OOC performance of the Case U c-chart when (i) $(m,c)=(20,20)$ and conditioning on the 90^{th} percentile of $V\,|\,c=20 \sim Poi(mc=400)$, and (ii) $(m,c)=(15,20)$ and we condition on the 90^{th} percentile of $V\,|\,c=20 \sim Poi(15\times 20=300)$ i.e. $V=322$ so that $\bar{c}=21.47$ (which corresponds to an decrease of 7.35% in c from 21.47 to 20), are similar.

Table 3.13: The false alarm rate (*FAR*), the average run-length (*ARL*) and the standard deviation of the run-length (*SDRL*) of the conditional run-length distribution of the *c*-chart for m = 10, 15, 20, 25, 50, 75, 100, 150, 200 when $c=5$ and 10

<table>
<tr><th></th><th colspan="9">c = 5</th><th colspan="9">c = 10</th></tr>
<tr><th>Percentile</th><th>m = 10</th><th>15</th><th>20</th><th>25</th><th>50</th><th>75</th><th>100</th><th>150</th><th>200</th><th>10</th><th>15</th><th>20</th><th>25</th><th>50</th><th>75</th><th>100</th><th>150</th><th>200</th></tr>
<tr><td>10th
(OOC)</td><td colspan="4">0.0204
48.94
48.44</td><td colspan="5">0.0122
82.03
81.53</td><td colspan="2">0.0143
69.82
69.32</td><td colspan="6">0.0072
138.28
137.78</td><td>0.0035
285.74
285.23</td></tr>
<tr><td>25th
(OOC)</td><td>0.0204
48.94
48.44</td><td colspan="8">0.0122
82.03
81.53</td><td colspan="4">0.0072
138.28
137.78</td><td colspan="5">0.0035
285.74
285.23</td></tr>
<tr><td>50th
(IC)</td><td colspan="9">CFAR = 0.0122
CARL = 82.03
CSDRL = 81.53</td><td colspan="9">CFAR = 0.0035
CARL = 285.74
CSDRL = 285.23</td></tr>
<tr><td>75th
(OOC)</td><td colspan="5">0.0088
114.20
113.70</td><td colspan="4">0.0122
82.03
81.53</td><td colspan="4">0.0016
612.12
611.62</td><td colspan="5">0.0035
285.74
285.23</td></tr>
<tr><td>90th
(OOC)</td><td>0.0074
134.48
133.98</td><td colspan="8">0.0088
114.20
113.70</td><td colspan="2">0.0012
833.99
833.49</td><td colspan="5">0.0016
612.12
611.62</td><td colspan="2">0.0035
285.74
285.23</td></tr>
</table>

Table 3.14: The false alarm rate (*FAR*), the average run-length (*ARL*) and the standard deviation of the run-length (*SDRL*) of the conditional run-length distribution of the *c*-chart for m = 10, 15, 20, 25, 50, 75, 100, 150, 200 when $c=15$ and 20

<table>
<tr><th></th><th colspan="9">c = 15</th><th colspan="9">c = 20</th></tr>
<tr><th>Percentile</th><th>m = 10</th><th>15</th><th>20</th><th>25</th><th>50</th><th>75</th><th>100</th><th>150</th><th>200</th><th>10</th><th>15</th><th>20</th><th>25</th><th>50</th><th>75</th><th>100</th><th>150</th><th>200</th></tr>
<tr><td>10th
(OOC)</td><td colspan="2">0.0112
89.25
88.75</td><td colspan="3">0.0062
160.66
160.16</td><td colspan="2">0.0064
156.34
155.84</td><td colspan="2">0.0035
283.83
283.33</td><td>0.0135
73.82
73.32</td><td colspan="3">0.0082
122.49
121.99</td><td colspan="5">0.0050
200.70
200.20</td></tr>
<tr><td>25th
(OOC)</td><td colspan="2">0.0062
160.66
160.16</td><td colspan="2">0.0064
156.34
155.84</td><td colspan="5">0.0035
283.83
283.33</td><td>0.0048
208.36
207.86</td><td colspan="5">0.0050
200.70
200.20</td><td colspan="3">0.0029
339.72
339.22</td></tr>
<tr><td>50th
(IC)</td><td colspan="9">CFAR = 0.0035
CARL = 283.83
CSDRL = 283.33</td><td colspan="9">CFAR = 0.0029
CARL = 339.72
CSDRL = 339.22</td></tr>
<tr><td>75th
(OOC)</td><td colspan="6">0.0019
518.90
518.40</td><td colspan="3">0.0035
283.83
283.33</td><td colspan="3">0.0023
440.99
440.49</td><td>0.0017
573.34
572.84</td><td colspan="5">0.0029
339.72
339.22</td></tr>
<tr><td>90th
(OOC)</td><td colspan="3">0.0017
582.29
581.79</td><td>0.0026
388.74
388.24</td><td colspan="5">0.0019
518.90
518.40</td><td colspan="3">0.0016
632.01
631.51</td><td colspan="3">0.0023
440.99
440.49</td><td colspan="2">0.0017
573.34
572.84</td><td>0.0029
339.72
339.22</td></tr>
</table>

Table 3.15: The false alarm rate (*FAR*), the average run-length (*ARL*) and the standard deviation of the run-length (*SDRL*) of the conditional run-length distribution of the *c*-chart for *m* = 10, 15, 20, 25, 50, 75, 100, 150, 200 when *c* = 30

<table>
<tr><th></th><th colspan="9">c = 30</th></tr>
<tr><th>Percentile</th><th>m = 10</th><th>15</th><th>20</th><th>25</th><th>50</th><th>75</th><th>100</th><th>150</th><th>200</th></tr>
<tr><td>10th
(OOC)</td><td>0.0098
102.05
101.55</td><td colspan="3">0.0064
155.37
154.87</td><td colspan="2">0.0041
242.41
241.91</td><td colspan="3">0.0044
229.10
228.60</td></tr>
<tr><td>25th
(OOC)</td><td>0.0064
155.37
154.87</td><td colspan="2">0.0041
242.41
241.91</td><td colspan="4">0.0044
229.10
228.60</td><td colspan="2">0.0029
349.94
349.44</td></tr>
<tr><td>50th
(IC)</td><td colspan="9">CFAR = 0.0029
CARL = 349.94
CSDRL = 349.44</td></tr>
<tr><td>75th
(OOC)</td><td colspan="4">0.0024
415.11
414.61</td><td>0.0019
527.54
527.04</td><td colspan="4">0.0029
349.94
349.44</td></tr>
<tr><td>90th
(OOC)</td><td>0.0025
405.45
404.95</td><td colspan="3">0.0018
553.19
552.69</td><td colspan="3">0.0024
415.11
414.61</td><td colspan="2">0.0019
527.54
527.04</td></tr>
</table>

3.2.4 Unconditional Phase II run-length distributions and characteristics

The conditional run-length distribution and the associated characteristics of the conditional run-length distribution present the performance of a chart only for one particular realization of the point estimator and a supposed value for the parameter. For each individual realization of $\overline{p} = U / mn$ or $\overline{c} = V / m$ and the true p or c value the performance of the chart will be different – some charts performing acceptable and others poorly.

In case of the *p*-chart the variable U can take on any value between and including 0 and mn i.e. $U \in \{0,1,\ldots,mn\}$, so that there is a finite number $mn+1$ possible values on which we can condition. For the *c*-chart the variable V can be any positive integer greater or equal to zero i.e. $V \in \{0,1,2,\ldots\}$, and so the number of possible values on which we can condition is infinite.

To avoid calculating the conditional performance of the charts for each realization of the point estimator and to asses the overall performance of the charts, the influence of a single realization should ideally be removed. The unconditional run-length distribution and its associated characteristics serve this purpose and better represent the overall performance of the charts when the parameters are estimated and let one see the bigger picture.

The unconditional characteristics of the charts can be found from the conditional run-length distribution by averaging over the distributions of U and V respectively, and allow us to look at the marginal (or the unconditional) run-length distribution. This incorporates the additional variation introduced to the run-length through the estimation of p and c by taking into account all possible realizations of the random variables on which we condition. In particular, we derive expressions for the:

(i) unconditional run-length distribution,

(ii) unconditional average run-length, and

(iii) unconditional variance of the run-length

of each chart.

Unconditional run-length distribution: *p*-chart and *c*-chart (Case U)

Because:

(i) the observations in the Phase I reference sample are assumed to be independent and identically distributed, that is, $X_i \sim iidBin(n,p)$ and $Y_i \sim iidPoi(c)$ for $i = 1,2,...,m$, and

(ii) we assume that the Phase I X_i's and Y_i's are independent from the Phase II observations i.e. $X_i \sim iidBin(n,p_1)$ and $Y_i \sim iidPoi(c_1)$ for $i = m+1, m+2,...$,

the joint probability distribution of

(i) the Phase I point estimator $U = mn\bar{p}$ and the Phase II run-length random variable N_p, and

(ii) the Phase I point estimator $V = m\bar{c}$ and the Phase II run-length random variable N_c

can straightforwardly be obtained (see e.g. Definition 4.2.1 in Casella and Berger, (2002) p. 148) as

$$\Pr(N_p = j, U = u; p_1, m, n \mid p) = \Pr(N_p = j; p_1, m, n \mid U = u, p).\Pr(U = u \mid p) \qquad (3\text{-}46)$$

and

$$\Pr(N_c = j, V = v; c_1, m \mid c) = \Pr(N_c = j; c_1, m \mid V = v, c).\Pr(V = v \mid c) \qquad (3\text{-}47)$$

for $j = 1,2,...$, $u = 0,1,...,mn$ and $v = 0,1,2,...$ where

$$\Pr(N_p = j; p_1, m, n \mid U = u, p) \qquad \text{and} \qquad \Pr(N_c = j; c_1, m \mid V = v, c)$$

are the conditional run-length distributions of the *p*-chart and the *c*-chart given in Tables 3.3 and 3.4, respectively, and

$$\Pr(U = u \mid p) = \binom{mn}{u} p^u (1-p)^{mn-u} \quad \text{for} \quad u = 0,1,2,...,mn$$

and

$$\Pr(V = v \mid c) = \frac{e^{-mc}(mc)^v}{v!} \quad \text{for} \quad v = 0,1,2,...$$

are the probability distributions of the estimators U and V, which depend on the unknown parameters p and c, respectively.

The marginal or unconditional run-length distributions are then found from the joint probability distributions and given by

$$\begin{aligned}\Pr(N_p = j; p_1, m, n \mid p) &= \sum_{u=0}^{mn} \Pr(N_p = j, U = u; p_1, m, n \mid p) \\ &= \sum_{u=0}^{mn} \Pr(N_p = j; p_1, m, n \mid U = u, p).\Pr(U = u \mid p) \\ &= \sum_{u=0}^{mn} \hat{\beta}(p_1, m, n \mid u, p)^{j-1}[1 - \hat{\beta}(p_1, m, n \mid u, p)].\binom{mn}{u} p^u (1-p)^{mn-u}\end{aligned} \tag{3-48}$$

and

$$\begin{aligned}\Pr(N_c = j; c_1, m \mid c) &= \sum_{v=0}^{\infty} \Pr(N_c = j, V = v; c_1, m \mid c) \\ &= \sum_{v=0}^{\infty} \Pr(N_c = j; c_1, m \mid V = v, c).\Pr(V = v \mid c) \\ &= \sum_{v=0}^{\infty} \hat{\beta}(c_1, m \mid v, c)^{j-1}[1 - \hat{\beta}(c_1, m \mid v, c)].\frac{e^{-mc}(mc)^v}{v!}\end{aligned} \tag{3-49}$$

for $j = 1,2,...$ (see e.g. Definition 4.2.1 in Casella and Berger, (2002) p. 148).

One can think of these unconditional distributions as weighted averages i.e. the conditional distributions averaged over all possible values of the parameter estimators, where a weight is the probability of obtaining a particular realization of the point estimator which is given by $\Pr(U = u \mid p)$ or $\Pr(V = v \mid p)$.

It is important to note that the unconditional run-length distributions in (3-48) and (3-49) are unconditional only with respect to the random variables U and V; the unconditional run-length distributions still depend on the parameters p and c. This means that when we evaluate the unconditional run-length distributions and the associated characteristics of the unconditional run-length distributions, the results apply only for those particular values of p and c that are used.

The unconditional average run-length and the unconditional variance of the run-length distributions

Apart from the unconditional run-length distributions we can also compute higher order moments of the unconditional run-length distribution.

The unconditional k^{th} non-central moments, for example, are

$$E(N_p^k \mid p) = E_U(E(N_p^k \mid U, p)) \quad \text{and} \quad E(N_c^k \mid c) = E_V(E(N_c^k \mid V, c))$$

where

$$E(N_p^k \mid U, p) \quad \text{and} \quad E(N_c^k \mid V, c)$$

are the k^{th} non-central moments of the conditional run-length distributions of the *p*-chart and *c*-chart, respectively (see e.g. Theorem 5.4.4 in Bain and Engelhardt, (1992) p. 183).

In particular, when $k = 1$ we have that the unconditional average run-length, denoted by *UARL*, which are

$$UARL_p = E(N_p \mid p) = E_U(E(N_p \mid U, p)) \quad \text{and} \quad UARL_c = E(N_c \mid c) = E_V(E(N_c \mid V, c))$$

where

$$E(N_p \mid U, p) = (1 - \hat{\beta}(p_1, m, n \mid U, p))^{-1} \quad \text{and} \quad E(N_c \mid V, c) = ((1 - \hat{\beta}(c_1, m \mid V, c))^{-1}$$

are the conditional *ARL*'s (conditioned on particular observations of the random variables U and V), respectively.

Hence, it follows that

$$\begin{aligned} UARL(p_1, m, n \mid p) &= \sum_{u=0}^{mn} (1 - \hat{\beta}(p_1, m, n \mid u, p))^{-1} \Pr(U = u \mid p) \\ &= \sum_{u=0}^{mn} (1 - \hat{\beta}(p_1, m, n \mid u, p))^{-1} \binom{mn}{u} p^u (1-p)^{mn-u} \end{aligned} \tag{3-50}$$

and

$$\begin{aligned} UARL(c_1, m \mid c) &= \sum_{v=0}^{\infty} (1 - \hat{\beta}(c_1, m \mid v, c))^{-1} \Pr(V = v \mid c) \\ &= \sum_{v=0}^{\infty} (1 - \hat{\beta}(c_1, m \mid v, c))^{-1} \frac{e^{-mc}(mc)^v}{v!}. \end{aligned} \tag{3-51}$$

Similarly, the unconditional variance of the run-length, denoted by *UVARL*, can be found using

(i) the conditional variance of the run-length ($CVAR$),

(ii) the conditional average run-length ($CARL$), and

(iii) the unconditional average run-length ($UARL$),

and is given by

$$UVARL = E_Z(CVARL) + E_Z(CARL^2) - UARL^2 . \quad \text{(3-52a)}$$

where Z plays the role of U and/or V.

Result (3-52a) follows from the fact that, in general, the unconditional variance can be obtained from the expected value of the conditional variance and the variance of the conditional expected value i.e.

$$\begin{aligned} \text{var}(N) &= E_Z(\text{var}(N \mid Z)) + \text{var}_Z(E(N \mid Z)) \\ &= E_Z(\text{var}(N \mid Z)) + \{E_Z[(E(N \mid Z))^2] - [E_Z(E(N \mid Z))]^2\} \quad \text{(3-52b)} \\ UVARL &= E_Z(CVARL) + E_Z(CARL^2) - UARL^2 \end{aligned}$$

where $\text{var}(N)$ is the unconditional variance of the run-length ,

$$CVARL = \text{var}(N \mid Z) = \hat{\beta}/(1-\hat{\beta})^2$$

denotes the conditional variance of the run-length,

$$CARL = E(N \mid Z) = 1/(1-\hat{\beta})$$

denotes the conditional average run-length, $\hat{\beta}$ denotes (in general) the conditional probability of a no-signal and Z plays the role of U and/or V, which is the random variable on which we condition in the particular case (see e.g. Theorem 5.4.3 in Bain and Engelhardt, (1992) p. 182).

In case of the p-chart, using (3-52a), the unconditional variance of the run-length is

$$\begin{aligned}
UVARL_p &= E_U\left(CVARL\right)+\left\{E_U\left(CARL^2\right)-[UARL]^2\right\} \\
&= E_U\left(\frac{\hat{\beta}(p_1,m,n\,|\,U,p)}{(1-\hat{\beta}(p_1,m,n\,|\,U,p))^2}\right)+\left\{E_U\left(\frac{1}{(1-\hat{\beta}(p_1,m,n\,|\,U,p))^2}\right)-\left[E_U\left(\frac{1}{1-\hat{\beta}(p_1,m,n\,|\,U,p)}\right)\right]^2\right\} \\
&= E_U\left(\frac{1+\hat{\beta}(p_1,m,n\,|\,U,p)}{(1-\hat{\beta}(p_1,m,n\,|\,U,p))^2}\right)-\left[E_U\left(\frac{1}{1-\hat{\beta}(p_1,m,n\,|\,U,p)}\right)\right]^2 \\
&= \sum_{u=0}^{mn}\left(\frac{1+\hat{\beta}(p_1,m,n\,|\,u,p)}{(1-\hat{\beta}(p_1,m,n\,|\,u,p))^2}\right)\binom{mn}{u}p^u(1-p)^{mn-u}-\left[\sum_{u=0}^{mn}\left(\frac{1}{1-\hat{\beta}(p_1,m,n\,|\,u,p)}\right)\binom{mn}{u}p^u(1-p)^{mn-u}\right]^2
\end{aligned}$$

whilst for the c-chart we have

$$\begin{aligned}
UVARL_c &= E_V\left(CVARL\right)+\left\{E_V\left(CARL^2\right)-[UARL]^2\right\} \\
&= E_V\left(\frac{\hat{\beta}(c_1,m\,|\,V,c)}{(1-\hat{\beta}(c_1,m\,|\,V,c))^2}\right)+\left\{E_V\left(\frac{1}{(1-\hat{\beta}(c,c_1,m\,|\,V,c))^2}\right)-\left[E_V\left(\frac{1}{1-\hat{\beta}(c,c_1,m\,|\,V,c)}\right)\right]^2\right\} \\
&= E_V\left(\frac{1+\hat{\beta}(c_1,m\,|\,V,c)}{(1-\hat{\beta}(c_1,m\,|\,V,c))^2}\right)-\left[E_V\left(\frac{1}{1-\hat{\beta}(c_1,m\,|\,V,c)}\right)\right]^2 \\
&= \sum_{v=0}^{\infty}\left(\frac{1+\hat{\beta}(c_1,m\,|\,v,c)}{(1-\hat{\beta}(c_1,m\,|\,v,c))^2}\right)\frac{e^{-mc}(mc)^v}{v!}-\left[\sum_{v=0}^{\infty}\left(\frac{1}{1-\hat{\beta}(c_1,m\,|\,v,c)}\right)\frac{e^{-mc}(mc)^v}{v!}\right]^2.
\end{aligned}$$

The unconditional standard deviation of the run-length follows by taking the square root of the unconditional variance of the run-length i.e. $USDRL=\sqrt{UVARL}$.

The unconditional probability mass function (u.p.m.f), the unconditional cumulative distribution function (u.c.d.f), the unconditional false alarm rate (*UFAR*), the unconditional average run-length (*UARL*), and the unconditional variance of the run-length (*UVARL*) for the p-chart and the c-chart are summarized in Tables 3.16 and 3.17, respectively.

These characteristics, as mentioned before, are important as they help us understand the full impact of estimating the unknown parameters on the performance of the charts. Note, however, that when evaluating the unconditional distributions and the unconditional characteristics in Tables 3.16 and 3.17

one still has to select values for p and c; hence, the results are only applicable to the particular values of p and c that are selected.

Table 3.16: The unconditional probability mass function (u.p.m.f), the unconditional cumulative distribution function (u.c.d.f), the unconditional false alarm rate (*UFAR*), the unconditional average run-length (*UARL*) and the unconditional variance of the run-length (*UVARL*) of the *p*-chart in Case U

u.p.m.f	$\Pr(N_p = j; p_1, m, n \mid p) = \sum_{u=0}^{mn} (\hat{\beta}(p_1, m, n \mid u, p))^{j-1}(1-\hat{\beta}(p_1, m, n \mid u, p))\binom{mn}{u} p^u (1-p)^{mn-u}$	(3-53)
u.c.d.f	$\Pr(N_p \le j; p_1, m, n \mid p) = \sum_{u=0}^{mn} (1-(\hat{\beta}(p_1, m, n \mid u, p))^j)\binom{mn}{u} p^u (1-p)^{mn-u} \qquad j = 1,2,....$	(3-54)
UFAR	$UFAR(p_1, m, n \mid p = p_1) = \sum_{u=0}^{mn} (1-\hat{\beta}(p_1, m, n \mid u, p = p_1))\binom{mn}{u} p_1^u (1-p_1)^{mn-u}$	(3-55)
UARL	$UARL(p_1, m, n \mid p) = \sum_{u=0}^{mn} (1-\hat{\beta}(p_1, m, n \mid u, p))^{-1}\binom{mn}{u} p^u (1-p)^{mn-u}$	(3-56)
UVARL	$UVARL_p = \sum_{u=0}^{mn}\left(\frac{1+\hat{\beta}(p_1, m, n \mid u, p)}{(1-\hat{\beta}(p_1, m, n \mid u, p))^2}\right)\binom{mn}{u} p^u (1-p)^{mn-u} - \left[\sum_{u=0}^{mn}\left(\frac{1}{1-\hat{\beta}(p_1, m, n \mid u, p)}\right)\binom{mn}{u} p^u (1-p)^{mn-u}\right]^2$	(3-57)

Table 3.17: The unconditional probability mass function (u.p.m.f), the unconditional cumulative distribution function (u.c.d.f.), the unconditional false alarm rate (*UFAR*), the unconditional average run-length (*UARL*) and the unconditional variance of the run-length (*UVARL*) of the *c*-chart in Case U

u.p.m.f	$\Pr(N_c = j; c_1, m \mid c) = \sum_{v=0}^{\infty} \hat{\beta}(c_1, m \mid v, c)^{j-1}(1-\hat{\beta}(c_1, m \mid v, c))\frac{e^{-mc}(mc)^v}{v!} \qquad j = 1,2,....$	(3-58)
u.c.d.f	$\Pr(N_c \le j; c_1, m \mid c) = \sum_{v=0}^{\infty} 1-(\hat{\beta}(c_1, m \mid v, c))^j \frac{e^{-mc}(mc)^v}{v!} \qquad j = 1,2,....$	(3-59)
UFAR	$UFAR(c_1, m \mid c = c_1) = \sum_{v=0}^{\infty} 1-\hat{\beta}(c_1, m \mid v, c = c_1)\frac{e^{-mc_1}(mc_1)^v}{v!}$	(3-60)
UARL	$UARL(c_1, m \mid c) = \sum_{v=0}^{\infty} (1-\hat{\beta}(c_1, m \mid v, c))^{-1}\frac{e^{-mc}(mc)^v}{v!}$	(3-61)
UVARL	$UVARL_c = \sum_{v=0}^{\infty}\left(\frac{1+\hat{\beta}(c_1, m \mid v, c)}{(1-\hat{\beta}(c_1, m \mid v, c))^2}\right)\frac{e^{-mc}(mc)^v}{v!} - \left[\sum_{v=0}^{\infty}\left(\frac{1}{1-\hat{\beta}(c_1, m \mid v, c)}\right)\frac{e^{-mc}(mc)^v}{v!}\right]^2$	(3-62)

3.2.4.1 Unconditional characteristics of the *p*-chart

The necessary steps and calculations to obtain a numerical value for a particular unconditional characteristic of the Phase II run-length distribution of the *p*-chart are explained via the examples shown in Tables 3.18 and 3.19; these tables are essentially the same as Tables 3.5 and 3.6 that we used to illustrate the mechanics for calculating the FAR , the ARL and the $SDRL$ of the conditional run-length distribution. However, here, we go a step further and calculate the unconditional characteristics of the run-length distribution, that is, the unconditional FAR ($UFAR$), the unconditional ARL ($UARL$) and the unconditional $SDRL$ ($USDRL$). In addition, note that, although we still assume that $p_1 = p = 0.5$ we now assume that $T = mn = 15$ with $(m,n) = (1,15)$ and $(m,n) = (3,5)$ individual Phase I reference observations are used to estimate p .

T = 15 with m = 1 and n = 15

First consider Table 3.18 which assumes that $(m,n) = (1,15)$. Recall that to calculate the conditional properties we begin in column 1 and sequentially move to the right-hand side of the table up to column 9. To illustrate the concept once more, assume that we observe nine nonconforming items from the entire fifteen reference observations i.e. suppose that $U = 9$, so that we get a point estimate of $\bar{p}_{\text{obs}} = 9/15 = 0.6$ for the unknown true fraction nonconforming $0 < p < 1$ in column 2. Thus, using (3-26), we find that the estimated control limits are $L\hat{C}L_p = 0.22$ and $U\hat{C}L_p = 0.98$; these values are listed in columns 3 and 4, respectively. Then, making use of (3-31) we find that the charting constants are $\hat{a} = 3$ and $\hat{b} = 14$ (which are listed in columns 5 and 6, respectively) so that (3-36) yields a conditional false alarm rate of $CFAR(p_1 = 0.5, m = 1, n = 15 \mid U = 9, p = 0.5) = 0.0176$ which leads to a conditional average run-length and a conditional variance of the run-length (found from (3-37) and (3-38)) of

$$CARL(p_1 = 0.5, m = 1, n = 15 \mid U = 9, p = 0.5) = 56.79$$

and

$$CVARL = [CSDRL(p_1 = 0.5, m = 1, n = 15 \mid U = 9, p = 0.5)]^2 = 3168.35;$$

these values are displayed in columns 8 and 9, respectively.

To calculate the unconditional properties of the *p*-chart we calculate a weighted average of all the values (rows) for each of columns 7, 8 and 9, respectively. The weights are found from the probability distribution of the random variable $U \mid p = 0.5 \sim Bin(15,0.5)$ which is given in column 10 and calculated from evaluating $\Pr(U = u \mid p = 0.5) = \binom{15}{u} 0.5^{15}$ for $u = 0,1,...,15$.

Table 3.18: The conditional and unconditional characteristics of the run-length distribution for $m = 1$ and $n = 15$ when $p = p_1 = 0.5$

Phase I						Phase II : Conditional Properties				Phase II : Unconditional Properties			
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)=(7)x(10)	(12)=(8)x(10)	(13)=(8)^{2}x(10)	(14)=(9)x(10)
U	$\bar{p}_{obs}$	$L\hat{C}L_p$	$U\hat{C}L_p$	$\hat{a}$	$\hat{b}$	*CFAR*	*CARL*	*CVARL*	Pr(*U*=*u*|*p*)	*CFAR*.Pr(*U*=*u*|*p*)	*CARL*.Pr(*U*=*u*|*p*)	*CARL*^{2}.Pr(*U*=*u*|*p*)	*CVARL*.Pr(*U*=*u*|*p*)
0	0.00	0.00	0.00	NA	NA	1.0000	1.00	0.00	0.0000	0.00003	0.00003	0.0000	0.0000
1	0.07	-0.13	0.26	NA	3	0.9824	1.02	0.02	0.0005	0.00045	0.00047	0.0005	0.0000
2	0.13	-0.13	0.40	NA	5	0.8491	1.18	0.21	0.0032	0.00272	0.00377	0.0044	0.0007
3	0.20	-0.11	0.51	NA	7	0.5000	2.00	2.00	0.0139	0.00694	0.02777	0.0555	0.0278
4	0.27	-0.08	0.61	NA	9	0.1509	6.63	37.30	0.0417	0.00629	0.27609	1.8299	1.5538
5	0.33	-0.03	0.70	NA	10	0.0592	16.88	268.12	0.0916	0.00543	1.54714	26.1189	24.5717
6	0.40	0.02	0.78	0	11	0.0176	56.79	3168.35	0.1527	0.00269	8.67418	492.60905	483.9349
7	0.47	0.08	0.85	1	12	0.0042	239.18	56969.08	0.1964	0.00082	46.97080	11234.5932	11187.6224
8	0.53	0.15	0.92	2	13	0.0042	239.18	56969.08	0.1964	0.00082	46.97080	11234.5932	11187.6224
9	0.60	0.22	0.98	3	14	0.0176	56.79	3168.35	0.1527	0.00269	8.67418	492.6091	483.9349
10	0.67	0.30	1.03	4	15	0.0592	16.88	268.12	0.0916	0.00543	1.54714	26.1189	24.5717
11	0.73	0.39	1.08	5	15	0.1509	6.63	37.30	0.0417	0.00629	0.27609	1.8299	1.5538
12	0.80	0.49	1.11	7	15	0.5000	2.00	2.00	0.0139	0.00694	0.02777	0.0555	0.0278
13	0.87	0.60	1.13	9	15	0.8491	1.18	0.21	0.0032	0.00272	0.00377	0.0044	0.0007
14	0.93	0.74	1.13	11	15	0.9824	1.02	0.02	0.0005	0.00045	0.00047	0.0005	0.0000
15	1.00	1.00	1.00	NA	NA	1.0000	1.00	0.00	0.0000	0.00003	0.00003	0.0000	0.0000
										0.05074	**115.00**	**23510.42**	**23395.42**
										UFAR	***UARL***	***USDRL* = 183.52**	

Unconditional false alarm rate

To obtain the unconditional false alarm rate (*UFAR*), we need the conditional false alarm rate and the related probability $\Pr(U = u \mid p = 0.5)$ for $u = 0,1,...,15$, which are listed in columns 7 and 10, respectively. Multiplying corresponding row entries of column 7 and column 10, we end up with column 11, that is,

$$CFAR(p_1 = 0.5,1,15 \mid U = u, p = 0.5) \times \Pr(U = u \mid p = 0.5) = (1 - \hat{\beta}(0.5,1,15 \mid U, p)) \times \Pr(U = u \mid p = 0.5)$$

for $u = 0,1,...,15$ so that adding up all the entries in column 11 yields the unconditional false alarm rate i.e.

$$UFAR(p_1 = 0.5,1,15 \mid p = 0.5) = \sum_{u=0}^{15} CFAR(0.5,1,15 \mid U = u, p = 0.5) \Pr(U = u \mid p = 0.5) = 0.05074$$

(see e.g. (3-55) in Table 3.16). The unconditional *FAR* value implies that the probability of a signal on any new incoming Phase II sample, for any practitioner, while the process is in-control at a fraction nonconforming of 0.5, is expected to be 0.05074.

Unconditional average run-length

Like the unconditional *FAR*, the unconditional *ARL* is found by multiplying each of the conditional average run-length values listed in column 8 with the corresponding probability $\Pr(U = u \mid p = 0.5)$ listed in column 10 and then adding up all the resultant products.

To this end, column 12 lists all the values of

$$CARL(p_1 = 0.5,1,15 \mid U = u, p = 0.5) \times \Pr(U = u \mid p = 0.5) \quad \text{for} \quad u = 0,1,\ldots,15$$

so that by totalling the values of column 12 we find the unconditional average run-length to be

$$UARL(p_1 = 0.5,1,15 \mid p = 0.5) = \sum_{u=0}^{15} CARL(p_1 = 0.5,1,15 \mid U = u, p = 0.5)\Pr(U = u \mid p = 0.5) = 115.00$$

(see Table 3.16, (3-56)).

An unconditional *ARL* of 115.00 means that a practitioner that estimates p using $\bar{p} = U / mn$, (which is based on a Phase I reference sample that consists of a total of $T = 15$ individual observations from 1 sample of size 15) can expect that his Phase II *p*-chart would, on average, signal on the 115th sample when the process remains in-control at a fraction nonconforming of 0.5.

Unconditional variance of the run-length

Using expression (3-52a) to calculate the unconditional variance of the run-length we note that, $E_U(CVARL) = 23395.42$ (listed in column 14), $E_U(CARL^2) = 23510.42$ (listed in column 13) so that the unconditional standard deviation of the run-length is found to be

$$USDRL = \sqrt{E_U(CVARL) + E_U(CARL^2) - UARL^2} = \sqrt{23395.42 + 23510.42 - (115.00)^2} = 183.52\,.$$

The unconditional standard deviation is the same for all the users and measures the overall variation in the run-length distribution.

Remark 7

In particular, note that, for $T = 15$ where $(m, n) = (1, 15)$:

(i) The unconditional average run-length is not equal to the reciprocal of the unconditional false alarm rate i.e. $UARL \neq (UFAR)^{-1}$. The reason is that the unconditional run-length distribution is not geometric (see e.g. expression (3-53) in Table 3.16).

This is unlike in Case K where $ARL = (FAR)^{-1}$ (see e.g. expression (3-12) in Table 3.1), which makes both the average run-length and false alarm rate popular measures of a control chart's performance.

(ii) The unconditional average run-length is smaller than the unconditional standard deviation of the run-length; this is not the situation in Case K where $ARL > SDRL = \sqrt{ARL(ARL-1)}$ (see e.g. Appendix 3A, section 3.4.2.2) and is due to extra variation introduced to the run-length distribution when estimating p.

(iii) The unconditional FAR is greater than the FAR of 0.0010 of Case K whilst the unconditional ARL and the unconditional $SDRL$ is less than the ARL of 1024.00 and $SDRL$ of 1023.50 of Case K, respectively.

This implies that a Phase II p-chart in Case U, based on an estimate of p using $T = 15$ observations, will signal more often than the Case K p-chart with a known standard.

T = 15 with m = 3 and n = 5

To study the effect of choosing a smaller value of n relative to m (i.e. changing the composition of the reference sample while keeping the total number of Phase I observation the same) on the unconditional characteristics of the run-length distribution, Table 3.19 shows the calculations necessary to obtain the unconditional *FAR*, the unconditional *ARL* and the unconditional *SDRL* when $T = 15$ with $m = 3$ and $n = 5$.

Although the steps in calculating the values in Table 3.19 are similar to that of Table 3.18, we note that the finer points where the $CFAR = 0$, are somewhat lost when we look at the unconditional FAR, which is found by averaging the conditional FAR (given in column (7)) over all fifteen values of U and their associated probabilities (as given in column (10)). For example, from column (11) in Table 3.19 an unconditional *FAR* equal to 0.01726 is found, which is more than six times the nominal false alarm rate of 0.0027; in spite of this, the unconditional ARL and the unconditional $SDRL$ are still undefined. One can therefore deduce that three subgroups each consisting of five in-control observations do not work satisfactorily in practice.

Table 3.19: The conditional and unconditional characteristics of the run-length distribution for $m = 3$ and $n = 5$ when $p = p_1 = 0.5$

Phase I						Phase II : Conditional Properties				Phase II : Unconditional Properties			
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)=(7)x(10)	(12)=(8)x(10)	(13)=(8)^{2}x(10)	(14)=(9)x(10)
U	$\bar{p}_{obs}$	$\hat{LCL}_p$	$\hat{UCL}_p$	$\hat{a}$	$\hat{b}$	*CFAR*	*CARL*	*CVARL*	Pr(U=u|p)	*CFAR*.Pr(U=u|p)	*CARL*.Pr(U=u|p)	*CARL*^{2}.Pr(U=u|p)	*CVARL*.Pr(U=u|p)
0	0.00	0	0	NA	NA	1.0000	1.00	0.00	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.07	-0.27	0.40	NA	2	0.5000	2.00	2.00	0.0005	0.0002	0.0009	0.0018	0.0009
2	0.13	-0.32	0.59	NA	2	0.5000	2.00	2.00	0.0032	0.0016	0.0064	0.0128	0.0064
3	0.20	-0.34	0.74	NA	3	0.1875	5.33	23.11	0.0139	0.0026	0.0741	0.3950	0.3209
4	0.27	-0.33	0.86	NA	4	0.0313	32.00	992.00	0.0417	0.0013	1.3330	42.6563	41.3232
5	0.33	-0.30	0.97	NA	4	0.0313	32.00	992.00	0.0916	0.0029	2.9326	93.8438	90.9111
6	0.40	-0.26	1.06	NA	5	0.0000	∞	∞	0.1527	0.0000	∞	∞	∞
7	0.47	-0.20	1.14	NA	5	0.0000	∞	∞	0.1964	0.0000	∞	∞	∞
8	0.53	-0.14	1.20	NA	5	0.0000	∞	∞	0.1964	0.0000	∞	∞	∞
9	0.60	-0.06	1.26	NA	5	0.0000	∞	∞	0.1527	0.0000	∞	∞	∞
10	0.67	0.03	1.30	0	5	0.0313	32.00	992.00	0.0916	0.0029	2.9326	93.8438	90.9111
11	0.73	0.14	1.33	0	5	0.0313	32.00	992.00	0.0417	0.0013	1.3330	42.6563	41.3232
12	0.80	0.26	1.34	1	5	0.1875	5.33	23.11	0.0139	0.0026	0.0741	0.3950	0.3209
13	0.87	0.41	1.32	2	5	0.5000	2.00	2.00	0.0032	0.0016	0.0064	0.0128	0.0064
14	0.93	0.60	1.27	2	5	0.5000	2.00	2.00	0.0005	0.0002	0.0009	0.0018	0.0009
15	1.00	1.00	1.00	NA	NA	1.0000	1.00	0.00	0.0000	0.0000	0.0000	0.0000	0.0000
										0.01726	∞	∞	∞
										UFAR	***UARL***	***USDRL* = ∞**	

To illustrate and help understand the overall effects of parameter estimation on the properties of the *p*-chart in more detail, some results (similar to those in Tables 3.18 and 3.19) are presented in Tables 3.20, 3.21 and 3.22 for $T = 10, 20, 25, 30, 50, 75, 100, 200, 250, 300, 500, 750, 1000$ and 1500, each time considering several combinations of m and n values so that $T = mn$. Thus, we look at what happens to the unconditional characteristics (in particular the *UFAR* and the *UARL*) when:

(a) T increases, and

(b) when the composition of the Phase I sample changes i.e. varying m and n.

The resulting unconditional *FAR*'s and the unconditional in-control *ARL*'s are listed under *UFAR* and $UARL_0$, respectively. Also shown is the percentage difference of the unconditional *FAR* and in-control unconditional *ARL* of Case U versus

(a) the *FAR* and *ARL* of Case K (see e.g. Tables A3.4 and A3.5 in Appendix 3A), and

(b) the nominal *FAR* of 0.0027 and the nominal *ARL* of 370.

Several interesting facts emerge from an examination of the results in Tables 3.20, 3.21 and 3.22:

(i) A lot of reference data is needed before the *UFAR* is anywhere near the nominal value of 0.0027 implicitly expected in a typical application of the *p*-chart. In addition, the choice of the number of subgroups m and the subgroup size n are both seen to be important.

For example, the calculations show that unlike in the case with variables data, when studying attributes data the subgroup size n needs to be much larger than the number of subgroups m, to ensure that the *UFAR* is reasonably close to the nominal value and (at the same time) ensure that the *UARL* is not undefined (see e.g. Table 3.21 where $T = 300$ with $m = 10$ and $n = 30$).

(ii) There is great variation in the *UFAR* values and it could be hundreds of percents off from its nominal value of 0.0027 and/or its Case K value for many combinations of m and n that are typically used in practice.

For example, when $T = 100$, we find that

(a) with $m = 4$ and $n = 25$ the *UFAR* is 191.5% above the nominal value and 92% above its Case K value of 0.0041, and

(b) with $m = 20$ and $n = 5$ the *UFAR* is 95.9% lower than the nominal value but close to its Case K value of zero.

(iii) Unless one is careful about the choice of *m* and *n*, the unconditional in-control average run-length of the chart can be undefined particularly when $m >> n$, which is undesirable in practice. This is due to the fact that the conditional probability of a false alarm can be zero for certain values of m and n, since although U can take on any integer value between 0 and mn (including both) with a non-zero probability, the binomial distribution (for the number of nonconforming items within each monitored group) assigns zero probability to any value greater than n.

(iv) The effect of the discreteness of the binomial distribution is also seen to be substantial on both the *FAR* and *ARL* values. For example, unlike in the variables case, with attributes data, only a certain number of ARL_0 values are attainable depending on the combination of values of m, n and p the user has at hand.

(v) As mentioned before, unlike in Case K, the unconditional *ARL* is not equal to the reciprocal of the unconditional *FAR* nor is it smaller than the unconditional *SDRL* (not listed here); this is an important effect of estimating the unknown parameter p.

(vi) For the (m,n) combinations where $n \leq 8$ the Case K *FAR* is zero and the associated Case K *ARL* is undefined (see e.g. Tables A3.4 and A3.5 in Appendix 3A).

In these cases, it is not practical to calculate the percentage difference and therefore indicated by an asterisk. In addition, for those (m,n) combinations where the *UFAR* is zero and/or the *UARL* is undefined it is impractical to calculate the percentage difference from the nominal values and thus indicated by the hash sign.

The aforementioned results suggest that there is a need for a large amount of reference data, with a larger amount of data in each subgroup than the number of subgroups i.e. $n >> m$. For example, when $T = 200$ with $m = 8$ and $n = 25$, the *UFAR* is 0.00447 which is 65.5% above the nominal value, whereas when $T = 500$, both $(m,n) = (25,20)$ and $(m,n) = (20,25)$ lead to an unconditional false alarm rate close to the nominal. This suggests one would need at least 400-500 in-control reference data points to achieve any meaningful control of the false alarm rate near the nominal 0.0027. An examination of the *UARL* values also lead to similar conclusions, in the sense that the combination of the number of subgroups and the size of the subgroup play an important role in dictating the (stable) properties of the *p*-chart.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Table 3.20: The unconditional false alarm rate ($UFAR$) and the unconditional in-control average run-length ($UARL_0$) values for the p-chart for various values of m and n such that $T = mn$ when $p = p_1 = 0.5$

	m	n	$UFAR$	$UARL_0$	% difference from Case K FAR[1]	% difference from Case K ARL[2]	% difference from nominal FAR=0.0027[3]	% difference from nominal ARL=370[4]
T = 10	**1**	**10**	0.06896	168.73	3348.1	203.5	2454.2	-54.4
	2	**5**	0.03552	∞	*	*	1215.6	#
	5	**2**	0.00684	∞	*	*	153.2	#
	10	**1**	0.01172	∞	*	*	334.0	#
T = 20	**1**	**20**	0.04553	135.62	1651.0	186.1	1586.2	-63.3
	2	**10**	0.01913	455.94	856.4	12.3	608.5	23.2
	4	**5**	0.01021	∞	*	*	278.1	#
	5	**4**	0.00787	∞	*	*	191.4	#
	10	**2**	0.00065	∞	*	*	-76.1	#
	20	**1**	0.00020	∞	*	*	-92.5	#
T = 25	**1**	**25**	0.04567	171.89	1014.0	42.7	1591.7	-53.5
	5	**5**	0.00405	∞	*	*	50.1	#
	25	**1**	0.00001	∞	*	*	-99.6	#
T = 30	**1**	**30**	0.04287	235.16	2962.0	197.2	1487.7	-36.4
	2	**15**	0.01765	288.78	1665.1	254.6	553.7	-22.0
	3	**10**	0.01119	605.83	459.4	-15.5	314.3	63.7
	5	**6**	0.00724	∞	*	*	168.2	#
	6	**5**	0.00333	∞	*	*	23.2	#
	10	**3**	0.00066	∞	*	*	-75.7	#
	15	**2**	0.00008	∞	*	*	-97.0	#
	30	**1**	0	∞	*	*	#	#
T = 50	**1**	**50**	0.03686	140.47	1317.6	173.6	1265.1	-62.0
	2	**25**	0.01838	171.32	348.2	43.2	580.6	-53.7
	5	**10**	0.00600	553.53	200.1	-7.5	122.3	49.6
	10	**5**	0.00104	∞	*	*	-61.5	#
	25,2	**50,1**	0	∞	*	*	#	#
T = 75	**1**	**75**	0.04094	105.69	1606.0	287.1	1416.4	-71.4
	3	**25**	0.01022	254.50	149.2	-3.6	278.5	-31.2
	5	**15**	0.00612	492.82	512.3	107.8	126.8	33.2
	15	**5**	0.00033	∞	*	*	-87.7	#
	25,3	**75,1**	0	∞	*	*	#	#
T = 100	**1**	**100**	0.04006	108.45	1044.5	162.1	1383.6	-70.7
	2	**50**	0.01475	234.72	467.3	63.7	446.3	-36.6
	4	**25**	0.00787	246.68	92.0	-0.6	191.5	-33.3
	5	**20**	0.00577	348.72	122.0	11.3	113.7	-5.8
	10	**10**	0.00332	647.93	65.9	-21.0	22.9	75.1
	20	**5**	0.00011	∞	*	*	-95.9	#
	25,4 100,1	**50,2**	0	∞	*	*	#	#

[1] % deviation $= 100(UFAR/FAR_{\text{Case K}} - 1)$; [2] % deviation $= 100(UARL_0/ARL_{\text{Case K}} - 1)$;
[3] % deviation $= 100(UFAR/0.0027 - 1)$; [4] % deviation $= 100(UARL_0/370 - 1)$

Table 3.21: The unconditional false alarm rate (*UFAR*) and the unconditional in-control average run-length ($UARL_0$) values for the *p*-chart for various values of m and n such that $T = mn$ when $p = p_1 = 0.5$

	m	*n*	*UFAR*	$UARL_0$	% difference from Case K *FAR*[1]	% difference from Case K *ARL*[2]	% difference from nominal *FAR*=0.0027[3]	% difference from nominal *ARL*=370[4]
***T* = 200**	**1**	**200**	0.03328	162.34	1347.0	170.2	1132.6	-56.1
	2	**100**	0.01587	164.72	353.3	72.6	487.6	-55.5
	4	**50**	0.00708	259.93	172.2	47.8	162.1	-29.7
	5	**40**	0.00593	275.48	169.4	63.4	119.5	-25.5
	8	**25**	0.00447	312.51	9.0	-21.5	65.5	-15.5
	10	**20**	0.00374	409.26	43.8	-5.2	38.5	10.6
	20	**10**	0.00207	683.52	3.7	-25.1	-23.2	84.7
	25	**8**	0.00171	∞	*	*	-36.7	#
	40,5 100,2	**50,4 200,1**	0	∞	*	*	#	#
***T* = 250**	**1**	**250**	0.03546	131.33	1122.7	164.5	1213.3	-64.5
	2	**125**	0.01440	187.77	554.7	139.2	433.4	-49.3
	5	**50**	0.00616	285.03	137.0	34.8	128.3	-23.0
	10	**25**	0.00392	330.65	-4.4	-25.8	45.1	-10.6
	25	**10**	0.00181	697.63	-9.4	46.8	-32.9	88.5
	50,5 250,1	**125,2**	0	∞	*	*	#	#
***T* = 300**	**1**	**300**	0.03725	120.40	1064.0	162.1	1279.6	-67.5
	2	**150**	0.01488	177.10	520.1	134.7	451.2	-52.1
	3	**100**	0.01006	197.37	187.3	44.0	272.5	-46.7
	4	**75**	0.00765	223.40	218.9	83.1	183.5	-39.6
	5	**60**	0.00602	282.98	122.8	32.3	122.8	-23.5
	6	**50**	0.00532	276.87	104.7	38.8	97.2	-25.2
	10	**30**	0.00387	369.27	176.1	89.3	43.2	-0.2
	12	**25**	0.00362	343.82	-11.7	-28.7	34.0	-7.1
	15	**20**	0.00303	444.15	16.4	-12.6	12.1	20.0
	20	**15**	0.00245	670.29	144.9	52.8	-9.3	81.2
	25	**12**	0.00228	988.67	355.3	107.1	-15.7	167.2
	30	**10**	0.00175	693.34	-12.5	-26.2	-35.2	87.4
	50	**6**	0.00001	∞	*	*	-99.6	#
	60,5 100,3 300,1	**75,4 150,2**	0	∞	*	*	#	#
***T* = 500**	**1**	**500**	0.03406	139.83	1161.6	165.2	1161.6	-62.2
	2	**250**	0.01416	187.50	388.3	85.3	424.4	-49.3
	4	**125**	0.00735	245.93	234.2	82.6	172.3	-33.5
	5	**100**	0.00638	236.94	82.3	20.0	136.3	-36.0
	10	**50**	0.00398	328.92	53.0	16.8	47.3	-11.1
	20	**25**	0.00296	373.74	-27.8	-34.4	9.7	1.0
	25	**20**	0.00258	470.72	-0.9	-17.6	-4.6	27.2
	50	**10**	0.00175	626.47	-12.5	-18.3	-35.2	69.3
	100,5 250,2	**125,4 500,1**	0	∞	*	*	#	#

[1] $\% \text{ deviation} = 100(UFAR/FAR_{\text{Case K}} - 1)$; [2] $\% \text{ deviation} = 100(UARL_0/ARL_{\text{Case K}} - 1)$;

[3] $\% \text{ deviation} = 100(UFAR/0.0027 - 1)$; [4] $\% \text{ deviation} = 100(UARL_0/370 - 1)$

Table 3.22: The unconditiona **nconditional in-control average run-length ($UARL_0$) values** **f m and n such that $T = mn$ when $p = p_1 = 0.5$**

	m	n	$UFAR$	$UARL_0$	% difference from Case K FAR[1]	% difference from Case K ARL[2]	% difference from nominal FAR=0.0027[3]	% difference from nominal ARL=370[4]
T = 750	1	750	0.03373	154.68	1305.5	167.4	1149.3	-58.2
	2	375	0.01429	195.58	429.2	89.7	429.2	-47.1
	3	250	0.00927	220.66	219.5	57.4	243.2	-40.4
	5	150	0.00628	239.35	161.6	73.7	132.6	-35.3
	6	125	0.00543	259.56	147.0	73.0	101.3	-29.8
	10	75	0.00417	291.37	73.6	40.4	54.3	-21.3
	15	50	0.00339	351.05	30.3	9.5	25.5	-5.1
	25	30	0.00282	426.66	101.5	63.8	4.5	15.3
	30	25	0.00270	388.13	-34.3	-36.8	-0.2	4.9
	50	15	0.00180	822.08	79.8	24.6	-33.4	122.2
	75	10	0.00180	590.93	-9.8	-13.4	-33.2	59.7
	125,6 250,3 750,1	150,5 375,2	0	∞	*	*	#	#
T = 1000	1	1000	0.03362	142.52	1193.2	165.2	1145.3	-61.5
	2	500	0.01434	193.35	431.0	91.8	431.0	-47.7
	4	250	0.00746	221.21	157.4	57.0	176.5	-40.2
	5	200	0.00613	253.58	166.6	73.0	127.1	-31.5
	8	125	0.00456	286.10	107.5	57.0	69.1	-22.7
	10	100	0.00422	285.01	20.6	-0.3	56.3	-23.0
	20	50	0.00312	364.70	20.0	5.4	15.6	-1.4
	25	40	0.00290	374.32	31.7	20.3	7.3	1.2
	40	25	0.00260	393.72	-36.7	-37.7	-3.8	6.4
	50	20	0.00232	475.03	-10.7	-18.3	-14.0	28.4
	100	10	0.00186	559.96	-6.9	-8.6	-31.0	51.3
	125	8	0.00024	∞	*	*	-91.0	#
	200,5 500,2	250,4 1000,1	0	∞	*	*	#	#
T = 1500	1	1500	0.03315	149.71	1226.1	166.3	1127.9	-59.5
	2	750	0.01401	189.14	483.9	118.7	419.0	-48.9
	3	500	0.00943	205.42	249.4	80.5	249.4	-44.5
	4	375	0.00716	233.25	165.1	59.0	165.1	-37.0
	5	300	0.00629	235.43	96.4	34.0	132.8	-36.4
	6	250	0.00538	258.84	85.4	34.2	99.1	-30.0
	10	150	0.00418	287.19	74.3	44.8	55.0	-22.4
	12	125	0.00382	307.31	73.8	46.2	41.6	-16.9
	15	100	0.00361	308.79	3.3	-7.9	33.9	-16.5
	20	75	0.00327	331.32	36.3	23.5	21.1	-10.5
	25	60	0.00303	360.64	12.4	3.8	12.4	-2.5
	30	50	0.00287	380.90	10.5	0.9	6.4	2.9
	50	30	0.00253	466.77	81.0	49.7	-6.2	26.2
	60	25	0.00254	397.09	-38.1	-38.2	-6.0	7.3
	75	20	0.00235	453.58	-9.8	-14.4	-13.1	22.6
	100	15	0.00135	933.58	34.6	9.7	-50.2	152.3
	125	12	0.00109	1686.74	118.2	21.4	-59.6	355.9
	150	10	0.00191	533.16	-4.4	-4.0	-29.2	44.1
	250,6 375,4 750,2	300,5 500,3 1500,1	0	∞	*	*	#	#

[1] % deviation $= 100(UFAR/FAR_{\text{Case K}} - 1)$; [2] % deviation $= 100(UARL_0/ARL_{\text{Case K}} - 1)$;
[3] % deviation $= 100(UFAR/0.0027 - 1)$; [4] % deviation $= 100(UARL_0/370 - 1)$

3.2.4.2 Unconditional characteristics of the *c*-chart

The unconditional characteristics of the *c*-chart can be calculated in the same manner as that of the *p*-chart. To this end, the necessary steps are shown in Table 3.23 where we assume that $c = c_1 = 1$ and $m = 5$ individual and independent Phase I inspection units are used to estimate c.

First, we calculate the conditional characteristics in columns 7, 8 and 9 (based on the observed value u or $\bar{c}_{\text{obs}}$ and the estimated control limits and resulting chart constants listed in columns 1 to 6) and then we calculate the unconditional properties of the run-length distribution (in particular, the *UFAR*, the *UARL* and the *USDRL* using expressions (3-59), (3-61) and (3-52)) by means of the results of columns 11 to 14. Note, however, that although theoretically $V \in \{0,1,2,...\}$, Table 3.23 only shows the conditional properties for $V \in \{0,1,2,...,20\}$ in order to save space.

Table 3.23: The conditional and unconditional characteristics of the run-length distribution for $m = 5$ when $c = 1$

Phase I						Phase II : Conditional Properties				Phase II : Unconditional Properties			
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)=(7)x(10)	(12)=(8)x(10)	(13)=(8)^{2}x(10)	(14)=(9)x(10)
v	$\bar{c}_{\text{obs}}$	$L\hat{C}L_c$	$U\hat{C}L_c$	$\hat{d}$	$\hat{f}$	*CFAR*	*CARL*	*CVARL*	Pr(*V*=*v*|*c*)	*CFAR*.Pr(*V*=*v*|*c*)	*CARL*.Pr(*V*=*v*|*c*)	*CARL*^{2}.Pr(*V*=*v*|*c*)	*CVARL*.Pr(*V*=*v*|*c*)
0	0.0	0.00	0.00	0	0	1.0000	1.0000	0.0000	0.00674	0.00674	0.00674	0.0067	0.0000
1	0.2	-1.14	1.54	0	1	0.6321	1.5820	0.9207	0.03369	0.02130	0.05330	0.0843	0.0310
2	0.4	-1.50	2.30	0	2	0.4482	2.2312	2.7472	0.08422	0.03775	0.18792	0.4193	0.2314
3	0.6	-1.72	2.92	0	2	0.4482	2.2312	2.7472	0.14037	0.06291	0.31321	0.6988	0.3856
4	0.8	-1.88	3.48	0	3	0.3869	2.5849	4.0967	0.17547	0.06788	0.45356	1.1724	0.7188
5	1.0	-2.00	4.00	0	3	0.3869	2.5849	4.0967	0.17547	0.06788	0.45356	1.1724	0.7188
6	1.2	-2.09	4.49	0	4	0.3715	2.6915	4.5527	0.14622	0.05433	0.39356	1.0593	0.6657
7	1.4	-2.15	4.95	0	4	0.3715	2.6915	4.5527	0.10444	0.03881	0.28111	0.7566	0.4755
8	1.6	-2.19	5.39	0	5	0.3685	2.7139	4.6513	0.06528	0.02405	0.17716	0.4808	0.3036
9	1.8	-2.22	5.82	0	5	0.3685	2.7139	4.6513	0.03627	0.01336	0.09842	0.2671	0.1687
10	2.0	-2.24	6.24	0	6	0.3680	2.7177	4.6680	0.01813	0.00667	0.04928	0.1339	0.0846
11	2.2	-2.25	6.65	0	6	0.3680	2.7177	4.6680	0.00824	0.00303	0.02240	0.0609	0.0385
12	2.4	-2.25	7.05	0	7	0.3679	2.7182	4.6704	0.00343	0.00126	0.00933	0.0254	0.0160
13	2.6	-2.24	7.44	0	7	0.3679	2.7182	4.6704	0.00132	0.00049	0.00359	0.0098	0.0062
14	2.8	-2.22	7.82	0	7	0.3679	2.7182	4.6704	0.00047	0.00017	0.00128	0.0035	0.0022
15	3.0	-2.20	8.20	0	8	0.3679	2.7183	4.6707	0.00016	0.00006	0.00043	0.0012	0.0007
16	3.2	-2.17	8.57	0	8	0.3679	2.7183	4.6707	0.00005	0.00002	0.00013	0.0004	0.0002
17	3.4	-2.13	8.93	0	8	0.3679	2.7183	4.6707	0.00001	0.00001	0.00004	0.0001	0.0001
18	3.6	-2.09	9.29	0	9	0.3679	2.7183	4.6708	0.00000	0.00000	0.00001	0.0000	0.0000
19	3.8	-2.05	9.65	0	9	0.3679	2.7183	4.6708	0.00000	0.00000	0.00000	0.0000	0.0000
20	4.0	-2.00	10.00	0	9	0.3679	2.7183	4.6708	0.00000	0.00000	0.00000	0.0000	0.0000
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
										0.40672	**2.51**	**6.35**	**3.85**
										UFAR	***UARL***	***USDRL* = 1.98**	

To obtain the unconditional false alarm rate, for instance, we need the conditional false alarm rate and the related probability $\Pr(V=v\,|\,c=1)$ for $v=0,1,2,...$ listed in columns 7 and 10, respectively. Multiplying the corresponding row entries of columns 7 and 10, we end up with column 11, that is,

$$CFAR(c_1=1\,,m=5\,|\,V=v,c=1)\times\Pr(V=v\,|\,c=1)=(1-\hat{\beta}(c_1=1\;,\;m=5\,|\,V=v,c=1))\times\Pr(V=v\,|\,c=1)$$

$$\text{for}\quad v=0,1,2,...$$

so that summing the entries in column 11 yields the unconditional false alarm rate i.e.

$$UFAR(c_1=1,m=5\,|\,c=1)=\sum_{v=0}^{\infty}CFAR(c_1=1,m=5\,|\,V=v,c=1)\Pr(V=v\,|\,c=1)=0.40672$$

(see e.g. (3-61) in Table 3.17). Similarly, we find an unconditional *ARL* and unconditional *SDRL* of 2.51 and 1.98, respectively. Note that, in the calculation of the unconditional characteristics in Table 3.23 the summation was done until $P(V=v\,|\,c=1)\approx 0$.

Compared to the Case K *FAR*, *ARL* and *SDRL* of 0.3869, 2.58 and 2.02, respectively (see e.g. Table A3.12 in Appendix 3A) we see that the unconditional values are not far off. However, the unconditional values do not measure up to the nominal *FAR*, *ARL* and *SDRL* values of 0.0027, 370 and 369 typically expected from a 3-sigma control chart; the reason for this big discrepancy is twofold:

(i) the normal approximation to the $Poi(c)$, for small c, is inaccurate so that the charting formula (mean $\pm$ 3 standard deviations) may be inaccurate, and

(ii) due to the discrete nature of the Poisson distribution only certain (conditional) *FAR*, *ARL* and *SDRL* values can be attained.

Note that, from Table 3.23 it is clear that, unlike in case of the *p*-chart, none of the *CFAR* values of the *c*-chart are zero and thus none of the moments, such as the *UFAR* and the *USDRL*, are undefined.

To illustrate the effect of parameter estimation on the overall performance of the *c*-chart, Table 3.24 displays the *UFAR*, the *UARL* and the *USDRL* for various values of m when $c=c_1=1,2,4,6,8,10,20\text{ and }30$. Also shown are the *FAR*, the *ARL* and the *SDRL* for Case K and the nominal values – given in the last two rows of the table.

We observe that, in general:

(i) As the size of the Phase I reference sample m becomes larger, the unconditional properties gets closer to the Case K values, regardless the value of c.

For instance, when $c = c_1 = 8$ and $m = 20$, the $UFAR = 0.0054$, the $UARL = 315.32$ and the $USDRL = 468.24$ but, when the Phase I sample increase to $m = 500$ inspection units, the $UFAR = 0.0041$, the $UARL = 246.81$ and the $USDRL = 247.68$, which is close to the *FAR*, the *ARL* and the *SDRL* of Case K i.e. 0.0041, 246.70 and 246.20, respectively;

(ii) Unless c and m are both large the *UFAR*, the *UARL* and the *USDRL* are nowhere near the nominally expected values of 0.0027, 370.0 and 369.9.

For instance, when $c = 6$ and $m = 25$, the $UFAR = 0.0079$, the $UARL = 156.49$ and the $USDRL = 181.41$ but, when $c = 20$ and $m = 200$ the $UFAR = 0.0032$, the $UARL = 333.40$ and the $USDRL = 352.01$ gets closer but still not equal to the nominal values. Although this could be a reason for concern for the practitioner, the nominal values are not entirely appropriate given the fact that the Poisson distribution is discrete ;

(iii) The unconditional *ARL* is not equal to the reciprocal of the unconditional *FAR* nor is it smaller than the unconditional *SDRL* (for all combinations of m and c).

This is unlike the situation of the *c*-chart in Case K and is a result of estimating the unknown parameter c; this was also observed in the case of the p-chart with an unknown standard.

Table 3.24: The unconditional false alarm rate (*UFAR*), the unconditional in-control average run-length ($UARL_0$) and the unconditional in-control standard deviation of the run-length ($USDRL_0$) values for the *c*-chart for various values of m when $c = c_1 = 1, 2, 4, 6, 8, 10, 20$ and 30 [1]

	c							
m	**1**	**2**	**4**	**6**	**8**	**10**	**20**	**30**
	0.4067	0.1603	0.0325	0.0136	0.0104	0.0095	0.0078	0.0072
5	2.51	6.54	38.49	166.91	436.17	399.00	303.41	269.39
	1.98	6.22	42.31	226.85	855.24	664.34	420.94	345.01
	0.3901	0.1485	0.0272	0.0097	0.0069	0.0062	0.0052	0.0048
10	2.58	6.82	40.34	162.21	370.41	378.91	330.91	307.82
	2.03	6.37	42.30	205.76	653.10	577.22	427.50	369.61
	0.3845	0.1463	0.0259	0.0087	0.0060	0.0053	0.0045	0.0042
15	2.61	6.88	41.04	159.53	326.93	356.59	333.40	321.49
	2.05	6.40	42.33	194.19	525.47	520.67	416.14	376.15
	0.3824	0.1448	0.0252	0.0082	0.0054	0.0048	0.0041	0.0038
20	2.62	6.93	41.48	157.90	315.32	353.51	338.79	328.11
	2.06	6.44	42.35	187.07	468.24	489.44	412.20	377.23
	0.3813	0.1446	0.0248	0.0079	0.0052	0.0045	0.0039	0.0037
25	2.63	6.94	41.74	156.49	298.67	343.85	336.93	330.19
	2.07	6.44	42.37	181.41	425.57	460.87	403.04	373.39
	0.3807	0.1439	0.0247	0.0077	0.0050	0.0044	0.0038	0.0036
30	2.63	6.96	41.78	155.76	290.10	333.52	334.53	332.60
	2.07	6.46	42.32	177.97	395.40	438.50	392.87	372.15
	0.3799	0.1434	0.0241	0.0073	0.0047	0.0040	0.0035	0.0033
50	2.63	6.99	42.22	154.09	276.24	322.48	335.16	336.25
	2.08	6.48	42.42	169.72	344.81	401.00	379.88	366.80
	0.3796	0.1424	0.0239	0.0070	0.0044	0.0038	0.0033	0.0032
100	2.64	7.03	42.40	154.12	261.79	308.18	334.20	339.49
	2.08	6.52	42.44	162.69	300.06	360.11	363.95	360.06
	0.3795	0.1413	0.0239	0.0066	0.0042	0.0037	0.0032	0.0031
200	2.64	7.08	42.45	156.83	252.11	295.09	333.40	341.43
	2.08	6.57	42.48	160.31	268.28	320.34	352.01	355.01
	0.3794	0.1407	0.0238	0.0064	0.0041	0.0036	0.0031	0.0030
300	2.64	7.11	42.47	159.17	248.61	289.82	332.90	342.34
	2.08	6.59	42.49	160.92	255.61	302.17	344.89	352.47
	0.3794	0.1402	0.0238	0.0062	0.0041	0.0035	0.0031	0.0030
500	2.64	7.13	42.48	161.91	246.81	285.81	331.23	339.38
	2.08	6.61	42.51	162.24	247.68	287.77	334.07	342.30
	0.3869	0.1399	0.0264	0.0061	0.0041	0.0035	0.0029	0.0029
Case K	2.58	7.15	37.81	163.74	246.70	285.74	339.72	349.94
	2.02	6.63	37.31	163.24	246.20	285.23	339.22	349.44
Nominal	0.0027 , 370.0 , 369.9							

[1]The three rows of each cell shows the *UFAR*, the $UARL_0$ and the $USDRL_0$, respectively

Example 1 continued: A Phase II *p*-chart

Phase I and Phase II (conditional)

Recall that the final Phase I data consisted of $m = 28$ samples each of size $n = 50$ (see pages 160-161). Based on these data, it was found that $\bar{p} = 301/1400 = 0.215$ so that the estimated Phase II control limits were set at $U\hat{C}L_p = 0.3893$ and $L\hat{C}L_p = 0.0407$. Given the particular Phase I data, it was shown that the resultant Phase II p-chart has a conditional false alarm rate of $CFAR = 0.002218$ and a conditional average run-length of $CARL_0 = 450.89$.

To get an idea of the general performance of a Phase II p-chart based on $m = 28$ samples each of size $n = 50$ (even prior to collecting the data) one has to look at the unconditional properties of the Phase II p-chart; the unconditional properties takes into account all the possible realizations of $\bar{p} = \frac{U}{mn} \in \{0, \frac{1}{mn}, \frac{2}{mn}, ..., \frac{mn-1}{mn}, 1\}$.

Phase II (unconditional)

Using (3-56) and averaging over all $mn + 1 = 28 \times 50 + 1 = 1401$ possible values and the corresponding binomial probabilities of U, the in-control unconditional *ARL* is found to be

$$UARL_0(p_1 = 0.2, m = 28, n = 50 \mid p = 0.2) = \sum_{u=0}^{1400} (1 - \hat{\beta}(0.2, 28, 50 \mid u, 0.2))^{-1} \binom{1400}{u} 0.2^u (0.8)^{1400-u} = 401.51$$

which is about 11% smaller than the in-control conditional *ARL* for the given data,

$$CARL_0(p_1 = 0.2, 28, 50 \mid U = 301, p = 0.2) = 450.89.$$

Perhaps more importantly, it is seen that when p is estimated from Phase I data, the in-control unconditional *ARL* is 8.5% higher than the corresponding in-control *ARL* of 370 as obtained in the standard known case. Thus, when p is estimated, the in-control *ARL* can be much larger than the nominal value.

Example 2 continued: A Phase II *c*-chart

Phase I and Phase II (conditional)

The final Phase I data consisted of $m = 24$ inspection units each of 100 individual items of product; this resulted in a point estimate $\bar{c} = 472/24 = 19.67$ so that the estimated Phase II control limits were set at $U\hat{C}L_c = 32.97$ and $L\hat{C}L_c = 6.36$ (see pages 162 – 163). Given the particular Phase I data, it was shown that the resultant Phase II c-chart has a conditional false alarm rate of $CFAR = 0.004983$; it follows that the conditional average run-length is $CARL_0 = 1/0.004983 = 200.68$.

Like in the case of the Phase II p-chart, one can get an idea of the general performance of a Phase II c-chart based on $m = 24$ inspection units each of 100 individual items of product (even prior to collecting the data) by looking at the unconditional properties of the Phase II c-chart; the unconditional properties take into account all the possible realizations of $\bar{c} = \frac{V}{m} \in \{0, \frac{1}{m}, \frac{2}{m}, ...\}$.

Phase II (unconditional)

Using (3-60) and (3-61), and averaging over all the possible values and the corresponding probabilities of $V \mid c = 20 \sim Poi(mc = 480)$, the unconditional false alarm rate ($UFAR$) is found to be 0.0039 and the in-control unconditional average run length ($UARL_0$) is found to be 335.30.

The $UFAR$ is 20% less than the $CFAR$ of 0.004983 and the $UARL_0$ is 67% larger than the $CARL$ of 200.68; both these conditional properties are based on an observed value of V equal to 472.

Note that, with regards to the unconditional chart properties, the in-control unconditional average run-length ($UARL_0$) is 1.3% less than the in-control average run-length of 339.72 one would have obtained in Case K for $c_0 = 20$ and the unconditional false alarm rate $UFAR$ is 34.5% larger than the FAR of 0.0029 obtainable in Case K (see e.g. Table A3.12 in Appendix 3A); we can thus expect more false alarms (given the Phase I data at hand) than what would be the case if it is known that $c = 20$.

3.3 Concluding remarks: Summary and recommendations

The false alarm rate (FAR) and the in-control average run length (ARL_0) of the *p*-chart and the *c*-chart are substantially affected by the estimation of the unknown true fraction of nonconforming items p and/or the unknown true average number of nonconformities in an inspection unit c. Calculations show that when p and c are estimated:

(i) The unconditional FAR, unlike in Case K, is not equal (not even close) to the reciprocal of the unconditional ARL_0 and vice versa;

(ii) The unconditional ARL_0 is, unlike in Case K, smaller than the unconditional $SDRL_0$;

(iii) Unless m and/or n are rather large, the unconditional false alarm rates and the in-control unconditional average run-lengths can be substantially different from the nominal values of 0.0027 and 370;

(iv) Even if more Phase I data is available, neither the $UARL_0$ nor the $UFAR$ will necessarily be exactly equal to the commonly used nominally expected values (primarily due to the discreteness of the underlying distributions);

(v) The typical recommendation of using between $m = 10$ and 25 subgroups of size 5 appears to be inadequate and can be very problematic with attributes charts with regard to a true FAR or true ARL_0 ; and

(vi) Since one deals with a discrete (binomial or Poisson) distribution in the case of attributes charts, it is rather unlikely to be able to guarantee an exact false alarm rate as is typical for a variables control chart.

For the *p*-chart, in particular, even with a large amount of reference data, if *m* is (much) larger than *n* (as is the case in a typical variables charting situation) the false alarm rate can be too small and the in-control average run-length can be undefined, which are, of course, undesirable. In practice, at least $T \geq 200$ reference data points are recommended, in 10 subgroups of 20 observations each; a general rule is $n/m \geq 0.5$. To this end, Table 3.20, 3.21 and 3.22 can provide valuable guidance in the process of choosing *m* and *n*. Similarly, for the *c*-chart, Table 3.24 can be used as a guide to select an appropriate number m of Phase I inspection units.

If the necessary amount of reference data is not available in a given situation, the user can calculate the exact unconditional false alarm rates and the exact in-control unconditional *ARL* values using the formulas given in this chapter for the specific *m* and/or *n* values at hand and get an idea of the ramifications of estimating p and/or c.

Another alternative would be to adjust the control limits by finding the value of the charting constant $k > 0$ (which is equal to 3 in routine applications) so that the unconditional FAR equals a specified FAR^* or the unconditional ARL equals a particular ARL_0^*, say. This would mean either expanding or contracting the control limits and entails, for example, in case of the Phase II *p*-chart, solving for k from

$$UARL_0(p_1 = p_*, m, n, k \mid p = p_*) = \sum_{u=0}^{mn} (1 - \hat{\beta}(p_1 = p_*, m, n \mid u, p = p_*))^{-1} \binom{mn}{u} p_*^u (1 - p_*)^{mn-u} = ARL_0^*$$

where m, n and $p_1 = p = p_*$ for some $0 < p_* < 1$ and

$$\hat{\beta}(p_1, m, n, k \mid u, p) = \Pr(L\hat{C}L_p < p_i < U\hat{C}L_p \mid \overline{p}, p)$$

$$= \Pr(\overline{p} - k\sqrt{\overline{p}(1-\overline{p})/n} < p_i < \overline{p} + k\sqrt{\overline{p}(1-\overline{p})/n} \mid \overline{p}, p)$$

$$= \Pr(\tfrac{u}{mn} - k\sqrt{\tfrac{u}{mn}(1-\tfrac{u}{mn})/n} < p_i < \tfrac{u}{mn} + k\sqrt{\tfrac{u}{mn}(1-\tfrac{u}{mn})/n} \mid u, p)$$

(see e.g. expression (3-30) where $k = 3$).

However, note that, in solving the above equation the user has to, as in the preceding examples, specify a value of p - the same parameter that is unknown! This implies that the practitioner has to know the process that is monitored quite well because the charting constant(s) found from solving the above equation would only be appropriate for the particular p that is selected.

To overcome the predicament of choosing a specific value for p (denoted p_*) one can, for instance, make use of the idea of mixture distributions and assume a particular distribution for p, say $f(p; \underline{\boldsymbol{\theta}})$ for $0 < p < 1$ where $\underline{\boldsymbol{\theta}}$ are the (known or specified) parameters of the distribution (which handles our uncertainty about the parameter p by treating it as a random variable rather than a fixed value) and then solve for k from

$$\int_0^1 \left(\sum_{u=0}^{mn} (1-\hat{\beta}(p_1 = p, m, n, k \mid u, p))^{-1} \binom{mn}{u} p^u (1-p)^{mn-u} \right) f(p;\underline{\boldsymbol{\theta}})dp = ARL_0^*.$$

Again, the exact equations given in this chapter can be helpful in this regard, but the practitioner still needs to select and substantiate, from a practical point of view, a distribution $f(p;\underline{\boldsymbol{\theta}})$ and provide the parameter(s) for this distribution.

If the idea of mixture distributions is to be followed, we suggest that $f(p;\underline{\boldsymbol{\theta}})$ and its parameters be chosen in such a way, that best conveys the practitioner's believe about the unknown true fraction nonconforming. For example, because we know that $0 < p < 1$, one possibility is to use the type I (or standard) Beta distribution with parameters (α, β), which has the interval (0,1) as support, as a prior distribution. But which beta distribution should we use? If it is believed that p is in the neighborhood of 0.25 (say) we may, for instance, choose a $\text{Beta}(\alpha = 1, \beta = 3)$ distribution which has a mean of 0.25 and a variance of 0.0375. Other options are certainly also available.

A third approach that one can use to obtain the appropriate Phase II control limits is a Bayesian procedure. As an example, we briefly outline the Bayes approach for the Phase II *p*-chart. According to Bayes' theorem the posterior distribution, g, is proportional to the likelihood function, L, times the prior distribution, f.

For the *p*-chart the likelihood is

$$L(p \mid \text{Phase I data}) \propto p^{\sum_{i=1}^{m} x_i} (1-p)^{mn - \sum_{i=1}^{m} x_i}$$

where the Phase I data are the observed values of X_i, $i = 1,2,...,m$ and denoted by x_i, $i = 1,2,...,m$. The Jeffreys' prior (which is the best noninformative prior for the unknown parameter p) is given by

$$f(p) \propto p^{-\frac{1}{2}} (1-p)^{-\frac{1}{2}}.$$

From the likelihood and the prior it follows that the posterior distribution of p is a beta distribution i.e.

$$g(p \mid \text{Phase I data}) = \left[B\left(\sum_{i=1}^{m} x_i + \frac{1}{2}, mn - \sum_{i=1}^{m} x_i + \frac{1}{2} \right) \right]^{-1} p^{\sum_{i=1}^{m} x_i - \frac{1}{2}} (1-p)^{mn - \sum_{i=1}^{m} x_i - \frac{1}{2}} \quad , \quad 0 < p < 1.$$

If the process remains in-control during Phase II monitoring, the control limits for a Phase II sample of n independent Bernoulli trials (units), which results in Y_i, $i = m+1, m+2,...$ successes (nonconforming units), can be derived using a predictive distribution, h. The conditional distribution of Y_i, $i = m+1, m+2,...$, given the sample size n and p, is binomial (n, p) and the unconditional predictive distribution of Y_i is a beta-binomial distribution i.e.

$$h(y \mid \text{Phase I data}) = \int_0^1 \binom{n}{y} p^y (1-p)^{n-y} g(p \mid \text{Phase I data}) dp$$

$$= \binom{n}{y} \frac{B\left(\sum_{i=1}^{m} x_i + \frac{1}{2} + y, mn - \sum_{i=1}^{m} x_i + \frac{1}{2} + n - y \right)}{B\left(\sum_{i=1}^{m} x_i + \frac{1}{2}, mn - \sum_{i=1}^{m} x_i + \frac{1}{2} \right)}$$

where $y = 0,1,2,...,n$ and $i = m+1, m+2,...$.

The Phase II control limits, via a Bayes approach, are then derived (using the unconditional predictive distribution) from the resulting rejection region of size α, that is, $R(\alpha)$, which is defined as

$$\alpha = \sum_{R(\alpha)} h(y \mid \text{Phase I data}).$$

Because there is currently no evidence to suggest that the one approach (i.e. either assuming that p is deterministic and unknown and therefore specifying a value for p or using a Bayes approach) is superior and none of the approaches is without any obstacles, more research is needed to find suitable charting constants for the Phase II attributes charts.

3.4 Appendix 3A: Characteristics of the *p*-chart and the *c*-chart in Case K

The characteristics of the *p*-chart and the *c*-chart in Case K are important because it

(i) helps us understand the operation and the performance of the charts in the simplest of cases (when the parameters are known), and

(ii) provide us with benchmark values that we can use to determine the effect of estimating the parameters on the operation and the performance of the charts in Case U (when the parameters are unknown).

We compute and examine the characteristics of the *p*-chart and that of the *c*-chart in two different sections. For each chart we give an example that shows

(i) the calculations that are needed to implement the chart, and

(ii) how to determine the chart's characteristics via its run-length distribution.

Each example is followed by a general discussion of the results which were obtained from an analysis of the chart's in-control (IC) and out-of-control (OOC) characteristics listed in Tables 3.1 and 3.2, respectively.

To the author's knowledge none of the standard textbooks and/or articles currently available in the literature give such a detailed elucidation of the *p*-chart's or the *c*-chart's characteristics as is done here.

3.4.1 The *p*-chart in Case K: An example

We first look at an example of a *p*-chart in Case K in order to illustrate the typical application of the chart. In other words, we investigate the properties of the chart for a specific combination of p_0 (the specified value of p) and n while varying $0 < p < 1$ (the true fraction nonconforming). In later sections the performance of the chart is then further studied by considering multiple (various) combinations of p_0 and n.

Example A1: A Case K *p*-chart

Assume that the sample size $n = 50$ and suppose that the true fraction nonconforming p is known or specified to be $p_0 = 0.2$. The upper control limit, the centerline and the lower control limit are then set at

$$UCL_p = 0.2 + 3\sqrt{0.2(0.8)/50} = 0.3697 \qquad CL_p = 0.20 \qquad LCL_p = 0.2 - 3\sqrt{0.2(0.8)/50} = 0.0303.$$

Twelve X_i values (or counts) that were simulated from a $Bin(50,0.2)$ distribution are shown in Table A3.1. Without any loss of generality these counts may be regarded as the number of nonconforming items in twelve independent random samples each of size 50 from a process with a true fraction nonconforming of 0.2. The corresponding sample fraction nonconforming $p_i = X_i / 50$ for each sample is also shown; these are the charting statistics of our *p*-chart.

The *p*-chart is shown in Figure A3.1. The chart displays the two control limits (UCL and LCL), the centerline (CL), and the sample fraction nonconforming p_i from each sample. Because none of the points plot outside the limits we continue to monitor the process. Once a point does plot outside the limits the charting procedure will stop and a search for assignable causes (i.e. additional and/or unwanted sources of variation) will begin.

Table A3.1: Data for the *p*-chart in Case K

Sample number / Time: i	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Counts: X_i	12	8	6	14	8	12	9	7	13	16	11	8
Sample fraction nonconforming: $p_i = X_i / 50$	0.24	0.16	0.12	0.28	0.16	0.24	0.18	0.14	0.26	0.32	0.22	0.16

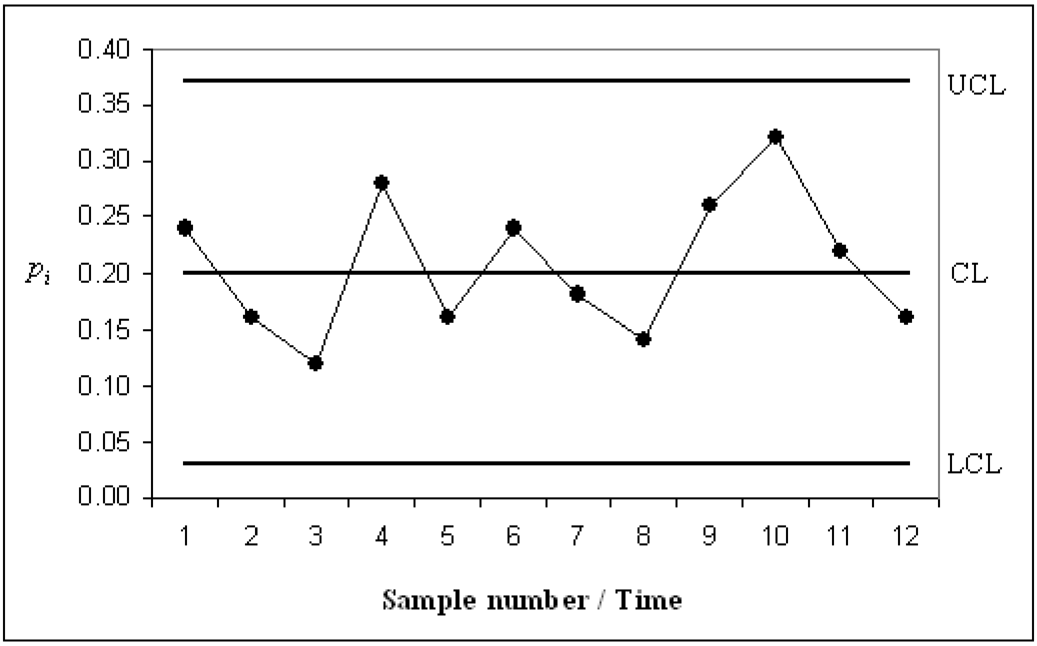


Figure A3.1: A *p*-chart in Case K

Given the operation of the chart it is natural to ask 'How long before the chart signals?' or 'What is the probability for a point to plot between or outside the control limits?' etc. These performance based questions are relevant while the process remains in-control and even more so when a shift occurs.

To answer these questions we study the run-length distribution of the chart. The run-length distribution, as mentioned earlier in section 3.1.1, is characterized entirely by the probability of a no-signal $\beta(p, p_0, n)$ or, equivalently, the probability of a signal $1-\beta(p, p_0, n)$ (see e.g. Table 3.1).

Our starting point when analyzing the performance of the chart is therefore to find the probability of a no-signal. Once we have the probability of a no-signal both the in-control and the out-of-control properties of the *p*-chart in Case K are easily found.

Performance of the *p*-chart

For our particular combination of the parameters i.e. $n = 50$ and $p_0 = 0.2$, the control limits are set at $UCL_p = 0.3697$ and $LCL_p = 0.0303$, so that we proceed as follows to find the probability of a no-signal:

First, we calculate the two charting constants a and b defined in (3-5), which gives

$a = [nLCL_p] = [(50)(0.0303)] = [1.52] = 1$ and $b = \min\{[nUCL_p], n\} = \min\{[18.49], 50\} = 18$.

Using (3-4) shows that the probability of a no-signal is

$$\beta(p, p_0 = 0.2, 50) = I_p(1,48) - I_p(18,31) \quad \text{for} \quad 0 < p < 1$$

so that substituting values for p allow us to study the in-control (when $p = 0.2$) and the out-of-control (when $p \neq 0.2$) properties and performance of the chart.

In-control properties

While the true fraction nonconforming p remains constant and equal to $p_0 = 0.2$ we have the in-control scenario. The probability of a no-signal is then

$$\beta(p = 0.2, 0.2, 50) = I_{0.2}(1,48) - I_{0.2}(18,31) = 0.9973$$

and the probability of a signal, or the false alarm rate, is

$$FAR(0.2, 50) = 1 - \beta(0.2, 0.2, 50) = 0.0027.$$

The in-control run-length distribution is therefore geometric with probability of success 0.0027, which we write as $N_0 \sim Geo(0.0027)$.

Expressions (3-15) and (3-16) in Table 3.1 show that the in-control ARL and the in-control $SDRL$ can be calculated as

$$ARL_0 = 1/(0.0027) = 370.4 \quad \text{and} \quad SDRL_0 = \sqrt{0.9973}/(0.0027) = 369.9$$

respectively.

An in-control *ARL* of 370.4 means that while the process remains in-control we could expect the chart to issue a false alarm or an erroneous signal (on average) every 370th sample. However, with the large standard deviation of 369.9 we could expect a phase (or cycle) during which the chart signals frequently i.e. many false signals occurring one after the other within a relatively short period of time, which is then followed by a phase where the chart hardly ever signals.

Out-of-control properties

When the true fraction nonconforming changes it implies that p is no longer equal to its specified value of $p_0 = 0.2$ and then we deal with the out-of-control case. Since p can increase or decrease we consider both situations.

Increase in p : Upward shift

Suppose that p increases by 12.5% from 0.2 to 0.225. The probability of a no-signal of 0.9973 then becomes

$$\beta(p = 0.225, 0.2, 50) = I_{0.225}(1,48) - I_{0.225}(18,31) = 0.0097$$

so that the probability of a signal becomes $1-\beta(0.225, 0.2, 50) = 0.9903$.

Assuming that the change in p is permanent so that all future samples that we collect come from a process with a fraction nonconforming equal to $p = 0.225$, the out-of-control run-length distribution of the *p*-chart is $N_1 \sim Geo(0.9903)$. The out-of-control average run-length is then calculated using (3-15) as

$$ARL_1 = 1/(0.9903) = 1.01$$

and implies that (on average) we could expect the chart to signal on approximately the 1st sample following an increase from 0.2 to 0.225. The out-of-control *SDRL* of the run-length distribution is

$$SDRL_1 = \sqrt{0.0097}\,/(0.9903) = 0.01$$

and is calculated using (3-16).

Decrease in p : Downward shift

Suppose that the true fraction nonconforming permanently decreased by 25% from 0.2 to 0.15. The probability of a no-signal then changes from its in-control value of 0.9973 to

$$\beta(p = 0.15, 0.2, 50) = I_{0.15}(1,48) - I_{0.15}(18,31) = 0.003$$

so that the out-of-control run-length distribution is geometric with probability of success equal to the probability of a signal $1-\beta(0.15, 0.2, 50) = 0.997$. We could thus expect the chart to signal (on average) on the 1st sample following the change (decrease) in p.

The OC-curve

The OC-curve is the probability of a no-signal $\beta(p, p_0, n)$ plotted as a function of p for a known (specified) value of p_0 and a given (selected) sample size *n*.

The OC-curve for $n = 50$ and $p_0 = 0.2$, that is, $\beta(p, p_0 = 0.2, n = 50)$ for $0 < p \leq 0.55$ is displayed in Figure A3.2. The probability of a signal $1-\beta(p, p_0 = 0.2, n = 50)$ as a function of p is also shown. These two probabilities are plotted on the vertical axis for a given value of p on the horizontal axis. Table A3.2 displays values of the OC and the probability of a signal for selected values of $p = 0.025(0.025)0.550$; it also shows the average run-length and the standard deviation of the run-length associated with each value of p.

Examining Figure A3.2 we begin at the in-control value of $p = 0.2$ where the probability of a no-signal is $\beta(0.2,0.2,50) = 0.9973$ and the probability of a signal is equal to $1-\beta(0.2,0.2,50) = 0.0027$; these two points are indicated on the graphs. We observe that:

(i) As we move in either direction away from $p = 0.2$ (i.e. either to the left or to the right) the probability of a no-signal, in general, decreases whereas the probability of a signal, in general, increases.

This indicates that as p changes (moves away) from the known or specified value of 0.2 the likelihood of a signal that the process is out-of-control increases. We can therefore expect the chart to signal more often (sooner) when the process is out-of-control than when the process is in-control; which is good and confirms that using a control chart is an effective tool in detecting changes in a process.

(ii) The values of $\beta(p,0.2,50)$ and $1-\beta(p,0.2,50)$ vary between zero and one, and happens since both functions compute a probability.

In particular, as the process moves further out-of-control the probability of a no-signal approaches zero whereas the probability of a signal approaches one.

(iii) Neither the probability of a no-signal nor the probability of a signal is symmetric about $p = 0.2$.

This implies, for example, that the rate at which $\beta(p,0.2,50)$ changes as p decreases or increases (i.e. moves to the left or to the right away from 0.2) is not the same. A decrease and an increase of 10% (say) in p from 0.2 to 0.18 and 0.22 (respectively) would therefore not result in the same decrease in $\beta(p,0.2,50)$. The same is true for the probability of a signal and happens since the binomial (50,0.2) distribution is not symmetric.

(iv) As p decreases from 0.2 the probability of a no-signal increases slightly until it reaches a maximum and then decreases (as mentioned in (i)). Similarly, the probability of a signal first decreases a little as p decreases until it reaches a minimum and then it increases again.

This tendency is also seen in Table A3.2. For instance, at $p = 0.2$ we have $\beta(p = 0.2,0.2,50) = 0.9973$ which is less than the probability of a no-signal at $p = 0.175$ of $\beta(p = 0.175,0.2,50) = 0.9988$. For a detailed discussion on this phenomenon see e.g. Acosta-Mejia (1999).

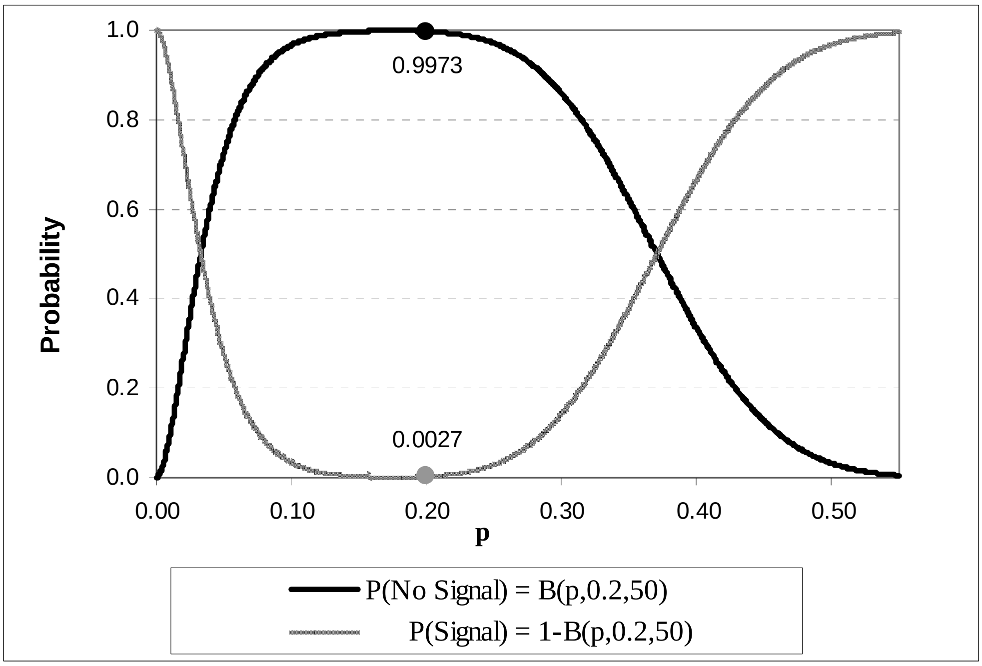


Figure A3.2: The OC-curve and the probability of a signal as a function of p when $p_0 = 0.2$ and $n = 50$

Table A3.2: The Probability of a no-signal, the Probability of a signal, the *ARL* and the *SDRL* for $p = 0.025(0.025)0.550$ when $p_0 = 0.2$ and $n = 50$

p	Pr(No Signal | process OOC)	Pr(Signal | process OOC)	*ARL*	*SDRL*
0.025	0.3565	0.6435	1.55	0.93
0.050	0.7206	0.2794	3.58	3.04
0.075	0.8975	0.1025	9.76	9.24
0.100	0.9662	0.0338	29.60	29.09
0.125	0.9897	0.0103	97.42	96.92
0.150	0.9970	0.0030	337.26	336.76
0.175	0.9988	0.0012	802.13	801.63
0.200	0.9973	0.0027	369.84	369.34
0.225	0.9903	0.0097	103.13	102.63
0.250	0.9713	0.0287	34.79	34.29
0.275	0.9306	0.0694	14.42	13.91
0.300	0.8594	0.1406	7.11	6.60
0.325	0.7544	0.2456	4.07	3.54
0.350	0.6216	0.3784	2.64	2.08
0.375	0.4758	0.5242	1.91	1.32
0.400	0.3356	0.6644	1.51	0.87
0.425	0.2167	0.7833	1.28	0.59
0.450	0.1273	0.8727	1.15	0.41
0.475	0.0678	0.9322	1.07	0.28
0.500	0.0325	0.9675	1.03	0.19
0.525	0.0139	0.9861	1.01	0.12
0.550	0.0053	0.9947	1.01	0.07

Average run-length

The average run-length is the expected number of samples that must be collected before the chart signals.

To quickly detect changes in a process it is desirable that the average run-length

$$ARL(p, p_0, n) = 1/(1-\beta(p, p_0, n))$$

is at its maximum when the process is in-control i.e. when $p = p_0$. This is not always the case for the *p*-chart. For a *p*-chart based on a charting statistic that has a (positively) skewed distribution such as the $Bin(50, 0.2)$ distribution the value of $ARL(p, p_0 = 0.2, n = 50)$ increases initially as p decreases; this causes the *p*-chart to have poor performance in detecting small to moderate decreases in p.

Figure A3.3 displays the average run-length $ARL(p, 0.2, 50)$ as a function of p for $0.05 \leq p \leq 0.3$. The value of $ARL(p, 0.2, 50)$ is plotted on the vertical axis for a specific value of p on the horizontal axis. The average run-length is much higher for values of p slightly less than 0.2 than at 0.2 i.e. the point that indicates the in-control average run-length of 369.84. In particular, at $p = 0.175$ the average run-length is 802.13 (see e.g. Table A3.2).

This phenomenon, as mentioned before, is caused by the skewness of the binomial distribution and the smaller the value of p the greater the skewness and the larger the problem. For a detailed discussion on this phenomenon see e.g. Acosta-Mejia (1999).

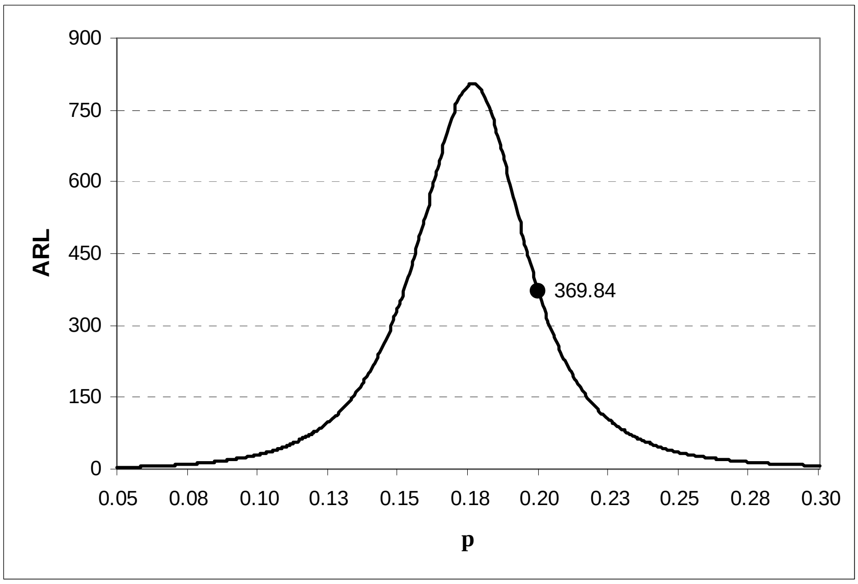


Figure A3.3: The average run-length (*ARL*) as a function of p when $n = 50$ and $p_0 = 0.2$

Run-length distributions

In Case K the in-control and the out-of-control run-length distributions are both geometric (see e.g. Table 3.1).

A graphical display of the in-control and the out-of-control run-length distributions is useful since it helps us (better) see the effect of a change in the process parameter on the entire run-length distribution.

We consider two types of displays: Boxplot-like graphs and probability mass functions (p.m.f's). The former (visually) reveals more about the change in the run-length distribution than do the p.m.f's.

Boxplot-like graphs

Figure A3.4 shows boxplot-like graphs (i.e. the minimum value is replaced by the 1st percentile of the run-length distribution and the maximum value is replaced by the 99th percentile of the run-length distribution) of the in-control as well as the out-of-control run-length distributions. Figure A3.4 is accompanied by Table A3.3 which summarizes some of the properties of the in-control and the out-of-control run-length distributions.

Studying Figure A3.4 and Table A3.3 we note that:

(i) The run-length distributions are severely positively skewed i.e. the spread (variation) in the upper 25% of the distribution between the 75th percentile (or Q_3) and the 99th percentile, is much larger than the spread in the lower 25% of the distribution between the 1st percentile and the 25th percentile (or Q_1).

The skewness of the run-length distribution is confirmed by the fact that the average run-length (indicated by the diamond symbol) is larger than the median run-length (indicated by the circle) in all three the distributions. The exact numerical values of the average run-lengths and the median run-lengths are also indicated. The skewness follows from the fact that the run-length distributions are geometric.

(ii) The run-length distribution is considerably altered following a process change.

Compare, for example, the boxplot-like graph associated with the run-length distribution of the out-of-control process (when $p = 0.225$) to the boxplot-like graph of the in-control run-length distribution (when $p = 0.2$). In particular we see that both the average run-length of 103.1 and the median run-length of 72 of the out-of-control run-length distribution is far less than the average run-length of 369.8 and the median run-length of 257 associated with the in-control run-length distribution. A comparison of the percentiles and the standard deviation of the run-length leads to the same conclusion.

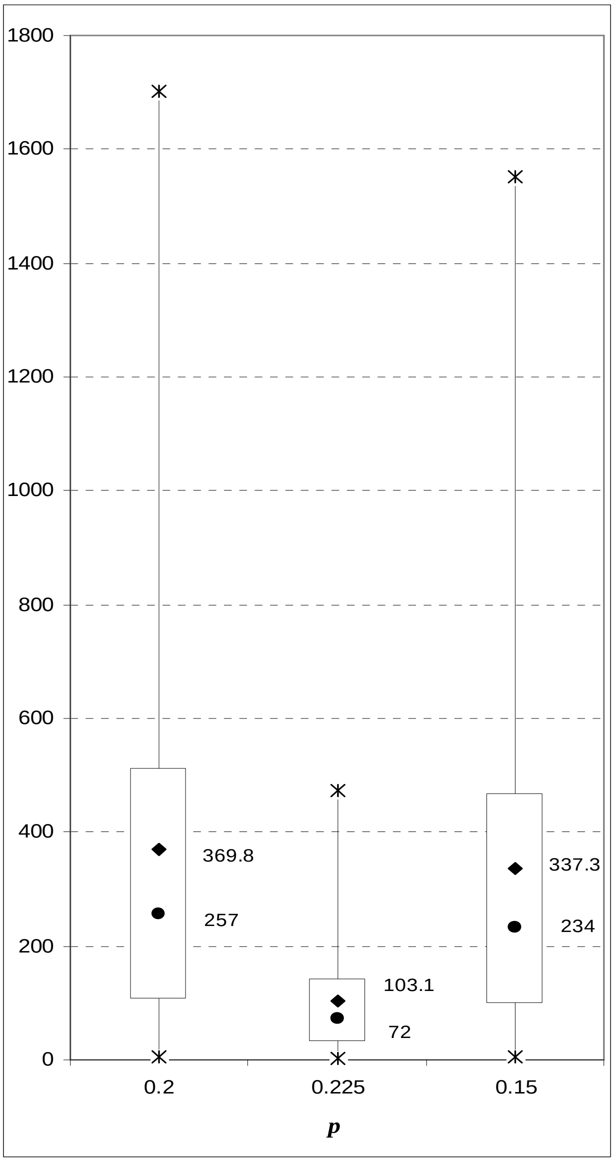


Figure A3.4: Boxplot-like graphs of the in-control and the out-of-control run-length distributions of the *p*-chart in Case K

Table A3.3: Summary measures of the in-control (IC) and the out-of-control (OOC) run-length distributions of the *p*-chart when $n = 50$ and $p_0 = 0.2$ in Case K

	IC	OOC (increase in *p*)	OOC (decrease in *p*)
p	**0.2**	**0.225**	**0.15**
Pr(No Signal)	0.9973	0.9903	0.997
Pr(Signal)	0.0027	0.0097	0.003
ARL	369.84	103.13	337.26
SDRL	369.34	102.63	336.76
1st percentile	4	2	4
5th percentile	19	6	18
10th percentile	39	11	36
25th (Q_1)	107	30	97
50th (*MDRL*)	257	72	234
75th (Q_3)	513	143	467
90th percentile	851	237	776
95th percentile	1107	308	1009
99th percentile	1701	473	1551

Probability mass functions

Studying the p.m.f's of the run-length distributions is another way to look at the effect of a change in the process on the performance of the chart.

Figure A3.5 displays the p.m.f's of the in-control and the out-of-control run-length distributions, that is,

$$\Pr(N_0 = j;0.2,0.2,50) = 0.9973^{j-1}0.0027 \quad \text{and} \quad \Pr(N_1 = j;0.15,0.2,50) = 0.003^{j-1}0.997$$

for $j = 1,2,....$. The former is the in-control p.m.f and the latter the out-of-control p.m.f which corresponds to a decrease by 25% in the fraction of nonconforming p from 0.2 to 0.15.

For values of j less than approximately 370 the likelihood of obtaining these shorter run-lengths is larger following a decrease in the fraction non-conforming. We can write this as $\Pr(N_1 = j) > \Pr(N_0 = j)$ for $j < 370$. The converse also holds, that is, for values of j larger than approximately 370 we see that $\Pr(N_1 = j) < \Pr(N_0 = j)$. This means that the *p*-chart will signal sooner when the process moves out-of-control than when it is in-control; which is good.

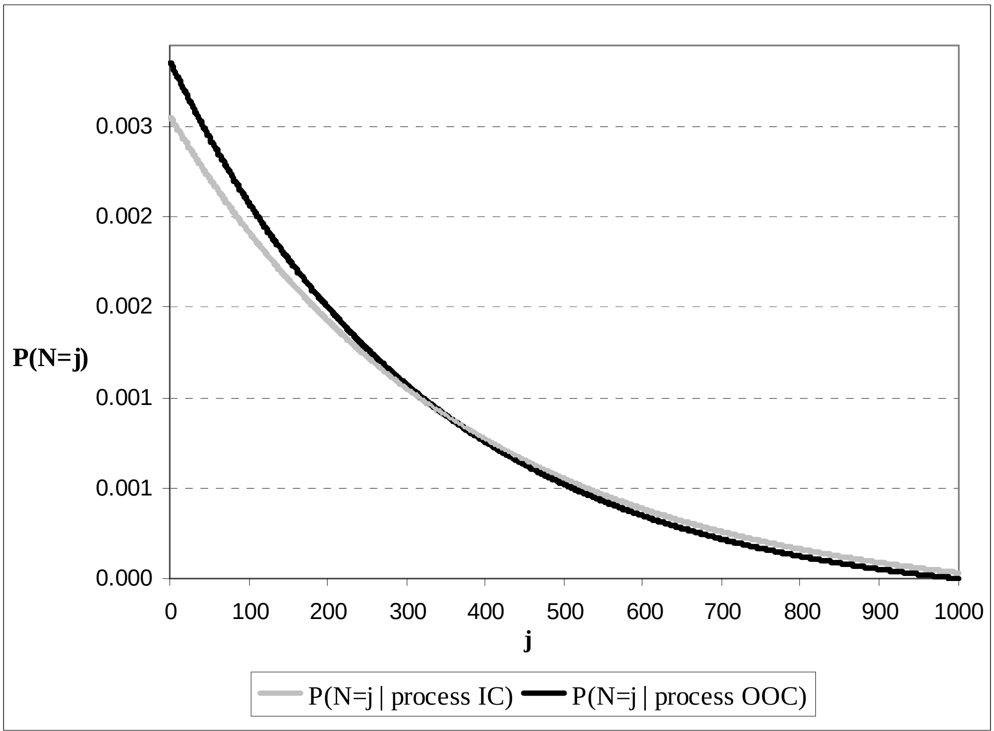


Figure A3.5: The probability distributions of N_0 (when $p = p_0 = 0.2$) and N_1 (when $p = 0.2$ with $p_0 = 0.15$)[1]

[1] *Note*: Instead of displaying the usual histograms, the tops of the bars of the histograms have been joined to better display the shapes of these distributions, and the bars of the histograms have been deleted.

3.4.2 The *p*-chart in Case K: Characteristics of the in-control run-length distribution

The preceding example focused on only one particular combination of n and p_0 i.e. $n = 50$ and $p_0 = 0.2$. Other combinations of n and p_0 are also of interest and gives us an idea of the *p*-chart's performance over a wider range of the parameters.

The false alarm rate and the average run-length are two well-known characteristics of the run-length distribution and most often used to measure a chart's performance. More recently other characteristics of the run-length distribution, such as the standard deviation and the percentiles (quartiles), have also been used and supplemented the false alarm rate and the average run-length.

We study all the abovementioned performance measures for the *p*-chart.

3.4.2.1 False alarm rate

The false alarm rate (FAR) is the probability of a signal when the process is truly in-control and is given by $1-\beta(p = p_0, p_0, n)$ where $\beta(p = p_0, p_0, n)$ is found from (3-4). We can calculate the FAR by substituting different combinations of values for n and p_0 into $1-\beta(p = p_0, p_0, n)$.

Table A3.4 lists the FAR-values (rounded to 4 decimal places) for $p_0 = 0.01$, 0.025, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.40 and 0.50 when the sample size $n =$ 1(1)10, 12, 15(5)30, 40, 50, 75, 100, 125, 150(50)300, 375, 500, 750, 1000 and 1500.

For some combinations of n and p_0 (especially for small values of n and large values of p_0) we observe that the false alarm rate is zero. Although we typically expect (desire) a small false alarm rate, zero is not practical since all moments (such as the average and the standard deviation) of the run-length distribution will be undefined (see e.g. Tables A3.5 and A3.6).

Table A3.4: The false alarm rate (*FAR*) of the p-chart as a function of the sample size n and the known or the specified true fraction nonconforming p_0 in Case K

Sample size n	The known or the specified true fraction nonconforming p_0									
	0.01	**0.025**	**0.05**	**0.10**	**0.15**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**
1	0.0100	0.0250	0.0500	0.1000	0.0	0.0	0.0	0.0	0. 0	0.0
2	0.0199	0.0494	0.0025	0.0100	0.0225	0.0	0.0	0.0	0.0	0.0
3	0.0297	0.0731	0.0073	0.0280	0.0034	0.0080	0.0156	0.0	0.0	0.0
4	0.0394	0.0036	0.0140	0.0037	0.0120	0.0016	0.0039	0.0081	0.0	0.0
5	0.0490	0.0059	0.0226	0.0086	0.0022	0.0067	0.0010	0.0024	0.0	0.0
6	0.0585	0.0088	0.0328	0.0158	0.0059	0.0016	0.0046	0.0007	0.0041	0.0
7	0.0679	0.0121	0.0038	0.0027	0.0121	0.0047	0.0013	0.0038	0.0016	0.0
8	0.0773	0.0158	0.0058	0.0050	0.0029	0.0104	0.0042	0.0013	0.0007	0.0
9	0.0865	0.0200	0.0084	0.0083	0.0056	0.0031	0.0013	0.0043	0.0003	0.0039
10	0.0043	0.0246	0.0115	0.0128	0.0099	0.0064	0.0035	0.0016	0.0017	0.0020
12	0.0062	0.0349	0.0196	0.0043	0.0046	0.0039	0.0028	0.0017	0.0028	0.0005
15	0.0096	0.0057	0.0055	0.0127	0.0036	0.0042	0.0042	0.0037	0.0024	0.0010
20	0.0169	0.0130	0.0159	0.0024	0.0059	0.0026	0.0039	0.0013	0.0021	0.0026
25	0.0258	0.0238	0.0072	0.0095	0.0021	0.0056	0.0034	0.0019	0.0016	0.0041
30	0.0361	0.0064	0.0033	0.0078	0.0029	0.0031	0.0029	0.0024	0.0012	0.0014
40	0.0075	0.0174	0.0034	0.0051	0.0043	0.0031	0.0019	0.0030	0.0018	0.0022
50	0.0138	0.0081	0.0032	0.0032	0.0019	0.0027	0.0031	0.0031	0.0021	0.0026
60	0.0224	0.0039	0.0028	0.0057	0.0024	0.0022	0.0017	0.0029	0.0022	0.0027
75	0.0069	0.0113	0.0041	0.0027	0.0028	0.0025	0.0036	0.0024	0.0030	0.0024
100	0.0184	0.0037	0.0043	0.0049	0.0034	0.0040	0.0038	0.0031	0.0029	0.0035
125	0.0087	0.0043	0.0040	0.0032	0.0031	0.0026	0.0029	0.0033	0.0025	0.0022
150	0.0042	0.0047	0.0036	0.0020	0.0030	0.0031	0.0025	0.0032	0.0034	0.0024
200	0.0043	0.0048	0.0027	0.0034	0.0022	0.0035	0.0025	0.0026	0.0030	0.0023
250	0.0040	0.0046	0.0042	0.0024	0.0027	0.0034	0.0021	0.0030	0.0024	0.0029
300	0.0036	0.0041	0.0027	0.0030	0.0028	0.0031	0.0027	0.0030	0.0026	0.0032
375	0.0051	0.0034	0.0031	0.0035	0.0032	0.0024	0.0029	0.0023	0.0026	0.0027
500	0.0052	0.0047	0.0032	0.0023	0.0033	0.0030	0.0023	0.0029	0.0030	0.0027
750	0.0044	0.0031	0.0027	0.0029	0.0030	0.0030	0.0024	0.0028	0.0025	0.0024
1000	0.0033	0.0036	0.0030	0.0027	0.0030	0.0030	0.0024	0.0027	0.0027	0.0026
1500	0.0034	0.0031	0.0026	0.0030	0.0027	0.0027	0.0026	0.0026	0.0029	0.0025

Figure A3.6 displays the *FAR*-values for $n = 10, 25$ and 50 on the vertical axis for selected values of p_0 on the horizontal axis. Also shown is the nominal *FAR* of 0.0027, which is the *FAR* on a 3-sigma Shewhart X-bar control chart when the charting statistics follow a normal distribution.

Figure A3.6 shows that for small values of p_0 the *FAR* of the *p*-chart is considerably larger than the nominal value of 0.0027. For larger values of p_0 (or, values nearer to 0.5) the *FAR* is closer to the nominal of 0.0027 but still not equal. This illustrates that even for known values of the true fraction nonconforming there is no guarantee that the *FAR* of the *p*-chart will be equal to the nominal 0.0027.

There are two reasons for these discrepancies:

(i) when p is small the normal approximation to the binomial distribution is poor so both the charting constant $k = 3$ and the charting formula (mean ± 3 standard deviations) may be inaccurate, and

(ii) due to the discrete nature of the binomial distribution only certain *FAR* values can be attained.

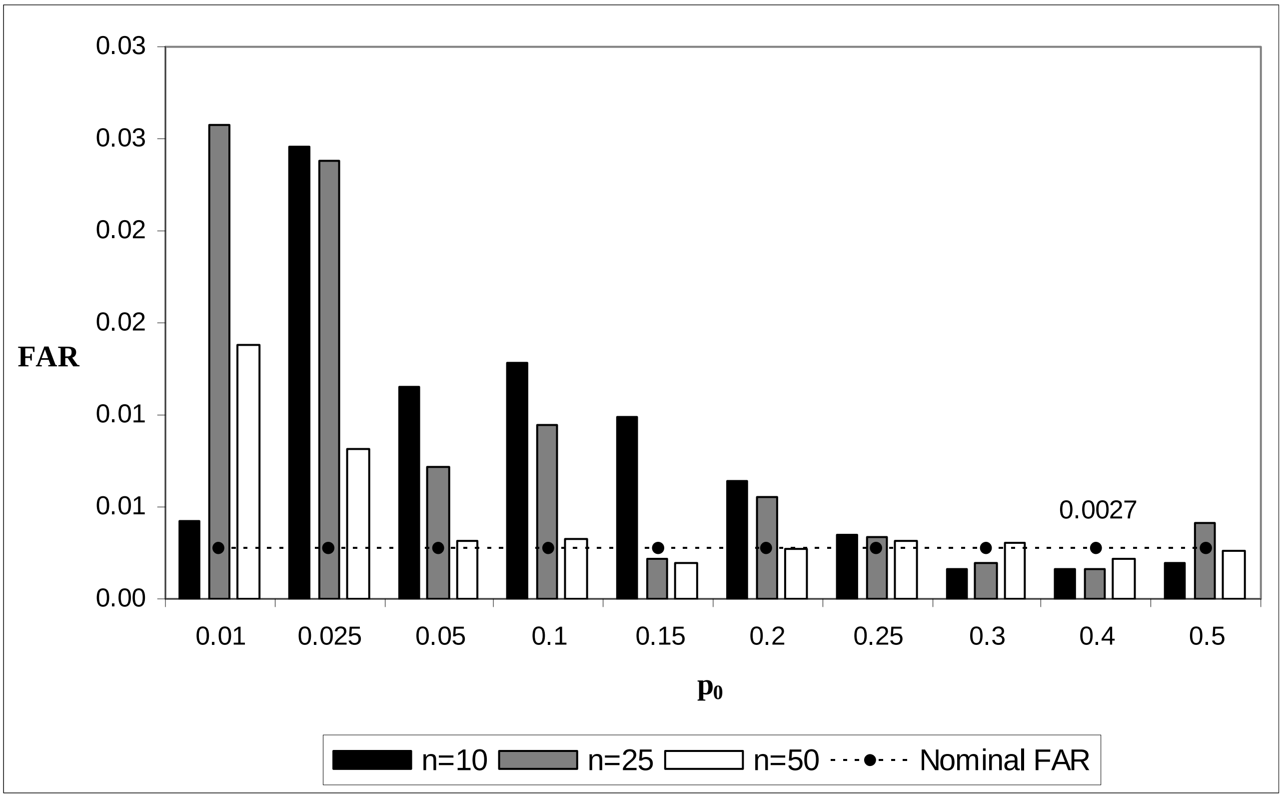


Figure A3.6: The false alarm rate (*FAR*) of the *p*-chart for $n = 10, 25$ and 50 when $p_0 = 0.01, 0.025, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4$ and 0.5 in Case K compared to the nominal *FAR* of 0.0027

3.4.2.2 Average run-length and standard deviation of the run-length

The average run-length (ARL) is the expected value or the mean of the run-length distribution and is equal to the reciprocal of the probability of a signal, that is,

$$ARL(p, p_0, n) = 1/(1-\beta(p, p_0, n)).$$

The in-control ARL is found by replacing p with p_0 in $ARL(p, p_0, n)$ and is the reciprocal of the false alarm rate, that is,

$$ARL_0 = ARL(p_0, p_0, n) = 1/(1-\beta(p_0, p_0, n)) = 1/FAR.$$

The ARL is a measure of how fast (or slow) the control chart signals and is therefore typically used for out-of-control performance comparisons of the charts.

Since the geometric distribution is (severely) positively skewed the ARL becomes questionable as the sole metric for a chart's performance and we therefore need to look at the standard deviation of the run-length ($SDRL$) too.

The $SDRL$ measures the variation or the spread in the run-length distribution and is given by

$$SDRL(p, p_0, n) = \sqrt{\beta(p, p_0, n)}/(1-\beta(p, p_0, n)).$$

The in-control $SDRL$ is found by substituting p_0 for p in $SDRL(p, p_0, n)$ which gives

$$SDRL_0 = SDRL(p_0, p_0, n) = \sqrt{\beta(p_0, p_0, n)}/(1-\beta(p_0, p_0, n)) = \sqrt{1-FAR}/FAR;$$

this shows that the $SDRL_0$ is (like the ARL_0) a function of the FAR.

The values of the ARL_0 and the $SDRL_0$ that correspond to the FAR-values of Table A3.4 are shown in Tables A3.5 and A3.6 (rounded to 2 decimal places), respectively. We can also calculate the ARL_0 and the $SDRL_0$ for different combinations of n and p_0 not shown in Tables A3.5 and Table A3.6 and is carried out by direct evaluation of expressions (3-15) and (3-16).

For example, to find the in-control ARL and the in-control $SDRL$ when $p = 0.25$, $p_0 = 0.25$, and $n = 11$ we proceed as follow:

We first calculate the control limits. These are given by (3-1) as

$$UCL_p = 0.25 + 3\sqrt{0.25(0.75)/11} = 0.6417 \quad \text{and} \quad LCL_p = 0.25 - 3\sqrt{0.25(0.75)/11} = -0.1417.$$

Then (3-5) shows that $b = \min\{[7.0584], 11\} = 7$. The constant a need not be calculated since the lower control limits is negative, that is, $LCL_p < 0$. Using (3-4) we find that the probability of a no-signal is $\beta(0.25, 0.25, 11) = 1 - I_{0.25}(7,3) = 0.9988$ so that the false alarm rate is

$$FAR(0.25, 0.25, 7) = 1 - \beta(0.25, 0.25, 11) = I_{0.25}(7,3) = 0.0012.$$

The in-control ARL is therefore $ARL_0 = (1 - 0.9988)^{-1} = 841.6$ and the in-control standard deviation is $SDRL_0 = \sqrt{0.9988}/(0.0012) = 846.1$.

The calculations for the out-of-control ARL and the out-of-control $SDRL$ are similar; we simply replace p in $\beta(p, 0.25, 11) = 1 - I_p(7,3)$ with a value other than $p_0 = 0.25$ and proceed along the same lines.

Table A3.5: The in-control average run-length (ARL_0) of the p-chart as a function of the sample size n and the known or the specified true fraction nonconforming p_0 in Case K

Sample size n	The known (specified) true fraction nonconforming p_0									
	0.01	**0.025**	**0.05**	**0.10**	**0.15**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**
1	100.00	40.00	20.00	10.00	∞	∞	∞	∞	∞	∞
2	50.25	20.25	400.00	100.00	44.44	∞	∞	∞	∞	∞
3	33.67	13.67	137.93	35.71	296.30	125.00	64.00	∞	∞	∞
4	25.38	275.77	71.33	270.27	83.46	625.00	256.00	123.46	∞	∞
5	20.40	168.26	44.26	116.82	448.93	148.81	1024.00	411.52	∞	∞
6	17.09	114.06	30.51	63.09	169.92	625.00	215.58	1371.74	244.14	∞
7	14.72	82.84	266.17	366.57	82.62	214.04	744.73	263.80	610.35	∞
8	12.94	63.17	172.76	199.03	350.40	96.09	236.59	775.00	1525.88	∞
9	11.56	49.96	119.60	120.03	177.66	326.12	744.73	233.05	3814.70	256.00
10	234.40	40.63	86.93	78.15	101.28	157.00	285.25	628.78	596.05	512.00
12	161.96	28.63	51.10	230.98	215.44	256.20	359.52	591.14	355.85	2048.00
15	103.84	176.24	182.91	78.61	277.35	235.86	238.49	273.78	417.02	1024.00
20	59.31	77.19	62.89	419.10	168.89	385.38	253.67	781.93	468.26	388.07
25	38.82	41.96	139.57	105.53	467.01	180.02	296.70	522.87	611.72	245.26
30	27.66	157.04	304.65	128.47	339.86	321.44	341.52	410.34	854.91	698.86
40	133.38	57.31	294.82	197.51	231.84	325.83	539.81	331.42	550.59	450.16
50	72.37	122.96	313.64	310.57	512.93	369.84	320.92	323.37	469.25	384.29
60	44.60	259.52	351.05	176.03	411.27	446.91	585.24	347.13	457.45	374.47
75	144.51	88.38	242.82	368.47	351.24	404.72	280.73	424.38	336.52	409.13
100	54.42	270.11	233.96	203.98	294.90	250.93	265.00	324.31	344.84	284.28
125	114.61	230.59	248.37	312.50	322.82	392.14	349.00	303.11	405.93	449.14
150	237.46	212.87	277.54	488.03	329.49	325.75	398.29	313.45	293.42	415.71
200	232.80	206.23	370.42	294.04	449.57	284.28	401.99	389.85	333.58	438.70
250	248.43	219.07	240.23	415.64	376.32	291.56	467.00	338.68	424.89	347.38
300	277.57	244.39	365.86	335.28	354.65	324.53	373.71	330.57	384.63	315.53
375	197.63	296.17	325.93	284.05	314.43	413.51	343.76	431.32	381.72	370.96
500	192.01	213.20	316.36	429.94	306.11	328.38	434.37	345.98	336.29	370.81
750	227.35	323.23	367.35	343.32	329.48	332.59	418.21	358.28	397.20	413.68
1000	300.16	279.22	327.92	370.18	331.16	330.18	410.94	374.21	374.59	378.00
1500	297.89	323.23	384.88	332.36	370.33	372.32	385.02	389.48	345.82	398.62

It is straightforward to show using (3-15) and (3-16) that $SDRL = \sqrt{ARL(ARL-1)}$ and implies that the standard deviation is always less than the average run-length i.e. $SDRL < ARL$, and holds whether the process is in-control or out-of-control.

This relationship between the $SDRL$ and the ARL is clearly visible from Tables A3.5 and A3.6. For example, for $n=5$ and $p_0 = 0.025$ the in-control ARL equals 168.26 whereas the in-control $SDRL$ equals 167.67. We also looked at this relationship between the $SDRL$ and the ARL of the run-length distribution in Case U when the process parameters are unknown.

Note that, as mentioned before, the in-control average run-length in Table A3.5 and the in-control standard deviation of the run-length in Table A3.6 are undefined for the same combinations of n and p_0 for which the false alarm rate in Table A3.4 is zero. This is undesirable and shows that for some combinations of n and p_0 the p-chart would not perform satisfactorily in practice.

Table A3.6: The in-control standard deviation of the run-length ($SDRL_0$) of the p-chart as a function of the sample size n and the known or the specified fraction nonconforming p_0 in Case K

Sample size n	The known or the specified fraction nonconforming p_0									
	0.01	0.025	0.05	0.10	0.15	0.20	0.25	0.30	0.40	0.50
1	99.50	39.50	19.49	9.49	∞	∞	∞	∞	∞	∞
2	49.75	19.75	399.50	99.50	43.94	∞	∞	∞	∞	∞
3	33.17	13.16	137.43	35.21	295.80	124.50	63.50	∞	∞	∞
4	24.87	275.27	70.83	269.77	82.96	624.50	255.50	122.96	∞	∞
5	19.90	167.76	43.76	116.32	448.43	148.31	1023.50	411.02	∞	∞
6	16.58	113.56	30.01	62.59	169.42	624.50	215.08	1371.24	243.64	∞
7	14.21	82.34	265.67	366.07	82.12	213.54	744.23	263.30	609.85	∞
8	12.43	62.67	172.26	198.53	349.90	95.59	236.09	774.50	1525.38	∞
9	11.05	49.45	119.10	119.53	177.16	325.62	744.23	232.55	3814.20	255.50
10	233.90	40.13	86.43	77.65	100.77	156.50	284.75	628.28	595.55	511.50
12	161.45	28.13	50.60	230.48	214.94	255.70	359.02	590.64	355.35	2047.50
15	103.34	175.74	182.41	78.11	276.85	235.36	237.99	273.28	416.52	1023.50
20	58.81	76.69	62.39	418.60	168.39	384.88	253.17	781.43	467.76	387.57
25	38.32	41.45	139.07	105.02	466.51	179.52	296.20	522.37	611.22	244.76
30	27.16	156.53	304.15	127.97	339.36	320.93	341.02	409.84	854.41	698.36
40	132.88	56.81	294.32	197.01	231.34	325.33	539.31	330.92	550.09	449.66
50	71.87	122.46	313.14	310.07	512.43	369.34	320.42	322.87	468.75	383.79
60	44.10	259.02	350.55	175.52	410.77	446.41	584.74	346.63	456.95	373.97
75	144.01	87.88	242.32	367.97	350.74	404.22	280.23	423.88	336.02	408.63
100	53.92	269.61	233.46	203.48	294.40	250.43	264.50	323.81	344.34	283.78
125	114.11	230.09	247.87	312.00	322.32	391.64	348.50	302.61	405.43	448.64
150	236.96	212.37	277.03	487.53	328.99	325.25	397.79	312.95	292.92	415.21
200	232.30	205.73	369.92	293.54	449.07	283.78	401.49	389.35	333.08	438.20
250	247.93	218.57	239.72	415.14	375.82	291.06	466.50	338.18	424.39	346.88
300	277.07	243.89	365.36	334.78	354.15	324.03	373.21	330.07	384.13	315.03
375	197.13	295.67	325.43	283.55	313.92	413.01	343.26	430.82	381.22	370.46
500	191.51	212.70	315.86	429.44	305.61	327.88	433.87	345.48	335.79	370.31
750	226.85	322.73	366.85	342.82	328.98	332.09	417.71	357.78	396.70	413.18
1000	299.66	278.72	327.42	369.68	330.66	329.68	410.44	373.71	374.09	377.50
1500	297.39	322.73	384.38	331.86	369.83	371.82	384.52	388.98	345.32	398.12

3.4.2.3 Run-length distribution

Figure A3.6 showed the discrepancy between the false alarm rate (FAR) of the *p*-chart in Case K and the nominal FAR of 0.0027 i.e. the FAR associated with a 3-sigma X-bar chart for a normal process. Because the run-length distribution holds more information than the FAR it is instructive to also look at graphs of the run-length distribution of the *p*-chart in Case K compared to the run-length distribution of the 3-sigma Shewhart X-bar chart.

Figure A3.7 displays boxplot-like graphs of the run-length distributions of the *p*-chart in Case K for $n = 10, 25 \text{ and } 50$ when $p_0 = 0.05, 0.1, 0.2, 0.3 \text{ and } 0.5$. Also shown in Figure A3.7 is the boxplot-like graph of the 3-sigma Shewhart X-bar chart, which has a FAR of 0.0027, an in-control ARL of 370.4 and an in-control $SDRL$ of 369.9.

The properties of the 3-sigma Shewhart X-bar chart are the nominally expected values for a 3-sigma chart such as the *p*-chart. We therefore typically use the performance characteristics of the X-bar chart as benchmark values for that of the *p*-chart (or any other Shewhart-type chart) in Case K.

Table A3.7 accompanies Figure A3.7 and shows the false alarm rate (FAR), the average run-length (ARL), the standard deviation of the run-length ($SDRL$) as well as the 1st, the 5th, the 10th, the 25th, the 50th, the 75th, the 95th and the 99th percentiles of all the run-length distributions displayed in Figure A3.7. The 25th percentile is the 1st quartile (typically denoted by Q_1), the 50th percentile is the 2nd quartile (also denoted by Q_2 and called the median run-length, or simply the $MDRL$), whereas the 75th percentile is the 3rd quartile (in some cases denoted by Q_3). These percentiles are all important descriptive statistics. For example, the inter-quartile range (IQR) is calculated as the difference between the 3rd and 1st quartiles, that is, $IQR = Q_3 - Q_1$. The IQR measures the spread of the middle 50% in the run-length distribution. The median run-length ($MDRL$) is a robust measure of the central tendency (location) of the run-length distribution and sometimes preferred instead of the average run-length.

All the abovementioned characteristics of the *p*-chart were computed using expressions (3-12) through (3-17) in Table 3.1. The properties of the 3-sigma Shewhart X-bar chart were calculated using expressions available in the literature (see e.g. Chakraborti, (2000)).

We assume that the 1st percentile is the minimum possible run-length and that the 99th percentile is the maximum achievable run-length and therefore compute the range (R) of the run-length distribution as the difference between the 99th percentile and the 1st percentile, that is, $R = \max - \min =$ 99th percentile – 1st percentile.

Figure A3.7 shows that for $n = 10$ and $n = 25$ the run-length distribution of the *p*-chart is much different from that of the X-bar chart. For example, for $p_0 = 0.05$, 0.1 and 0.2 the ARL and the $SDRL$ are both far less than the ARL of 370.4 and the $SDRL$ of 369.9 of the X-bar chart (see e.g. Table A3.7). The range of the run-length distributions are also less. For $p_0 = 0.3$ and 0.5 the converse holds. In other words, the ARL, the $SDRL$ and the range of the run-length distribution of the *p*-chart are all larger than what we would nominally expect from a 3-sigma Shewhart-type control chart.

For $n = 50$, the run-length distribution is more like that of the X-bar chart in that the ARL is approximately equal to 370.4, the $SDRL$ is almost 369.9 and the range of the run-length distribution is close to being between 4 (the 1st percentile of the X-bar chart) and 1704 (the 99th percentile of the X-bar chart). However, the run-length distribution is still not exactly the same. This shows that even if the true fraction nonconforming is specified (known) and n is large, the *p*-chart still does not perform as (nominally) expected.

Table A3.7: Properties of the in-control (IC) run-length distribution of the *p*-chart for $n = 10, 25$ and 50 when $p_0 = 0.05, 0.1, 0.2, 0.3$ and 0.5 in Case K, and that of the 3-sigma Shewhart X-bar chart

					Percentiles / Quartiles								
n	p_0	***FAR***	***ARL***	***SDRL***	**1st**	**5th**	**10th**	**25th (Q_1)**	**50th (*MDRL*)**	**75th (Q_3)**	**90th**	**95th**	**99th**
***n* = 10**	**0.05**	0.0115	86.9	86.4	1	5	10	25	60	120	200	259	399
	0.10	0.0128	78.2	77.7	1	4	9	23	54	108	179	233	358
	0.20	0.0064	157.0	156.5	2	9	17	46	109	217	361	469	721
	0.30	0.0016	628.8	628.3	7	33	67	181	436	871	1447	1883	2894
	0.50	0.0020	512.0	511.5	6	27	54	148	355	710	1178	1533	2356
***n* = 25**	**0.05**	0.0072	139.6	139.1	2	8	15	41	97	193	321	417	641
	0.10	0.0095	105.5	105.0	2	6	12	31	73	146	242	315	484
	0.20	0.0056	180.0	179.5	2	10	19	52	125	249	414	538	827
	0.30	0.0019	522.9	522.4	6	27	56	151	363	725	1203	1565	2406
	0.50	0.0041	245.3	244.8	3	13	26	71	170	340	564	734	1128
***n* = 50**	**0.05**	0.0032	313.6	313.1	4	17	33	91	218	435	722	939	1443
	0.10	0.0032	310.6	310.1	4	16	33	90	215	430	714	929	1428
	0.20	0.0027	369.8	369.3	4	19	39	107	257	513	851	1107	1701
	0.30	0.0031	323.4	322.9	4	17	35	93	224	448	744	968	1487
	0.50	0.0026	384.3	383.8	4	20	41	111	267	533	884	1150	1768
3-sigma X-bar		0.0027	370.4	369.9	4	19	39	107	257	513	852	1109	1704

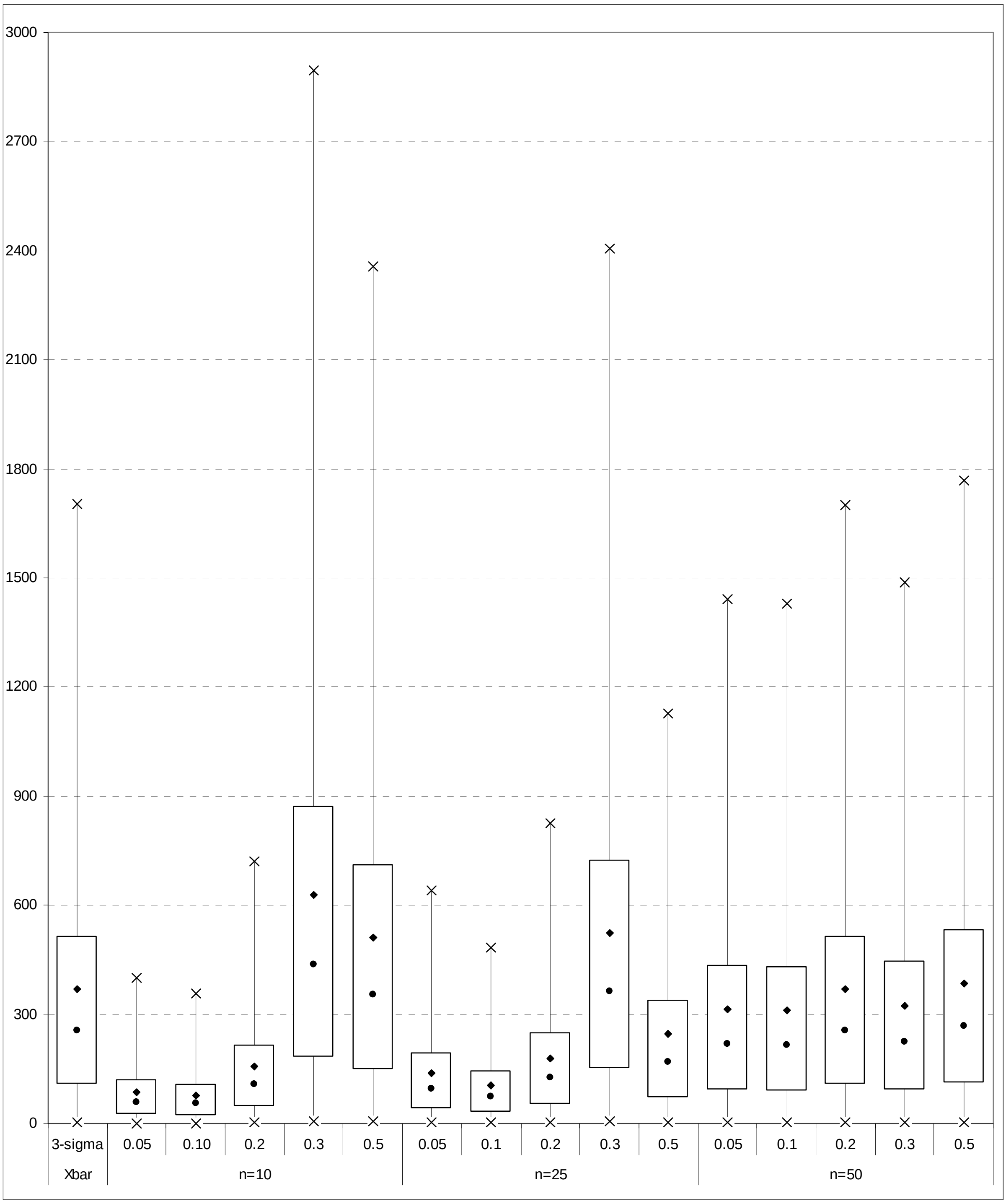


Figure A3.7: Boxplot-like graphs of the in-control (IC) run-length distribution of the p-chart for $n = 10, 25$ and 50 when $p_0 = 0.05, 0.1, 0.2, 0.3$ and 0.5 in Case K compared to the run-length distribution of the 3-sigma Shewhart X-bar chart

The foregoing discussion focused on the performance of the *p*-chart as measured by the false alarm rate, the average run-length, the standard deviation of the run-length and the percentiles of the run-length distribution and compared the *p*-chart's performance to that of the well-known 3-sigma Shewhart X-bar chart. It is also useful and important to know how to design a *p*-chart. The design of the *p*-chart in Case K is therefore looked at next.

3.4.2.4 The OC-curves and *ARL* curves

When designing a *p*-chart in Case K we need to choose a sample size n and while doing so keep in mind the size of the shift we are interested in detecting i.e. by how much the true fraction nonconforming p will differ from its specified value p_0 once a shift occurs.

Choosing the appropriate sample size is typically carried out by looking at a family of OC-curves or a family of *ARL*-curves, which are obtained by plotting multiple (at least two) OC-curves or multiple *ARL*-curves on the same set of axis.

Recall that an OC-curve is a graph (plot) of the probability of a no-signal $\beta(p, p_0, n)$ on the vertical axis for some values of $0 < p < 1$ on the horizontal axis. Hence, a family of two OC-curves is obtained by plotting $\beta(p, p_0, n = n_1)$ and $\beta(p, p_0, n = n_2)$, where n_1 and n_2 denote two different sample sizes, on the same set of axes; hence, each OC-curve corresponds to a specific sample size (in this case n_1 or n_2) but the value of p_0 is the same for each curve. Similarly, an *ARL*-curve is a graph (plot) of the average run-length $ARL(p, p_0, n)$ on the vertical axis for some values of $0 < p < 1$ on the horizontal axis so that family of two *ARL*-curves is obtained by plotting $ARL(p, p_0, n = n_1)$ and $ARL(p, p_0, n = n_2)$ on the same set of axes.

Suppose that we would like to compare and decide between two control charting plans to monitor the specified fraction nonconforming of $p_0 = 0.5$. Further, assume that both plans use a *p*-chart with 3-sigma control limits; the first plan uses $n = 25$ items per sample whereas the second plan uses double that i.e. $n = 50$; the question is then what the effect of sampling twice as many items is.

To assist us with our choice between the two control charting plans Figure A3.8 shows the OC-curve of each of the control charting plans. In other words, Figure A3.8 shows a family of two OC-curves where $\beta(p, p_0 = 0.5, n = 25)$ and $\beta(p, p_0 = 0.5, n = 50)$ are plotted on the vertical axis versus $0 < p < 1$ on the horizontal axis. In addition, Table A3.8 lists some values of $\beta(p,0.5,25)$ and $\beta(p,0.5,50)$ for values of $p = 0.05(0.05)0.95$.

Figure A3.8 shows that the plan that uses $n = 50$ items per sample has a consistently lower β-risk or OC. Thus, if the objective is to detect a shift in the fraction nonconforming as soon as possible and we can afford the extra cost of sampling, this plan will be preferred. In the language of hypothesis testing, this shows that with all other things being equal, the power of test to detect a shift is higher for a larger sample size.

Figure A3.9 displays a family of two *ARL*-curves which corresponds to the OC-curves of Figure A3.8, that is, Figure A3.9 shows the average run-lengths $ARL(p, p_0 = 0.5, n = 25)$ and $ARL(p, p_0 = 0.5, n = 50)$ as functions of $0 < p < 1$. A decision based on the OC-curves of Figure A3.8 and a decision based on the *ARL*-curves of Figure A3.9 will therefore be exactly the same; this is so since the relationship between the average run-length and the probability of a no-signal is one-to-one and given by $ARL(p, p_0, n) = (1-\beta(p, p_0, n))^{-1}$. Table A3.8 also shows the exact numerical values of $ARL(p,0.5,n)$ for $n = 50$ and 25 at $p = 0.05(0.05)0.95$.

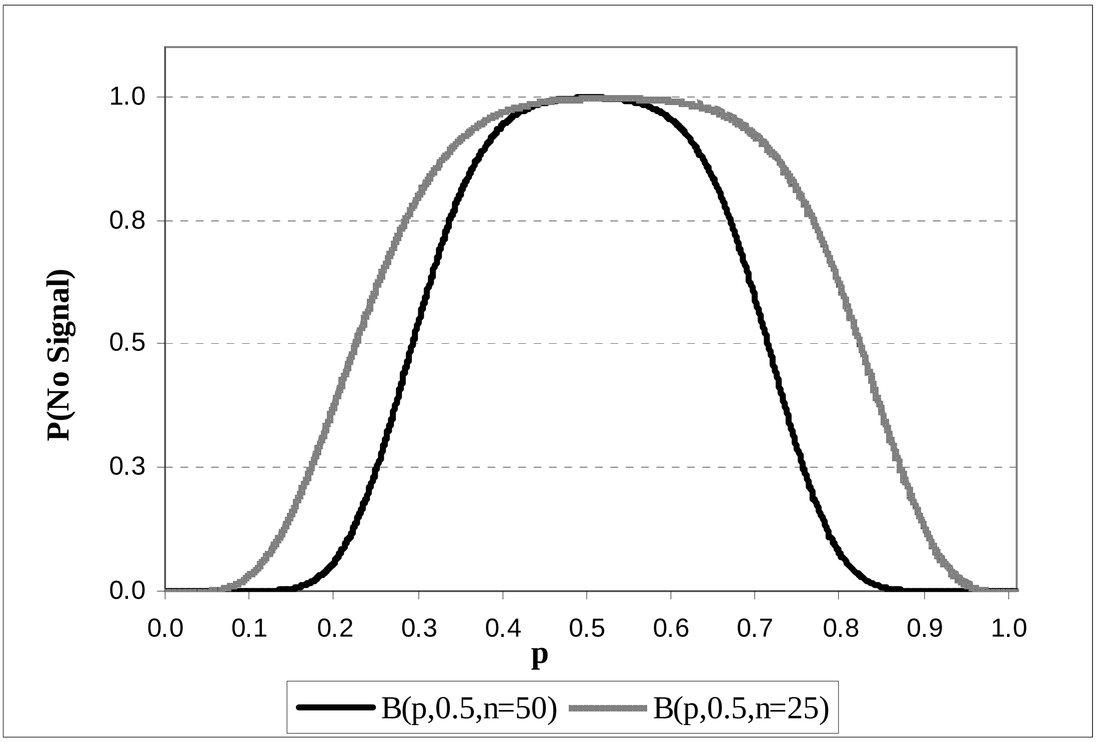


Figure A3.8: Family of Operating Characteristic (OC) Curves for the *p*-chart for a specified fraction nonconforming of $p_0 = 0.5$ when $n = 50$ and 25

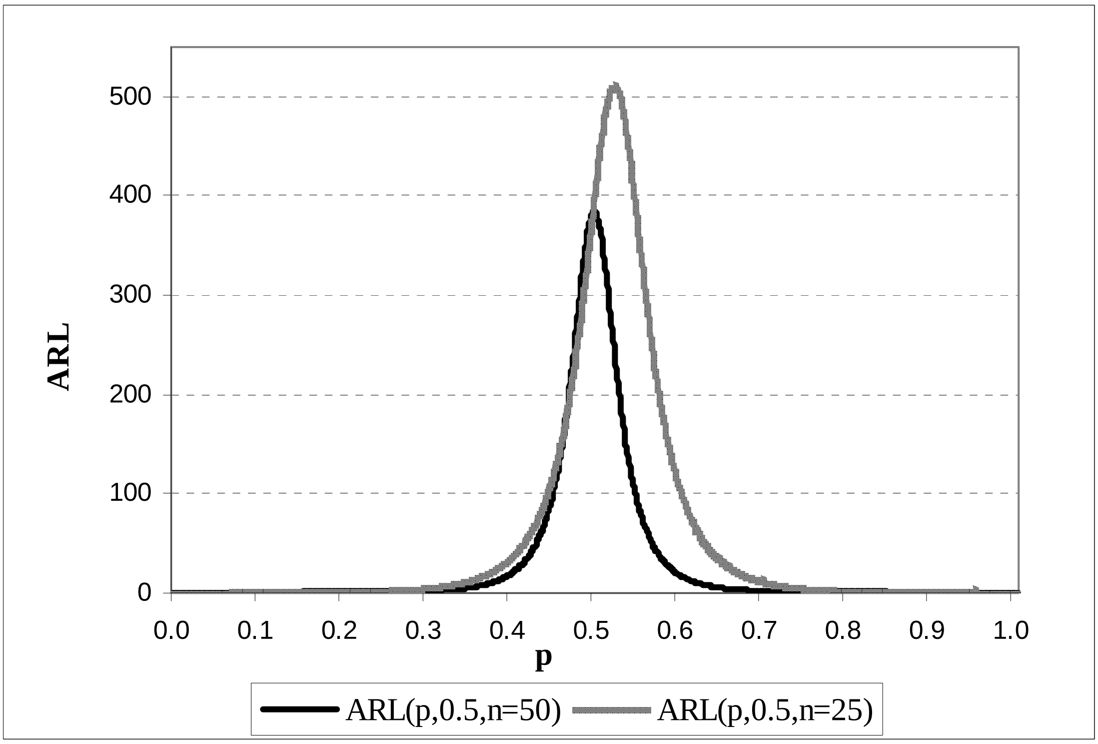


Figure A3.9: Family of Average Run-Length (*ARL*) Curves for the *p*-chart for a specified fraction nonconforming of $p_0 = 0.5$ when $n = 50$ and 25

Table A3.8: The values of the OC and the *ARL* of a *p*-chart n Case K when $p_0 = 0.5$ and $n = 50$ and 25

	Probability of a no-signal / Operating Characteristic (OC)		Average Run-Length (*ARL*)	
p	$\beta(p, 0.5, n = 50)$	$\beta(p, 0.5, n = 25)$	$ARL(p, 0.5, n = 50)$	$ARL(p, 0.5, n = 25)$
0.05	0.0000	0.0012	1.00	1.00
0.10	0.0001	0.0334	1.00	1.03
0.15	0.0053	0.1615	1.01	1.19
0.20	0.0607	0.3833	1.06	1.62
0.25	0.2519	0.6217	1.34	2.64
0.30	0.5532	0.8065	2.24	5.17
0.35	0.8122	0.9174	5.33	12.10
0.40	0.9460	0.9706	18.53	34.05
0.45	0.9895	0.9913	95.37	115.35
0.50	0.9974	0.9975	384.29	400.98
0.55	0.9895	0.9973	95.37	371.83
0.60	0.9460	0.9905	18.53	104.99
0.65	0.8122	0.9679	5.33	31.20
0.70	0.5532	0.9095	2.24	11.05
0.75	0.2519	0.7863	1.34	4.68
0.80	0.0607	0.5793	1.06	2.38
0.85	0.0053	0.3179	1.01	1.47
0.90	0.0001	0.0980	1.00	1.11
0.95	0.0000	0.0072	1.00	1.01

Summary

The *p*-chart is well-known, easy to use and its' applications is based on the implicit assumption that the binomial distribution is well approximated by the normal distribution, which, as one might expect, is not always the case. For example, as the preceding discussion shows, in some cases (especially for small values of n) the FAR is zero which implies that the ARL, the $SDRL$ and other moments are undefined. Moreover, the performance of the *p*-chart with a known or given or specified value for p might not be anything like that of the 3-sigma X-bar chart.

The *p*-chart is used to monitor the fraction nonconforming in a sample. The *c*-chart on the other hand is used to monitor the number of nonconformities in an inspection unit and is based on the Poisson distribution. We study the Case K *c*-chart in the next sections.

3.4.3 The *c*-chart in Case K: An example

We first look at an example of a *c*-chart in Case K to illustrate the typical application of the chart and investigate the characteristics of the chart for a specific value of c_0 (the specified value of c) while varying $c > 0$ (true average number of nonconformities in an inspection). The performance of chart is then further studied in subsequent sections by considering multiple (various) values of c_0.

Example A2: A Case K *c*-chart

Suppose that the true average number of nonconformities in an inspection unit c is known or specified to be $c_0 = 14$. The 3-sigma control limits for the *c*-chart are

$$UCL_c = 14 + 3\sqrt{14} = 25.22 \qquad CL_c = 14 \qquad LCL_c = 14 - 3\sqrt{14} = 2.78$$

and are calculated using (3-2).

Table A3.9 shows ten values simulated from a $Poi(14)$ distribution. We can assume without loss of generality that the values (counts) are the charting statistics of the *c*-chart; we therefore denote them by Y_i for $i = 1,2,...,10$. The *c*-chart is shown in Figure A3.10. The chart displays the upper control limit (*UCL*), the center line (*CL*), the lower control (*LCL*) and the Y_i's from each inspection unit plotted on the vertical axis versus the inspection unit number (time) on the horizontal axis. We see from Figure A3.10 that none of the 10 points plot out-of-control.

As long as no point plots outside the control limits we continue to monitor the process; this involves obtaining independent successive inspection units, calculating the charting statistic (i.e. the number of nonconformities) for each new inspection unit, and then plotting these one at a time on the chart. Once a point plots outside the limits it is taken as evidence that c is no longer equal to its specified value of $c_0 = 14$. A search for assignable causes is then started.

Table A3.9: Data for the *c*-chart in Case K

Inspection unit number / Time: *i*	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Counts: Y_i	17	9	17	12	16	16	9	21	15	11

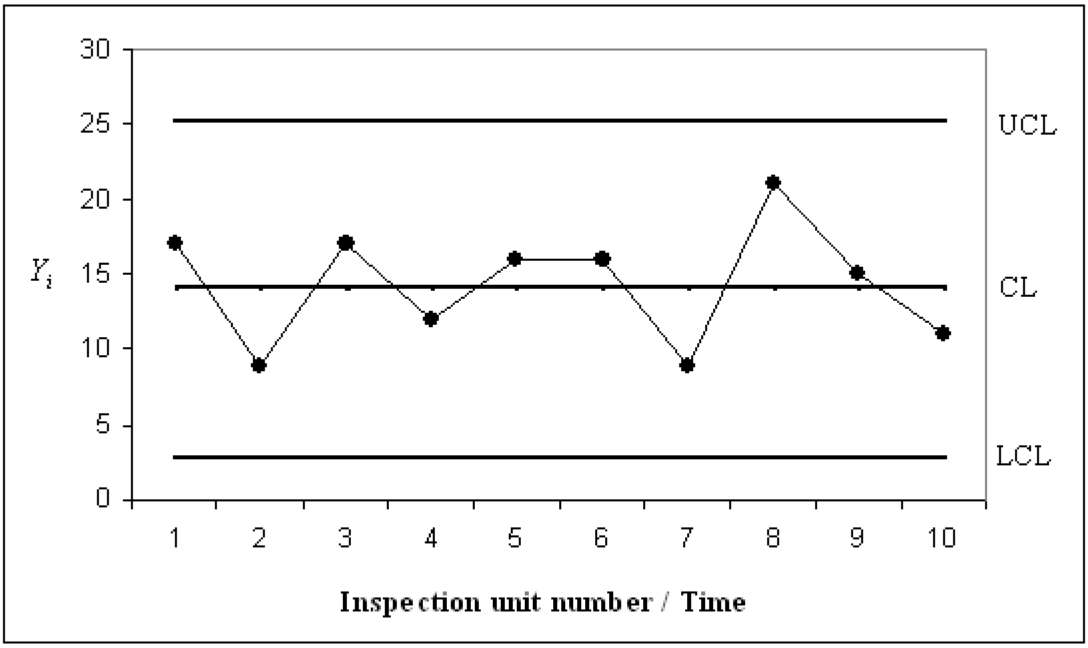


Figure A3.10: A *c*-chart in Case K

Performance of the *c*-chart

To study the performance of the aforementioned *c*-chart we analyze its in-control and out-of-control properties for which we need the probability of a no-signal, or equivalently, the probability of a signal. The probability of a no-signal completely characterizes the run-length distribution of the chart.

For $c_0 = 14$ it was shown that the upper control limit is $UCL_c = 25.22$ and the lower control limit is $LCL_c = 2.78$. Expression (3-8) shows that $d = [2.78] = 2$ and $f = [25.22] = 25$; these constants are needed to calculate the probability of a no-signal. We can study the in-control and the out-of-control performance of the chart by substituting values for c in the probability of a no-signal which is $\beta(c,14) = \Gamma_{26}(c) - \Gamma_3(c)$ and is found using (3-7).

In-control properties

As long as the true average number of nonconformities c remains unchanged and equal to its specified value of $c_0 = 14$ we deal with an in-control process. The probability of a no signal is then

$$\beta(c = 14{,}14) = \Gamma_{26}(14) - \Gamma_3(14) = 0.9973$$

so that the false alarm rate is $FAR = 1 - \beta(14{,}14) = 0.0027$.

The in-control run-length distribution is therefore geometric with probability of success equal to 0.0027, which we write as $N_0 \sim Geo(0.0027)$.

Out-of-control properties

When the true average number of nonconformities in an inspection unit changes, c is no longer equal to $c_0 = 14$ and implies that we have the out-of-control scenario. We look at the scenario when c increases; a decrease in c can be handled in a similar fashion.

Increase in c: Upward shift

Suppose c increases from 14 to 15; this is approximately a 7.14% increase in c. The probability of a no-signal decreases from 0.9973 (when the process was in-control) to

$$\beta(c = 15{,}14) = \Gamma_{26}(15) - \Gamma_3(15) = 0.9938$$

whereas the probability of a signal increases from 0.0027 to $1 - \beta(c = 15{,}14) = 0.0062$. The increase in the probability of a signal is good since the likelihood of detecting the shift increases.

The out-of-control run-length distribution is geometric with probability of success equal to 0.0062. Expression (3-21) shows that the out-of-control average run-length is $ARL_1 = 1/0.0062 = 160.66$. So, if it happens that c increases from 14 to 15 (and stays fixed at 15) one would expect the chart to detect such a shift (and signal) on approximately the 161^{st} sample following the shift.

The OC-curve

The OC-curve and the probability of a signal as functions of c for $0 < c \leq 42$ are shown in Figure A3.11. In addition, Table A3.10 shows values of the probability of a no-signal $\beta(c,14) = \Gamma_{26}(c) - \Gamma_3(c)$ and the probability of a signal $1-\beta(c,14) = 1-\Gamma_{26}(c) + \Gamma_3(c)$ for values of $c = 2(2)42$.

Studying the OC-curve and the probability of a signal as function of c helps us see what the performance of our c-chart would be when a shift occurs. For example, if c was to decrease from $c = 14$ to $c = 8$ (which may be interpreted as an improvement in the process as approximately 42.9% less nonconformities (on average) in an inspection unit will in future be observed) we see from Table A3.10 that $1-\beta(c = 8, c_0 = 14) = 0.0138$ so that the $ARL = 72.70$ and the $SDRL = 72.20$.

Note that, the two curves of Figure A3.11 are very similar to that of the p-chart considered earlier (see e.g. Figure A3.2); this is so because the values of n and p_0 (for the p-chart) and c_0 (in case of the c-chart) is such that the false alarm rate (FAR) of both the charts are 0.0027, and so, the IC run-length distributions of these charts and all other performance measures (including the OOC performance measures) are roughly the same.

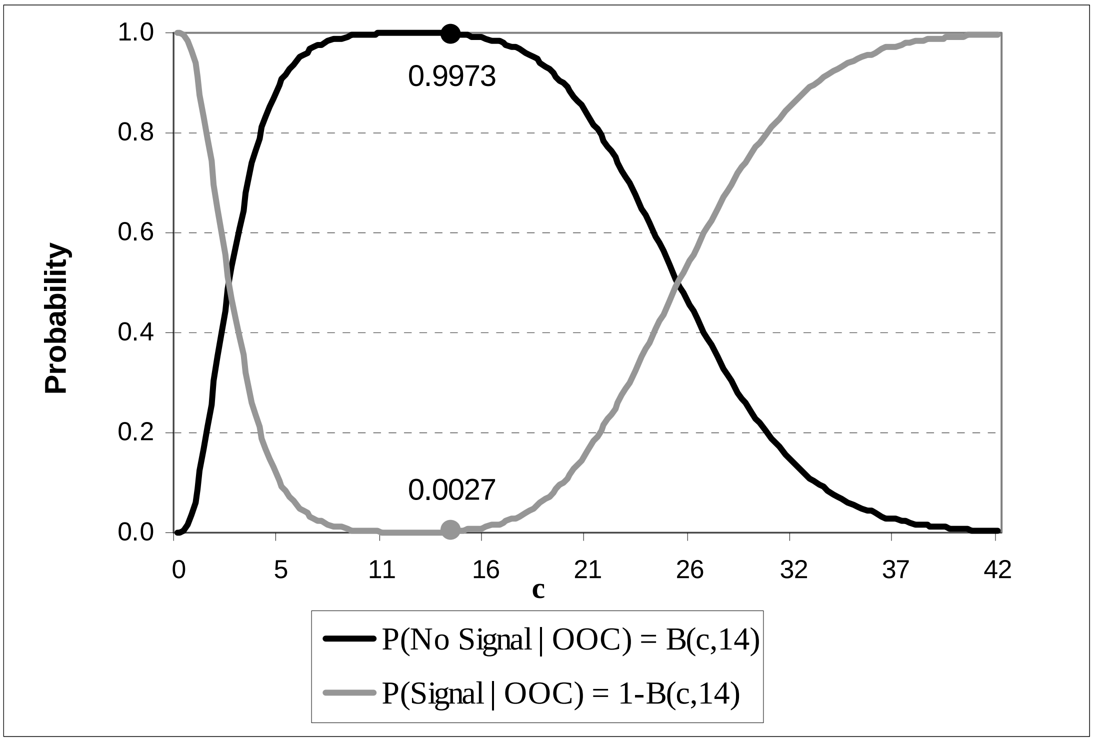


Figure A3.11: The OC-curve and the probability of a signal as a function of c when $c_0 = 14$

Table A3.10: The Probability of a no-signal and the Probability of a signal for $c = 2(2)42$ when $c_0 = 14$

c	P(No Signal | process OOC)	P(Signal | process OOC)	*ARL*	*SDRL*
2	0.3233	0.6767	1.48	0.84
4	0.7619	0.2381	4.20	3.67
6	0.9380	0.0620	16.14	15.63
8	0.9862	0.0138	72.70	72.20
10	0.9972	0.0028	358.80	358.30
12	0.9992	0.0008	1204.80	1204.30
14	0.9973	0.0027	370.16	369.66
16	0.9869	0.0131	76.13	75.63
18	0.9554	0.0446	22.42	21.91
20	0.8878	0.1122	8.91	8.40
22	0.7771	0.2229	4.49	3.95
24	0.6319	0.3681	2.72	2.16
26	0.4739	0.5261	1.90	1.31
28	0.3272	0.6728	1.49	0.85
30	0.2084	0.7916	1.26	0.58
32	0.1228	0.8772	1.14	0.40
34	0.0674	0.9326	1.07	0.28
36	0.0345	0.9655	1.04	0.19
38	0.0166	0.9834	1.02	0.13
40	0.0076	0.9924	1.01	0.09
42	0.0033	0.9967	1.00	0.06

Run-length distributions

Figure A3.12 displays boxplot-like graphs of the in-control and the out-of-control run-length distributions of the *c*-chart with the average run-lengths (*ARL*’s) and the median run-lengths (*MDRL*’s) indicated (the former by diamond symbols and the latter by circles). The exact numerical values of the *ARL*’s and the *MDRL*’s are also shown in Figure A3.12 and listed in Table A3.11 together with the probability of a no-signal, the probability of a signal and some percentiles (quartiles) of the in-control and the out-of-control run-length distributions.

The *ARL* and the *MDRL* measures the central tendency (location) of the run-length distribution. The *MDRL* however is more robust and outlier resistant than the *ARL*. In both the in-control and the out-of-control run-length distributions the *ARL* is larger than the *MDRL* and indicates that the in-control and the out-of-control run-length distributions are non-normal and positively skewed. The skewness of the run-length distributions is also observed by comparing the upper and the lower tails of each of the distributions, that is, the distance between the 99th percentile and the 75th percentile to the distance between the 25th percentile and the 1st percentile; this comparison between the upper and lower tails is done separately for each distribution.

For example, for the in-control run-length distribution (when $c = c_0 = 14$) Table A3.11 shows that the distance between the 99th percentile and the 75th percentile is $1703 - 513 = 1190$ whereas the distance between the 25th and the 1st percentiles of the in-control run-length distribution is $107 - 4 = 103$. The latter is much larger (approximately 1190/103 = 11.5 times) than the former and shows, as mentioned before, that the in-control run-length distribution is positively skewed.

Most importantly however Figure A3.12 shows the overall difference between the in-control ($c = 14$) and the out-of-control ($c = 15$) run-length distributions. For example, the out-of-control average run-length is 160.7 compared to the in-control average run-length of 370.2. Similarly, the out-of-control median run-length is 112 versus the in-control median run-length of 257. Furthermore, the range (*R*) and the inter-quartile range (*IQR*) of the in-control and the out-of-control run-length distributions differ somewhat. Both the range and the inter-quartile range measure the spread (variation) in the run-length distributions. The range measures the overall spread and is the distance between the 99th percentile (maximum) and the 1st percentile (minimum). The *IQR*, on the other hand, is the distance between the 3rd and the 1st quartile, that is, $IQR = Q_3 - Q_1$ and measures the variation in the middle 50% of the distribution. For the in-control run-length distribution Table A3.11 shows that the range of the in-control run-length distribution is $R_0 = 1703 - 4 = 1699$ and that the inter-quartile range of the in-control run-length distribution is $IQR_0 = 513 - 107 = 406$. The values of R_0 and IQR_0 are both larger than that of the out-of-control run-length distribution. For the out-of-control run-length distribution we have that $R_1 = 738 - 2 = 736$ and $IQR_1 = 223 - 47 = 176$. This big discrepancy between the range and the inter-quartile range of the in-control and the out-of-control run-length distributions emphasizes that once a shift occurs, the run-length distribution is severely altered in that we can expect the chart to signal (detect the shift) sooner, which is of course good.

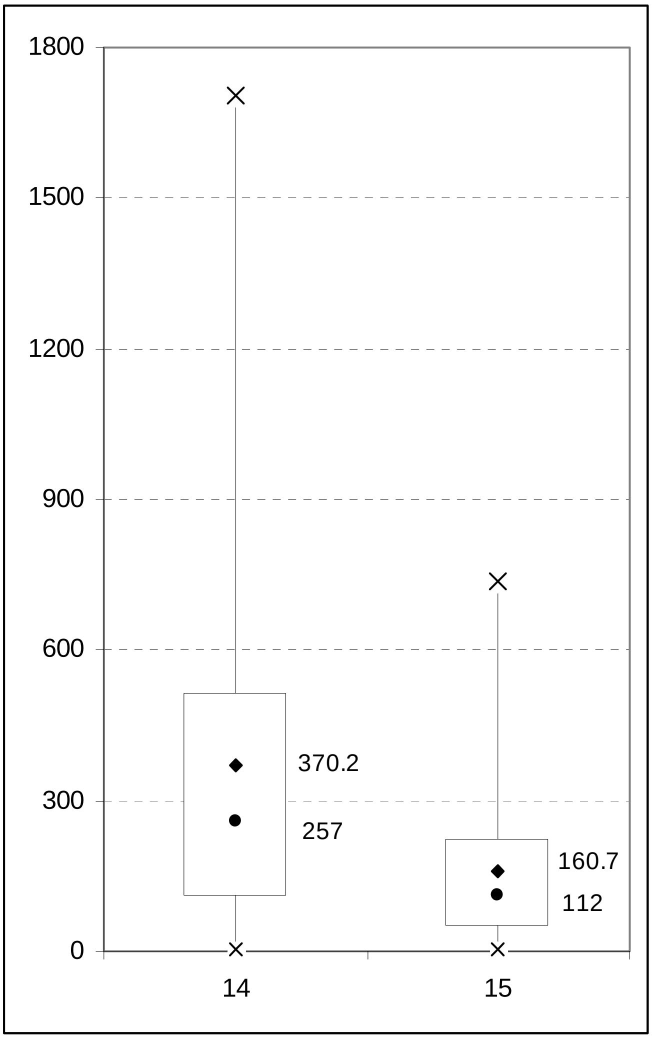


Figure A3.12: Boxplot-like graphs of the in-control (IC) and the out-of-control (OOC) run-length distributions of the *c*-chart in Case K

Table A3.11: Summary measures of the in-control (IC) and the out-of-control (OOC) run-length distributions of the *c*-chart when $c_0 = 14$ in Case K

	IC	OOC (increase in *c*)
c	**14**	**15**
Pr(No Signal)	0.9973	0.9938
Pr(Signal)	0.0027	0.0062
ARL	370.16	160.66
SDRL	369.66	160.16
1st percentile	4	2
5th percentile	19	9
10th percentile	39	17
25th (Q_1)	107	47
50th (*MDRL*)	257	112
75th (Q_3)	513	223
90th percentile	852	369
95th percentile	1108	480
99th percentile	1703	738

The preceding discussion focused on the properties of the *c*-chart for one particular value of c i.e. $c_0 = 14$. Other values of c_0 are also of interest in order to get an idea of the overall performance of the *c*-chart and can only be obtained by studying the characteristics of the *c*-chart for a wider range of values for c_0.

3.4.4 The *c*-chart in Case K: Characteristics of the in-control run-length distribution

To get a better idea of the overall performance of the *c*-chart in Case K we look at the run-length distribution and its characteristics for a range of values for c_0. We consider both small and large values of c_0.

To this end, Table A3.12 shows the control limits (LCL_c and UCL_c), the charting constants d and f, the probability of a no signal when the process is in-control, the false alarm rate (*FAR*), the in-control average run-length (ARL_0), and the in-control standard deviation of the run-length ($SDRL_0$) when $c_0 = 1(1)10(5)50$, 75 and 100, respectively.

Table A3.12 is accompanied by Table A3.13 which shows the percentiles of the run-length distributions of the *c*-chart for the same values of c_0. The values in columns (2) through (9) of Table A3.12 were computed using expressions (3-2), (3-7), (3-8) and the expressions in Table 3.2. The percentiles were calculated using expression (3-23) in Table 3.2.

For illustration purposes, consider the *c*-chart with $c_0 = 35$. Table A3.12 shows that

$$LCL_c = 35 - 3\sqrt{35} = 17.25 \quad \text{and} \quad UCL_c = 35 + 3\sqrt{35} = 52.75$$

so that $d = [17.25] = 17$ and $f = [52.75] = 52$. It thus follows that $\beta(c = 35, c_0 = 35) = 0.9967$, $ARL_0(c = 35, c_0 = 35) = 301.42$ and $SDRL_0(c = 35, c_0 = 35) = 300.92$. In addition Table A3.13 shows that the $MDRL_0 = 209$ and that the $Q_3 = 418$ and the $Q_1 = 87$ so that the $IQR = Q_3 - Q_1 = 402$.

However, note that, since the *FAR* and the *ARL* are most often used in OOC performance comparisons we primarily focus on the *FAR* and the in-control *ARL* in our discussion on the performance of the *c*-chart in Case K. In particular, we compare the *FAR* and the *ARL* of the *c*-chart in Case K to that of the well-known 3-sigma X-bar chart.

Table A3.12: Characteristics of the in-control run-length distribution of the *c*-chart for $c_0 = 1(1)10(5)50$, 75 and 100

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
c_0	LCL_c	UCL_c	d	f	Pr(No Signal | IC)	*FAR*	ARL_0	$SDRL_0$
1	-2.00	4.00	0	3	0.6131	0.3869	2.58	2.02
2	-2.24	6.24	0	6	0.8601	0.1399	7.15	6.63
3	-2.20	8.20	0	8	0.9464	0.0536	18.66	18.15
4	-2.00	10.00	0	9	0.9736	0.0264	37.81	37.31
5	-1.71	11.71	0	11	0.9878	0.0122	82.03	81.53
6	-1.35	13.35	0	13	0.9939	0.0061	163.74	163.24
7	-0.94	14.94	0	14	0.9934	0.0066	150.85	150.35
8	-0.49	16.49	0	16	0.9959	0.0041	246.70	246.20
9	0.00	18.00	0	17	0.9946	0.0054	183.72	183.22
10	0.51	19.49	0	19	0.9965	0.0035	285.74	285.23
15	3.38	26.62	3	26	0.9965	0.0035	283.83	283.33
20	6.58	33.42	6	33	0.9971	0.0029	339.72	339.22
25	10.00	40.00	10	39	0.9960	0.0040	248.14	247.64
30	13.57	46.43	13	46	0.9971	0.0029	349.94	349.44
35	17.25	52.75	17	52	0.9967	0.0033	301.42	300.92
40	21.03	58.97	21	58	0.9964	0.0036	275.36	274.86
45	24.88	65.12	24	65	0.9976	0.0024	413.04	412.54
50	28.79	71.21	28	71	0.9975	0.0025	396.70	396.20
75	49.02	100.98	49	100	0.9967	0.0033	299.77	299.27
100	70.00	130.00	70	129	0.9967	0.0033	307.36	306.86

Table A3.13: Percentiles of the in-control run-length distribution of the *c*-chart for $c_0 = 1(1)10(5)50$, 75 and 100

	Percentiles of the run-length distribution								
c_0	1st	5th	10th	25th (Q_1)	50th (*MDRL*)	75th (Q_3)	90th	95th	99th
1	2	2	2	2	2	3	5	7	10
2	2	2	2	2	5	10	16	20	31
3	2	2	2	6	13	26	42	55	84
4	2	2	4	11	26	52	86	112	172
5	5	5	9	24	57	114	188	245	376
6	2	9	18	47	114	227	376	490	752
7	2	8	16	44	105	209	347	451	693
8	3	13	26	71	171	342	567	738	1134
9	2	10	20	53	128	255	422	549	844
10	3	15	31	83	198	396	657	855	1314
15	3	15	30	82	197	393	653	849	1305
20	4	18	36	98	236	471	782	1017	1563
25	3	13	27	72	172	344	571	742	1141
30	4	18	37	101	243	485	805	1047	1610
35	4	16	32	87	209	418	693	902	1386
40	3	15	29	80	191	382	633	824	1266
45	5	22	44	119	286	572	950	1236	1900
50	4	21	42	114	275	550	913	1187	1825
75	4	16	32	87	208	415	690	897	1379
100	4	16	33	89	213	426	707	920	1414

3.4.4.1 False alarm rate and average run-length

Figure A3.13 shows the percentage difference between the false alarm rate (*FAR*) of the *c*-chart in Case K and the nominal *FAR* of 0.0027 i.e. the *FAR* of a 3-sigma Shewhart X-bar chart. The percentage difference is seen to be mostly positive; only for $c_0 = 45$ and 50 is the percentage difference negative. It is also clear that, in general, the *FAR* is far from 0.0027; especially for small values of c_0 i.e. less than or equal to 10, say.

In particular, Figure A3.13 shows, in general, that (i) the *FAR* is hundreds of percents larger than 0.0027, and (ii) as c_0 increases the percentage difference gets smaller. A *c*-chart based on $c_0 = 6$, for instance, has a *FAR* of 0.0061, which is 126% larger than 0.0027 whereas a *c*-chart based on $c_0 = 10$ has a *FAR* of 0.0035 (which is 30% larger than 0.0027) and a *c*-chart based on $c_0 = 35$ has a *FAR* equal to 0.0033 (which is only 23% larger).

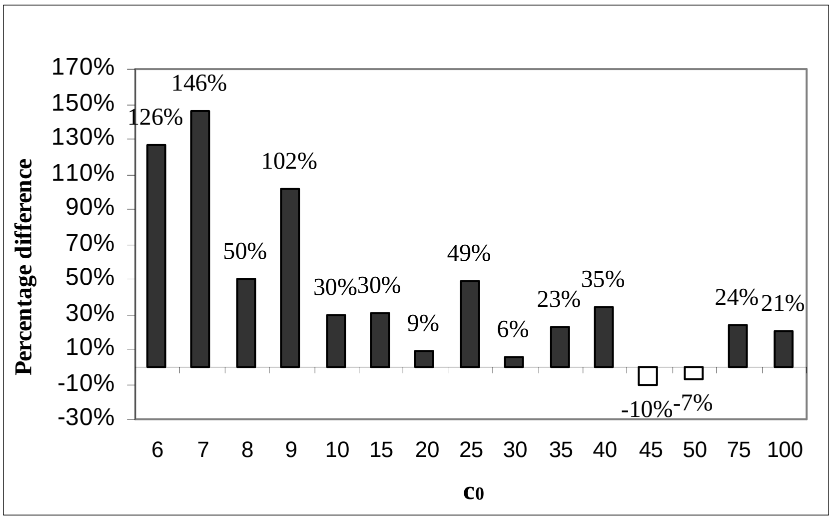


Figure A3.13: Percentage difference between the false alarm rate (*FAR*) of the *c*-chart and the nominal *FAR* of 0.0027 for $c_0 = 6(1)10(5)50$, 75 and 100

Figure A3.14 shows the percentage difference between the average run-length (*ARL*) of the *c*-chart in Case K and that of the nominal *ARL* of 370.4, which is the *ARL* of a 3-sigma Shewhart X-bar chart. The percentage difference is seen to be mostly negative and implies shorter in-control *ARL*’s than nominally expected from a 3-sigma chart like the *c*-chart. Thus, we can deduce that, unless the specified value c_0 of c is reasonably large, the *c*-chart will erroneously signal more often than what is nominally expected from a 3-sigma chart.

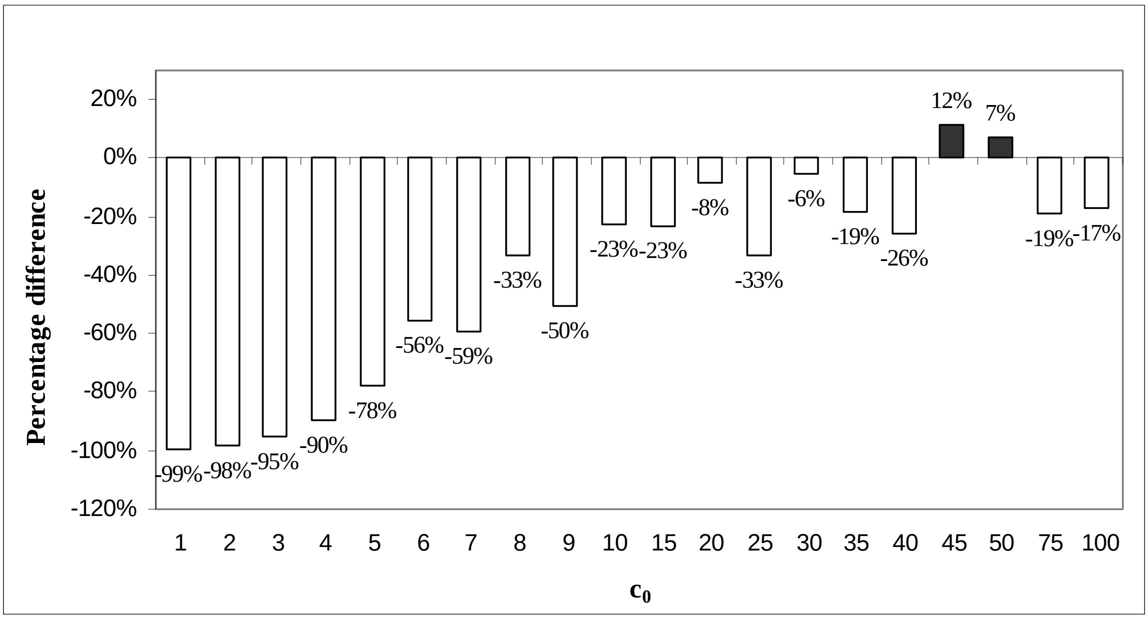


Figure A3.14: Percentage difference between the average run-length (*ARL*) of the *c*-chart and the nominal *ARL* of 370.4 for $c_0 = 1(1)10(5)50$, 75 and 100.

3.4.4.2 The run-length distribution

It is good to make a visual comparison of the run-length distributions since it gives us an overall idea of just how different (or similar) the run-length distribution of the *c*-chart is to that of the 3-sigma X-bar chart. Figure A3.15 displays boxplot-like graphs of the run-length distribution of the *c*-chart when $c_0 = 6(1)10(5)50$ and also shows a boxplot-like graph of the run-length distribution of the 3-sigma X-bar chart.

We see that, in general, for small values of c_0 the run-length distribution of the c-chart differs substantially from that of the 3-sigma X-bar chart in that the ARL_0 and the $MDRL_0$ are considerably smaller and the spread (as measured by the range *R*) in the run-length distribution of the *c*-chart is noticeably less than that of the X-bar chart. Only for larger values of c_0 does the run-length distribution of the c-chart become more like that of the 3-sigma X-bar chart.

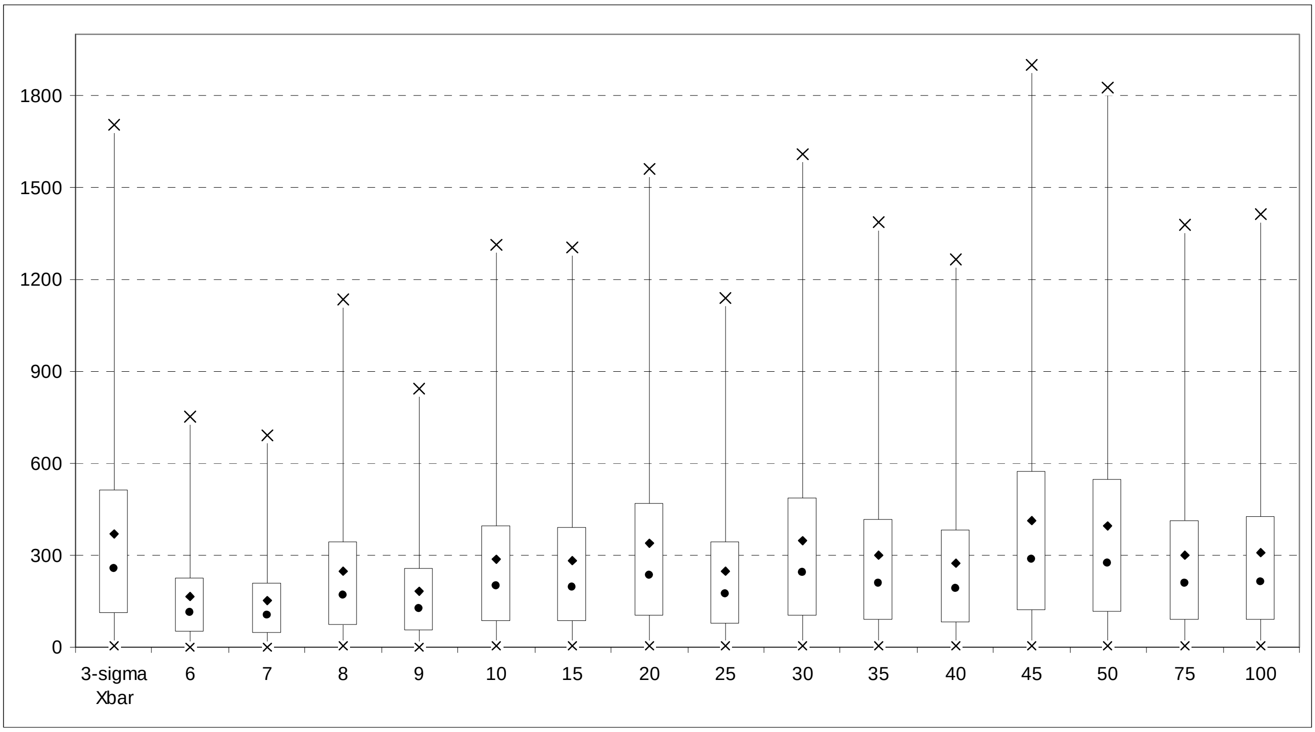


Figure A3.15: Boxplot-like graphs of the in-control (IC) run-length distribution of the *c*-chart for $c_0 = 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 75$ and 100 in Case K compared to the run-length distribution of the 3-sigma Shewhart X-bar chart

Summary

Like the *p*-chart, the *c*-chart is well-known and easy to apply but, even in Case K, the *c*-chart does not perform anything like the 3-sigma Shewhart X-bar chart. The discrepancy is due to the facts that

(i) when *c* is small the normal approximation to the Poisson distribution is poor so both the charting constant $k = 3$ and the charting formula (mean ± 3 standard deviations) may be inaccurate, and

(ii) due to the discrete nature of the Poisson distribution only certain *FAR* values can be attained.