Thailand Statistician
October 2025; 23(4): 732-744
http://statassoc.or.th
Contributed paper

Performance of Phase II Control Chart for Location When One Parameter is Estimated in Terms of Run Length Distribution and Percentiles

Samson Offorma Ugwu* **and Akaninyene Udo Udom**
Department of Statistics, University of Nigeria Nsukka, Enugu-State, Nigeria.
*Corresponding author; e-mail: offorma.ugwu@unn.edu.ng

Received: 17 November 2021
Revised: 1 August 2022
Accepted: 30 September 2022

Abstract

It is well known that the median run length measures chart performance better than the average run-length. Some authors have advocated more representative measures like the percentiles for assessing charts performance and called for an examination of the percentiles of the entire run-length distribution. An earlier work studied the percentiles of Shewhart $\bar{x}$-chart when the process mean and variance are known (Case KK), later, another when the process mean and variance are unknown (Case UU). Here, we consider when only the process mean is unknown (Case UK) and when only the process variance is unknown (Case KU) by evaluating and plotting their exact run-length cumulative distribution functions for some reference samples (m) of size 5 at a given false alarm rate. We compare the results with those for Cases KK and UU. Unlike in Case UU, the cumulative distribution function curves for Case UK for small to moderate m are stochastically ordered relative to that of the geometric distribution and dominance is a function of δ, however, in line with Case UU, in Case KU, the curves cross that for the geometric distribution at some points and for at least $m = 500$ and $n = 5$, the curves for Cases UK and KU converge with that for Cases KK and UU.

Keywords: average run length (ARL), false alarm rate (FAR), median run length (MRL), parameter estimation, stochastic ordering

1. Introduction

Control charts are essential tools in ascertaining if a process is in a state of statistical control or not, (see in Montgomery, 1991, p. 186). The Shewhart $\bar{x}$-chart is one of the most frequently used charts for monitoring and controlling for the mean of a process, Jardim et al. (2019). To design this chart, the process mean and standard deviation are needed and they are either specified or estimated from m phase I samples taken from the process during an in-control state. Therefore, four situations are involved and they are; when both the process mean and variance are specified, when the process mean is specified but the variance is unknown, when the process mean is unspecified but the variance is specified and when both the process mean and variance are unspecified. These are referred to as Cases KK, KU, UK and UU respectively. The performance of this chart like every other chart is an important consideration in application and is always judged in terms of certain measures or characteristics associated with its run-length distribution, Chakraborti (2000). The run-length, a random variable, is the number of subgroups that must be collected or put in another way; the number of plotting statistics that must be plotted until the first signal suggesting a change from in-control process is observed, Chakraborti (2007). It is well known that the run-length distribution is right-skewed and in a right-skewed distribution, median is a better measure of the central tendency. Hence, the median run length (MRL) is a better measure of a chart performance than the average run length (ARL).

When the parameters used in the chart's design are specified or known, that is, the ideal but a rare practical Case KK, a well-known fact is that the run-length distribution of the Shewhart $\bar{x}$-chart is geometric, Chakraborti (2006). This is simply because, in this case, the signaling events are mutually independent. However, in many applications of the Shewhart $\bar{x}$-charts, the process parameters such as the mean and the variance are not known, instead, their estimates are obtained and used in the chart's design. Plugging in these estimates in its design has significant effects on the properties of the chart as the signaling events are no longer independent and hence, the run-length distribution is no longer geometric, Quesenberry (1993). Several works have been done in the literature to study the effect of the estimated parameter(s) on the performance of the Shewhart $\bar{x}$-chart especially as it concerns its probability of signal and the run-length distribution; see Quesenberry (1993), Chen (1997), Chakraborti (2000), Chakraborti (2006) and Goedhart et al. (2016). These studies focused on the marginal distribution of the in-control (IC) run-length, most especially on its expected value, the so-called unconditional in-control average run length (ARL_0), Jardim et al. (2018). In line with what has been said about the run-length distribution, Chakraborti (2007) and some earlier works stated that since the run-length random variable takes on only positive integer values, the distributional shape is significantly right-skewed. Also, Chakraborti (2006) categorically stated that because of the skewed nature of the run-length distribution, one might prefer the median (or some other quantile) and not the mean (ARL) as a measure of typical chart performance. Chakraborti (2007) went further to say that some studies have advocated the use of other more representative measures other than the ARL for the assessment of chart performance, noting that one such measure is the percentile of the run-length distribution. The percentiles provide wider information about a distribution and hence on the performance of a chart not provided by the mean or the expected value. Works exist in the literature on the examination of the percentiles of the run-length distribution of the Shewhart $\bar{x}$-chart. This can be seen in Shmueli and Cohen (2003), Khoo (2004) and Radson and Boyd (2005). However, these papers considered the examination based on the standard known case (Case KK) only. Not only was Case KK considered alone but there is also a lack in the provision of broader details in the examination of the percentiles even as it concerns the in-control and out-of-control states coverage. Chakraborti (2007) at m = 20, 30, 50, 100, 500 and ∞ provided the percentiles of the Shewhart $\bar{x}$-chart when the process mean is unknown but the variance is known referred to as Case UK. The same paper under the same conditions of phase I samples (m) considered the examination when both mean and standard deviation are unknown referred to as Case UU. However, only the 5th, 50th (median), and 95th percentiles were considered in the work at $\delta = 0$, that is, at an in-control state only. As a way of filling this gap, Chakraborti (2007) examined the run-length distribution and demonstrated a more detailed examination of the percentiles at "13" different percentage points beginning from the 5th to the 95th percentiles for the chart in Case UU at different values of m, each of size (n = 5) when the nominal false alarm rate, α = 0.0027 and δ = 0, 0.2, 0.6 and 1.0 with 0 as the in-control state and 0.2, 0.6 and 1.0 as out-of-control states of the respective sizes. Some details of the work will be provided here during the presentation and discussion of the results for this work. Although it was stated in Chakraborti (2007) that the same methodology in the paper can be adapted to obtain similar results for Case UK and Case KU, they were never provided. To the best of our knowledge, these are currently unavailable in the literature. Motivated by this and by the sense of examination of all estimation cases to ensure completeness, the purpose and contribution of this paper, therefore, is to fill this gap by providing a detailed examination of the run-length distribution and the percentiles of Shewhart $\bar{x}$-chart for Cases UK and KU and comparing the results with those in Cases KK and UU as it concerns the chart performance from the c.d.fs and percentiles standpoints.

The remainder of this paper is structured as follows. We present the review of the probability of signal, false alarm rate (FAR), c.d.f, and the percentile distribution of Cases KK in Section 2. In Section 3, we present the derivations of the c.d.fs and the percentile distributions of the Cases UK and KU respectively. The results of the evaluations in simulation of the given and derived expressions will be presented in Section 4 and finally, the conclusion of the work is presented in Section 5.

2. Mean and Standard Deviation Both Known (Case KK)

Suppose that m subgroups (phase I samples) each of size n, say; X_{i1}, X_{i2}, ..., X_{in}, $i = 1, 2, \ldots, m$ are available from an in-control state with the process mean (μ_0) and standard deviation (σ_0), both known or specified. Then, the general expressions for the upper and lower control limits (UCL and LCL) of a $Z_{\alpha/2}$-Sigma Shewhart $\overline{x}$-chart are given by UCL=μ_0+ $Z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}$ and LCL=μ_0- $Z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}$, respectively, where the charting constant, $Z_{\alpha/2}$ = $\Phi^{-1}(1-\alpha/2)$ and $\Phi\,(.)$ is the standard normal cumulative density function. The probability of signal by the chart is given by $1 - P(LCL \le \mu_1 \le UCL)$, which can be expressed as

$$1 - \beta(\delta, n) = 1 - \left\{\Phi\left(Z_{\alpha/2} - \delta\sqrt{n}\right) - \Phi\left(-Z_{\alpha/2} - \delta\sqrt{n}\right)\right\}$$

where $\delta = \frac{(\mu_1 - \mu_0)}{\sigma_0}$, μ_1 is taken to be the mean of any subgroup sample in phase II and $\Phi\,(.)$ remains as already defined in Chakraborti (2000). It is well-known that in Case KK, the run length, N follows a geometric distribution, (see in Montgomery, 2004, p. 190-208). Therefore, the characteristics of the chart here can be inferred by the characteristics of a geometric distribution. Consequently, the in-control average run length, ARL_0 of the chart is the expected value of a geometric distribution defined as

$$\alpha^{-1} \text{ or equivalently } \left[1 - \left\{\Phi\left(Z_{\alpha/2}\right) - \Phi\left(-Z_{\alpha/2}\right)\right\}\right]^{-1} \tag{1}$$

where α is the false alarm rate (the probability of signal at δ=0). The out-of-control average run length, ARL_δ is defined for some values of $\delta \neq 0$ as

$$\left[1 - \left\{\Phi\left(Z_{\alpha/2} - \delta\sqrt{n}\right) - \Phi\left(-Z_{\alpha/2} - \delta\sqrt{n}\right)\right\}\right]^{-1} \tag{2}$$

Run length distribution

Therefore, the cumulative distribution function, c.d.f of the run length, N is defined as

$$\begin{cases} P(N \le s) = 1 - \left\{\Phi\left(Z_{\alpha/2} - \delta\sqrt{n}\right) - \Phi\left(-Z_{\alpha/2} - \delta\sqrt{n}\right)\right\}^s, & \delta = 0 \\ P(N \le s) = 1 - (1-\beta)^s, & s = 1, 2, \ldots \end{cases} \tag{3}$$

where β also is the FAR always set at 0.0027 in many applications. The results of the evaluation of expressions (1) and (2) are presented in Table 1 specifically at the Table section (standards known case) under, ARL.

Percentiles of run length

The $100pth$ $(0 < p < 1)$ percentile is defined as the smallest integer, 's' so that the c.d.f in expression (3) at 's' is at least equal to p, Chakraborti (2007). Therefore, the expressions in (3) can be redefined for this purpose as

$$1 - \left\{\Phi\left(Z_{\alpha/2} - \delta\sqrt{n}\right) - \Phi\left(-Z_{\alpha/2} - \delta\sqrt{n}\right)\right\}^s \ge p \text{ or } 1 - (1-\beta)^s \ge p, \tag{4}$$

where $s = 1, 2, \ldots$,and can be used to study the statistical properties of the control chart, including the various performance characteristics. Since the first part of the expressions in (3) can easily be evaluated at different values of δ, it will be used in this work for the evaluations for all possible values of p within the range. The results of the evaluations at different values of δ and p are presented in Table1 for comparison with other cases. It was also evaluated graphically at $\delta = 0$ and 0.5 for all possible values of p as the plots (standard known case plots) and included in all the figures for comparison.

3. Process Mean Unknown but the Standard Deviation is Known (Case UK)

Suppose that the process mean (μ_0) is unknown but that the standard deviation (σ_0) is known. Then, the mean (μ_0) is estimated by the grand mean ($\bar{\bar{X}}$) from the m subgroup means and the control limits of the Shewhart $\bar{x}$-chart becomes; $\bar{\bar{X}} \pm Z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}$. Quesenberry (1993) and Chakraborti (2000) stated that once estimated control limits are used, the properties of the control chart fundamentally change as the run-length distribution, N is no longer geometric. However, conditionally on $\bar{\bar{X}}$, the random variable N has a geometric distribution with probability of a success (a signal) given as; $1-\beta(\delta, n\ \bar{\bar{X}})$ where

$$\beta\left(\delta, n, \bar{\bar{X}}\right) = \Phi\left(-\delta\sqrt{n} + Z_{\frac{\alpha}{2}} + \frac{\sqrt{n}\left(\bar{\bar{X}} - \mu_0\right)}{\sigma_0}\right) - \Phi\left(-\delta\sqrt{n} - Z_{\frac{\alpha}{2}} + \frac{\sqrt{n}\left(\bar{\bar{X}} - \mu_0\right)}{\sigma_0}\right)$$

and δ is as defined before, Chakraborti (2000). This is called the conditional probability of signal, CPS and at $\delta = 0$, it is then called the conditional false alarm rate, CFAR of the Shewhart $\bar{x}$-chart when the process mean is estimated. Since $\bar{\bar{X}}$ has a normal distribution with mean, μ_0 and variance σ_0/mn, the CPS can be expressed in the alternative canonical form as

$$1 - \left\{\Phi\left(-\delta\sqrt{n} + Z_{\frac{\alpha}{2}} + \frac{Z}{\sqrt{m}}\right) - \Phi\left(-\delta\sqrt{n} - Z_{\frac{\alpha}{2}} + \frac{Z}{\sqrt{m}}\right)\right\}$$

also called CFAR at $\delta = 0$, where $Z = \sqrt{mn}(\bar{\bar{X}} - \mu_0)/\sigma_0$ has a standard normal distribution, Chakraborti (2000). Therefore, the exact in-control and out-of-control average run lengths are defined respectively by

$$ARL_0 = \int_{-\infty}^{\infty} \left[1 - \left\{\Phi\left(Z_{\frac{\alpha}{2}} + \frac{Z}{\sqrt{m}}\right) - \Phi\left(-Z_{\frac{\alpha}{2}} + \frac{Z}{\sqrt{m}}\right)\right\}\right]^{-1} \Phi\left(z\right) d\left(z\right) \tag{5}$$

and

$$ARL_\delta = \int_{-\infty}^{\infty} \left[1 - \left\{\Phi\left(-\delta\sqrt{n} + Z_{\frac{\alpha}{2}} + \frac{Z}{\sqrt{m}}\right) - \Phi\left(-\delta\sqrt{n} - Z_{\frac{\alpha}{2}} + \frac{Z}{\sqrt{m}}\right)\right\}\right]^{-1} \Phi\left(z\right) d\left(z\right), \delta \neq 0 \tag{6}$$

The ARL Equations in (5) and (6) were evaluated by simulation at 10^6 replications and the same simulation size is maintained throughout this work. The results are given in Table 1 for $m = 20, 30, 50$ and 100 and for $\delta = 0.0, 0.2, 0.5, 0.6,$ and 1.0 with $n = 5$ and $\alpha = 0.0027$. Note that Equation (6) was evaluated at different shift sizes (δ).

Distribution of the run length

Since the distribution of N is no longer geometric with estimated parameter, $\bar{\bar{X}}$. It can be deduced from Chakraborti (2000) and Chakraborti (2007) that the exact c.d.f of N is given by

$$P(N \leq s) = 1 - \int_{-\infty}^{\infty} \left[\Phi\left(-\delta\sqrt{n} + Z_{\frac{\alpha}{2}} + \frac{Z}{\sqrt{m}}\right) - \Phi\left(-\delta\sqrt{n} - Z_{\frac{\alpha}{2}} + \frac{Z}{\sqrt{m}}\right)\right]^{s} \Phi\left(z\right) d\left(z\right), \tag{7}$$

where $s = 1, 2, \ldots, \Phi(.)$ and Z are as already defined.

Percentiles of the run length

As already stated, the $100pth$ percentile can be formulated from the definition that it is the smallest positive integer, 's' such that the c.d.f at 's' is at least equal to p. Therefore, the equation in (7) is redefined as in (4) for the purpose. That is

$$1 - \int_{-\infty}^{\infty} \left[\Phi\left(-\delta\sqrt{n} + Z_{\frac{\alpha}{2}} + \frac{Z}{\sqrt{m}} \right) - \Phi\left(-\delta\sqrt{n} - Z_{\frac{\alpha}{2}} + \frac{Z}{\sqrt{m}} \right) \right]^{s} \Phi(z)\, d(z) \geq p, \qquad (8)$$

where $s = 1, 2, \ldots$. By evaluating this equation at different values of p and δ , one can study the performance characteristics of the Shewhart $\bar{x}$-chart with estimated process mean, $\bar{\bar{X}}$. Aside evaluating the percentiles from equation (8), it was also evaluated and plotted for all possible values of p, first, at $m = 30, 100$ and 500 when $\delta = 0$ and at $m = 30, 100$ and 500 when $\delta = 0.5$. The choice of these values is in line with Chakraborti (2007) as the both results will be compared in this work . The graphs for $m = 30, 100$ and 500 when $\delta = 0$ and at $m = 30, 100$ and 500 when $\delta = 0.5$ are presented in Figures 1 and 2 respectively and like already hinted, the graph of the standards known case (Case KK) will be included in each case for reference purposes. Each figure in each case will be discussed and compared with the graphs in Figures 5 and 6 in this work which show Shewhart $\bar{x}$-charts operating under the same conditions of m, n and δ when the mean and variance are unknown and estimated by $\bar{\bar{X}}$and S as reported by Chakraborti (2007).

3.1. Process mean known but the standard deviation unknown (Case KU)

When the mean standard $\mu = \mu_0$ is given but the standard deviation σ_0 is unknown, the process standard deviation is typically estimated from m reference (phase I) samples each of size n when the process is in-control from a phase 1 analysis, Jardim et al. (2018). Mahmoud et al. (2010) analyzed several estimators of the standard deviation (σ_0) in terms of the mean squared error and recommended the use of $\widehat{\sigma_0} = S_p$=$\sqrt{\frac{1}{m}\sum_{i=1}^{m} S_i^2}$ where $S_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(X_{ij} - \bar{X}_i\right)^2$and $\bar{X}_i = 1/n\sum_{j=1}^{n} X_{ij}$ for $i = 1, 2, \ldots, m, j = 1, 2, \ldots, n$ and X_{ij} is the jth observation of the ith Phase 1 sample of size n. In this work, the recommended estimator will be considered in use. Then, the control limits of the chart becomes; $\mu_0 \pm Z_{\alpha/2}\frac{S_p}{\sqrt{n}}$. Because of this estimated limits plugged into the design of the chart, the run length random variable (N) is no longer geometric, Quesenberry (1993). But conditionally on S_p, N has a geometric distribution with the probability of success(signal) given by $1 - \beta(\delta, n, S_p)$ where

$$\beta(\delta, n, S_p) = \Phi\left(-\delta\sqrt{n} + Z_{\alpha/2}\frac{S_p}{\sqrt{n}} \right) - \Phi\left(-\delta\sqrt{n} - Z_{\alpha/2}\frac{S_p}{\sqrt{n}} \right).$$

Like before, this is called the conditional probability of signal and at $\delta = 0$, it is then called the conditional false alarm rate of the Shewhart $\bar{x}$-chart when the process variance is estimated. Using the fact that $Y = m(n-1)S_p^2/\sigma_0^2$ follows chi-square distribution with $m(n-1)$ degrees of freedom, that is, $Y = m(n-1)S_p^2/\sigma_0^2 \sim \chi^2_{m(n-1)}$ and letting $v = m(n-1)$, the probability of signal can be rewritten as

$$1 - \left\{ \Phi\left(-\delta\sqrt{n} + \frac{Z_{\frac{\alpha}{2}}}{\sqrt{v}}\sqrt{y} \right) - \Phi\left(-\delta\sqrt{n} - \frac{Z_{\frac{\alpha}{2}}}{\sqrt{v}}\sqrt{y} \right) \right\}.$$

Therefore, the exact in-control and out-of-control average run lengths are defined respectively by

$$ARL_0 = \int_0^{\infty} \left[1 - \left\{ \Phi\left(\frac{Z_{\frac{\alpha}{2}}}{\sqrt{v}}\sqrt{y} \right) - \Phi\left(-\frac{Z_{\frac{\alpha}{2}}}{\sqrt{v}}\sqrt{y} \right) \right\} \right]^{-1} f_{\chi_v}(y)\, dy \qquad (9)$$

and

$$ARL_\delta = \int_0^{\infty} \left[1 - \left\{ \Phi\left(-\delta\sqrt{n} + \frac{Z_{\frac{\alpha}{2}}}{\sqrt{v}}\sqrt{y} \right) - \Phi\left(-\delta\sqrt{n} - \frac{Z_{\frac{\alpha}{2}}}{\sqrt{v}}\sqrt{y} \right) \right\} \right]^{-1} f_{\chi^2_v}(y)\, dy, \delta \neq 0 \qquad (10)$$

where $f_{\chi^2_v}$ is the Chi-square probability density function (p.d.f) with v degrees of freedom, Chakraborti (2000). Equations (9) and (10) were evaluated and the results presented in Table 2. Note that equation (10) was evaluated at different shift sizes (δ).

Distribution of the run length

Again, since the distribution of N is no longer geometric with the estimated parameter, S_p. It can be deduced from Chakraborti (2000) and Chakraborti (2007) that the exact c.d.f of N is given by

$$P(N \leq s) = 1 - \int_0^\infty \left[\Phi\left(-\delta\sqrt{n} + \frac{Z_{\frac{\alpha}{2}}}{\sqrt{v}}\sqrt{y}\right) - \Phi\left(-\delta\sqrt{n} - \frac{Z_{\frac{\alpha}{2}}}{\sqrt{v}}\sqrt{y}\right)\right]^s f_{\chi_v}(y)\, dy, \tag{11}$$

where $s = 1, 2, \ldots$.

Percentiles of the run length

Following the argument before, the percentiles of the distribution can be written as

$$1 - \int_0^\infty \left[\Phi\left(-\delta\sqrt{n} + \frac{Z_{\frac{\alpha}{2}}}{\sqrt{v}}\sqrt{y}\right) - \Phi\left(-\delta\sqrt{n} - \frac{Z_{\frac{\alpha}{2}}}{\sqrt{v}}\sqrt{y}\right)\right]^s f_{\chi_v}(y)\, dy \geq p, \tag{12}$$

where $s = 1, 2, \ldots$.

Expression (12) was evaluated at different values of p and δ for the study of the chart performance characteristics with estimated process variance. Aside from evaluating the percentiles from equation (12), it was also evaluated and plotted for all possible values of p, first, at $m = 30, 100$ and 500 when $\delta = 0$ and at $m = 30, 100$ and 500 when $\delta = 0.5$. The graphs are presented in Figures 3 and 4 respectively alongside the graph of the standards known case (Case KK) included for reference purpose and which will be discussed and compared with the other graphs in Figures 1, 2, 5 and 6 for other cases.

4. Results and Discussion

As pointed out already, specifically, for comparison with the results in Chakraborti (2007), the c.d.fs of the run length distribution for the Shewhart $\bar{x}$-chart for Cases UK and KU are evaluated and plotted for $m = 30, 100, 500$, $n = 5$, and $\alpha = 0.0027$ for shift sizes ($\delta = 0$ and 0.5).

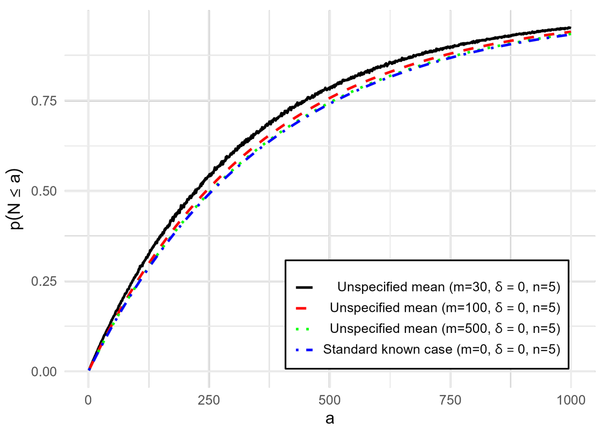

Figure 1 In-control $(\delta = 0)$ run length c.d.f of the Shewhart $\bar{x}$-chart of Case UK for m=30 ,100 and 500, n=5 in each and α=0.0027

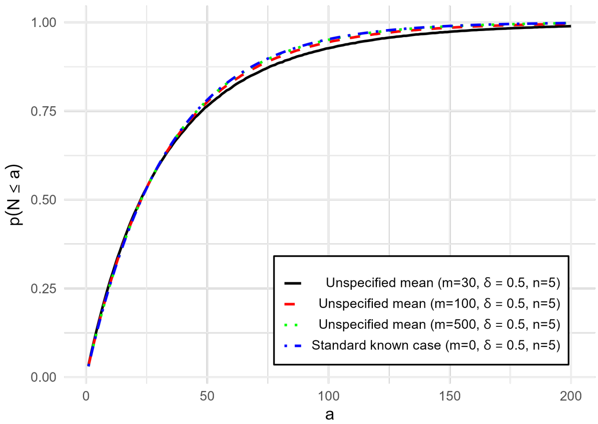

Figure 2 Out-of-control $(\delta = 0.5)$ run length c.d.f of the Shewhart $\bar{x}$-chart in Case UK for m=30,100 and 500, n=5 in each and α=0.0027

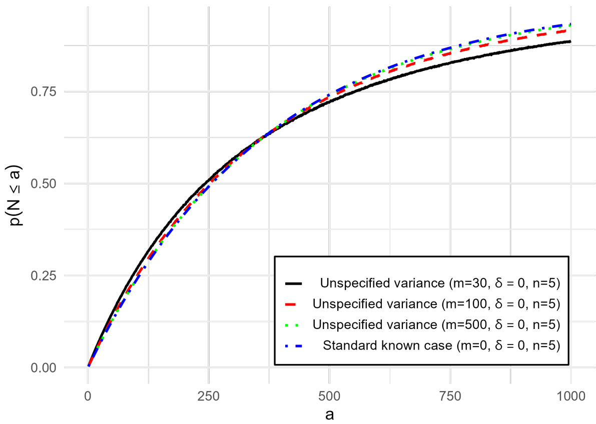

Figure 3 In-control $(\delta = 0)$ run length c.d.f of the Shewhart $\bar{x}$-chart of Case KU for m=30 ,100 and 500, n=5 in each and α=0.0027

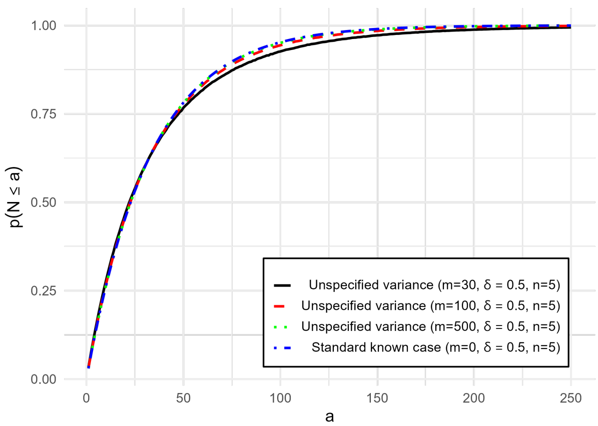

Figure 4 Out-of-control $(\delta = 0.5)$ run length c.d.f of the Shewhart $\bar{x}$-chart in Case KU for m=30,100 and 500, n=5 in each and α=0.0027

Before the discussions on the figures, it suffices to note that the information from the figures of each case (Cases UK, KU, and UU) is appreciated more by paying greater attention to the in-control $(\delta = 0)$ state of each case because with the out-of-control $(\delta = 0.5)$ size, the run-length distributions at small, moderate and large values of m tend to converge alongside that of the standards known case. Therefore, viewing the movement of a particular curve clearly might appear a bit clumsy but this is not so with the former.

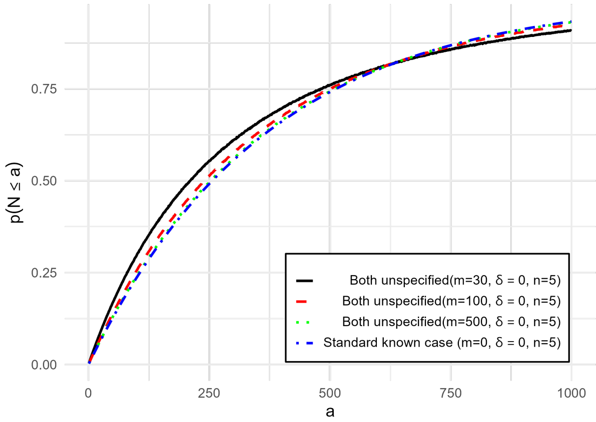

Figure 5 In-control $(\delta = 0)$ run length c.d.f of the Shewhart $\bar{x}$-chart in Case UU for m=30,100, and 500, n=5 in each and α=0.0027 reproduced for comparison from (Chakraorti, 2007)

.

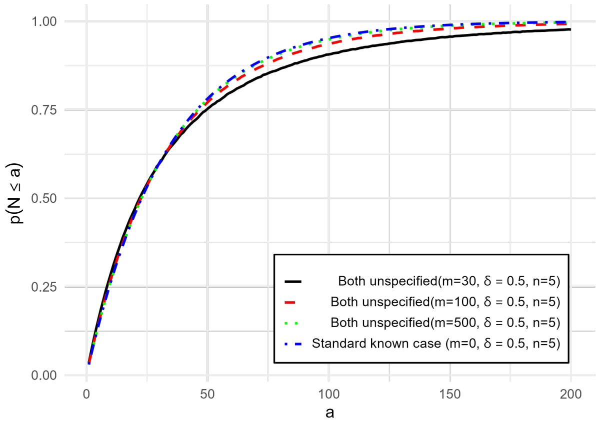

Figure 6 Out-of-control $(\delta = 0.5)$ run length c.d.f of the Shewhart $\bar{x}$-chart in Case UU for m=30,100 and 500, n=5 in each and α=0.0027 reproduced for comparison from (Chakraorti, 2007)

The graphs in Chakraborti (2007) for the Shewhart $\bar{x}$-chart when the parameters, the process mean and variance are unknown but estimated are reproduced in this work and presented in Figures 5 and 6 to concretize our comparative discussions by having the results side by side. As already mentioned, the graph of the c.d.f of the run-length distribution for the standards known case (geometric distribution) for nominal $(\alpha = 0.0027)$ was also included in each case for reference. In Figure 1 of the Case UK, the run length c.d.f curves differ significantly from that of the standards known case (geometric distribution) especially at small to moderate values of m but at large m say 500 and above, it was discovered that the distance between the curve and that of the standards known case thins off significantly as the two curves are seen to converge everywhere including at both tails of the distribution. Similar results were obtained in Figure 2 with the shift size of 0.5 from the in-control state except that the curve for the standards known case lies above the other curves unlike before, pointing to the fact that the curve for the standards known case raises more alarms to this shift. These observations are true with Case KU which can be seen in Figures 3 and 4 for the in-control and 0.5 out-of-control states respectively. All these reports thus far, agree with the remarks in Chakraborti (2007) for the Case UU as can be seen in Figures 5 and 6. Therefore, at large m, at least 500, the run-length c.d.f curves of Shewhart $\bar{x}$-chart when the process mean is unknown but estimated (Case UK), when the variance is unknown but estimated (Case KU) and when both the process mean and variance are unknown but estimated (Case UU) converge with that of the standards known case (Case KK).

Again, Chakraborti (2007) showed that in the in-control state, when m=30 and 100, the run-length c.d.fs for Case UU lie above that of the geometric distribution up to (approximately) the $82nd$ percentile but beyond that, they lie below it. It was noted in that paper again that similar behavior was observed in the out-of-control state ($\delta = 0.5$) but that the c.d.fs lie above up to (approximately) the $62nd$ percentiles and lie below that of the geometric distribution beyond the $62nd$ percentiles. This is evident in Figures 5 and 6 of this work as reproduced from Chakraborti (2007). Owing to this, it was further noted that when parameters are estimated and m is small to moderate, the run-length distributions neither dominate nor are dominated by the run-length distribution in the standard known case describing them are not being stochastically ordered relative to the geometric distribution. A random variable is said to be stochastically smaller (larger) than the second when the c.d.f of that random variable lies entirely above (below) the c.d.f of the second random variable, Gibbons and Chakraborti (2003). Note that a stochastically larger random variable has larger percentiles than the smaller one and vice versa. Very similar results are obtained in Figures 3 and 4 when a parameter (variance) is estimated (Case KU), only that in both shifts states (0 and 0.5), the c.d.f curves lie above that of geometric distribution up to (approximately) the 62nd percentiles beyond which they lie below it making them not to be stochastically ordered relative to the geometric distribution.

However, this is not true in the case when only the process mean is estimated (Case UK) as the c.d.f curves for $m = 30$ and 100 are seen in Figure 1 to lie above that of the standards known (geometric distribution) throughout their course. Thus, in line with this augment, the run-length distributions of the Shewhart $\bar{x}$-chart are dominated by the run-length distribution in the standard known case (geometric distribution) when only the process mean is estimated with small to moderate values of m. This shows that the run-length distributions here are stochastically ordered relative to the geometric distribution. Again, with a shift in the in-control state of $\delta = 0.5$, the c.d.f curves are still stochastically ordered relative to that of the standard known case which can be seen in Figure 2 except that the run-length distribution for the standard known is now dominated as it lies entirely above the curves in Case UK for $m = 30$ and 100. This implies smaller run-length percentiles for the standards known case at all times which means more out-of-control signal for this shift compared to Cases UK, making it more efficient. This is also different from what was reported in Chakraborti (2007) which was reproduced and presented in Figure 6 as no one c.d.f is seen to dominate the other for the shift size, $\delta = 0.5$.

Table 1 Average Run Lengths and the Percentiles of the Shewhart $\bar{x}$-chart, When the process mean is unknown and the variance is known Case UK for a Number of Phase I samples (m) and Shift Size n at n = 5 and α=0.0027

		Percentiles of the run length distribution of Case UK												
Shift (δ)	ARL	V	X	XX	XXV	XXX	XL	L(MRL)	LX	LXX	LXXV	LXXX	XC	XCV
							$m = XX$							
0	311.21	16	31	66	86	106	153	209	279	369	427	498	725	954
0.2	191.8	8	16	36	45	56	82	116	158	213	251	297	460	637
0.5	38.69	2	4	7	9	12	17	23	31	42	49	59	92	129
0.6	23.37	1	1	5	6	7	10	14	19	26	30	36	55	76
1	4.79	1	1	1	1	2	2	5	4	5	6	7	10	14
							$m = XXX$							
0	327.02	17	34	71	92	114	164	223	295	389	449	523	775	991
0.2	188.47	8	17	36	47	59	85	118	159	214	250	296	449	613
0.5	36.81	2	4	7	9	12	17	23	31	42	48	57	86	119
0.6	22.37	1	1	4	5	6	10	14	19	25	29	35	52	71
1	4.69	1	1	1	1	2	2	3	4	5	6	7	9	12
							$m = L$							
0	341.66	18	36	76	97	121	173	234	310	409	471	548	787	1027
0.2	185.18	9	18	38	49	61	87	120	160	213	249	293	436	585
0.5	35.39	1	4	7	9	12	17	23	31	41	47	56	82	111
0.6	21.62	1	1	4	5	7	10	13	18	24	29	34	50	66
1	4.41	1	1	1	1	1	2	3	3	5	6	7	10	13
							$m = C$							
0	354.6	18	38	79	102	126	181	245	324	426	490	569	816	1063
0.2	181.84	9	19	39	50	62	89	121	162	214	248	289	423	561
0.5	34.37	2	3	7	9	11	16	23	31	41	46	54	79	105
0.6	21.08	1	1	4	5	7	10	14	18	24	28	33	48	63
1	4.55	1	1	1	1	1	2	3	4	5	6	6	9	12
							Standards Known (Case KK)							
0	370.37	19	39	83	107	132	189	257	339	446	513	596	852	1109
0.2	177.72	10	19	40	40	64	91	123	163	214	214	286	409	531
0.5	33.4	2	4	7	10	11	17	23	30	39	45	53	76	99
0.6	20.56	2	3	5	5	8	11	14	19	25	25	33	47	61
1	4.49	1	1	1	1	2	3	3	5	5	6	7	10	12

Table 2 Average Run Lengths and the Percentiles of the Shewhart $\bar{x}$-chart, When the process mean is known and the variance is unknown (Case KU) for a Number of Phase I samples (m) and Shift Size n at n=5 and α=0.0027

		Percentiles of the run length distribution of Case KU												
Shift (δ)	ARL	V	X	XX	XXV	XXX	XL	L(MRL)	LX	LXX	LXXV	LXXX	XC	XCV
							$m = XX$							
0	511.46	14	30	67	87	111	165	238	332	477	571	706	1212	1507
0.2	229.12	7	15	32	43	55	81	114	161	223	268	327	539	811
0.5	39.15	1	3	6	8	10	15	21	29	41	48	57	91	130
0.6	23.52	1	2	4	5	6	10	13	18	24	29	35	54	77
1.0	4.79	1	1	1	1	1	1	2	2	4	6	7	10	13
							$m = XXX$							
0	456.85	15	33	72	92	117	172	243	336	464	553	671	1078	1586
0.2	209.65	8	17	35	46	57	84	118	161	220	260	315	496	910
0.5	37.04	2	3	6	9	11	16	22	300	40	47	56	86	118
0.6	22.45	1	2	4	6	7	9	13	18	24	29	34	51	71
1.0	4.69	1	1	1	1	1	2	3	4	5	6	6	9	13
							$m = L$							
0	418.95	17	35	75	97	122	178	248	337	456	537	639	987	1389
0.2	195.89	8	17	37	48	59	86	120	162	218	256	303	459	637
0.5	35.50	2	3	7	9	10	16	22	30	40	46	55	82	111
0.6	21.66	1	2	4	5	7	10	14	18	24	28	33	49	66
1.0	4.61	1	1	1	1	1	2	3	3	6	6	6	10	12
							$m = C$							
0	393.56	18	37	79	102	127	184	252	336	452	524	618	921	1243
0.2	186.46	8	18	38	49	61	88	121	161	216	250	294	434	582
0.5	34.42	2	3	7	9	11	16	22	30	39	46	54	79	104
0.6	21.10	1	2	4	5	7	10	13	18	24	28	32	48	63
1.0	4.55	1	1	1	1	1	1	2	3	5	5	6	9	12
							Standards Known Case (CaseKK)							
0	370.37	19	39	83	107	132	189	257	339	446	513	596	852	1109
0.2	177.72	10	19	40	40	64	91	123	163	214	214	286	409	531
0.5	33.4	2	4	7	10	11	17	23	30	39	45	53	76	99
0.6	20.56	2	3	5	5	8	11	14	19	25	25	33	47	61
1.0	4.49	1	1	1	1	2	3	3	5	5	6	7	10	12

As already stated, several percentiles of the run-length distribution for both in-control and out-of-control states were evaluated for $m = 20, 30, 50$, and 100 and at shift sizes of $\delta = 0.0, 0.2, 0.5, 0.6$ and 1.0 when the sample size and the FAR are fixed at 5 and 0.0027 respectively for Cases UK and KU. The evaluations of the ARLs and the c.d.fs involved numerical integration and in this paper, statistical software, R (R core, 2019) was used for the evaluations. Table 1 is the result of the detailed examination of the percentiles of the run- length distribution of the Shewhart $\bar{x}$-chart when the process mean is unknown but estimated from m phase I samples, (Case UK) while Table 2 is that of the Case KU. Juxtaposing the results with those in Chakraborti (2007) of Case UU and given the results of the standards known case, Case KK too, the following remarks can be made of them. First, at an in-control state ($\delta = 0$), Chakraborti (2007) reported that the $50th$ percentiles (MRL_0) for Case UU are $194, 211, 227$ and 241 for $m = 20, 30, 50$, and 100 respectively and compared it with the (MRL_0), 257 for the standard known case, Case KK. The paper further noted that the median run lengths, when parameters are estimated are shorter especially at small to moderate values of m compared to that of the Case KK. Of course, it implies more false alarms than what is nominally expected in Case KK with FAR of 0.0027. But in Case UK, the $50th$ percentiles (MRL_0) are $209, 223, 234$ and 245 for $m = 20, 30, 50$ and 100 respectively and when compared with that of the standards known case, the values imply shorter run-lengths and more false alarms as well, however, they mean longer run lengths with smaller false alarms when compared to Case UU.

However, in Table 2, it can be seen that at $\delta = 0$ (in-control state), the $50th$ percentiles (MRL_0) for Case KU are $238, 243, 248$ and 252 for $m = 20, 30, 50$ and 100 respectively. When compared with the MRL_0 for the standard known case (257), the values are all smaller in terms of the median run length implying more false alarms but when compared to the Case UU of Chakraborti (2007) and Case UK of this work, the values stand, instead, for larger median run length and smaller false alarms. This is a bit contradictory to what one might ordinarily think and expect on the basis that in Case UU, both parameters are estimated and therefore, more variation is expected in the process which implies more false alarms and a smaller median run length. A similar conclusion is reached in terms of the average run length (ARL_0)

Looking at columns 2 and 9 of Tables 1 and 2, which show the ARL and the $50th$ percentiles (MRL) values of the Cases UK and KU , it is clear that the ARL values are larger than that of the MRL values which points to the right-skewed nature of the run-length distribution and it is therefore misleading to settle for the ARL as a better or a sole measure of chart performance. Again, Chakraborti (2007) stated that the tails of the c.d.f of the run length when parameters are estimated behave differently relative to that of the geometric distribution, noting that in Case UU, while the percentiles for $p \leq 0.82$ (approximately) are smaller than those for the standard known case (geometric distribution), for $p > 0.82$, the percentiles are larger. Figure 5 helps to affirm this statement. A similar conclusion is reached for Case KU as the percentiles for $p \leq 0.62$ (approximately) are smaller than those for the standard known case, however, for $p > 0.62$, the percentiles are larger, see Figures 3 and 4. However, in Case UK, the c.d.fs for small to moderate values lie entirely above that of the geometric distribution (stochastically smaller), therefore, the percentiles over the entire range are smaller than those of the standards known case. The same paper had it that when $m = 500$, so that $500 \times 5 = 2,500$ phase I data points are used to estimate the mean and the variance, the c.d.f of the run length and hence the percentiles all converge to those of the standard known case. Similar results were obtained for Cases UK and KU

5. Conclusion

The c.d.fs of Case UK for small to moderate values of Phase I samples are stochastically ordered (specifically smaller) relative to that of the standards known case (geometric distribution) unlike in the Cases UU and KU where the c.d.fs cross with that of the geometric distribution at some points making them not to be stochastically ordered relative it. Smaller false alarms were witnessed in Case UU, with MRL_0 of 194, 211, 227, and 241 for m = 20, 30, 50, and 100 respectively, followed by Case UK with MRL_0 of 209, 223, 234 and 245 for m = 20, 30, 50 and 100 respectively and then Case

KU with MRL_0 of 238, 243, 248 and 252 for $m = 20$, 30, 50 and 100 respectively when compared to that of Case KK. These results imply that the Shewhart $\bar{x}$-chart gives out more false alarms when the process mean and variance are estimated, followed by only when the process mean is estimated, and finally when only the variance is estimated. The run-length c.d.fs of Case UU and KU for small to moderate values of m are not stochastically ordered relative to that of the Case KK but the run lengths c.d.fs of Case UK are. For small to moderate sizes of m, the c.d.fs of the Shewhart $\bar{x}$-chart obtained in all cases of parameters estimation are significantly different from that of the standards known case but, at large size of m (Phase I samples), say up to 500 each of size 5, the c.d.f.s in Cases UU, UK and KU converge to that of the standard known case as well as the percentiles.

Acknowledgements

We would like to thank Dr. T.E .Ugah for assisting us immensely in preparing the manuscript.

References

Chakraborti S. Run-length, average run length, and false alarm rate of Shewhart $\bar{X}$-chart: Exact derivations by conditioning. Commun Stat Simulat. 2000; 29(1): 61-81.

Chakraborti S. Parameter Estimation and Design Considerations in Prospective Applications of the $\bar{X}$-chart. Commun Stat Simulat. 2006; 33(4): 439-459.

Chakraborti S. Run-length distribution and percentiles: The Shewhart Chart with Unknown Parameters. Qual Eng. 2007; 19(2): 119-127.

Chen G. The run-length distributions of R, S and S^2-control charts when is estimated. Can J Stat. 1997; 26(2): 311-322.

Jardim FS, Chakraborti S, Epprecht EK. $\bar{X}$-Chart with Estimated Parameters: The Conditional ARL Distribution and New Insights. Prod Oper Manag. 2018; 28(6): 1545-1557.

Jardim FS, Chakraborti S, Epprecht EK. Two perspectives for designing a phase II control chart with estimated parameters: The case of the Shewhart Chart, J Qual Technol. 2019; 52(2): 198-217.

Gibbons JD, Chakraborti S. Nonparametric Statistical Inference. 4th ed. New York: Marcel Dekker; 2003.

Goedhart R, Silva MM, Schoonhoven M, Epprecht EK, Chakraborti S, Does RJ, Veiga A. Correction factors for Shewhart x and $\bar{X}$-control charts to achieve desired unconditional ARL. Int J Prod Res. 2016; 54(24): 7464-7479.

Khoo MB. CPerformance measures for the Shewhart $\bar{X}$-control chart. Qual Eng. 2004; 16(1): 585-590.

Mahmoud MA, Henderson GR, Epprecht EK, Woodall WH. Estimating the standard deviation in quality control applications. J Qual Technol. 2010; 42(4): 348-357.

Montgomery DC. Statistical Quality Control. New York: John Willey and Sons; 1991.

Montgomery DC. Introduction to Statistical Quality Control. New York: John Wiley and Sons; 2004.

Quesenberry CP. The effect of sample size on estimated limits for x and $\overline{X}$-control charts. J Qual Technol. 1993; 25(1): 237-247.

Radson D, Boyd AH. Graphical representation of run length distributions. Qual Eng. 2005; 17(1): 301-308.

Shmueli G, Cohen A. Run-length distribution for control charts with runs and scans rules. Commun Stat Theory. 2003; 32(1): 475-495.