A Note on the Robustness of the $\bar{X}$ Chart to Non-Normality

Sungim Lee[1,a]

[a]Department of Statistics, Dankook University

Abstract

These days the interest of quality leads to the necessity of control charts for monitoring the process in various fields of practical applications. The $\bar{X}$ chart is one of the most widely used tools for quality control that also performs well under the normality of quality characteristics. However, quality characteristics tend to have non-normal properties in real applications. Numerous recent studies have tried to find and explore the performance of $\bar{X}$ chart due to non-normality; however previous studies numerically examined the effects of non-normality and did not provide any theoretical justification. Moreover, numerical studies are restricted to specific type of distributions such as Burr or gamma distribution that are known to be flexible but can hardly replace other general distributions. In this paper, we approximate the false alarm rate(FAR) of the $\bar{X}$ chart using the Edgeworth expansion up to $1/n$-order with the fourth cumulant. This allows us to examine the theoretical effects of non-normality, as measured by the skewness and kurtosis, on $\bar{X}$ chart. In addition, we investigate the effect of skewness and kurtosis on $\bar{X}$ chart in numerical studies. We use a skewed-normal distribution with a skew parameter to comprehensively investigate the effect of skewness.

Keywords: Average run length, $\bar{X}$ chart, Edgeworth expansion, simulation.

1. Introduction

The $\bar{X}$-chart has been widely used to monitor an independent univariate process as a useful tool of quality control. In theory, the $\bar{X}$-chart assumes that the quality characteristic is from normal distribution or the number of its observations is large enough for the central limit theorem to hold. As previous authors, such as Bai and Choi (1995), Roes and Does (1995), Janacek and Meikle (1997), have remarked the observations are often from a non-normal distribution and its size is not large in practice. Moreover, the control parameters such as μ and σ^2 are usually unknown and therefore estimated. These situations influence the performance of the $\bar{X}$ chart. It is usually evaluated by the average run length(ARL) where the run length indicates the number of subgroups taken until an alarm is signaled by the control chart.

Several authors study the control limits of the $\bar{X}$ chart when the observations are not normal. Burr (1967) showed that the control limits in this chart are very robust to normal assumptions, that were examined by Burr distribution. Similarly, Schilling and Nelson (1976) investigated the effect of nonnormality on the control limits and they report that samples of size 4 or 5 in the $\bar{X}$ chart are sufficient to ensure reasonable robustness for the normal assumption. This is investigated by well-known distributions such as uniform or gamma distribution. Chan *et al.* (1988) concluded that the non-normality of data has little effect on the control limits of the $\bar{X}$-chart and this was studied by Tukey's λ-family of symmetric distribution. Wheeler and Chambers (1992) reported that the control

limits do not change appreciably even when the data are non-normal. These remarks all agree with Shewhart's original assumption in which the standard normal based-control limits are appropriate for $\bar{X}$ chart Shewhart (1931).

Others (*e.g.*, Burrows, 1962; Yourstone and Zimmer, 1992; Chakraborti *et al.*, 2004) have remarked that the $\bar{X}$ chart is significantly affected by non-normality in terms of the performance. Burrows (1962) studied the effect of skewed distributions on $\bar{X}$ charts and found that the ARLs of moderately skewed distributions significantly differ from the stated ARLs based on the normal distribution. He concluded that non-normal data may lead to asymmetric control limits rather than symmetric ones. Yourstone and Zimmer (1992) developed the asymmetric $\bar{X}$ charts based on probability limits using the Burr distribution and showed that Shewhart-type control chart can lead to inappropriate decisions especially for skewed distribution.

There has been a great deal of research on the performance of $\bar{X}$ chart; however, most studies are limited in several points. First, they are restricted to a specific type of distributions such as the Burr, gamma, and Weibull distributions. Even though some distributions are known to be flexible, they can hardly replace other general distributions. In particular, the skewness and kurtosis in theirs tend to confound each other and it is hard to investigate their unique effect on the ARLs. Second, previous works numerically investigate the effects of non-normality on the ARLs and do not provide any theoretical justification.

In this paper, we assume the information about the third and forth moments of underlying distribution and we will approximate the FAR of the $\bar{X}$ chart using the Edgeworth expansion up to $1/n$-order with the fourth cumulant. The approximation shows how the FAR and also the ARL depend on the third and the fourth cumulants (or the skewness and kurtosis) of the underlying distribution.

Most previous works focused on the ARL based on numerical studies with well-known distributions in order to examine the effect of departures from normality on $\bar{X}$ chart. For instance, Chakraborti *et al.* (2004) investigated the gamma, t or uniform distributions. However, as noted by their paper, the run length may go to infinity (depending on the distributional properties) and it may not be appropriate to simulate the run length directly. Hence, we provide numerical studies about the ARL via the FAR to see how the ARL is influenced by the skewness with a small subgroup. In particular, we consider the class of skew-normal distribution where the departure of the underlying distribution from the normal distribution can be relatively well regulated by the skewness parameter.

The paper is organized as follows. In Section 2, we approximate the false alarm rate of the $\bar{X}$ chart using the Edgeworth expansion. In Section 3, we numerically investigate the effect of the skewness on the in-control ARL of $\bar{X}$ chart. Here, we employ the class of the skew-normal distribution as true underlying distributions. We numerically compute the distribution of the ARL using the approximate FAR obtained in Section 3. Conclusions and discussions are given in Section 4.

2. Edgeworth Expansion of the False Alarm Rate

Suppose that the parameters μ and σ represent the in-control process mean and standard deviation, respectively. The upper and lower control limits, LCL and UCL, are given by

$$\text{UCL} = \mu + F^{-1}\left(1 - \frac{\alpha}{2}\right)\frac{\sigma}{\sqrt{n}} \tag{2.1}$$

and

$$\text{LCL} = \mu - F^{-1}\left(\frac{\alpha}{2}\right)\frac{\sigma}{\sqrt{n}}, \tag{2.2}$$

where the distribution $F(\cdot)$ represents the sampling distribution of the sample mean, $\bar{X}$, and is generally used with the standard normal distribution $\Phi(\cdot)$ according to the central limit theorem. To monitor and control the process mean μ, we will collect the test sample $x_{t1}, x_{t2}, \ldots, x_{tn}$ at each inspection time $t = 1, 2, 3, \ldots$ and compute the test statistic $\bar{X}_t$ from the sample. If a sample mean falls outside of these limits in (2.1) or (2.2), it indicates that the process is out of control. If it remains within the control limits, it is regarded under control. Let N be the run length of the control statistic $\bar{X}$ defined as

$$N = \min \left\{ t : \bar{X}_t < \text{LCL} \ \text{ or } \ \bar{X}_t > \text{UCL} \right\} \qquad (2.3)$$

for $t = 1, 2, \ldots$. Then the average run length is given by

$$E(N) = \sum_{l=0}^{\infty} P(N > l).$$

Here, by the definition of the run length N in (2.3),

$$P(N > l) = P\left(\text{LCL} < \bar{X}_1 < \text{UCL}, \ldots, \text{LCL} < \overline{X}_l < \text{UCL}\right)$$
$$= P\left(\text{LCL} < \bar{X}_1 < \text{UCL}\right)^l.$$

Thus, the ARL for the control statistic $\bar{X}_1$ becomes

$$E[N] = \sum_{l=0}^{\infty} P(N > l)$$
$$= \sum_{l=0}^{\infty} P\left(\text{LCL} < \bar{X}_1 < \text{UCL}\right)^l$$
$$= \frac{1}{1 - P\left(\text{LCL} \leq \bar{X}_1 \leq \text{UCL}\right)} \qquad (2.4)$$
$$= \frac{1}{1 + P\left(\bar{X}_1 \leq \text{LCL}\right) - P\left(\bar{X}_1 \leq \text{UCL}\right)}. \qquad (2.5)$$

As shown in (2.4), the ARL is given by the reciprocal of the FAR which means the probability that any point exceeds the control limits. To evaluate the performance of the control chart based on the ARL, we need to calculate the in-control ARL and out-of-control ARL that correspond to the conditional expectation of the run length; given that the process is under control and out-of-control, respectively. Efficient charts have the larger in-control ARL and smaller out-of-control ARL. Based on the relationship between ARL and FAR in (2.4), we will evaluate the ARL by improving FAR on the normal approximation of $\bar{X}$. To do this, we consider two cases for the Edgeworth expansion of the FAR: (i) the process parameters μ and σ are known (ii) the process parameters μ and σ are unknown.

2.1. The case where both μ and σ are known

Let us first consider the control limits when μ and σ are known. As discussed earlier, the previous works show that the in-control ARL is affected by the skewness and the kurtosis but they do not provide any theoretical justification. In order to evaluate the effect of the skewness and the kurtosis on the ARL, we approximate the ARL via improving the probability of a subgroup $\bar{X}$ falling outside

control limits, UCL and LCL. As displayed in (2.5), the probability exceeding lower limit, $P(\bar{X} \leq$ LCL), plays an important role in improving in the FAR. If we let this probability of exceeding lower limit with mean shift from μ_0 to $\mu = \mu_0 + \delta/\sqrt{n}$, it is represented as

$$A(k, \delta) = P\left(\bar{X} \leq \mu_0 - k\frac{\sigma}{\sqrt{n}} \,\Big|\, \mu = \mu_0 + \frac{\delta}{\sqrt{n}}\right).$$

When the process is under control, it is approximated by the Edgeworth expansion as

$$
\begin{aligned}
A(k, 0) &= P\left(\bar{X} \leq \mu_0 - k\frac{\sigma}{\sqrt{n}} \,\Big|\, \mu = \mu_0\right) \\
&= P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \leq -k\right) \\
&= \Phi(-k) - \frac{1}{6\sqrt{n}}\kappa_1\left\{(-k)^2 - 1\right\}\phi(-k) \\
&\quad - \left[\frac{1}{24n}\kappa_2\left\{(-k)^3 - 3(-k)\right\} + \frac{1}{72n}\kappa_1^2\left\{(-k)^5 - 10(-k)^3 + 15(-k)\right\}\right]\phi(-k) + O\left(n^{-\frac{3}{2}}\right), \quad (2.6)
\end{aligned}
$$

where $\kappa_1 = E\left[(X - \mu)/\sigma\right]^3$ and $\kappa_2 = E\left[(X - \mu)/\sigma\right]^4 - 3$ stand for skewness and kurtosis, respectively. This reveals how the probability of detecting mean shift is affected by the degrees of skewness and kurtosis when the sample mean is based on n observations. The term of order $n^{-1/2}$ in (2.6) corrects the basic normal approximation for the main effect of skewness, while the term of order n^{-1} corrects the main effect of kurtosis and the secondary effect of skewness. Similarly, we can get the Edgeworth expansion of $P(\bar{X} \geq \text{UCL})$ under control. Due to the symmetry of $\Phi(\cdot)$ and $\phi(\cdot)$, the in-control ARL is then approximated by

$$E\left[N\,|\,\mu = \mu_0\right] = \left[2\Phi(-k) + \frac{1}{n}\left\{\frac{1}{12} \times \kappa_2\left(k^3 - 3k\right) + \frac{1}{36} \times \kappa_1^2\left(k^5 - 10k^3 + 15k\right)\right\}\phi(-k) + O\left(n^{-\frac{3}{2}}\right)\right]^{-1}$$

showing how the ARL is affected by skewness and kurtosis theoretically. Interestingly, the ARL has no contribution from the skewness of order $n^{-1/2}$ but is affected by the skewness and kurtosis of order n^{-1}.

Suppose that the process is out-of-control, and the process mean has shifted to $\mu = \mu_0 + \delta/\sqrt{n}$. Then the probability $A(k, \delta)$ is approximated as

$$
\begin{aligned}
A(k, \delta) &= P\left(\bar{X} \leq \mu_0 - k\frac{\sigma}{\sqrt{n}} \,\Big|\, \mu = \mu_0 + \frac{1}{\sqrt{n}}\delta\right) \\
&= P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{\mu_0 - \{\mu_0 + (1/\sqrt{n})\delta\}}{\sigma/\sqrt{n}} - k\right) \\
&= P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq -k - \frac{\delta}{\sigma}\right) \\
&= \Phi\left(-k - \frac{\delta}{\sigma}\right) - \frac{1}{\sqrt{n}}\frac{1}{6}\kappa_1\left\{\left(-k - \frac{\delta}{\sigma}\right)^2 - 1\right\}\phi\left(-k - \frac{\delta}{\sigma}\right) - \frac{1}{n}\left[\frac{1}{24}\kappa_2\left\{\left(-k - \frac{\delta}{\sigma}\right)^3 - 3\left(-k - \frac{\delta}{\sigma}\right)\right\}\right. \\
&\quad \left. + \frac{1}{72}\kappa_1^2\left\{\left(-k - \frac{\delta}{\sigma}\right)^5 - 10\left(-k - \frac{\delta}{\sigma}\right)^3 + 15\left(-k - \frac{\delta}{\sigma}\right)\right\}\right]\phi\left(-k - \frac{\delta}{\sigma}\right) + O\left(n^{-\frac{3}{2}}\right).
\end{aligned}
$$

Therefore, we can get out-of-control ARL as

$$E\left[N\,\middle|\,\mu = \mu_0 + \frac{1}{\sqrt{n}}\delta\right] = \{1 + A(k, \delta) - A(-k, \delta)\}^{-1}. \tag{2.7}$$

Unlike the in-control ARL, the out-of-control ARL is affected by the skewness of order $n^{-1/2}$.

2.2. The case where μ is unknown but σ is known

In practice, the control limits in (2.1) and (2.2) are estimated from a reference sample to apply prospectively to monitor the changes of the process mean since the control parameters are generally unknown. Here we assume that the mean of the underlying distribution is unknown, but the variance is known. Suppose that a reference random sample of m subgroups, $x_{i1}, x_{i2}, \ldots, x_{in}$, for $i = 1, 2, \ldots, m$ is available from an in-control process. Then, the in-control process mean μ and standard deviation σ are estimated from reference samples. Here, $\bar{\bar{X}}$, is the mean of means of subgroups $\bar{X}_i$ in the m reference subgroups and $\bar{R}$ is the mean of ranges of subgroups R_i in the ith reference subgroup each containing n observations. With $\hat{\mu} = \bar{\bar{X}}$ in control limits, the probability $A(k, 0)$ is approximated by

$$A(k, 0) = P\left(\bar{X} \le \hat{\mu} - k\frac{\sigma}{\sqrt{n}}\right) = P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \le -k + \frac{1}{\sigma}\frac{U}{\sqrt{m}}\right)$$

$$= E_U\left[P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \le -k + \frac{1}{\sigma}\frac{u}{\sqrt{m}}\right)\,\middle|\,U = u\right].$$

where $\hat{\mu} = \mu_0 + U/\sqrt{mn}$ and $U = \mu_0 + O_p((mn)^{-1/2})$. Given $U = u$,

$$P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \le -k + \frac{1}{\sigma}\frac{u}{\sqrt{m}}\right)$$

$$= \Phi\left(-k + \frac{1}{\sigma}\frac{u}{\sqrt{m}}\right) - \frac{1}{6\sqrt{n}}\kappa_1\left\{\left(-k + \frac{1}{\sigma}\frac{u}{\sqrt{m}}\right)^2 - 1\right\}\phi\left(-k + \frac{1}{\sigma}\left(\frac{u}{\sqrt{m}}\right)\right)$$

$$- \frac{1}{24n}\kappa_2\left\{\left(-k + \frac{1}{\sigma}\frac{u}{\sqrt{m}}\right)^3 - 3\left(-k + \frac{1}{\sigma}\frac{u}{\sqrt{m}}\right)\right\} \times \phi\left(-k + \frac{1}{\sigma}\left(\frac{u}{\sqrt{m}}\right)\right)$$

$$- \frac{1}{72n}\kappa_1^2\left\{\left(-k + \frac{1}{\sigma}\frac{u}{\sqrt{m}}\right)^5 - 10\left(-k + \frac{1}{\sigma}\frac{u}{\sqrt{m}}\right)^3 + 15\left(-k + \frac{1}{\sigma}\frac{u}{\sqrt{m}}\right)\right\} \times \phi\left(-k + \frac{1}{\sigma}\frac{u}{\sqrt{m}}\right) + o\left(n^{-1}\right)$$

$$= \left\{\Phi(-k) + \phi(-k)\frac{1}{\sigma}\frac{u}{\sqrt{m}} + \frac{1}{2}\phi'(-k)\frac{1}{\sigma^2}\frac{u^2}{m} + O\left(m^{-\frac{3}{2}}\right)\right\} - \frac{1}{6\sqrt{n}}\kappa_1\left\{\left((-k)^2 - 1\right) + \frac{1}{\sigma^2}\frac{u^2}{m}\right\}$$

$$\times \left\{\phi(-k) + \phi'(-k)\frac{1}{\sigma}\frac{u}{\sqrt{m}} + \frac{1}{2}\phi''(-k)\frac{1}{\sigma^2}\frac{u^2}{m} + O\left(m^{-\frac{3}{2}}\right)\right\}$$

$$- \frac{1}{24n}\kappa_2\left\{\left((-k)^3 - 3(-k)\right) + 3(-k)\frac{1}{\sigma^2}\frac{u^2}{m} + O\left(m^{-\frac{3}{2}}\right)\right\}$$

$$\times \left\{\phi(-k) + \phi'(-k)\frac{1}{\sigma}\left(\frac{u}{\sqrt{m}}\right) + \frac{1}{2}\phi''(-k)\frac{1}{\sigma^2} \cdot \frac{u^2}{m} + O\left(m^{-\frac{3}{2}}\right)\right\}$$

$$- \frac{1}{72n} \kappa_1^2 \left\{ \left((-k)^5 - 10(-k)^3 + 15(-k) \right) + 10(-k)^3 \frac{1}{\sigma^2} \cdot \frac{u^2}{m} - 30(-k) \frac{1}{\sigma^2} \cdot \frac{u^2}{m} \right\}$$

$$\times \left\{ \phi(-k) + \phi'(-k) \frac{1}{\sigma} \frac{u}{\sqrt{m}} + \frac{1}{2} \phi''(-k) \frac{1}{\sigma^2} \cdot \frac{u^2}{m} + O\left(m^{-\frac{3}{2}} \right) \right\} + o\left(n^{-1} \right).$$

Taking expectation with respect to U and $A(k, 0)$ becomes

$$E_U \left[P\left(\frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \le -k + \frac{1}{\sigma} \frac{u}{\sqrt{m}} \, \middle| \, U = u \right) \right]$$

$$= \left\{ \Phi(-k) + \frac{1}{2} \phi'(-k) \frac{1}{\sigma^2} \left(\frac{\sigma_U^2}{m} \right) + O\left(m^{-\frac{3}{2}} \right) \right\} - \frac{1}{6\sqrt{n}} \kappa_1 \left\{ \left((-k)^2 - 1 \right) + \frac{1}{\sigma^2} \frac{\sigma_U^2}{m} + O\left(m^{-\frac{3}{2}} \right) \right\} \phi(-k)$$

$$- \frac{1}{24n} \kappa_2 \left\{ \left((-k)^3 - 3(-k) \right) + 3(-k) \frac{1}{\sigma^2} \frac{\sigma_U^2}{m} \right\} \phi(-k) + O\left(m^{-\frac{3}{2}} \right)$$

$$- \frac{1}{72n} \kappa_1^2 \left\{ \left((-k)^5 - 10(-k)^3 + 15(-k) \right) + 10(-k)^3 \frac{1}{\sigma^2} \frac{\sigma_U^2}{m} - 30(-k) \frac{1}{\sigma^2} \frac{\sigma_U^2}{m} \right\} \phi(-k) + O\left(m^{-\frac{3}{2}} \right)$$

$$= \Phi(-k) + \frac{1}{2} \phi'(-k) \frac{1}{\sigma^2} \frac{\sigma_U^2}{m} - \frac{1}{6\sqrt{n}} \kappa_1 \left\{ \left((-k)^2 - 1 \right) + \frac{1}{\sigma^2} \frac{\sigma_U^2}{m} \right\} \phi(-k)$$

$$- \frac{1}{24n} \kappa_2 \left\{ \left((-k)^3 - 3(-k) \right) + 3(-k) \frac{1}{\sigma^2} \frac{\sigma_U^2}{m} \right\} \times \phi(-k)$$

$$- \frac{1}{72n} \kappa_1^2 \left\{ \left((-k)^5 - 10(-k)^3 + 15(-k) \right) + \left(10(-k)^3 - 30(-k) \right) \times \frac{1}{\sigma^2} \frac{\sigma_U^2}{m} \right\} \times \phi(-k) + O\left(m^{-\frac{3}{2}} \right).$$

Here, take $m = n^\alpha$. In particular, for $\alpha > 2$, we can obtain $A(k, 0)$ as

$$A(k, 0) \approx \Phi(-k) - \frac{1}{6\sqrt{n}} \kappa_1 \left\{ (-k)^2 - 1 \right\} \phi(-k)$$

$$- \frac{1}{n} \left\{ \frac{1}{24} \kappa_2 \left((-k)^3 - 3(-k) \right) + \frac{1}{72} \kappa_1^2 \left((-k)^5 - 10(-k)^3 + 15(-k) \right) \right\} \phi(-k).$$

If a reference subgroup size is large enough satisfying $m = n^{2+\alpha}$ for $\alpha > 0$, this corresponds to the case where the process mean and variance are known. Similarly, the out-of-control ARL can be given as in (2.7) under the estimated control limits. Consequently, we examine that the in-control ARL is affected by the skewness and kurtosis of order n^{-1} while the out-of-control ARL is by the skewness of order $n^{-1/2}$.

3. Numerical Studies

In the previous section, we presented the asymptotic in-control ARL and out-of-control ARL based on the Edgeworth expansion of the false alarm rate. In this section, we consider the behavior of ARL in small subgroups under control based on simulation studies. Particularly we focus on the in-control ARL values according to the changes of a skewness for a skew-normal distribution. For numerical evaluation concerning out-of-control ARL, it needs to make and adjustment of control limits to a prespecified FAR; therefore, we do not deal with out-of-control ARL here. Depending on the knowledge of the process parameters μ and σ, we divide simulation studies into two cases: (1) both are known and (2) μ is unknown but σ is known.

To explain how the in-control ARL values are affected by the degrees of skewness, we assume that the process follows a skew-normal distribution where the departure of the underlying distribution from the normal distribution is purely regulated by the skewness. For these distributions, refer to Azzalini (1985, 1986), Azzalini and Dalla-Valle (1996), and Azzalini and Capitanio (1999). The skew-normal distribution, denoted by $SN(\mu, \sigma^2, \alpha)$, has the probability density function

$$f\left(x \mid \mu, \sigma^2, \alpha\right) = \left(\frac{2}{\sigma}\right) \phi\left(\frac{x - \mu}{\sigma}\right) \Phi\left(\alpha\left(\frac{x - \mu}{\sigma}\right)\right), \tag{3.1}$$

where μ, σ and α are the location, scale, and skewness parameters, respectively. The skewness parameter α regulates the skewness in $(-\infty, \infty)$. Note that, if $\alpha = 0$, the density of X in (3.1) reduces to the density of the standard normal distribution. For simplicity, we assume that $\mu = 0$ and $\sigma = 1$ throughout numerical studies in this paper. Some properties of this distribution include:

$$E(X) = \sqrt{\frac{2}{\pi}} \delta \sigma, \qquad \text{Var}(X) = 1 - \frac{2}{\pi} \delta^2$$

with skewness and kurtosis measures given by

$$\kappa_3 = \frac{1}{2}(4 - \pi) \left(\frac{E^2(X)}{\text{Var}(X)}\right)^{\frac{3}{2}}, \qquad \kappa_4 = 2(\pi - 3) \left(\frac{E^2(X)}{\text{Var}(X)}\right)^2$$

where $\delta = \alpha / \sqrt{1 + \alpha^2}$.

3.1. The case where both μ and σ are known

In this subsection, we assume that the process mean and standard deviation, μ and σ are known. We evaluate the performance of the $\bar{X}$ control chart with the changes of skewness while the sample size is set at $n = 5, 7, 9$ and 25. For this purpose, the quality characteristics are assumed to follow the skew-normal distribution for various skewness parameter α. The control limits are set by Shewhart type with 3σ that makes the false alarm rate equal to 0.0027; therefore, the ARL values correspond to 370. In order to estimate the false alarm rate, 100,000 subgroups of size n were generated from the assumed distribution and compute the number of samples whose means are out-of-control limits. We repeated this process based on the same distribution with an increasing skewness parameter. The in-control ARL values are finally computed by the reciprocal of mean value of FARs'.

In Figure 1, skewness parameter α varies from 0 to 8 by 0.5 for the skew-normal distribution to examine the effect of skewness on the in-control ARL for sample sizes $n = 5, 7, 9$ and 25. These skew parameters produce the skewness, κ_3, from 0.00 to 0.93 with kurtosis between 0.00 to 0.80. The kurtosis increases moderately as the data become skewed. Note that the skew-normal distribution with $\alpha = 0$ corresponds to a normal distribution. In Figure 1(a), the in-control ARL tends to fall considerably as the skewness parameter becomes larger than 1.5. We observe that the in-control ARL decreases greatly as the skewness parameter α goes up to suggest that the data tends to be increasingly skewed. The gap between ARL values and 370 shown in Figure 1(b)–(d); however, it is reduced as the sample size increases.
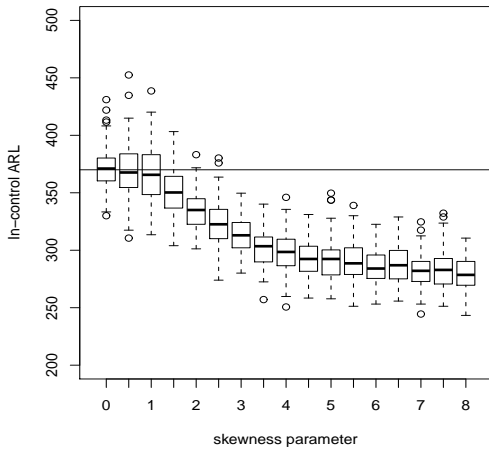
3.2. The case where μ is unknown but σ is known

In this subsection we consider the case that the process mean μ is unknown but standard deviation σ is known. To compute the control limits, we estimate the process mean μ based on two kinds of
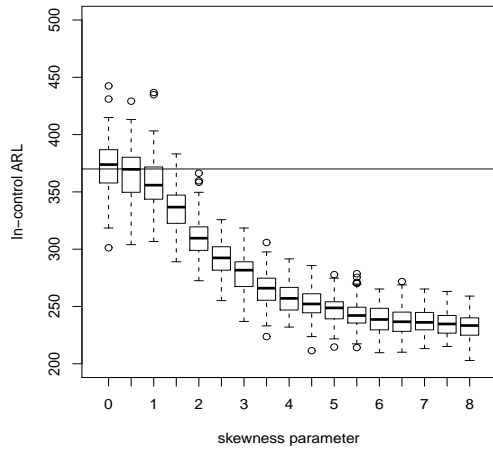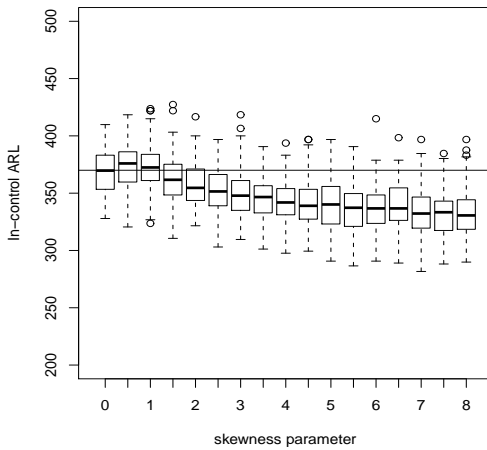
(a) In-control ARL with subgroup of size 5
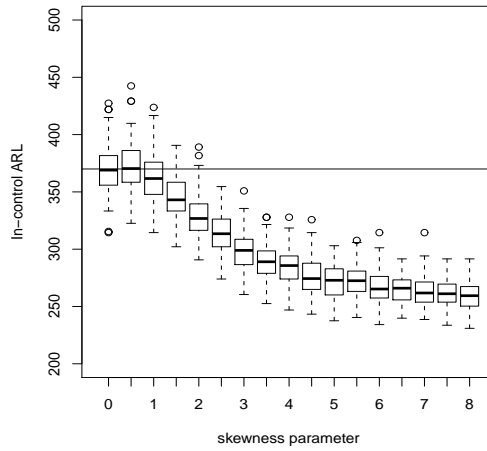
(b) In-control ARL with subgroup of size 7

(c) In-control ARL with subgroup of size 9

(d) In-control ARL with subgroup of size 25

Figure 1: *The in-control ARL values for $\bar{X}$ chart against the increase of skewness parameter α, $\alpha = 0, \ldots, 8$ by 0.5 that makes the skewness vary from 0.0 to 0.9.*

reference samples with $m = 20$ and $m = 200$ of size n. Let $\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_m$ be the average of each sample. Then $\bar{\bar{X}}$ are used as an estimate of the population μ. In this situation, the Shewhart-type control limits are written as

$$\text{LCL} = \bar{\bar{X}} - 3\frac{\sigma}{\sqrt{n}} \tag{3.2}$$

and

$$\text{UCL} = \bar{\bar{X}} + 3\frac{\sigma}{\sqrt{n}}. \tag{3.3}$$

(a) In-control ARL with subgroup of size 5

(b) In-control ARL with subgroup of size 7

(c) In-control ARL with subgroup of size 9
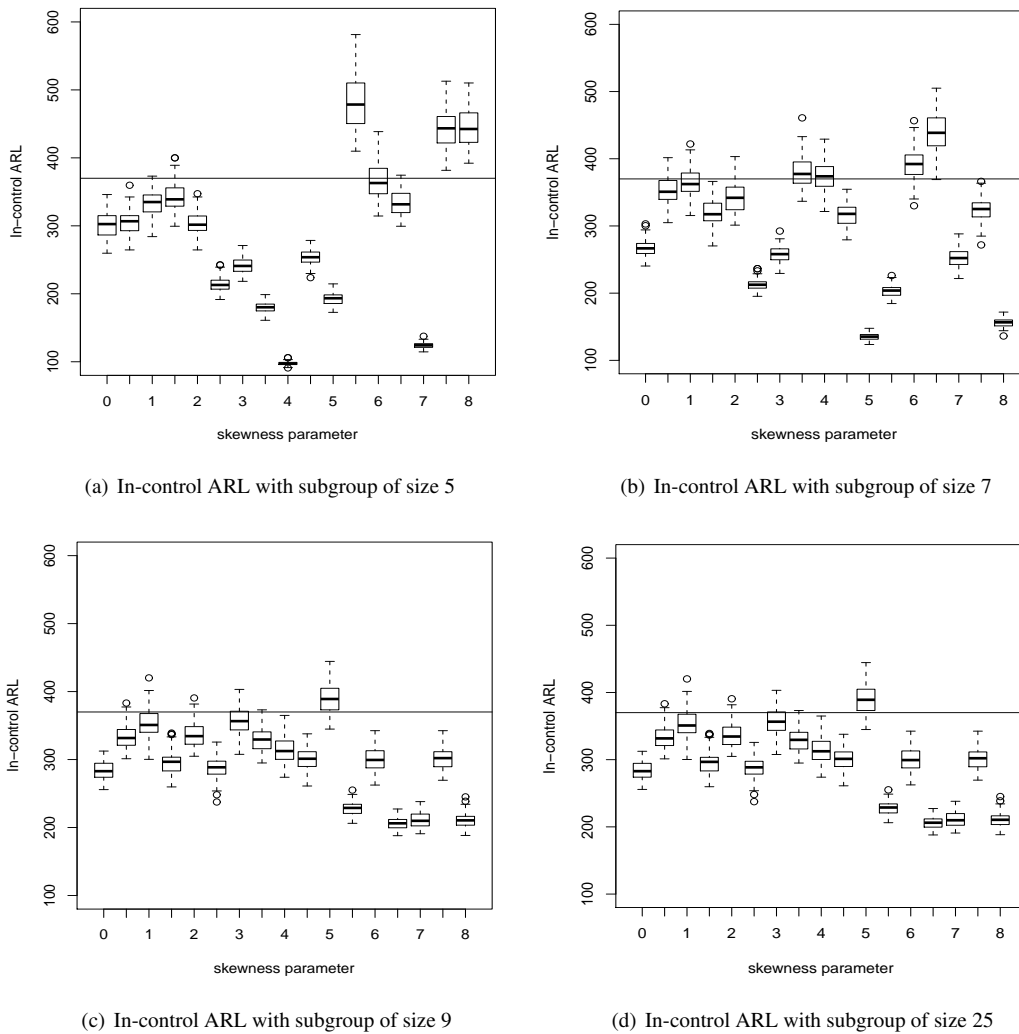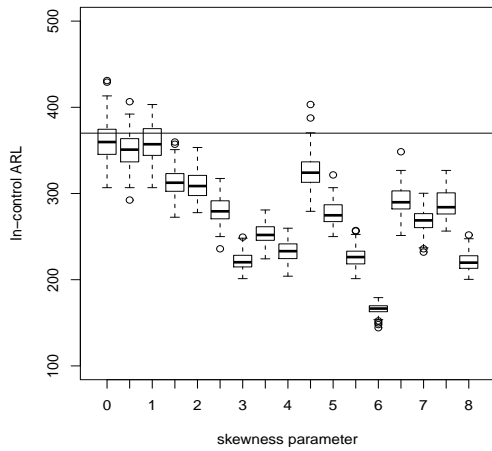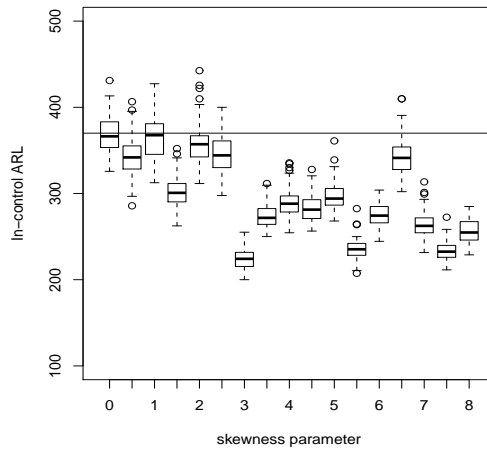
(d) In-control ARL with subgroup of size 25

Figure 2: *In-control ARL of $\bar{X}$ chart with the increase of skewness generated by the skew normal distribution with unknown μ but known σ. To compute control limits, the mean μ is estimated based on $m = 20$ reference group*

To estimate of the process mean μ with $\bar{\bar{X}}$, reference samples of size m are collected and then 100,000 samples of size n are generated from the same distribution to calculate the false alarm rate. We compute the number of samples whose means exceed the control limits in (3.2) and (3.3). We simulate this process for 100 kinds of reference samples in order to average out false alarm rates and this procedure is repeated by increasing skewness parameter.
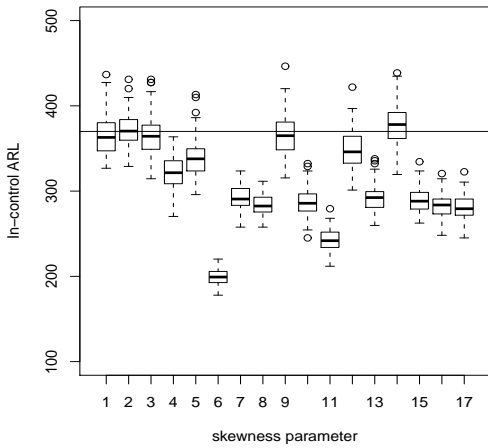
Figure 2 shows the in-control ARL with the skew-normal distribution with $m = 20$ reference samples for sample sizes $n = 5, 7, 9$ and 25. As shown in Figure 2, the in-control ARL values differ from 370 appreciably even when the skew parameter is equal to zero. This is same regardless of the sample size. This may result from the inconsistent estimator from the small reference. Interestingly, as displayed in Figure 3, these discrepancies disappear in a large reference sample of size 200. The
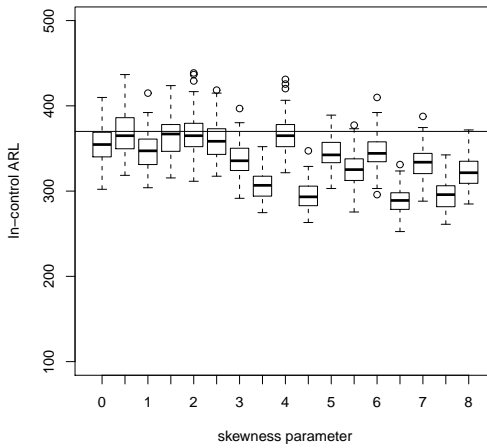
(a) In-control ARL with subgroup of size 5

(b) In-control ARL with subgroup of size 7

(c) In-control ARL with subgroup of size 9

(d) In-control ARL with subgroup of size 25

Figure 3: *In-control ARL of $\bar{X}$ chart with the increase of skewness generated by the skew normal distribution with unknown μ but known σ. To compute control limits, the mean μ is estimated based on $m = 200$ reference group.*

effects of the changes of skewness on the in-control ARL look similar those in Figure 1 even though the decreasing pattern of in-control ARL appears not smooth.

We therefore observe that the parameter estimation affects the chart performance. The in-control ARL is significantly lower regardless of the degree of skewness when the control limits are estimated based on small reference samples.

4. Conclusion

In this paper, we present how the ARL of $\bar{X}$ chart is affected by non-normality measured by skewness and kurtosis. They have an effect on the in-control ARL of order n^{-1}. Numerical studies with skew-

normal distribution also show the effect of skewness on the ARL when the subgroup size is small. Particularly, we found that the effect of skewed data is confounded by the estimation accuracy of the process parameters that are unknown. In such situation, the effect of deviation from normality on the ARL cannot be well-explained because the in-control ARL may be significantly affected by estimated control limits. To address the performance of $\bar{X}$ charts on the non-normality in practice (in which the process parameters are generally unknown) the efficient estimation of control limits (based on a large reference sample) needs to be developed as well as the chart design. We also observe that the in-control ARL tends to fall considerably with the changes of skewness; however, although its effect tends to disappear as the size of subgroup, n becomes large.

To sum up, the $\bar{X}$ control chart with usual subgroup size of $n = 4$ or 5 needs to be used carefully in real practice where the underlying distribution is usually unknown, although the in-control ARL have an effect in non-normality of order n^{-1}. In order to maintain in-control ARL in real practice to a prespecified level, the control limits need to be adjusted for the non-normality of underlying distribution.

Acknowledgements

References

Azzalini, A. (1985). A class of distributions which includes the normal ones, *Scandinavian Journal of Statistics*, **12**, 171–178.

Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones, *Statistics*, **44**, 199–208.

Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skewnormal distributions, *Journal of the Royal Statistical Society, Series B*, **61**, 579–602.

Azzalini, A. and Dalla-Valle, A. (1996). The multivariate skew-normal distribution, *Biometrika*, **83**, 715–726.

Bai, D. S. and Choi, I. S. (1995). $\bar{X}$ and R control charts for skewed populations, *Journal of Quality Technology*, **27**, 120–131.

Burr, I. W. (1967) The effect of non-normality on constants for $\bar{X}$ and R charts, *Industrial Quality Control*, **24**, 563–569.

Burrows, P. M. (1962). $\bar{X}$ control schemes for a production variable with skewed distribution, *Statistician*, **12**, 296–312.

Chakraborti, S., van der Laan, P. and van de Wiel, M. A. (2004). A class of distribution-free control charts, *Applied Statistics*, **53**, 443–462.

Chan, L. K., Hapuarachchi, K. P. and Macpherson, B. D. (1988). Robustness of X and R charts, *IEEE. Transactions on Reliability*, **37**, 117–123.

Janacek, G. J. and Meikle, S. E. (1997). Control charts based on medians, *The Statistician*, **46**, 19–31.

Roes, K. C. B. and Does, R. J. M. M. (1995). Shewhart-type charts in nonstandard situations(with discussions), *Technometrics*, **37**, 15–40.

Schilling, E. G. and Nelson, P. R. (1976). The effect of non-normality on the control limits of *X* chart, *Journal of Quality Technology*, **8**, 183–188.

Shewhart, W. A. (1931). *Economic Control of Quality of Manufactured Product*, D. Van Nostrand, New York.

Yourstone, S. A. and Zimmer, W. J. (1992). Non-normality and the design of control charts for averages, *Decision Sciences*, **23**, 1099–1113.

Wheeler, D. J. and Chambers, D. S. (1992). *Understanding Statistical Process Control*, SPC Press, Knoxville, TN.