Communications in Statistics - Simulation and Computation

Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597237

Robustness of R-Chart to Non Normality
Shih-Chou Kao a; Chuanching Ho b

a Institute of Operation and Management, Kao Yuan University, Taiwan, R.O.C.
b Chung-Shan Institute of Science and Technology, Taiwan, R.O.C.

Online Publication Date: 01 September 2007
To cite this Article: Kao, Shih-Chou and Ho, Chuanching (2007) 'Robustness of R-Chart to Non Normality', Communications in Statistics - Simulation and Computation, 36:5, 1089 - 1098
To link to this article: DOI: 10.1080/03610910701540003
URL: http://dx.doi.org/10.1080/03610910701540003

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf
This article maybe used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Quality Control

Robustness of R-Chart to Non Normality

SHIH-CHOU KAO¹ AND CHUANCHING HO²

¹Institute of Operation and Management, Kao Yuan University, Taiwan, R.O.C.
²Chung-Shan Institute of Science and Technology, Taiwan, R.O.C.

This study addresses the appropriate $d_3$ values for constructing range control charts (R-charts) when the distributions of the processes are the uniform, triangular, exponential, and Erlang. Comparisons of the range charts are based on Type I error probabilities obtained using simulations. The results reveal that inappropriate use of the $d_3$ values strongly affected the performance of the R-charts. Practitioners should be more careful in selecting suitable coefficients when using R-charts methods to process data. The distribution of the processes must be examined before the coefficients are chosen.

Keywords
$d_3$: False alarm rate (Type I error probability); Range control chart.

Mathematics Subject Classification
Primary 62P30; Secondary 62F10.

1. Introduction

Tippett (1925) provided the properties of sample ranges taken from a normal population. Patnak (1950) then proposed the use of the mean range to estimate population variance. Since then, sample ranges have been extensively used in industry to estimate process standard deviation and construct control charts limits. Doing so involves the derivation and use of coefficients such as $d_2$, $d_1$, $D_3$, and $D_4$. Notably, the coefficients thus derived are based on the normal distribution. However, in the real world, this assumption does not hold for various processes. This work determines the false alarm rate (Type I error probabilities) that applicable were the underlying distribution of the process to be far from normal.

Most previous work has studied the effects of the $d_2$ coefficient on range control charts. Recently, Mahoney (1998), who considered only the performance of $\bar{X}$ charts, investigated the effect of parent population distributions on $d_2$. The population distributions were uniform, triangular, exponential, Erlang, and normal.
He concluded that incorrect use of the $d_2$ values increased the false alarm rate of the range charts.

This work investigates the effects of another coefficient, $d_3$, on the $R$-charts. The method applied is as that of Mahoney (1998). The $d_3$ values for the uniform, triangular, exponential, and Erlang distributions were derived and compared with those for the normal distribution. The rest of this article is organized as follows. Section 2 presents the derivation and a general form of $\sigma_R$. Section 3 gives $\sigma_R$ and $d_3$ for the uniform, triangular, exponential, and Erlang distributions. Section 4 compares the false alarm rate of the range control charts given these population distributions. Finally, Sec. 5 draws the conclusions.

2. Derivation of $\sigma_R$

Define $f(x)$ as a probability density function of a random variable $x$, $a \leq x \leq b$, where $a$ and $b$ are the lower bound and upper bound, respectively, on the variable $x$. The cumulative distribution function is defined as

$$y = P(x) = 1 - Q(x) = \int_a^x f(t)dt.$$ 

The expectation and variance of $x$ are

$$\mu = E(x) = \int_a^b xf(x)dx,$$

$$\sigma^2 = E(x - \mu)^2 = E(x^2) - (E(x))^2.$$ 

A random sample of size $n$ is taken from the population and the range of this sample is defined as

$$R = x_{\text{max}} - x_{\text{min}}.$$ 

The expectation and standard deviation of the sample ranges (Appendix A) are

$$\mu_R = E(R) = d_2\sigma,$$

$$\sigma_R = \left\{2 \int_a^b \int_s^{y_L} \left[1 - y_L^n - (y_L - y_S)^n + (y_L - y_S)^n - y_L^n\right]dy_Sdx_L - \mu_R^2\right\}^{1/2},$$ (1)

where $x_L$ is the largest value of $x$ and $x_S$ is the smallest value of $x$. $y_L$, $y_S$, $Q_L$, and $Q_S$ are similarly defined. Mahoney (1998) provided a formula of $\mu_R$ as

$$\mu_R = \begin{cases} \sum_{i=1}^{n-1} \binom{n}{i} (-1)^{i+1} \int_a^b Q' dx & \text{if } n \text{ is odd} \\ \sum_{i=1}^{n-1} \binom{n}{i} (-1)^{i+1} \int_a^b Q' dx - 2 \int_a^b Q^2 dx & \text{if } n \text{ is even} \end{cases}.$$
Robustness of R-Chart to Non Normality

Substituting $\mu_R$ into (1) and applying a binomial transformation yields the following two forms of $\sigma_R$ (Appendix B).

$$\sigma_R = \left[ 2 \int_a^b \int_a^{x_L} \sum_{i=1}^{n-1} \binom{n}{i} \left\{ (-1)^{i+1} y_S^i + (-1)^i y_L^{n-i} y_S^i \right\} dx_L dx_L - \mu_R^2 \right]^{1/2},$$

or,

$$\sigma_R = \left[ 2 \int_a^b \int_a^{x_L} \sum_{i=1}^{n-1} \binom{n}{i} \left\{ (-1)^{i+1} Q'_L + (-1)^i Q''_L \right\} dx_L dx_L - \mu_R^2 \right]^{1/2}.$$  (3)

Equations (2) and (3) are equivalent since $y = 1 - Q(x)$. The values of the $d_3$ coefficient can be determined from $\sigma_R/\sigma$ because $\sigma_R = d_3 \sigma$.

3. $\sigma_R$ and $d_3$ for the Uniform, Triangular, Exponential, and Erlang Distribution

As stated above, the $d_3$ coefficient can be obtained by determining $\sigma_R/\sigma$. Accordingly, $\sigma_R$ is initially found for each distribution and then $d_3$ is determined.

3.1. Uniform Distribution

The probability density function of the uniform distribution of interest is defined as

$$f(x) = 1 \quad 0 \leq x \leq 1.$$  

The standard deviation of the sample range is

$$\sigma_R = \left[ 2 \int_0^1 \int_0^{x_L} \binom{n}{i} \left\{ (1 - x_L^n) - (1 - x_L^n) + (x_L - x_S^n) \right\} dx_S dx_L - \mu_R^2 \right]^{1/2}$$

$$= \left[ 1 - \frac{4n + 2}{(n + 1)(n + 2)} - \mu_R^2 \right]^{1/2}.$$  

The $d_3$ coefficients are obtained by substituting $\mu_R$ into $\sigma_R$ and then divided by the standard deviation of the population. This procedure also applies to the following triangular, exponential, and Erlang distributions.

3.2. Triangular Distribution

The probability density function of the Beta distribution is described as follows, where $\alpha = v_1/2$, $\beta = v_2/2$ and $v_1$ and $v_2$ are the numbers of degrees-of-freedom of an $F$ variable.

The distribution is right triangular when $\alpha = 2$ and $\beta = 1$:

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1 - x)^{\beta-1} \quad 0 < x < 1, \alpha, \beta > 0.$$
The standard deviation of the sample range is as follows when \( \beta = 1 \) and \( y = x^2 \):

\[
\sigma_R = \left[ 2 \sum_{i=1}^{n-1} \binom{n}{i} \left\{ \frac{(1)_{i+1}}{i!} \int_0^\infty \int_{x_L}^\infty (1 + x_L + \ldots + (n-1)_{i+1}x_L) \left( 1 + \frac{x_L + \ldots + (n-1)_{i+1}x_L}{(n-1)_{i+1}x_L} \right) dx_L dx_S \right\} \right]^{1/2}
\]

or:

\[
= \left[ 2 \sum_{i=1}^{n-1} \binom{n}{i} \left\{ \frac{(1)_{i+1}}{i!} \int_0^\infty \int_{x_L}^\infty (1 + x_L + \ldots + (n-1)_{i+1}x_L) \right\} dx_L dx_S \right]^{1/2} - \mu_R^2
\]

3.3. Exponential Distribution

The probability density function of the exponential distribution of interest and the standard deviation of the sample range are defined as follows, and \( \lambda \) is the reciprocal of the mean:

\[
f(x) = \lambda e^{-\lambda x}, \quad x \geq 0, \quad \lambda > 0
\]

\[
\sigma_R = \left[ 2 \sum_{i=1}^{n-1} \binom{n}{i} \left\{ \frac{(1)_{i+1}}{i!} \int_0^\infty \int_{x_L}^\infty (1 + x_L + \ldots + (n-1)_{i+1}x_L) \right\} dx_L dx_S \right]^{1/2} - \mu_R^2
\]

3.4. Erlang Distribution

The probability density function of the Erlang distribution is as follows:

\[
f(x) = \frac{x^{\alpha-1} e^{-x}}{\Gamma(\alpha)}, \quad x \geq 0, \quad \alpha = 1, 2, \ldots.
\]

Let \( \alpha = 2 \); the probability density function becomes

\[
f(x) = x e^{-x}, \quad 0 \leq x \leq \infty.
\]

The cumulative distribution function is \( y = 1 - (1 + x)e^{-x} \), and hence, \( Q = (1 + x)e^{-x} \). The expression for \( \sigma_R \) is as follows:

\[
\sigma_R = \left[ 2 \sum_{i=1}^{n-1} \binom{n}{i} \left\{ \frac{(1)_{i+1}}{i!} \int_0^\infty \int_{x_L}^\infty (1 + x_L + \ldots + (n-1)_{i+1}x_L) \right\} dx_L dx_S \right]^{1/2} - \mu_R^2
\]

or:

\[
= \left\{ 2 \sum_{i=1}^{n-1} \binom{n}{i} \left\{ \frac{(1)_{i+1}}{i!} \int_0^\infty \int_{x_L}^\infty (1 + x_L + \ldots + (n-1)_{i+1}x_L) \right\} dx_L dx_S \right\}^{1/2} - \mu_R^2
\]

or:

\[
= \left\{ 2 \sum_{i=1}^{n-1} \binom{n}{i} \left\{ \frac{(1)_{i+1}}{i!} \sum_{j=0}^{i-1} \frac{1}{j!} \sum_{k=0}^{i} \frac{1}{k!} \sum_{l=0}^{k} a_{i,j,k,l} \right\} \right\}^{1/2} - \mu_R^2
\]

where

\[
a_{i,j,k,l} = \frac{(n-i-j)^l}{l^{i+j-l}} \sum_{k=0}^{i} \frac{1}{k!} (n^{i+j-k} - 1)
\]
Robustness of R-Chart to Non Normality

Table 1 presents $d_3$ values for the population distributions considered herein. Values of $d_3$ for the normal distribution are widely available in quality control textbooks, such as that written by Grant and Leavenworth (1996). Let $S(n)$ be the smallest $d_3$ values from the uniform, triangular, exponential, and Erlang distributions, $L(n)$ be the largest such value, and $d_{3\text{Nor}}$ be the value for the normal distribution. Define $M_S$ and $M_L$ as follows:

$$M_S = \frac{S(n) - d_{3\text{Nor}}}{d_{3\text{Nor}}} \times 100\%$$

$$M_L = \frac{L(n) - d_{3\text{Nor}}}{d_{3\text{Nor}}} \times 100\%.$$  

The results show that:

1. The $d_3$ value decreases as $n$ increases for the uniform, triangular, and normal distributions, indicating that $\sigma_R$ decreases as $n$ increases for each of these three distributions. This fact indicates that, for these three distributions, from the perspective of reducing the variation in the sample ranges, large sample sizes are preferred in statistical process control applications to reduce the variation in the sample ranges. Notably, these three distributions are symmetric.

2. For the Erlang distribution, the $d_3$ value increases first, is stabilized from $n = 5–10$, and then decreases as $n$ increases. The results seem to suggest that using either smaller sample sizes or larger sample sizes in a statistical process control application for this particular distribution, is preferred to reduce the variation in the sample ranges. Sample sizes that are commonly used, such as $n = 5–10$, are inappropriate.

3. The $d_3$ value of the exponential distribution increases with $n$; it appears to approach asymptotically a value of approximately 1.283. Notably, the exponential distribution is highly skewed. This is regarded as the reason why the variation of the sample ranges increases with the sample sizes.

<table>
<thead>
<tr>
<th>$n$</th>
<th>Uniform</th>
<th>Triangular</th>
<th>Exponential</th>
<th>Erlang</th>
<th>Normal</th>
<th>$M_S$ (%)</th>
<th>$M_L$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.816</td>
<td>0.849</td>
<td>1.000</td>
<td>0.935</td>
<td>0.853</td>
<td>−4.34</td>
<td>17.23</td>
</tr>
<tr>
<td>3</td>
<td>0.775</td>
<td>0.849</td>
<td>1.118</td>
<td>1.019</td>
<td>0.888</td>
<td>−12.73</td>
<td>25.90</td>
</tr>
<tr>
<td>4</td>
<td>0.694</td>
<td>0.803</td>
<td>1.167</td>
<td>1.045</td>
<td>0.880</td>
<td>−21.14</td>
<td>32.61</td>
</tr>
<tr>
<td>5</td>
<td>0.619</td>
<td>0.752</td>
<td>1.194</td>
<td>1.055</td>
<td>0.864</td>
<td>−28.36</td>
<td>38.19</td>
</tr>
<tr>
<td>6</td>
<td>0.555</td>
<td>0.707</td>
<td>1.210</td>
<td>1.059</td>
<td>0.848</td>
<td>−34.55</td>
<td>42.69</td>
</tr>
<tr>
<td>7</td>
<td>0.500</td>
<td>0.668</td>
<td>1.221</td>
<td>1.061</td>
<td>0.833</td>
<td>−39.98</td>
<td>46.58</td>
</tr>
<tr>
<td>8</td>
<td>0.451</td>
<td>0.638</td>
<td>1.229</td>
<td>1.060</td>
<td>0.820</td>
<td>−45.00</td>
<td>49.88</td>
</tr>
<tr>
<td>9</td>
<td>0.420</td>
<td>0.609</td>
<td>1.236</td>
<td>1.059</td>
<td>0.808</td>
<td>−48.02</td>
<td>52.97</td>
</tr>
<tr>
<td>10</td>
<td>0.388</td>
<td>0.581</td>
<td>1.239</td>
<td>1.060</td>
<td>0.797</td>
<td>−51.32</td>
<td>55.46</td>
</tr>
<tr>
<td>20</td>
<td>0.219</td>
<td>0.425</td>
<td>1.262</td>
<td>1.047</td>
<td>0.729</td>
<td>−69.96</td>
<td>73.11</td>
</tr>
<tr>
<td>50</td>
<td>0.102</td>
<td>0.273</td>
<td>1.275</td>
<td>1.034</td>
<td>0.652</td>
<td>−84.36</td>
<td>95.52</td>
</tr>
</tbody>
</table>
The $d_3$ values for the uniform and the triangular distributions are always smaller than those for the normal distribution, while the reverse is true for the exponential and Erlang distributions, perhaps because the random variable of the uniform and triangular distributions are bounded, even though the distributions are symmetrical. The random variable of a normal distribution is unbounded. The $d_3$ values from the exponential and Erlang distributions exceed those of a normal distribution because the former two distributions are skewed.

Of the distributions studied herein, the uniform distribution always yielded the lowest $d_3$ values, while the exponential distribution always yielded the highest. The absolute values of both $M_S$ and $M_L$ increase with $n$. This fact implies that $d_3$ values for non-normal distributions deviate substantially from those of the normal distributions as $n$ increases.

### 4. Comparing False Alarm Rate Associated with R-Chart

A Monte Carlo simulation with a procedure similar to that presented by Bai and Choi (1995) is developed to examine the performance of $R$-charts for the distributions considered herein, based on their false alarm rates. The simulation is performed in two stages. In the first stage, the control limits of a range chart are established using 1,000 subgroup ranges each of size $n$. In the second stage, the false alarm rate ($\alpha$) is obtained by generating a million samples, calculating and comparing the ranges to the control limits established in the first stage. The simulation program is coded using the FORTRAN language and all of the random numbers are generated from its IMSL library throughout the simulation.

The performance of the $R$-charts is compared in two situations. Situation 1 uses ($d_2$, $d_3$) values from the normal distribution to construct the range chart control limits, independently of the original distribution of the population. Situation 2 uses suitable ($d_2$, $d_3$) values of the original population distribution to determine the control chart limits. Specifically, the exact $d_3$ values are those from Mahoney (1998). Notably, although the $d_3$ values of the uniform, triangular, exponential, and Erlang distributions differ from those of the normal distribution, the deviations are not considerable.

Table 2 provides the false alarm rates of the $R$-charts for the population distributions considered herein. Those of the normal distribution are calculated using Pearson and Hartley’s formula (1942).

Consider only the results in Table 2 concerning situation 1. The false alarm rates are all zero for the uniform and triangular distributions, and increase with $n$ for the exponential and Erlang distributions. A comparison with the false alarm rates from the normal distribution reveals that the false alarm rates of the uniform and triangular distributions are always underestimate, while the reverse holds for the exponential and Erlang distributions, when inappropriate ($d_2$, $d_3$) values are used. This phenomenon can also be explained as follows. Notably, the upper control limit of a $R$-chart is determined to be $D_4\overline{R}$ where $D_4 = 1 + 3(d_3/d_2)$. Suppose $d_3$ is a constant; then $D_4$ increases with $d_3$. The increase in $D_4$ widens the control limit, and so reduces the false alarm rate. This is exactly the case for the uniform and triangular distributions. Also, this observation strongly relates to the shape of the distribution and the restriction of its random variable, as stated above. For instance, the false alarm rates of the range charts for the uniform and triangular distributions...
Robustness of $R$-Chart to Non Normality

Table 2
The false alarm rates

<table>
<thead>
<tr>
<th>$n$</th>
<th>Uniform</th>
<th>Triangular</th>
<th>Exponential</th>
<th>Erlang</th>
<th>Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0383</td>
<td>0.0184</td>
</tr>
<tr>
<td>3</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0418</td>
<td>0.0155</td>
</tr>
<tr>
<td>4</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0451</td>
<td>0.0144</td>
</tr>
<tr>
<td>5</td>
<td>0.0000</td>
<td>0.0014</td>
<td>0.0000</td>
<td>0.0481</td>
<td>0.0138</td>
</tr>
<tr>
<td>6</td>
<td>0.0000</td>
<td>0.0035</td>
<td>0.0000</td>
<td>0.0504</td>
<td>0.0135</td>
</tr>
<tr>
<td>7</td>
<td>0.0000</td>
<td>0.0052</td>
<td>0.0000</td>
<td>0.0528</td>
<td>0.0132</td>
</tr>
<tr>
<td>8</td>
<td>0.0000</td>
<td>0.0066</td>
<td>0.0000</td>
<td>0.0548</td>
<td>0.0132</td>
</tr>
<tr>
<td>9</td>
<td>0.0000</td>
<td>0.0074</td>
<td>0.0000</td>
<td>0.0565</td>
<td>0.0130</td>
</tr>
<tr>
<td>10</td>
<td>0.0000</td>
<td>0.0080</td>
<td>0.0000</td>
<td>0.0586</td>
<td>0.0129</td>
</tr>
<tr>
<td>20</td>
<td>0.0000</td>
<td>0.0114</td>
<td>0.0000</td>
<td>0.0733</td>
<td>0.0123</td>
</tr>
<tr>
<td>50</td>
<td>0.0000</td>
<td>0.0132</td>
<td>0.0000</td>
<td>0.0935</td>
<td>0.0121</td>
</tr>
</tbody>
</table>

are zero because the distributions are symmetrical and that their variables are restricted. The use of inappropriate $d_3$ values broadens the upper control limit of the $R$-charts and makes the false alarm rate zero.

Consider the results in Table 2 regarding situation 2. Recall that this situation involves the suitable $(d_2, d_3)$ combination of coefficients. The false alarm rate increases with $n$ for the uniform and triangular distributions because the $d_3$ value decreases as $n$ increases for both of these two distributions, so $D_4$ decreases, increasing the false alarm rate with $n$. In contrast, the false alarm rate decreases as $n$ increases for the exponential and Erlang distributions because the $d_3$ value increases with $n$. This result may be treated as acceptable if the false alarm rate of the $R$-chart for the normal distribution is said to be approximately 0.005; the practice is to let the false alarm rate be approximately the same for both charts. Then, the results in Table 2 seem to indicate that the preferred sample size is 7 if the population distribution is uniform, 50 if it is exponential, and the-larger-the-better if the population distributions are skewed and the variables unbounded.

Comparing the false alarm rates across the distributions reveals that the false alarm rate is always the largest for the exponential distribution and always exceeds that for the normal distribution; also, the false alarm rate is always smallest for the triangular distribution, and always smaller than that for the normal distribution.

5. Conclusions

The results presented in this work demonstrate that for the four non normal distributions considered herein, the use of $d_3$ values computed based on an assumption of normality seriously affects the performance of range charts. Practitioners should be aware of this potential problem and should take care when applying $R$-charts to the process data. As in any statistical process control application, the underlying distribution of the process data should be examined initially to validate any assumptions. In the case of $R$-charts, the $d_3$ values for the
normal distribution should only be used when the data do not deviate substantially from normality. Otherwise, $d_3$ coefficients should be chosen with great care.

**Appendix A: Derivation of $\sigma_R^2$**

Let $f(x)$ be the probability density function of a random variable $x$, $a \leq x \leq b$. The cumulative distribution function of $x$ can be expressed as $P(x) = y$, $0 \leq y \leq 1$. A random sample of size $n$ is taken from the population of $x$ and is represented as $(x_1, \ldots, x_n)$. Each $x_i$ has one and only one corresponding $y_i$, $i = 1, \ldots, n$. Let $x_s$ be the smallest and $x_L$ be the largest values of $(x_1, \ldots, x_n)$. Also, let $y_s$ be the smallest and $y_L$ be the largest values of $y_i$, $i = 1, \ldots, n$.

Use Mahoney’s method (1998) to derive the expected value of the sample ranges $\mu_R$; the variance of the sample ranges $\sigma_R^2$ can be written as

$$\sigma_R^2 = n(n-1) \int_0^1 \int_0^{\gamma_L} (x_L - x_s)^2 (y_L - y_s)^{n-2} dy_s dy_L - \mu_R^2. \quad (A1)$$

Let

$$s = n(n-1) \int_0^1 \int_0^{\gamma_L} (x_L - x_s)^2 (y_L - y_s)^{n-2} dy_s dy_L$$

$$= n(n-1) \left\{ \int_0^1 \int_0^{\gamma_L} \frac{x_L^2(y_L - y_s)^{n-2} dy_s dy_L}{n-1} - 2 \int_0^1 \int_0^{\gamma_L} x_L x_s(y_L - y_s)^{n-2} dy_s dy_L + \int_0^1 \int_0^{\gamma_L} \frac{x_s^2(y_L - y_s)^{n-2} dy_s dy_L}{n-1} \right\}$$

$$= n(n-1)(s_1 - s_2 + s_3) \quad (A2)$$

where

$$s_1 = \int_0^1 \int_0^{\gamma_L} \frac{x_L^2(y_L - y_s)^{n-2} dy_s dy_L}{n-1} = \int_0^1 x_L \int_0^{\gamma_L} \frac{(y_L - y_s)^{n-2} dy_s dy_L}{n-1}$$

$$= \frac{1}{n-1} \int_0^1 \int_0^{\gamma_L} x_L^2 (y_L - y_s)^{n-2} dy_L$$

$$= \frac{1}{n(n-1)} \left[ b^2 - 2 \int_a^b x_L y_L^a dx_L \right]$$

$$= \frac{1}{n(n-1)} \left[ b^2 - 2 \int_a^b \frac{x_L^2 - x_L}{y_L - y_s} dy_L \right] \quad (A3)$$

$$s_2 = \int_0^1 \int_0^{\gamma_L} x_L x_s(y_L - y_s)^{n-2} dy_s dy_L$$

$$= \int_0^1 x_L dy_L \int_0^{\gamma_L} x_s(y_L - y_s)^{n-2} dy_s$$

$$= \int_0^1 x_L \left[ ay_L^{n-1} + \int_a^{y_L} (y_L - y_s)^{n-1} dy_s \right] dy_L$$

$$= \frac{2}{n-1} \left[ ab - a \int_a^b y_L^a dy_L + b \int_a^b (1 - y_s)^a dy_s - \int_a^b \int_a^{y_L} (y_L - y_s)^a dy_s dx_L \right] \quad (A4)$$

$$s_3 = \int_0^1 \int_0^{\gamma_L} \frac{x_s^2(y_L - y_s)^{n-2} dy_s dy_L}{n-1}$$

$$= \frac{1}{n(n-1)} \left[ a^2 - 2 \int_a^b \frac{x_L^2 - x_L}{y_L - y_s} dy_L \right]$$

$$= \frac{1}{n(n-1)} \left[ a^2 - 2 \int_a^b \frac{x_L^2 - x_L}{y_L - y_s} dy_L \right]$$

$$= \frac{1}{n(n-1)} \left[ a^2 - 2 \int_a^b \frac{x_L^2 - x_L}{y_L - y_s} dy_L \right]$$

$$= \frac{1}{n(n-1)} \left[ a^2 - 2 \int_a^b \frac{x_L^2 - x_L}{y_L - y_s} dy_L \right]$$

$$= \frac{1}{n(n-1)} \left[ a^2 - 2 \int_a^b \frac{x_L^2 - x_L}{y_L - y_s} dy_L \right]$$

$$= \frac{1}{n(n-1)} \left[ a^2 - 2 \int_a^b \frac{x_L^2 - x_L}{y_L - y_s} dy_L \right]$$
Robustness of R-Chart to Non Normality

\[ s_3 = \int_0^1 \int_0^y x_s^2 (y_L - y_S)^{n-2} dy_L dy_s = \int_0^1 x_s^2 \int_y^1 (y_L - y_S)^{n-2} dy_L dy_s \]
\[ = \frac{1}{n-1} \int_0^1 x_s^2 (1 - y_S)^{n-1} dy_S \]
\[ = \frac{1}{n(n-1)} \left[ a^2 - 2 \int_a^b \int_{s_x}^b (1 - y_S)^n dx_s dx_L + 2 \int_a^b b(1 - y_S)^n dx_s \right]. \quad (A5) \]

Substitute (A3), (A4), and (A5) into (A2):
\[ \varsigma = 2 \left[ \frac{b^2}{2} - ab + \frac{a^2}{2} - \int_a^b \int_{s_x}^b y_L^n dx_s dx_L \right. \]
\[ \left. - \int_a^b \int_{s_x}^b (1 - y_S)^n dx_s dx_L + \int_a^b \int_{s_x}^b (y_L - y_S)^n dx_s dx_L \right]. \]

Now,
\[ \int_a^b \int_{s_x}^b 1 dx_s dx_L = \frac{b^2}{2} - ab + \frac{a^2}{2}, \]
so
\[ \varsigma = 2 \int_a^b \int_{s_x}^b 1 - y_L^n - (1 - y_S)^n + (y_L - y_S)^n dx_s dx_L. \quad (A6) \]

Substituting (A6) into (A1) enables \( \sigma_R^2 \) to be expressed as follows:
\[ \sigma_R^2 = 2 \int_a^b \int_{s_x}^b 1 - y_L^n - (1 - y_S)^n + (y_L - y_S)^n dx_s dx_L - \mu_R^2. \quad (A7) \]

Appendix B: \( \sigma_R^2 \) with Binominal Transformation

This Appendix shows how the binomial transformation can be used to determine \( \sigma_R^2 \). The following is known:
\[ (a - b)^n = \sum_{i=0}^n \binom{n}{i} (-1)^i a^{n-i} b^i. \quad (B1) \]

Substituting (B1) into (A6) and letting
\[ \xi = 1 - y_L^n - (1 - y_S)^n + (y_L - y_S)^n \]
\[ \zeta = 1 - y_L^n - \sum_{i=0}^n \binom{n}{i} (-1)^i y_S^{n-i} + \sum_{i=0}^n \binom{n}{i} (-1)^i y_L^{n-i} \]
\[ = n \int_0^1 \left[ \sum_{i=1}^{n-1} \binom{n}{i} (-1)^{i+1} y_S^i + \sum_{i=1}^{n-1} \binom{n}{i} (-1)^i y_L^{n-i} y_S^i \right] dx_L \]
yield
\[ \varsigma = 2 \int_a^b \int_{s_x}^b 1 - y_L^n - (1 - y_S)^n + (y_L - y_S)^n dx_s dx_L. \]
\[
\sum_{i=1}^{n-1} \binom{n}{i} \left[ (-1)^{i+1} y_S^{i+1} + (-1)^{i} y_L^{i+1} y_S^i \right]. \tag{B3}
\]

Let \( y = 1 - Q(x) \), \( Q_L = Q(x_L) \), and \( Q_S = Q(x_S) \). Then, \( y_L = 1 - Q_L \) and \( y_S = 1 - Q_S \). Term \( \xi \) can be transformed into a function of \( Q(x) \), as follows:

\[
\xi = 1 - (1 - Q_L)^n - Q^n_S + (Q_S - Q_L)^n \\
= \sum_{i=1}^{n-1} \binom{n}{i} \left[ (-1)^{i+1} Q_L^i + (-1)^i Q_S^{n-i} Q_L^n \right]. \tag{B4}
\]

Substituting (B3) and (B4) into (A6) yields

\[
s = 2 \int_a^b \int_a^{x_L} \sum_{i=1}^{n-1} \binom{n}{i} \left[ (-1)^{i+1} y_S^{i+1} + (-1)^i y_L^{i+1} y_S^i \right] dx_S dx_L \tag{B5}
\]

and

\[
s = 2 \int_a^b \int_a^{x_L} \sum_{i=1}^{n-1} \binom{n}{i} \left[ (-1)^{i+1} Q_L^i + (-1)^i Q_S^{n-i} Q_L^n \right] dx_S dx_L. \tag{B6}
\]

References


