Section B: Monitoring the location of a process when the target location is unspecified or unknown (Case U)

Introduction

In Section A we focussed on monitoring the location of a chart when the location is specified (case K). This 'standard(s) known' case is when the underlying parameters of the process distribution are known or specified. In Section B we focus on monitoring the location of a chart when the location is unspecified or unknown (case U). This 'standard(s) unknown' case is when the parameters are unknown and need to be estimated.

Chapter 4: Sign-like control charts

4.1. The Shewhart-type control chart

4.1.1. Introduction

Janacek and Meikle (1997) proposed a Phase II nonparametric control chart useful in case U. The control limits of this chart are given by two selected order statistics of a Phase I reference sample. The charting statistic is the median M_i of the Phase II samples taken sequentially.

Chakraborti, Van der Laan and Van de Wiel (2004; hereafter CVV) generalized the work of Janacek and Meikle (1997). They considered using some order statistic of a Phase II sample as the charting statistic and control limits constructed from a Phase I reference sample. Their work involves a class of two-sample nonparametric statistics, called precedence statistics and their Shewhart-type charts are called precedence charts. The terms *precedence charts* and *sign-like charts* will be used interchangeably throughout this text.

Assume that a reference sample of size m, $X_1, X_2, ..., X_m$, is available from an in-control process with an unknown continuous cdf $F(x)$. The estimated control limits of the precedence chart are given by two reference sample order statistics, say, $L\hat{C}L = X_{(a:m)}$ and

$U\hat{C}L = X_{(b:m)}$, where $1 \le a < b \le m$. Let $Y_1^h, Y_2^h, ..., Y_{n_h}^h$, $h = 1,2,...,$ denote the h^{th} test sample of size n_h. The plotting statistic $Y_{(j:n_h)}^h$ is the j^{th} order statistic from the h^{th} Phase II sample of size n_h. Let $G^h(y)$ denote the cdf of the distribution of the h^{th} Phase II sample. $G^h(y) = G(y) \ \forall h$, since the Phase II samples are all assumed to be identically distributed. Assume that the Phase II samples are all of the same size, n, so that the subscript h can be suppressed. Under this assumption the plotting statistic is denoted by $Y_{(j:n)}$. For illustration purposes the plotting statistic is taken to be the median, but it can be any percentile of the Phase II sample. CVV provided recommendations and tables for the implementation of precedence charts and examined the chart performance in terms of the average run length. The overall conclusion is that the Shewhart-type precedence charts are more robust than their parametric counterparts, such as the Shewhart $\overline{X}$ chart. The precedence chart, being nonparametric, has the in-control robustness property (such as the same ARL_0 or the FAR for all continuous distributions), whereas as we noted earlier, the performance of the Shewhart $\overline{X}$ (and other parametric charts) is significantly (highly) degraded if the distributional form of the observations differs from normality.

4.1.2. Preliminary

Let W_j denote the number of X-observations that precede $Y_{(j:n)}$. The statistic W_j is called a precedence statistic and subsequently a test based on a precedence statistic is called a precedence test. Chakraborti and Van der Laan (1996, 1997; hereafter CV) gave an overview of some nonparametric procedures based on precedence statistics. CV's procedures included both hypothesis testing and confidence intervals. CV also highlighted the fact that precedence tests are simple and robust nonparametric procedures that are useful for comparing two or more distributions.

Let $P_C(W_j = w)$ denote the in-control probability distribution of W_j, where the subscript C refers to the in-control case. If $W_j = w$ it means that w X-observations precede $Y_{(j:n)}$. If w X-observations are less than or equal to $Y_{(j:n)}$, then $(m-w)$ X-observations are greater than $Y_{(j:n)}$. If we combine the reference sample (containing m X-observations) with the test sample (containing n Y-observations) we obtain a single sample consisting of $N = m + n$ observations. From this combined sample, w X-observations and j Y-observations are less than or equal to $Y_{(j:n)}$. On the other hand, $(m - w)$ X-observations and $(n - j)$ Y-observations are greater than $Y_{(j:n)}$. There are a total of $w + j - 1$ observations that are less than $Y_{(j:n)}$ and a total of $(m - w) + (n - j) = m + n - j - w$ observations that are greater than $Y_{(j:n)}$. The in-control distribution of W_j can be obtained by using combinatorics which allows one to count the number of experimental outcomes when the experiment involves selecting a number of objects, say r, from a larger set of objects, say R. The rule then states that the number of combinations of R objects taken r at a time is given by $\binom{R}{r}$. By using such combinatorial arguments the in-control distribution of W_j can be obtained and is given by

$$P_C(W_j = w) = \frac{\binom{w+j-1}{w}\binom{m+n-j-w}{m-w}}{\binom{m+n}{m}} \quad \text{with} \quad w = 0,1,2,\ldots,m. \qquad (4.1)$$

Note that the in-control probability distribution of W_j, i.e. when $F = G$, only depends on the number of observations in the reference sample m, the number of observations in each test sample n and the chosen percentile of the Phase II sample j. Thus, the in-control run length distribution of these precedence charts are distribution-free. The only condition is that the distribution of the reference sample and the distribution of the test sample be continuous and identical which is the case when the process is under control. It should be noted that this result is also given by Randles and Wolfe (1979), Theorem 11.4.4.

As illustration, let the number of observations in the Phase I reference sample be 25 ($m = 25$), the number of observations in each Phase II test sample be 15 ($n = 15$) and the chosen percentile of the Phase II sample be the median $\left(j = \frac{n+1}{2} = \frac{15+1}{2} = 8\right)$. The in-control distribution, i.e. when $F = G$, of W_j is then given by

$$P_C(W_j = w) = \frac{\binom{w+7}{w}\binom{32-w}{25-w}}{\binom{40}{25}} \quad \text{with} \quad w = 0,1,...,25. \qquad (4.2)$$

Figure 4.1 represents the in-control distribution of W_j when $m = 25$, $n = 15$ and $j = 8$. Note that the in-control probability distribution of W_j is symmetric. In general, the in-control probability distribution of W_j is symmetric when n is odd and the chosen percentile of the Phase II sample is the median of an odd Phase II sample.

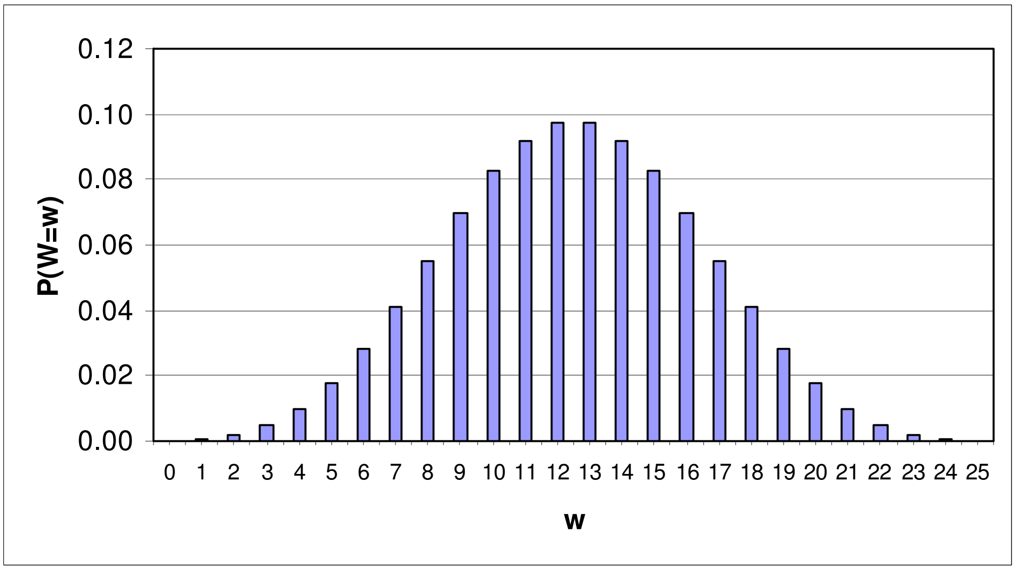

Figure 4.1. The in-control distribution of W_j when $m = 25$, $n = 15$ and $j = 8$.

Figure 4.2 represents the in-control probability distribution of W_j when the number of observations in the reference sample is kept at 25 ($m = 25$), the number of observations in each test sample is kept at 15 ($n = 15$), but the chosen percentile of the Phase II sample is not the median, i.e. $j \neq 8$. We take $j = 4$ for illustration purposes. Note that the in-control probability distribution of W_j is now asymmetric.

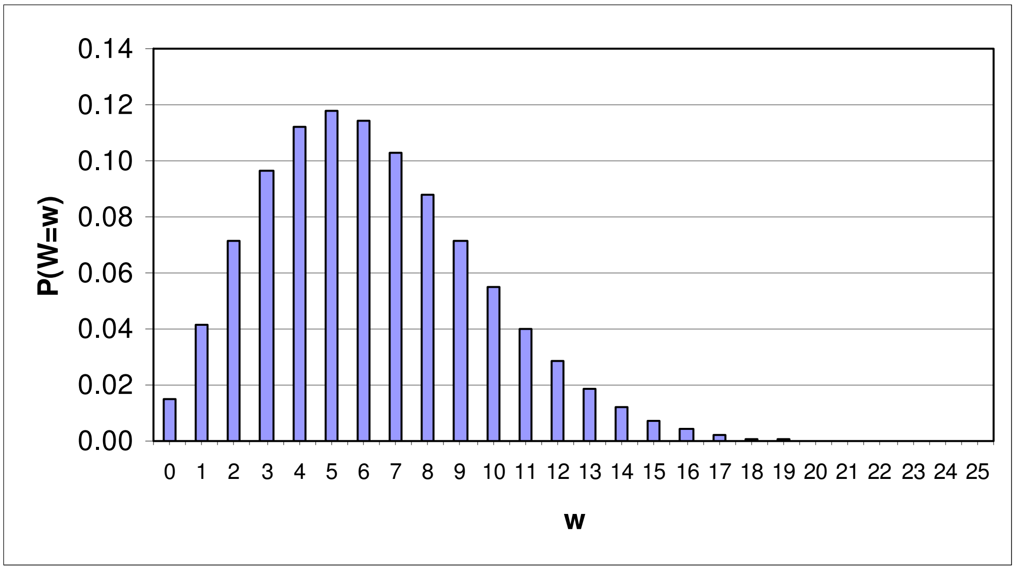

Figure 4.2. The in-control distribution of W_j when $m = 25$, $n = 15$ and $j = 4$.

4.1.3. Probability of no signal

Recall that $L\hat{C}L = X_{(a:m)}$ and $U\hat{C}L = X_{(b:m)}$. A non-signalling event in the case of the two-sided chart occurs when $X_{(a:m)} \le Y_{(j:n)} \le X_{(b:m)}$. Stated differently, a non-signalling event occurs when at least a X-observations precede $Y_{(j:n)}$ and at most $b-1$ X-observations precede $Y_{(j:n)}$, i.e. $a \le W_j \le b-1$. Let the probability of no signal be denoted by p. Then, the probability of no signal is given by

$$p = p(m,n,j;F,G) = P\left(X_{(a:m)} \le Y_{(j:n)} \le X_{(b:m)}\right) = P\left(a \le W_j \le b-1\right). \qquad (4.3)$$

From (4.3) it can be seen that the probability of no signal, p, can be expressed in terms of the precedence statistic W_j, thus simplifying the probability calculations (see Randles and Wolfe (1979), Example 11.4.19).

Let p_0 denote the in-control value of p. A process is said to be in-control when $G = F$. Therefore, the expression for p_0 can be obtained by simply substituting $G = F$ in expression (4.3). Thus,

$$p_0 = p(m,n,j;F,F) = P(\text{No Signal} \mid \text{In-control}) = P_C\left(a \le W_j \le b-1\right). \qquad (4.4)$$

Recall that a false alarm is given when a signaling event occurs, given that the process is actually in-control. Therefore, the probability of a false alarm (also referred to as the false alarm rate (*FAR*)) is given by

$$1-p_0=1-P(\text{No Signal}\mid\text{In - control})=P(\text{Signal}\mid\text{In - control})=FAR\,. \qquad (4.5)$$

4.1.4. Determination of chart constants

The charting constants a and b are typically selected so that a specified false alarm rate or a specified in-control average run-length is attained. The exact expression for the ARL_0 is derived later on in this chapter using a conditioning method. In this section we will focus on the *FAR*. Hence, the charting constants a and b are found by either setting the *FAR* (given by $1-p_0$) to a desirable small value, say $1-P_0$, or by setting p_0 to some desirable large value, say P_0. Take note that P_0 will usually be chosen to be a large value such as 0.95 or 0.99 and the desired or specified value of the *FAR*, given by $1-P_0$, will be a small value such as 0.05 or 0.01. The charting constants are found such that $p_0=P(\text{No Signal}\mid\text{In - control})$ is not smaller than the desired or specified value P_0, that is, $p_0=P(\text{No Signal}\mid\text{In - control})\geq P_0$ (this is due to the discrete nature of the distribution of W_j). Stated differently, the charting constants are selected such that $1-p_0=P(\text{Signal}\mid\text{In - control})$ is not larger than the desired or specified value $1-P_0$, that is, $1-p_0=P(\text{Signal}\mid\text{In - control})\leq 1-P_0$. Since the statistic W_j is discrete, not all desired or specified P_0 values are attainable for all combinations of m, n and j. The inequality sign in (4.6) ensures that we are conservative. The charting constants are found such that

$$P_C(a\leq W_j\leq b-1)=\sum_{w=a}^{b-1}\frac{\binom{w+j-1}{w}\binom{m+n-j-w}{m-w}}{\binom{m+n}{m}}\geq p_0\,. \qquad (4.6)$$

We can use any test sample order statistic (including the median) when implementing the two-sided precedence chart. If the plotting statistic is taken to be the median, the in-control probability distribution of W_j is symmetric (for odd sample sizes) and a reasonable choice for b is $m-a+1$. Once the charting constants a and b are found, the estimated control limits $L\hat{C}L = X_{(a:m)}$ and $U\hat{C}L = X_{(b:m)}$ can be determined. Therefore, when the plotting statistic is taken to be the median, we replace b by $m-a+1$ in (4.6) to obtain

$$P_C(a \le W_j \le m-a) = \sum_{w=a}^{m-a} \frac{\binom{w+j-1}{w}\binom{m+n-j-w}{m-w}}{\binom{m+n}{m}} \ge p_0 . \tag{4.7}$$

For example, let the number of observations in the reference sample be 125 ($m = 125$), the number of observations in each test sample be 5 ($n = 5$) and the chosen percentile of the Phase II sample be the median $\left(j = \frac{n+1}{2} = \frac{5+1}{2} = 3 \quad \text{(when } n \text{ is odd)} \right)$. By substituting $m = 125$, $n = 5$ and $j = 3$ into (4.7) we obtain

$$P_C(a \le W_3 \le 125-a) = \sum_{w=a}^{125-a} \frac{\binom{w+2}{w}\binom{127-w}{125-w}}{\binom{130}{125}} \ge p_0 . \tag{4.8}$$

Possible control limits were calculated using (4.8) and are shown in Table 4.1.

Table 4.1*. False alarm rate (FAR) and chart constant (a) values for the Shewhart sign-like chart when $m = 125$, $n = 5$ and $j = 3$.

$\boldsymbol{a}$	3	4	5	6	7	8	9	10
$\boldsymbol{FAR}$	0.000546	0.001079	0.001865	0.002948	0.004368	0.006164	0.008372	0.011025

From Table 4.1 we see that for a false alarm rate of 0.004368 one can take $a = 7$ so $b = m-a+1 = 125-7+1 = 119$ so that the control limits are the 7th and 119th ordered values of the reference sample. Thus, $L\hat{C}L = X_{(7:125)}$ and $U\hat{C}L = X_{(119:125)}$. For another example on exceedance statistics see Randles and Wolfe (1979), Example 11.4.19.

* The values in Table 4.1 were generated using Microsoft Excel. Table 4.1 is an extension of Table 3 given in Chakraborti, Eryilmaz and Human (2006).

4.1.5. The median chart

Let $n = 2s + 1$, where $s = 0,1,2,...,$ (so that n is odd). Therefore, the median is uniquely given by $j = \frac{n+1}{2} = \frac{(2s+1)+1}{2} = s+1$. The statistic W_{s+1} is called the median statistic of Mathisen (1943). The in-control probability distribution of W_{s+1} is found by substituting $n = 2s + 1$ and $j = s + 1$ into (4.1) and is given by

$$P_C(W_{s+1} = w) = \frac{\binom{w+s}{w}\binom{m+s-w}{m-w}}{\binom{m+2s+1}{m}} \tag{4.9}$$

(see Randles and Wolfe (1979), Example 11.4.5). Recall that the in-control distribution of W_j (in this case, W_{s+1}) is symmetric when n is odd and the chosen percentile of the Phase II sample is the median. In this case a reasonable choice for b is $m - a + 1$. The charting constant a is found by substituting $n = 2s + 1$, $j = s + 1$ and $b = m - a + 1$ into equation (4.6) and then solving for a such that (4.10) is satisfied.

$$P_C(a \le W_{s+1} \le m - a) = \sum_{w=a}^{m-a} \frac{\binom{w+s}{w}\binom{m+s-w}{m-w}}{\binom{m+2s+1}{m}} \ge p_0 \,. \tag{4.10}$$

Once the charting constant a is found using expression (4.10), the charting constant b is found from the relationship $b = m - a + 1$. Thereafter, the control limits $L\hat{C}L = X_{(a:m)}$ and $U\hat{C}L = X_{(b:m)}$ can be determined. By using symmetry we have that

$P_C(a \le W_j \le m - a)$

$= 1 - \left(P_C(0 \le W_j \le a-1) + P_C(m - a + 1 \le W_j \le m)\right)$

$= 1 - 2P_C(0 \le W_j \le a - 1)$

and by setting $1 - 2P_C(0 \le W_j \le a-1) \ge P_0$ we obtain

$$P_C(0 \le W_j \le a-1) \le \frac{1 - P_0}{2} \,. \tag{4.11}$$

Therefore, expression (4.10) can be re-written as expression (4.11) which is more convenient to work with.

For example, let the number of observations in the reference sample be 125 ($m = 125$) and $s = 2$ so that $n = 2s + 1 = 5$ and $j = s + 1 = 3$. By substituting $m = 125$ and $s = 2$ into (4.10) we obtain

$$P_C(a \le W_3 \le 125 - a) = \sum_{w=a}^{125-a} \frac{\binom{w+2}{w}\binom{127-w}{125-w}}{\binom{130}{125}} \ge p_0$$

which is equal to expression (4.8). Therefore, the FAR values given in Table 4.1 can be used in this example, meaning that one can take $a = 7$ so $b = m - a + 1 = 125 - 7 + 1 = 119$ so that the control limits are the 7th and 119th ordered values of the reference sample. Thus, $L\hat{C}L = X_{(7:125)}$ and $U\hat{C}L = X_{(119:125)}$, which yield a FAR of 0.004368.

4.1.6. Control charts for other percentiles

Since we could be interested in other percentiles than the median (see Radson and Boyd (2005) and Shmueli and Cohen (2003)), the distribution of W_j is not symmetric (in such cases) and finding the charting constants a and b is much more difficult.

Chakraborti, Van der Laan and Van de Wiel (2004) proposed the equal-tailed[*] procedure when the $100\rho^{th}$ percentile is of interest where $0 < \rho < 1$. The equal-tailed procedure is as follows:

Find the *largest* integer a $(1 \le a \le [m\rho])$ such that

$$P_C(0 \le W_j \le a - 1) \le \frac{1 - P_0}{2},$$

and the *smallest* integer b $(a < b \le m)$ such that

$$P_C(b + 1 \le W_j \le m) \le \frac{1 - P_0}{2}.$$

These a and b values are then substituted in the control limits $L\hat{C}L = X_{(a:m)}$ and $U\hat{C}L = X_{(b:m)}$.

[*] Note that in general $b \ne m - a + 1$ in this case so that the "equal-tailed" means equality in tail probabilities.

4.1.7. Properties of order statistics

The ordered values of a sample are known as the order statistics. Various authors have studied order statistics (see for example Randles and Wolfe (1979)). Our goal is to study the distribution of order statistics. In addition, we give some well-known properties and results of order statistics that will be used later on.

Suppose that $X_1, X_2, ..., X_n$ denotes a random sample of size n from a continuous pdf, $f(x)$. The pdf of the k^{th} order statistic $X_{(k:n)}$ is given by

$$g_k(x_{(k:n)}) = \frac{n!}{(k-1)!(n-k)!}\left(F(x_{(k:n)})\right)^{k-1}\left(1-F(x_{(k:n)})\right)^{n-k} f(x_{(k:n)}). \tag{4.12}$$

The joint pdf for $X_{(k:n)}$ and $X_{(l:n)}$ is given by

$$g_{kl}(x_{(k:n)}, x_{(l:n)}) = \frac{n!}{(k-1)!(l-k-1)!(n-l)!}\times$$
$$\left(F(x_{(k:n)})\right)^{k-1} f(x_{(k:n)})\left(F(x_{(l:n)}-x_{(k:n)})\right)^{l-k-1}\left(1-F(x_{(l:n)})\right)^{n-l} f(x_{(l:n)}). \tag{4.13}$$

Let $U_{(k:n)}$ denote the k^{th} order statistic of a sample of size n from the Uniform(0,1) distribution. The pdf of $U_{(k:n)}$ is given by

$$f_{U_{(k:n)}}(u) = \frac{1}{\beta(k, n-k+1)}u^{k-1}(1-u)^{n-k} \tag{4.14}$$

where $\beta(k, n-k+1) = \frac{\Gamma(k)\Gamma(n-k+1)}{\Gamma(n+1)} = \frac{(k-1)!(n-k)!}{n!}$.

The binomial series arises in connection with distributions of order statistics. The binomial theorem gives the expansion of $(a+b)^k$. Using the binomial theorem we obtain

$$(a-b)^k = \sum_{n=0}^{k}(-1)^n\binom{k}{n}a^{k-n}b^n \tag{4.15}$$

where a and b are any real numbers and k is a positive integer.

4.1.8. One-sided control charts

In this section the lower- and upper one-sided precedence control charts are considered. The lower one-sided chart will have a $L\hat{C}L$ equal to some constant value and an $U\hat{C}L = \infty$. In contrast, the upper one-sided chart will have an $U\hat{C}L$ equal to some constant and a $L\hat{C}L = -\infty$.

4.1.8.1. Lower one-sided control charts

For the lower one-sided chart we have the $L\hat{C}L = X_{(a:m)}$. Therefore, a non-signalling event occurs when $Y_{(j:n)} \geq X_{(a:m)}$.

Result 4.1: Probability of no signal - conditional

$$P\left(\text{No Signal} \mid X_{(a:m)} = x\right) = \int_{G(x)}^{1} \frac{1}{\beta(j, n-j+1)} u^{j-1}(1-u)^{n-j}\, du$$

Using the probability integral transformation (PIT) (see, for example, Gibbons and Chakraborti (2003)), we know that $Y_{(j:n)} = G^{-1}(U_{(j:n)})$ and $X_{(a:m)} = F^{-1}(U_{(a:m)})$ where F and G are both continuous cdf's.

$$\begin{aligned}
P\left(\text{No Signal} \mid X_{(a:m)} = x\right) &= p_L(x), \text{ say,} \\
&= P\left(Y_{(j:n)} \geq x \mid X_{(a:m)} = x\right) \\
&= P\left(G^{-1}(U_{(j:n)}) \geq x \mid X_{(a:m)} = x\right) \\
&= P\left(U_{(j:n)} \geq G(x) \mid X_{(a:m)} = x\right) \\
&= \int_{G(x)}^{1} \frac{1}{\beta(j, n-j+1)} u^{j-1}(1-u)^{n-j}\, du
\end{aligned}$$

since $\frac{1}{\beta(j, n-j+1)} u^{j-1}(1-u)^{n-j}$ is the pdf of $U_{(j:n)}$ (see equation (4.14)).

Result 4.2: Probability of no signal – unconditional

Let p_L denote the unconditional probability of no signal, then:

$$P(\text{No Signal}) = P\left(Y_{(j:n)} \geq X_{(a:m)}\right)$$
$$= \int_0^1 \left(\frac{1}{\beta(j, n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(1 - GF^{-1}(v)\right)^{j+h} \right) \frac{m!}{(a-1)!(m-a)!} v^{a-1} (1-v)^{m-a} dv$$

$$\begin{aligned}
&p_L \\
&= P(\text{No Signal}) \\
&= P\left(Y_{(j:n)} \geq X_{(a:m)}\right) \\
&= E_{X_{(a:m)}}\left(P(Y_{(j:n)} \geq X_{(a:m)} \mid X_{(a:m)})\right) \\
&= E_{X_{(a:m)}}\left(P(G(Y_{(j:n)}) \geq G(X_{(a:m)}) \mid X_{(a:m)})\right)
\end{aligned}$$

By the PIT we have that $U_{(j:n)} = G(Y_{(j:n)})$ where G is the continuous cdf of the Phase II sample $Y_1, Y_2, ..., Y_n$. Using this we obtain

$$= E_{X_{(a:m)}}\left(P(U_{(j:n)} \geq G(X_{(a:m)}) \mid X_{(a:m)})\right)$$

By the PIT we have that $U_{(a:m)} = F(X_{(a:m)})$ so that $X_{(a:m)} = F^{-1}(U_{(a:m)})$ where F is the continuous cdf of the reference sample $X_1, X_2, ..., X_m$. Using this we obtain

$$= E_{U_{(a:m)}}\left(P(U_{(j:n)} \geq GF^{-1}(U_{(a:m)}) \mid U_{(a:m)})\right)$$
$$= \int_0^1 P(U_{(j:n)} \geq GF^{-1}(v) \mid U_{(a:m)} = v) f(v) dv$$

where $f(v)$ is the pdf of $U_{(a:m)}$ which is given by $f(v) = \dfrac{m!}{(a-1)!(m-a)!} v^{a-1}(1-v)^{m-a}$

$$= \int_0^1 \left(\int_{GF^{-1}(v)}^1 \frac{1}{\beta(j, n-j+1)} u^{j-1} (1-u)^{n-j} du \right) f(v) dv$$

$$=\int_0^1\left(\int_{GF^{-1}(v)}^1 \frac{1}{\beta(j,n-j+1)}u^{j-1}(1-u)^{n-j}\,du\right)\frac{m!}{(a-1)!(m-a)!}v^{a-1}(1-v)^{m-a}\,dv$$

The term $(1-u)^{n-j}$ can be expanded to $(1-u)^{n-j}=\sum_{h=0}^{n-j}(-1)^h\binom{n-j}{h}1^{n-j-h}u^h=\sum_{h=0}^{n-j}(-1)^h\binom{n-j}{h}u^h$ by using a binomial expansion (see equation (4.15)) and we obtain

$$=\int_0^1\left(\int_{GF^{-1}(v)}^1 \frac{1}{\beta(j,n-j+1)}u^{j-1}\left(\sum_{h=0}^{n-j}(-1)^h\binom{n-j}{h}u^h\right)du\right)\frac{m!}{(a-1)!(m-a)!}v^{a-1}(1-v)^{m-a}\,dv$$

By taking all the constants out of the integral sign and simplifying by setting $u^{j-1}u^h=u^{j-1+h}$ we obtain

$$=\int_0^1\left(\frac{1}{\beta(j,n-j+1)}\sum_{h=0}^{n-j}(-1)^h\binom{n-j}{h}\int_{GF^{-1}(v)}^1 u^{j-1+h}\,du\right)\frac{m!}{(a-1)!(m-a)!}v^{a-1}(1-v)^{m-a}\,dv$$

Integrating $\int_{GF^{-1}(v)}^1 u^{j-1+h}\,du=\left.\frac{u^{j+h}}{j+h}\right|_{u=GF^{-1}(v)}^{u=1}=\frac{1^{j+h}}{j+h}-\frac{\left(GF^{-1}(v)\right)^{j+h}}{j+h}=\frac{1-\left(GF^{-1}(v)\right)^{j+h}}{j+h}$ we have

$$=\int_0^1\left(\frac{1}{\beta(j,n-j+1)}\sum_{h=0}^{n-j}(-1)^h\binom{n-j}{h}\left(\frac{1-\left(GF^{-1}(v)\right)^{j+h}}{j+h}\right)\right)\frac{m!}{(a-1)!(m-a)!}v^{a-1}(1-v)^{m-a}\,dv$$

which simplifies to

$$=\int_0^1\left(\frac{1}{\beta(j,n-j+1)}\sum_{h=0}^{n-j}\frac{(-1)^h}{j+h}\binom{n-j}{h}\left(1-GF^{-1}(v)\right)^{j+h}\right)\frac{m!}{(a-1)!(m-a)!}v^{a-1}(1-v)^{m-a}\,dv\,.$$

Result 4.3: Probability of a signal - conditional

A signalling event occurs when $Y_{(j:n)} < X_{(a:m)}$.

$$P\left(\text{Signal} \mid X_{(a:m)} = x\right) = 1 - \int_{G(x)}^{1} \frac{1}{\beta(j, n-j+1)} u^{j-1} (1-u)^{n-j} \, du$$

Result 4.4: Probability of a signal - unconditional

$$1 - p_L = P\left(\text{Signal}\right) =$$
$$1 - \int_0^1 \left(\frac{1}{\beta(j, n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(1 - GF^{-1}(v)\right)^{j+h} \right) \frac{m!}{(a-1)!(m-a)!} v^{a-1} (1-v)^{m-a} \, dv$$

Result 4.5: Probability of a false alarm - conditional

$$CFAR = 1 - \int_{F(x)}^{1} \frac{1}{\beta(j, n-j+1)} u^{j-1} (1-u)^{n-j} \, du$$

Follows immediately from Result 4.3, since $G = F$.

Result 4.6: Probability of a false alarm – unconditional

$$FAR = 1 - \int_0^1 \left(\frac{1}{\beta(j, n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(1 - v\right)^{j+h} \right) \frac{m!}{(a-1)!(m-a)!} v^{a-1} (1-v)^{m-a} \, dv$$

Follows immediately from Result 4.4, since $G = F$ and therefore $GF^{-1}(v) = FF^{-1}(v) = v$.

Result 4.7: Run-length distribution - conditional

$$P\left(N = k \mid X_{(a:m)} = x\right) = (p_L(x))^{k-1}(1 - p_L(x)) \quad \text{for} \quad k = 1,2,3,...$$

The conditional run length, denoted by $N \mid X_{(a:m)} = x$, will have a geometric distribution with parameter $1 - p_L(x)$, because all the signalling events are independent. Therefore we have that

$N \mid X_{(a:m)} = x \sim GEO(1 - p_L(x))$

$P\left(N = k \mid X_{(a:m)} = x\right) = (p_L(x))^{k-1}(1 - p_L(x)) \quad \text{for} \quad k = 1,2,3,...$

Consequently, the cumulative distribution function (cdf) is found from

$P\left(N \leq k \mid X_{(a:m)} = x\right)$

$= \sum_{i=1}^{k} (p_L(x))^{i-1}(1 - p_L(x)) = 1 - (p_L(x))^k \quad \text{for} \quad k = 1,2,3,...$

We also have that

$P\left(N > k \mid X_{(a:m)} = x\right) = 1 - \left(1 - (p_L(x))^k\right) = (p_L(x))^k$.

Result 4.8: Average run-length - conditional

$$CARL = E\left(N \mid X_{(a:m)} = x\right) = \frac{1}{1 - p_L(x)}$$

or

$$CARL = E\left(N \mid X_{(a:m)} = x\right) = \sum_{k=0}^{\infty} (p_L(x))^k$$

Since the conditional run length, denoted by $N \mid X_{(a:m)} = x$ has a geometric distribution with parameter $1 - p_L(x)$, the conditional average run length is given by

$$CARL = E\left(N \mid X_{(a:m)} = x\right) = \frac{1}{1 - p_L(x)}.$$

The second expression follows immediately from the geometric expansion of $(1 - p_L(x))^{-1}$ for $p_L(x) < 1$.

Result 4.9: Run-length distribution - unconditional

$$P(N=k)=D_L^*(k-1)-D_L^*(k) \quad \text{for} \quad k=1,2,3,\ldots, \quad D_L^*(0)=1$$

where

$$D_L^*(k)=\int_0^1\left(\frac{1}{\beta(j,n-j+1)}\sum_{h=0}^{n-j}\frac{(-1)^h}{j+h}\binom{n-j}{h}\left(1-GF^{-1}(v)\right)^{j+h}\right)^k\frac{m!}{(a-1)!(m-a)!}v^{a-1}(1-v)^{m-a}\,dv$$

$$\begin{aligned}
&P(N=k)\\
&=E_{X_{(a:m)}}\left(P\left(N=k \mid X_{(a:m)}=x\right)\right)\\
&=E_{X_{(a:m)}}\left((p_L(x))^{k-1}(1-p_L(x))\right)\\
&=E_{X_{(a:m)}}\left((p_L(x))^{k-1}-(p_L(x))^k\right)\\
&=E_{X_{(a:m)}}\left((p_L(x))^{k-1}\right)-E_{X_{(a:m)}}\left((p_L(x))^k\right)
\end{aligned}$$

By only focussing on $E_{X_{(a:m)}}\left((p_L(x))^k\right)$ we have

$$\begin{aligned}
&E_{X_{(a:m)}}\left((p_L(x))^k\right)\\
&=E_{U_{(a:m)}}\left(\left(\int_{GF^{-1}(v)}^1\frac{1}{\beta(j,n-j+1)}u^{j-1}(1-u)^{n-j}\,du\right)^k\right)\\
&=\int_0^1\left(\left(\int_{GF^{-1}(v)}^1\frac{1}{\beta(j,n-j+1)}u^{j-1}(1-u)^{n-j}\,du\right)^k\right)f(v)dv
\end{aligned}$$

where $f(v)$ is the pdf of $U_{(a:m)}$ which is given by $f(v)=\dfrac{m!}{(a-1)!(m-a)!}v^{a-1}(1-v)^{m-a}$

$$=\int_0^1\left(\left(\int_{GF^{-1}(v)}^1\frac{1}{\beta(j,n-j+1)}u^{j-1}(1-u)^{n-j}\,du\right)^k\right)\frac{m!}{(a-1)!(m-a)!}v^{a-1}(1-v)^{m-a}\,dv$$

The term $(1-u)^{n-j}$ can be expanded to $(1-u)^{n-j}=\sum_{h=0}^{n-j}(-1)^h\binom{n-j}{h}1^{n-j-h}u^h=\sum_{h=0}^{n-j}(-1)^h\binom{n-j}{h}u^h$ by using a binomial expansion (see equation (4.15)) and we obtain

$$= \int_0^1 \left(\left(\int_{GF^{-1}(v)}^{1} \frac{1}{\beta(j, n-j+1)} u^{j-1} \left(\sum_{h=0}^{n-j} (-1)^h \binom{n-j}{h} u^h \right) du \right)^k \right) \frac{m!}{(a-1)!(m-a)!} v^{a-1} (1-v)^{m-a} dv$$

By taking all the constants out of the integral sign and simplifying by setting $u^{j-1}u^h = u^{j-1+h}$ we obtain

$$= \int_0^1 \left(\frac{1}{\beta(j, n-j+1)} \sum_{h=0}^{n-j} (-1)^h \binom{n-j}{h} \int_{GF^{-1}(v)}^{1} u^{j-1+h} du \right)^k \frac{m!}{(a-1)!(m-a)!} v^{a-1} (1-v)^{m-a} dv$$

Integrating $\int_{GF^{-1}(v)}^{1} u^{j-1+h} du = \left. \frac{u^{j+h}}{j+h} \right|_{u=GF^{-1}(v)}^{u=1} = \frac{1^{j+h}}{j+h} - \frac{\left(GF^{-1}(v)\right)^{j+h}}{j+h} = \frac{1-\left(GF^{-1}(v)\right)^{j+h}}{j+h}$ we have

$$= \int_0^1 \left(\frac{1}{\beta(j, n-j+1)} \sum_{h=0}^{n-j} (-1)^h \binom{n-j}{h} \left(\frac{1-\left(GF^{-1}(v)\right)^{j+h}}{j+h} \right) \right)^k \frac{m!}{(a-1)!(m-a)!} v^{a-1} (1-v)^{m-a} dv$$

which simplifies to

$$= \int_0^1 \left(\frac{1}{\beta(j, n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(1-GF^{-1}(v)\right)^{j+h} \right)^k \frac{m!}{(a-1)!(m-a)!} v^{a-1} (1-v)^{m-a} dv$$

$= D_L^*(k)$,

say.

Recall that $P(N=k) = E_{X_{(a:m)}}\left((p_L(x))^{k-1}\right) - E_{X_{(a:m)}}\left((p_L(x))^{k}\right)$.

Therefore, $P(N=k) = D_L^*(k-1) - D_L^*(k)$ for $k = 1,2,3,\ldots,$ and $D_L^*(0) = 1$.

$D_L^*(0) = 1$ since

$$
\begin{aligned}
D_L^*(0) &= \int_0^1 \left(\frac{1}{\beta(j,n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(1 - GF^{-1}(v)\right)^{j+h} \right)^0 \frac{m!}{(a-1)!(m-a)!} v^{a-1}(1-v)^{m-a} dv \\
&= \int_0^1 \frac{m!}{(a-1)!(m-a)!} v^{a-1}(1-v)^{m-a} dv \\
&= \int_0^1 f(v) dv \\
&= 1.
\end{aligned}
$$

This equals one, because in general, $\int_{-\infty}^{\infty} f(x)dx = 1$ for real x.

Result 4.10: In-control run-length distribution

$$P_C(N = k) = D_L(k-1) - D_L(k) \quad \text{for} \quad k = 1,2,3,..., \quad D_L(0) = 1$$

and

$$D_L(k) = \int_0^1 \left(\frac{1}{\beta(j,n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} (1-v)^{j+h} \right)^k \frac{m!}{(a-1)!(m-a)!} v^{a-1}(1-v)^{m-a} dv$$

Recall that the reference sample of size m, $X_1, X_2, ..., X_m$, is available from an in-control process with a continuous cdf, $F(x)$. The plotting statistic $Y_{(j:n)}^h$ is the j^{th} order statistic from the h^{th} Phase II sample of size n_h. Let $G^h(y)$ denote the cdf of the distribution of the h^{th} Phase II sample. A process is said to be in-control at stage h when $G^h = F$. Assume that the Phase II samples are all of the same size, n, so that the subscript h can be suppressed. Therefore, a process is said to be in-control when $G = F$. Therefore, the in-control run length distribution is obtained by setting $G = F$ into the equation for the out-of-control run length distribution.

The out-of-control run length distribution for the lower one-sided chart is given by

$$P(N = k) = D_L^*(k-1) - D_L^*(k) \quad \text{for} \quad k = 1,2,3,..., \quad D_L^*(0) = 1$$

and (4.16)

$$D_L^*(k)=\int_0^1\left(\frac{1}{\beta(j,n-j+1)}\sum_{h=0}^{n-j}\frac{(-1)^h}{j+h}\binom{n-j}{h}\left(1-GF^{-1}(v)\right)^{j+h}\right)^k\frac{m!}{(a-1)!(m-a)!}v^{a-1}(1-v)^{m-a}dv.$$

Therefore, the in-control run length distribution for the lower one-sided chart is obtained by setting $G=F$ into equation (4.16) and we obtain

$$P_C(N=k)=D_L(k-1)-D_L(k)\quad\text{for}\quad k=1,2,3,\ldots,\quad D_L(0)=1$$

and (4.17)

$$D_L(k)=\int_0^1\left(\frac{1}{\beta(j,n-j+1)}\sum_{h=0}^{n-j}\frac{(-1)^h}{j+h}\binom{n-j}{h}(1-v)^{j+h}\right)^k\frac{m!}{(a-1)!(m-a)!}v^{a-1}(1-v)^{m-a}dv\,.$$

Result 4.11: Out-of-control average run-length - unconditional

$$UARL_{L,\delta}=\int_0^1\frac{1}{1-S_L(v,j,n,F,G)}f(v)dv$$

with

$$S_L(v,j,n,F,G)=\frac{1}{\beta(j,n-j+1)}\sum_{h=0}^{n-j}\frac{(-1)^h}{j+h}\binom{n-j}{h}\left(1-GF^{-1}(v)\right)^{j+h}$$

Let $UARL_{L,\delta}$ denote the unconditional average run length, where δ refers to the out-of-control case. To derive an expression for the $UARL_{L,\delta}$, recall that

$$E_{X_{(a:m)}}\left((p_L(x))^k\right)$$
$$=D_L^*(k)$$
$$=\int_0^1\left(\frac{1}{\beta(j,n-j+1)}\sum_{h=0}^{n-j}\frac{(-1)^h}{j+h}\binom{n-j}{h}\left(1-GF^{-1}(v)\right)^{j+h}\right)^k\frac{m!}{(a-1)!(m-a)!}v^{a-1}(1-v)^{m-a}dv$$
$$=\int_0^1\left(\frac{1}{\beta(j,n-j+1)}\sum_{h=0}^{n-j}\frac{(-1)^h}{j+h}\binom{n-j}{h}\left(1-GF^{-1}(v)\right)^{j+h}\right)^k f(v)dv$$

For simplicity let $S_L(v,j,n,F,G)=\frac{1}{\beta(j,n-j+1)}\sum_{h=0}^{n-j}\frac{(-1)^h}{j+h}\binom{n-j}{h}\left(1-GF^{-1}(v)\right)^{j+h}$, therefore we obtain

$$= \int_0^1 \left(S_L(v, j, n, F, G)\right)^k f(v)dv$$

Finally, we have that (from the second expression in Result 4.8)

$$\begin{aligned}
&UARL_{L,\delta} \\
&= \sum_{k=0}^{\infty} E_{X_{(a:m)}}\left((p_L(x))^k\right) \\
&= \sum_{k=0}^{\infty} D_L^*(k) \\
&= \sum_{k=0}^{\infty} \int_0^1 \left(S_L(v, j, n, F, G)\right)^k f(v)dv \\
&= \int_0^1 \sum_{k=0}^{\infty} \left(S_L(v, j, n, F, G)\right)^k f(v)dv \\
&= \int_0^1 \frac{1}{1 - S_L(v, j, n, F, G)} f(v)dv \text{ from the geometric expansion of } (1 - S_L(v, j, n, F, G))^{-1}.
\end{aligned}$$

Result 4.12: In-control average run-length - unconditional

$$UARL_{L,0} = \int_0^1 \left(\frac{1}{1 - S_L(v, j, n)}\right) \frac{m!}{(a-1)!(m-a)!} v^{a-1}(1-v)^{m-a} dv$$

with

$$S_L(v, j, n) = \frac{1}{\beta(j, n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} (1-v)^{j+h}$$

Let $UARL_{L,0}$ denote the unconditional average run length, where 0 refers to the in-control case. To derive an expression for the $UARL_{L,0}$, recall that the in-control run length distribution for the lower one-sided chart is given by

$P_C(N = k) = D_L(k-1) - D_L(k) \quad \text{for} \quad k = 1,2,3,..., \quad D_L(0) = 1$

with

$$D_L(k) = \int_0^1 \left(\frac{1}{\beta(j, n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} (1-v)^{j+h}\right)^k \frac{m!}{(a-1)!(m-a)!} v^{a-1}(1-v)^{m-a} dv$$

$$=\int_0^1\left(\frac{1}{\beta(j,n-j+1)}\sum_{h=0}^{n-j}\frac{(-1)^h}{j+h}\binom{n-j}{h}(1-v)^{j+h}\right)^k f(v)dv$$

For simplicity let $S_L(v,j,n)=\dfrac{1}{\beta(j,n-j+1)}\sum_{h=0}^{n-j}\dfrac{(-1)^h}{j+h}\dbinom{n-j}{h}(1-v)^{j+h}$, therefore we obtain

$$D_L(k)=\int_0^1 (S_L(v,j,n))^k f(v)dv$$

and, finally, we have that

$$UARL_{L,0}$$
$$=\sum_{k=0}^{\infty}D_L(k)$$
$$=\sum_{k=0}^{\infty}\int_0^1 (S_L(v,j,n))^k f(v)dv$$
$$=\int_0^1\sum_{k=0}^{\infty}(S_L(v,j,n))^k f(v)dv$$
$$=\int_0^1\frac{1}{1-S_L(v,j,n)}f(v)dv\,.$$

4.1.8.2. Upper one-sided control charts

For the upper one-sided chart we have $U\hat{C}L=X_{(b:m)}$. Therefore, a non-signalling event occurs when $Y_{(j:n)}\le X_{(b:m)}$.

Result 4.13: Probability of no signal - conditional

$$P\left(\text{No Signal}\mid X_{(b:m)}=z\right)=\int_0^{G(z)}\frac{1}{\beta(j,n-j+1)}u^{j-1}(1-u)^{n-j}du$$

Using the PIT, we know that $Y_{(j:n)}=G^{-1}(U_{(j:n)})$ and $X_{(b:m)}=F^{-1}(U_{(b:m)})$ where F and G are both continuous cdf's.

$$
\begin{aligned}
&P\left(\text{No Signal} \mid X_{(b:m)} = z\right) = p_U(z)\text{, say,} \\
&= P\left(Y_{(j:n)} \le z \mid X_{(b:m)} = z\right) \\
&= P\left(G^{-1}(U_{(j:n)}) \le z \mid X_{(b:m)} = z\right) \\
&= P\left(U_{(j:n)} \le G(z) \mid X_{(b:m)} = z\right) \\
&= \int_0^{G(z)} \frac{1}{\beta(j, n-j+1)} u^{j-1}(1-u)^{n-j}\, du
\end{aligned}
$$

since $\frac{1}{\beta(j, n-j+1)} u^{j-1}(1-u)^{n-j}$ is the pdf of $U_{(j:n)}$ (see equation (4.14)).

Result 4.14: Probability of no signal – unconditional

Let p_U denote the unconditional probability of no signal, then:

$$
\boxed{
\begin{aligned}
&P\left(\text{No Signal}\right) = P(Y_{(j:n)} \le X_{(b:m)}) \\
&= \int_0^1 \left(\frac{1}{\beta(j, n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(GF^{-1}(v)\right)^{j+h} \right) \frac{m!}{(b-1)!(m-b)!} v^{b-1}(1-v)^{m-b}\, dv
\end{aligned}
}
$$

$$
\begin{aligned}
&p_U \\
&= P\left(\text{No Signal}\right) \\
&= P(Y_{(j:n)} \le X_{(b:m)}) \\
&= E_{X_{(b:m)}}\left(P\left(Y_{(j:n)} \le X_{(b:m)} \mid X_{(b:m)}\right)\right) \\
&= E_{X_{(b:m)}}\left(P\left(G(Y_{(j:n)}) \le G(X_{(b:m)}) \mid X_{(b:m)}\right)\right)
\end{aligned}
$$

By the PIT we have that $U_{(j:n)} = G(Y_{(j:n)})$ where G is the continous cdf of the Phase II sample $Y_1, Y_2, \ldots, Y_n$. Using this we obtain

$$
= E_{X_{(b:m)}}\left(P\left(U_{(j:n)} \le G(X_{(b:m)}) \mid X_{(b:m)}\right)\right)
$$

By the PIT we have that $U_{(b:m)} = F(X_{(b:m)})$ so that $X_{(b:m)} = F^{-1}(U_{(b:m)})$ where F is the continous cdf of the reference sample $X_1, X_2, \ldots, X_m$. Using this we obtain

$$
\begin{aligned}
&= E_{U_{(b:m)}}\left(P\left(U_{(j:n)} \le GF^{-1}(U_{(b:m)}) \mid U_{(b:m)}\right)\right) \\
&= \int_0^1 P\left(U_{(j:n)} \le GF^{-1}(v) \mid U_{(b:m)} = v\right) f(v)\, dv
\end{aligned}
$$

where $f(v)$ is the pdf of $U_{(b:m)}$ which is given by $f(v)=\frac{m!}{(b-1)!(m-b)!}v^{b-1}(1-v)^{m-b}$

$$=\int_0^1\left(\int_0^{GF^{-1}(v)}\frac{1}{\beta(j,n-j+1)}u^{j-1}(1-u)^{n-j}du\right)\frac{m!}{(b-1)!(m-b)!}v^{b-1}(1-v)^{m-b}dv$$

The term $(1-u)^{n-j}$ can be expanded to $(1-u)^{n-j}=\sum_{h=0}^{n-j}(-1)^h\binom{n-j}{h}1^{n-j-h}u^h=\sum_{h=0}^{n-j}(-1)^h\binom{n-j}{h}u^h$ by using a binomial expansion (see equation (4.15)) and we obtain

$$=\int_0^1\left(\int_0^{GF^{-1}(v)}\frac{1}{\beta(j,n-j+1)}u^{j-1}\left(\sum_{h=0}^{n-j}(-1)^h\binom{n-j}{h}u^h\right)du\right)\frac{m!}{(b-1)!(m-b)!}v^{b-1}(1-v)^{m-b}dv$$

By taking all the constants out of the integral sign and simplifying by setting $u^{j-1}u^h=u^{j-1+h}$ we obtain

$$=\int_0^1\left(\frac{1}{\beta(j,n-j+1)}\sum_{h=0}^{n-j}(-1)^h\binom{n-j}{h}\int_0^{GF^{-1}(v)}u^{j-1+h}du\right)\frac{m!}{(b-1)!(m-b)!}v^{b-1}(1-v)^{m-b}dv$$

By integrating $\int_0^{GF^{-1}(v)}u^{j-1+h}du=\left.\frac{u^{j+h}}{j+h}\right|_{u=0}^{u=GF^{-1}(v)}=\frac{\left(GF^{-1}(v)\right)^{j+h}}{j+h}-0=\frac{\left(GF^{-1}(v)\right)^{j+h}}{j+h}$ we obtain

$$=\int_0^1\left(\frac{1}{\beta(j,n-j+1)}\sum_{h=0}^{n-j}(-1)^h\binom{n-j}{h}\frac{\left(GF^{-1}(v)\right)^{j+h}}{j+h}\right)\frac{m!}{(b-1)!(m-b)!}v^{b-1}(1-v)^{m-b}dv$$

which simplifies to

$$=\int_0^1\left(\frac{1}{\beta(j,n-j+1)}\sum_{h=0}^{n-j}\frac{(-1)^h}{j+h}\binom{n-j}{h}\left(GF^{-1}(v)\right)^{j+h}\right)\frac{m!}{(b-1)!(m-b)!}v^{b-1}(1-v)^{m-b}dv\,.$$

Result 4.15: Probability of a signal - conditional

A signalling event occurs when $Y_{(j:n)} > X_{(b:m)}$.

$$P\left(\text{Signal} \mid X_{(b:m)} = z\right) = 1 - \int_0^{G(z)} \frac{1}{\beta(j, n-j+1)} u^{j-1}(1-u)^{n-j}\, du$$

Result 4.16: Probability of a signal – unconditional

$$1 - p_U = P\left(\text{Signal}\right) =$$

$$1 - \int_0^1 \left(\frac{1}{\beta(j, n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(GF^{-1}(v)\right)^{j+h} \right) \frac{m!}{(b-1)!(m-b)!} v^{b-1}(1-v)^{m-b}\, dv$$

Result 4.17: Probability of a false alarm - conditional

$$CFAR = 1 - \int_0^{F(z)} \frac{1}{\beta(j, n-j+1)} u^{j-1}(1-u)^{n-j}\, du$$

Follows immediately from Result 4.15, since $G = F$.

Result 4.18: Probability of a false alarm - unconditional

$$FAR = 1 - \left(\frac{1}{\beta(j, n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} v^{j+h} \right) \frac{m!}{(b-1)!(m-b)!} v^{b-1}(1-v)^{m-b}\, dv$$

Follows immediately from Result 4.16, since $G = F$ and therefore $GF^{-1}(v) = FF^{-1}(v) = v$.

Result 4.19: Run-length distribution - conditional

$$P\left(N = k \mid X_{(b:m)} = z\right) = (p_U(z))^{k-1}(1 - p_U(z)) \quad \text{for} \quad k = 1,2,3,...$$

The conditional run length, denoted by $N \mid X_{(b:m)} = z$, will have a geometric distribution with parameter $1 - p_U(z)$, because all the signalling events are independent. Therefore we have that

$N \mid X_{(b:m)} = z \sim GEO(1 - p_U(z))$

$P\left(N = k \mid X_{(b:m)} = z\right) = (p_U(z))^{k-1}(1 - p_U(z)) \quad \text{for} \quad k = 1,2,3,...$

Consequently, the cumulative distribution function (cdf) is found from

$P\left(N \le k \mid X_{(b:m)} = z\right)$

$= \sum_{i=1}^{k}(p_U(z))^{i-1}(1 - p_U(z)) = 1 - (p_U(z))^k \quad \text{for} \quad k = 1,2,3,...$

We also have that

$P\left(N > k \mid X_{(b:m)} = z\right) = 1 - \left(1 - (p_U(z))^k\right) = (p_U(z))^k$.

Result 4.20: Average run-length - conditional

$$CARL = E\left(N \mid X_{(b:m)} = z\right) = \frac{1}{1 - p_U(z)}$$

or

$$CARL = E\left(N \mid X_{(b:m)} = z\right) = \sum_{k=0}^{\infty}(p_U(z))^k$$

Since the conditional run length, denoted by $N \mid X_{(b:m)} = z$ has a geometric distribution with parameter $1 - p_U(z)$, the conditional average run length is given by

$$CARL = E\left(N \mid X_{(b:m)} = z\right) = \frac{1}{1 - p_U(z)}.$$

The second expression follows immediately from the geometric expansion of $(1 - p_U(z))^{-1}$ for $p_U(z) < 1$.

Result 4.21: Run-length distribution - unconditional

$$P(N=k)=D_U^*(k-1)-D_U^*(k) \quad \text{for} \quad k=1,2,3,\dots, \quad D_U^*(0)=1$$

and

$$D_U^*(k)=\int_0^1\left(\frac{1}{\beta(j,n-j+1)}\sum_{h=0}^{n-j}\frac{(-1)^h}{j+h}\binom{n-j}{h}\left(GF^{-1}(v)\right)^{j+h}\right)^k\frac{m!}{(b-1)!(m-b)!}v^{b-1}(1-v)^{m-b}\,dv$$

$$\begin{aligned}
&P(N=k)\\
&=E_{X_{(b:m)}}\left(P\left(N=k \mid X_{(b:m)}=z\right)\right)\\
&=E_{X_{(b:m)}}\left((p_U(z))^{k-1}(1-p_U(z))\right)\\
&=E_{X_{(b:m)}}\left((p_U(z))^{k-1}-(p_U(z))^{k}\right)\\
&=E_{X_{(b:m)}}\left((p_U(z))^{k-1}\right)-E_{X_{(b:m)}}\left((p_U(z))^{k}\right)
\end{aligned}$$

By only focussing on $E_{X_{(b:m)}}\left((p_U(z))^k\right)$ we have

$$\begin{aligned}
&E_{X_{(b:m)}}\left((p_U(z))^k\right)\\
&=E_{U_{(b:m)}}\left(\left(\int_0^{GF^{-1}(v)}\frac{1}{\beta(j,n-j+1)}u^{j-1}(1-u)^{n-j}\,du\right)^k\right)
\end{aligned}$$

$$=\int_0^1\left(\left(\int_0^{GF^{-1}(v)}\frac{1}{\beta(j,n-j+1)}u^{j-1}(1-u)^{n-j}\,du\right)^k\right)f(v)dv$$

where $f(v)$ is the pdf of $U_{(b:m)}$ which is given by $f(v)=\frac{m!}{(b-1)!(m-b)!}v^{b-1}(1-v)^{m-b}$

$$=\int_0^1\left(\left(\int_0^{GF^{-1}(v)}\frac{1}{\beta(j,n-j+1)}u^{j-1}(1-u)^{n-j}\,du\right)^k\right)\frac{m!}{(b-1)!(m-b)!}v^{b-1}(1-v)^{m-b}\,dv$$

The term $(1-u)^{n-j}$ can be expanded to $(1-u)^{n-j} = \sum_{h=0}^{n-j}(-1)^h \binom{n-j}{h} 1^{n-j-h}u^h$ = $\sum_{h=0}^{n-j}(-1)^h \binom{n-j}{h} u^h$ by using a binomial expansion (see equation 4.15)) and we obtain

$$= \int_0^1 \left(\left(\int_0^{GF^{-1}(v)} \frac{1}{\beta(j,n-j+1)} u^{j-1} \left(\sum_{h=0}^{n-j}(-1)^h \binom{n-j}{h} u^h \right) du \right)^k \right) \frac{m!}{(b-1)!(m-b)!} v^{b-1}(1-v)^{m-b}\, dv$$

By taking all the constants out of the integral sign and simplifying by setting $u^{j-1}u^h = u^{j-1+h}$ we obtain

$$= \int_0^1 \left(\frac{1}{\beta(j,n-j+1)} \sum_{h=0}^{n-j}(-1)^h \binom{n-j}{h} \int_0^{GF^{-1}(v)} u^{j-1+h}\, du \right)^k \frac{m!}{(b-1)!(m-b)!} v^{b-1}(1-v)^{m-b}\, dv$$

By integrating $\int_0^{GF^{-1}(v)} u^{j-1+h}\, du = \left. \frac{u^{j+h}}{j+h} \right|_{u=0}^{u=GF^{-1}(v)} = \frac{\left(GF^{-1}(v)\right)^{j+h}}{j+h} - 0 = \frac{\left(GF^{-1}(v)\right)^{j+h}}{j+h}$ we obtain

$$= \int_0^1 \left(\frac{1}{\beta(j,n-j+1)} \sum_{h=0}^{n-j}(-1)^h \binom{n-j}{h} \frac{\left(GF^{-1}(v)\right)^{j+h}}{j+h} \right)^k \frac{m!}{(b-1)!(m-b)!} v^{b-1}(1-v)^{m-b}\, dv$$

which simplifies to

$$= \int_0^1 \left(\frac{1}{\beta(j,n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(GF^{-1}(v)\right)^{j+h} \right)^k \frac{m!}{(b-1)!(m-b)!} v^{b-1}(1-v)^{m-b}\, dv$$

$= D_U^*(k)$,

say.

Recall that $P(N=k) = E_{X_{(b:m)}}\left((p_U(z))^{k-1}\right) - E_{X_{(b:m)}}\left((p_U(z))^k\right)$.

Therefore, $P(N=k) = D_U^*(k-1) - D_U^*(k)$ for $k = 1,2,3,\ldots$, and $D_U^*(0) = 1$.

$D_U^*(0) = 1$ since

$$
\begin{aligned}
D_U^*(0) &= \int_0^1 \left(\frac{1}{\beta(j, n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(GF^{-1}(v)\right)^{j+h} \right)^0 \frac{m!}{(b-1)!(m-b)!} v^{b-1}(1-v)^{m-b}\, dv \\
&= \int_0^1 \frac{m!}{(b-1)!(m-b)!} v^{b-1}(1-v)^{m-b}\, dv \\
&= \int_0^1 f(v)dv \\
&= 1.
\end{aligned}
$$

This equals one, because in general, $\int_{-\infty}^{\infty} f(x)dx = 1$ for real x.

Result 4.22: In-control run-length distribution

$$P_C(N = k) = D_U(k-1) - D_U(k) \quad \text{for} \quad k = 1,2,3,\ldots, \quad D_U(0) = 1$$

and

$$D_U(k) = \int_0^1 \left(\frac{1}{\beta(j, n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} (v)^{j+h} \right)^k \frac{m!}{(b-1)!(m-b)!} v^{b-1}(1-v)^{m-b}\, dv$$

Recall that the reference sample of size m, $X_1, X_2, \ldots, X_m$, is available from an in-control process with a continuous cdf, $F(x)$. The plotting statistic $Y_{(j:n)}^h$ is the j^{th} order statistic from the h^{th} Phase II sample of size n_h. Let $G^h(y)$ denote the cdf of the distribution of the h^{th} Phase II sample. A process is said to be in-control at stage h when $G^h = F$. Assume that the Phase II samples are all of the same size, n, so that the subscript h can be suppressed. Therefore, a process is said to be in-control when $G = F$. Therefore, the in-control run length distribution is obtained by setting $G = F$ into the equation for the out-of-control run length distribution.

The out-of-control run length distribution for the upper one-sided chart is given by

$$P(N = k) = D_U^*(k-1) - D_U^*(k) \quad \text{for} \quad k = 1,2,3,\ldots, \quad D_U^*(0) = 1$$

and (4.18)

$$D_U^*(k) = \int_0^1 \left(\frac{1}{\beta(j,n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(GF^{-1}(v)\right)^{j+h} \right)^k \frac{m!}{(b-1)!(m-b)!} v^{b-1}(1-v)^{m-b}\,dv\,.$$

Therefore, the in-control run length distribution for the upper one-sided chart is obtained by setting $G = F$ into equation (4.18) and we obtain

$$P_C(N = k) = D_U(k-1) - D_U(k) \quad \text{for} \quad k = 1,2,3,\ldots, \quad D_U(0) = 1$$

and (4.19)

$$D_U(k) = \int_0^1 \left(\frac{1}{\beta(j,n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} (v)^{j+h} \right)^k \frac{m!}{(b-1)!(m-b)!} v^{b-1}(1-v)^{m-b}\,dv\,.$$

Result 4.23: Out-of-control average run-length - unconditional

$$UARL_{U,\delta} = \int_0^1 \frac{1}{1 - S_U(v,j,n,F,G)} f(v)dv$$

with

$$S_U(v,j,n,F,G) = \frac{1}{\beta(j,n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(GF^{-1}(v)\right)^{j+h}$$

Let $UARL_{U,\delta}$ denote the unconditional average run length, where δ refers to the out-of-control case. To derive an expression for the $UARL_{U,\delta}$, recall that

$$\begin{aligned}
&E_{X_{(b:m)}}\left((p_U(z))^k\right) \\
&= D_U^*(k) \\
&= \int_0^1 \left(\frac{1}{\beta(j,n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(GF^{-1}(v)\right)^{j+h} \right)^k \frac{m!}{(b-1)!(m-b)!} v^{b-1}(1-v)^{m-b}\,dv \\
&= \int_0^1 \left(\frac{1}{\beta(j,n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(GF^{-1}(v)\right)^{j+h} \right)^k f(v)dv
\end{aligned}$$

For simplicity let $S_U(v,j,n,F,G) = \frac{1}{\beta(j,n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(GF^{-1}(v)\right)^{j+h}$, therefore we obtain

$$D_U^*(k) = \int_0^1 (S_U(v, j, n, F, G))^k f(v)dv$$

Finally, we have that (from the second expression in Result 4.20)

$$\begin{aligned}
&UARL_{U,\delta} \\
&= \sum_{k=0}^{\infty} E_{X_{(b:m)}}\left((p_U(z))^k\right) \\
&= \sum_{k=0}^{\infty} D_U^*(k) \\
&= \sum_{k=0}^{\infty} \int_0^1 (S_U(v, j, n, F, G))^k f(v)dv \\
&= \int_0^1 \sum_{k=0}^{\infty} (S_U(v, j, n, F, G)^k f(v)dv \\
&= \int_0^1 \frac{1}{1 - S_U(v, j, n, F, G)} f(v)dv\,.
\end{aligned}$$

Result 4.24: In-control average run-length - unconditional

$$UARL_{U,0} = \int_0^1 \left(\frac{1}{1 - S_U(v, j, n)}\right) \frac{m!}{(b-1)!(m-b)!} v^{b-1}(1-v)^{m-b}\, dv$$

with

$$S_U(v, j, n) = \frac{1}{\beta(j, n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} (v)^{j+h}$$

Let $UARL_{U,0}$ denote the unconditional average run length, where 0 refers to the in-control case. To derive an expression for the $UARL_{U,0}$, recall that the in-control run length distribution for the upper one-sided chart is given by

$P_C(N = k) = D_U(k-1) - D_U(k)$ for $k = 1,2,3,\ldots,$ $D_U(0) = 1$

with

$$D_U(k) = \int_0^1 \left(\frac{1}{\beta(j, n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} (v)^{j+h}\right)^k \frac{m!}{(b-1)!(m-b)!} v^{b-1}(1-v)^{m-b}\, dv$$

$$=\int_0^1\left(\frac{1}{\beta(j,n-j+1)}\sum_{h=0}^{n-j}\frac{(-1)^h}{j+h}\binom{n-j}{h}(v)^{j+h}\right)^k f(v)dv\ .$$

For simplicity let $S_U(v,j,n)=\frac{1}{\beta(j,n-j+1)}\sum_{h=0}^{n-j}\frac{(-1)^h}{j+h}\binom{n-j}{h}(v)^{j+h}$, therefore we obtain

$$D_U(k)=\int_0^1\left(S_U(v,j,n)\right)^k f(v)dv$$

Finally, we have that

$$\begin{aligned}
&UARL_{U,0}\\
&=\sum_{k=0}^{\infty}D_U(k)\\
&=\sum_{k=0}^{\infty}\int_0^1\left(S_U(v,j,n)\right)^k f(v)dv\\
&=\int_0^1\sum_{k=0}^{\infty}\left(S_U(v,j,n)\right)^k f(v)dv\\
&=\int_0^1\frac{1}{1-S_U(v,j,n)}f(v)dv\ .
\end{aligned}$$

4.1.9. Two-sided control charts

For the two-sided chart we have $L\hat{C}L=X_{(a:m)}$ and $U\hat{C}L=X_{(b:m)}$. Therefore, a non-signalling event occurs when $X_{(a:m)}\le Y_{(j:n)}\le X_{(b:m)}$.

Result 4.25: Probability of no signal - conditional

$$\boxed{P\left(\text{No Signal}\mid X_{(a:m)}=x,X_{(b:m)}=z\right)=\int_{G(x)}^{G(z)}\frac{1}{\beta(j,n-j+1)}u^{j-1}(1-u)^{n-j}du}$$

Using the PIT, we know that $Y_{(j:n)}=G^{-1}(U_{(j:n)})$, $X_{(a:m)}=F^{-1}(U_{(a:m)})$ and $X_{(b:m)}=F^{-1}(U_{(b:m)})$ where F and G are both continuous cdf's.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

$P\left(\text{No Signal} \mid X_{(a:m)} = x, X_{(b:m)} = z\right) = p(x,z)$, say,
$= P\left(x \le Y_{(j:n)} \le z \mid X_{(a:m)} = x, X_{(b:m)} = z\right)$
$= P\left(x \le G^{-1}(U_{(j:n)}) \le z \mid X_{(a:m)} = x, X_{(b:m)} = z\right)$
$= P\left(G(x) \le U_{(j:n)} \le G(z) \mid X_{(a:m)} = x, X_{(b:m)} = z\right)$

$$= \int_{G(x)}^{G(z)} \frac{1}{\beta(j, n-j+1)} u^{j-1}(1-u)^{n-j}\, du$$

since $\dfrac{1}{\beta(j,n-j+1)} u^{j-1}(1-u)^{n-j}$ is the pdf of $U_{(j:n)}$ (see equation 4.14).

Result 4.26: Probability of no signal – unconditional

Let p denote the unconditional probability of no signal, then:

$$P\left(\text{No Signal}\right) = P(X_{(a:m)} \le Y_{(j:n)} \le X_{(b:m)})$$
$$\int_0^1 \int_0^t \left(\frac{1}{\beta(j,n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(\left(GF^{-1}(t)\right)^{j+h} - \left(GF^{-1}(s)\right)^{j+h} \right) \right) \times$$
$$\frac{m!}{(a-1)!(b-a-1)!(m-b)!} s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b}\, dsdt$$

p
$= P\left(\text{No Signal}\right)$
$= P(X_{(a:m)} \le Y_{(j:n)} \le X_{(b:m)})$
$= E_{X_{(a:m)}, X_{(b:m)}}\left(P(X_{(a:m)} \le Y_{(j:n)} \le X_{(b:m)}) \mid X_{(a:m)}, X_{(b:m)}\right)$
$= E_{X_{(a:m)}, X_{(b:m)}}\left(P(G(X_{(a:m)}) \le G(Y_{(j:n)}) \le G(X_{(b:m)})) \mid X_{(a:m)}, X_{(b:m)}\right)$

By the PIT we have that $U_{(j:n)} = G(Y_{(j:n)})$ where G is the continous cdf of the Phase II sample $Y_1, Y_2, ..., Y_n$. Using this we obtain

$= E_{X_{(a:m)}, X_{(b:m)}}\left(P(G(X_{(a:m)}) \le U_{(j:n)} \le G(X_{(b:m)})) \mid X_{(a:m)}, X_{(b:m)}\right)$

By the PIT we have that $U_{(a:m)} = F(X_{(a:m)})$ so that $X_{(a:m)} = F^{-1}(U_{(a:m)})$ and $U_{(b:m)} = F(X_{(b:m)})$ so that $X_{(b:m)} = F^{-1}(U_{(b:m)})$ where F is the continous cdf of the reference sample $X_1, X_2, ..., X_m$. Using this we obtain

$$= E_{U_{(a:m)},U_{(b:m)}}\left(P(GF^{-1}(U_{(a:m)}) \le U_{(j:n)} \le GF^{-1}(U_{(b:m)})) \mid U_{(a:m)}, U_{(b:m)}\right)$$

$$= \int_0^1 (GF^{-1}(s) \le U_{(j:n)} \le GF^{-1}(t)) \mid U_{(a:m)} = s, U_{(b:m)} = t) f(s,t) dsdt$$

where $f(s,t)$ is the joint pdf of $U_{(a:m)}$ and $U_{(b:m)}$ which is given by $f(s,t) = \frac{m!}{(a-1)!(b-a-1)!(m-b)!} s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b}$

$$= \int_0^1\int_0^t \left(\int_{GF^{-1}(s)}^{GF^{-1}(t)} \frac{1}{\beta(j,n-j+1)} u^{j-1}(1-u)^{n-j} du \right) f(s,t) dsdt$$

$$= \int_0^1\int_0^t \left(\left(\int_{GF^{-1}(s)}^{GF^{-1}(t)} \frac{1}{\beta(j,n-j+1)} u^{j-1}(1-u)^{n-j} du \right)^k \right) \frac{m!}{(a-1)!(b-a-1)!(m-b)!} s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b} dsdt$$

The term $(1-u)^{n-j}$ can be expanded to $(1-u)^{n-j} = \sum_{h=0}^{n-j}(-1)^h \binom{n-j}{h} 1^{n-j-h} u^h = \sum_{h=0}^{n-j}(-1)^h \binom{n-j}{h} u^h$ by using a binomial expansion (see equation 4.15)) and we obtain

$$= \int_0^1\int_0^t \left(\left(\int_{GF^{-1}(s)}^{GF^{-1}(t)} \frac{1}{\beta(j,n-j+1)} u^{j-1} \left(\sum_{h=0}^{n-j}(-1)^h \binom{n-j}{h} u^h \right) du \right)^k \right) \frac{m!}{(a-1)!(b-a-1)!(m-b)!} \times$$
$$s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b} dsdt$$

By taking all the constants out of the integral sign and simplifying by setting $u^{j-1}u^h = u^{j-1+h}$ we obtain

$$= \int_0^1\int_0^t \left(\frac{1}{\beta(j,n-j+1)} \sum_{h=0}^{n-j}(-1)^h \binom{n-j}{h} \int_{GF^{-1}(s)}^{GF^{-1}(t)} u^{j-1+h} du \right)^k \frac{m!}{(a-1)!(b-a-1)!(m-b)!} \times$$
$$s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b} dsdt$$

By integrating $\int_{GF^{-1}(s)}^{GF^{-1}(t)} u^{j-1+h} du = \left. \frac{u^{j+h}}{j+h} \right|_{u=GF^{-1}(s)}^{u=GF^{-1}(t)} = \frac{\left(GF^{-1}(t)\right)^{j+h} - \left(GF^{-1}(s)\right)^{j+h}}{j+h}$ we obtain

$$= \int_0^1 \int_0^t \left(\frac{1}{\beta(j,n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(\left(GF^{-1}(t)\right)^{j+h} - \left(GF^{-1}(s)\right)^{j+h} \right) \right) \times$$
$$\frac{m!}{(a-1)!(b-a-1)!(m-b)!} s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b} \, dsdt \, .$$

Result 4.27: Probability of a signal - conditional

A signalling event occurs when $Y_{(j:n)} < X_{(a:m)}$ or $Y_{(j:n)} > X_{(b:m)}$.

$$P\left(\text{Signal} \mid X_{(a:m)} = x, X_{(b:m)} = z\right) = 1 - \int_{G(x)}^{G(z)} \frac{1}{\beta(j,n-j+1)} u^{j-1}(1-u)^{n-j} \, du$$

Result 4.28: Probability of a signal - unconditional

$$1 - p = P(\text{Signal}) =$$
$$1 - \int_0^1 \int_0^t \left(\frac{1}{\beta(j,n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(\left(GF^{-1}(t)\right)^{j+h} - \left(GF^{-1}(s)\right)^{j+h} \right) \right) \times$$
$$\frac{m!}{(a-1)!(b-a-1)!(m-b)!} s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b} \, dsdt$$

Result 4.29: Probability of a false alarm - conditional

$$CFAR = 1 - \int_{F(x)}^{F(z)} \frac{1}{\beta(j,n-j+1)} u^{j-1}(1-u)^{n-j} \, du$$

Follows immediately from Result 4.27, since $G = F$.

Result 4.30: Probability of a false alarm - unconditional

$$FAR = 1 - \int_0^1 \int_0^t \left(\frac{1}{\beta(j, n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(t^{j+h} - s^{j+h} \right) \right) \times \frac{m!}{(a-1)!(b-a-1)!(m-b)!} s^{a-1} (t-s)^{b-a-1} (1-t)^{m-b} \, ds dt$$

Follows immediately from Result 4.28, since $G = F$ and therefore $GF^{-1}(s) = FF^{-1}(s) = s$ and $GF^{-1}(t) = FF^{-1}(t) = t$.

Result 4.31: Run-length distribution - conditional

$$P\left(N = k \mid X_{(a:m)} = x, X_{(b:m)} = z\right) = (p(x,z))^{k-1}(1 - p(x,z)) \quad \text{for} \quad k = 1,2,3,...$$

The conditional run length, denoted by $N \mid X_{(a:m)} = x, X_{(b:m)} = z$, will have a geometric distribution with parameter $1 - p(x,z)$, because all the signalling events are independent. Therefore we have that

$N \mid X_{(a:m)} = x, X_{(b:m)} = z \sim GEO(1 - p(x,z))$

$P\left(N = k \mid X_{(a:m)} = x, X_{(b:m)} = z\right) = (p(x,z))^{k-1}(1 - p(x,z))$

for $k = 1,2,3,...$

Consequently, the cumulative distribution function (cdf) is found from

$P\left(N \le k \mid X_{(a:m)} = x, X_{(b:m)} = z\right)$

$= \sum_{i=1}^{k} (p(x,z))^{i-1}(1 - p(x,z)) = 1 - (p(x,z))^k$ for $k = 1,2,3,...$

We also have that

$P\left(N > k \mid X_{(a:m)} = x, X_{(b:m)} = z\right) = 1 - \left(1 - (p(x,z))^k\right) = (p(x,z))^k$.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Result 4.32: Average run-length - conditional

$$CARL = E\left(N \mid X_{(a:m)} = x, X_{(b:m)} = z\right) = \frac{1}{1 - p(x,z)}$$

or

$$CARL = E\left(N \mid X_{(a:m)} = x, X_{(b:m)} = z\right) = \sum_{k=0}^{\infty} (p(x,z))^k$$

Since the conditional run length, denoted by $N \mid X_{(a:m)} = x, X_{(b:m)} = z$ has a geometric distribution with parameter $1 - p(x,z)$, the conditional average run length is given by

$$CARL = E\left(N \mid X_{(a:m)} = x, X_{(b:m)} = z\right) = \frac{1}{1 - p(x,z)}.$$

The second expression follows immediately from the geometric expansion of $(1 - p(x,z))^{-1}$ for $p(x,z) < 1$.

Result 4.33: Run-length distribution - unconditional

$$P(N = k) = D^*(k-1) - D^*(k) \quad \text{for} \quad k = 1,2,3,\ldots, \quad D^*(0) = 1$$

and

$$D^*(k) = \int_0^1 \int_0^t \left(\frac{1}{\beta(j, n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(\left(GF^{-1}(t)\right)^{j+h} - \left(GF^{-1}(s)\right)^{j+h} \right) \right)^k \times$$
$$\frac{m!}{(a-1)!(b-a-1)!(m-b)!} s^{a-1} (t-s)^{b-a-1} (1-t)^{m-b} \, ds dt$$

$$\begin{aligned}
&P(N = k) \\
&= E_{X_{(a:m)}, X_{(b:m)}}\left(P\left(N = k \mid X_{(a:m)} = x, X_{(b:m)} = z\right)\right) \\
&= E_{X_{(a:m)}, X_{(b:m)}}\left((p(x,z))^{k-1}(1 - p(x,z))\right) \\
&= E_{X_{(a:m)}, X_{(b:m)}}\left((p(x,z))^{k-1} - (p(x,z))^k\right) \\
&= E_{X_{(a:m)}, X_{(b:m)}}\left((p(x,z))^{k-1}\right) - E_{X_{(a:m)}, X_{(b:m)}}\left((p(x,z))^k\right)
\end{aligned}$$

By only focussing on $E_{X_{(a:m)}, X_{(b:m)}}\left((p(x,z))^k\right)$ we have

$$E_{X_{(a:m)},X_{(b:m)}}\left((p(x,z))^k\right)$$

$$= E_{U_{(a:m)},U_{(b:m)}}\left(\left(\int_{GF^{-1}(s)}^{GF^{-1}(t)}\frac{1}{\beta(j,n-j+1)}u^{j-1}(1-u)^{n-j}du\right)^k\right)$$

$$= \int_0^1\int_0^t\left(\left(\int_{GF^{-1}(s)}^{GF^{-1}(t)}\frac{1}{\beta(j,n-j+1)}u^{j-1}(1-u)^{n-j}du\right)^k\right)f(s,t)dsdt$$

where $f(s,t)$ is the joint pdf of $U_{(a:m)}$ and $U_{(b:m)}$ which is given by $f(s,t)=\frac{m!}{(a-1)!(b-a-1)!(m-b)!}s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b}$

$$= \int_0^1\int_0^t\left(\left(\int_{GF^{-1}(s)}^{GF^{-1}(t)}\frac{1}{\beta(j,n-j+1)}u^{j-1}(1-u)^{n-j}du\right)^k\right)\frac{m!}{(a-1)!(b-a-1)!(m-b)!}s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b}dsdt$$

The term $(1-u)^{n-j}$ can be expanded to $(1-u)^{n-j}=\sum_{h=0}^{n-j}(-1)^h\binom{n-j}{h}1^{n-j-h}u^h=\sum_{h=0}^{n-j}(-1)^h\binom{n-j}{h}u^h$ by using a binomial expansion (see equation (4.15)) and we obtain

$$= \int_0^1\int_0^t\left(\left(\int_{GF^{-1}(s)}^{GF^{-1}(t)}\frac{1}{\beta(j,n-j+1)}u^{j-1}\left(\sum_{h=0}^{n-j}(-1)^h\binom{n-j}{h}u^h\right)du\right)^k\right)\frac{m!}{(a-1)!(b-a-1)!(m-b)!}\times$$
$$s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b}dsdt$$

By taking all the constants out of the integral sign and simplifying by setting $u^{j-1}u^h=u^{j-1+h}$ we obtain

$$= \int_0^1\int_0^t\left(\frac{1}{\beta(j,n-j+1)}\sum_{h=0}^{n-j}(-1)^h\binom{n-j}{h}\int_{GF^{-1}(s)}^{GF^{-1}(t)}u^{j-1+h}du\right)^k\frac{m!}{(a-1)!(b-a-1)!(m-b)!}\times$$
$$s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b}dsdt$$

By integrating $\int\limits_{GF^{-1}(s)}^{GF^{-1}(t)} u^{j-1+h} du = \left.\frac{u^{j+h}}{j+h}\right|_{u=GF^{-1}(s)}^{u=GF^{-1}(t)} = \frac{\left(GF^{-1}(t)\right)^{j+h} - \left(GF^{-1}(s)\right)^{j+h}}{j+h}$ we obtain

$$= \int_0^1\int_0^t \left(\frac{1}{\beta(j,n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h}\binom{n-j}{h}\left(\left(GF^{-1}(t)\right)^{j+h} - \left(GF^{-1}(s)\right)^{j+h}\right)\right)^k \times$$
$$\frac{m!}{(a-1)!(b-a-1)!(m-b)!} s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b} dsdt$$

$= D^*(k)$,

say.

Recall that $P(N=k) = E_{X_{(a:m)},X_{(b:m)}}\left((p(x,z))^{k-1}\right) - E_{X_{(a:m)},X_{(b:m)}}\left((p(x,z))^{k}\right)$

Therefore, $P(N=k) = D^*(k-1) - D^*(k)$ for $k = 1,2,3,\ldots,$ and $D^*(0) = 1$.

$D^*(0) = 1$ since

$$D^*(0) = \int_0^1\int_0^t \left(\frac{1}{\beta(j,n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h}\binom{n-j}{h}\left(\left(GF^{-1}(t)\right)^{j+h} - \left(GF^{-1}(s)\right)^{j+h}\right)\right)^0 \times$$
$$\frac{m!}{(a-1)!(b-a-1)!(m-b)!} s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b} dsdt$$

$$= \int_0^1\int_0^t \frac{m!}{(a-1)!(b-a-1)!(m-b)!} s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b} dsdt$$
$$= \int_0^1\int_0^t f(s,t) dsdt$$
$$= 1.$$

Result 4.34: In-control run-length distribution

$$P_C(N=k) = D(k-1) - D(k) \quad \text{for} \quad k = 1,2,3,\ldots, \quad D(0) = 1$$

and

$$D(k) = \int_0^1\int_0^t \left(\frac{1}{\beta(j,n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h}\binom{n-j}{h}\left(t^{j+h} - s^{j+h}\right)\right)^k \frac{m!}{(a-1)!(b-a-1)!(m-b)!} \times$$
$$s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b} dsdt$$

Recall that the reference sample of size m, $X_1, X_2, ..., X_m$, is available from an in-control process with a continuous cdf, $F(x)$. The plotting statistic $Y_{(j:n)}^h$ is the j^{th} order statistic from the h^{th} Phase II sample of size n_h. Let $G^h(y)$ denote the cdf of the distribution of the h^{th} Phase II sample. A process is said to be in-control at stage h when $G^h = F$. Assume that the Phase II samples are all of the same size, n, so that the subscript h can be suppressed. Therefore, a process is said to be in-control when $G = F$. Therefore, the in-control run length distribution is obtained by setting $G = F$ into the equation for the out-of-control run length distribution.

The out-of-control run length distribution for the two-sided chart is given by

$$P(N = k) = D^*(k-1) - D^*(k) \quad \text{for} \quad k = 1,2,3,..., \quad D^*(0) = 1$$

and (4.20)

$$D^*(k) = \int_0^1 \int_0^t \left(\frac{1}{\beta(j, n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(\left(GF^{-1}(t)\right)^{j+h} - \left(GF^{-1}(s)\right)^{j+h} \right) \right)^k \times$$

$$\frac{m!}{(a-1)!(b-a-1)!(m-b)!} s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b} \, ds dt \, .$$

Therefore, the in-control run length distribution for the two-sided chart is obtained by setting $G = F$ into equation (4.20) and we obtain

$$P_C(N = k) = D(k-1) - D(k) \quad \text{for} \quad k = 1,2,3,..., \quad D(0) = 1$$

and (4.21)

$$D(k) = \int_0^1 \int_0^t \left(\frac{1}{\beta(j, n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(t^{j+h} - s^{j+h} \right) \right)^k \frac{m!}{(a-1)!(b-a-1)!(m-b)!} \times$$

$$s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b} \, ds dt \, .$$

Result 4.35: Out-of-control average run-length - unconditional

$$UARL_\delta = \int_0^1 \int_0^t \frac{1}{1 - S(s,t,j,n,F,G)} f(s,t)dsdt$$

with

$$S(s,t,j,n,F,G) = \frac{1}{\beta(j,n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(\left(GF^{-1}(t)\right)^{j+h} - \left(GF^{-1}(s)\right)^{j+h} \right)$$

Let $UARL_\delta$ denote the unconditional average run length, where δ refers to the out-of-control case. To derive an expression for the $UARL_\delta$, recall that

$$\begin{aligned}
&E_{X_{(a:m)},X_{(b:m)}}\left((p(x,z))^k\right) \\
&= D^*(k) \\
&= \int_0^1 \int_0^t \left(\frac{1}{\beta(j,n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(\left(GF^{-1}(t)\right)^{j+h} - \left(GF^{-1}(s)\right)^{j+h} \right) \right)^k \times \\
&\quad \frac{m!}{(a-1)!(b-a-1)!(m-b)!} s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b} dsdt\,. \\
&= \int_0^1 \int_0^t \left(\frac{1}{\beta(j,n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(\left(GF^{-1}(t)\right)^{j+h} - \left(GF^{-1}(s)\right)^{j+h} \right) \right)^k f(s,t)dsdt
\end{aligned}$$

Let $S(s,t,j,n,F,G) = \frac{1}{\beta(j,n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(\left(GF^{-1}(t)\right)^{j+h} - \left(GF^{-1}(s)\right)^{j+h} \right)$

$$= \int_0^1 \int_0^t \left(S(s,t,j,n,F,G)\right)^k f(s,t)dsdt\,.$$

Finally, we have that (from the second expression in Result 4.32)

$$\begin{aligned}
&UARL_\delta \\
&= \sum_{k=0}^{\infty} E_{X_{(a:m)},X_{(b:m)}}\left((p(x,z))^k\right) \\
&= \sum_{k=0}^{\infty} D^*(k)
\end{aligned}$$

$$= \sum_{k=0}^{\infty} \int_0^1 \int_0^t \left(S(s,t,j,n,F,G)\right)^k f(s,t)\,dsdt$$

$$= \int_0^1 \int_0^t \sum_{k=0}^{\infty} \left(S(s,t,j,n,F,G)\right)^k f(s,t)\,dsdt$$

$$= \int_0^1 \int_0^t \frac{1}{1-S(s,t,j,n,F,G)} f(s,t)\,dsdt\,.$$

Result 4.36: In-control average run-length - unconditional

$$UARL_0 = \int_0^1 \int_0^t \left(\frac{1}{1-S(s,t,j,n)}\right) \frac{m!}{(a-1)!(b-a-1)!(m-b)!} s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b}\,dsdt$$

with

$$S(s,t,j,n) = \frac{1}{\beta(j,n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(t^{j+h} - s^{j+h}\right)$$

Let $UARL_0$ denote the unconditional average run length, where 0 refers to the in-control case. To derive an expression for the $UARL_0$, recall that the in-control run length distribution for the two-sided chart is given by

$P_C(N=k) = D(k-1) - D(k)$ for $k = 1,2,3,\ldots$, $D(0)=1$

with

$$D(k) = \int_0^1 \int_0^t \left(\frac{1}{\beta(j,n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(t^{j+h} - s^{j+h}\right)\right)^k \frac{m!}{(a-1)!(b-a-1)!(m-b)!} \times$$
$$s^{a-1}(t-s)^{b-a-1}(1-t)^{m-b}\,dsdt$$

$$= \int_0^1 \int_0^t \left(\frac{1}{\beta(j,n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(t^{j+h} - s^{j+h}\right)\right)^k f(s,t)\,dsdt$$

For simplicity let $S(s,t,j,n) = \frac{1}{\beta(j,n-j+1)} \sum_{h=0}^{n-j} \frac{(-1)^h}{j+h} \binom{n-j}{h} \left(t^{j+h} - s^{j+h}\right)$, therefore we obtain

$$D(k) = \int_0^1 \int_0^t \left(S(s,t,j,n)\right)^k f(s,t)\,dsdt$$

Finally, we have that

$$
\begin{aligned}
&UARL_0 \\
&= \sum_{k=0}^{\infty} D(k) \\
&= \sum_{k=0}^{\infty} \int_0^1 \int_0^t (S(s,t,j,n))^k f(s,t)\,ds\,dt \\
&= \int_0^1 \int_0^t \sum_{k=0}^{\infty} (S(s,t,j,n))^k f(s,t)\,ds\,dt \\
&= \int_0^1 \int_0^t \frac{1}{1-S(s,t,j,n)} f(s,t)\,ds\,dt\,.
\end{aligned}
$$

4.1.10. Run-length distribution and *ARL* under some alternatives

In the nonparametric setting, we consider, more generally, monitoring the center value or the location parameter and/or a scale parameter of a process. The location parameter represents a typical value and could be the mean or the median or some other percentile of the distribution; the latter two are especially attractive when the underlying distribution is skewed. When the underlying distribution is symmetric, the mean and the median are the same. Also in the nonparametric setting, the processes are implicitly assumed to follow (i) a location model, with a cdf $F(x-\theta)$, where θ is the location parameter or (ii) a scale model, with a cdf $F\left(\frac{x}{\tau}\right)$, where $\tau(>0)$ is the scale parameter. Even more generally, one might consider (iii) the location-scale model with cdf $F\left(\frac{x-\theta}{\tau}\right)$, where θ and τ are the location and the scale parameter, respectively.

Recall that the reference sample is available from an in-control process with a continuous cdf, $F(x)$, and that $G(y)$ denotes the cdf of the distribution of the Phase II sample. The run length distribution depends on F and G, through the function $\psi = GF^{-1}$. A process is said to be in-control when $G = F$. In this case $\psi(u) = G\left(F^{-1}(u)\right) = F\left(F^{-1}(u)\right) = u$.

4.1.10.1. Location alternatives

$F(x) = H(x-\theta_1)$ and $G(x) = H(x-\theta_2)$, where H is a continuous cdf, $x \in \Re$ and $\theta_1, \theta_2 \in \Re$, $\psi(u) = H(\theta_1 - \theta_2 + H^{-1}(u))$. For example, let both F and G be normally distributed with a change in the mean, i.e. $F(x) = \Phi(x)$ and $G(x) = \Phi(x-\theta)$. But $\psi(u)$ $= G(F^{-1}(u))$ (by definition) and therefore $\psi(u) = \Phi(\Phi^{-1}(u) - \theta)$.

4.1.10.2. Scale alternatives

$F(x) = H\left(\frac{x}{\gamma_1}\right)$ and $G(x) = H\left(\frac{x}{\gamma_2}\right)$, where H is a continuous cdf, $x \in \Re$ and $\gamma_1, \gamma_2 \in \Re^+$, $\psi(u) = H\left(\frac{\gamma_1}{\gamma_2} H^{-1}(u)\right)$. For example, let both F and G be normally distributed with a change in the spread, i.e. $F(x) = \Phi(x)$ and $G(x) = \Phi\left(\frac{x}{\gamma}\right)$. But $\psi(u) = G(F^{-1}(u))$ (by definition) and therefore $\psi(u) = \Phi\left(\frac{\Phi^{-1}(u)}{\gamma}\right)$.

4.1.10.3. Location-scale alternatives

$F(x) = H\left(\frac{x-\theta_1}{\gamma_1}\right)$ and $G(x) = H\left(\frac{x-\theta_2}{\gamma_2}\right)$, where H is a continuous cdf, $x \in \Re$, $\theta_1, \theta_2 \in \Re$, and $\gamma_1, \gamma_2 \in \Re^+$, $\psi(u) = H\left(\frac{\theta_1 - \theta_2}{\gamma_2} + \frac{\gamma_1}{\gamma_2} H^{-1}(u)\right)$. For example, let both F and G be normally distributed with a change in the mean and spread, i.e. $F(x) = \Phi(x)$ and $G(x) = \Phi\left(\frac{x-\theta}{\gamma}\right)$. But $\psi(u) = G(F^{-1}(u))$ (by definition) and therefore $\psi(u) =$ $\Phi\left(\frac{\Phi^{-1}(u) - \theta}{\gamma}\right) = \Phi\left(\frac{\Phi^{-1}(u)}{\gamma} - \frac{\theta}{\gamma}\right)$.

4.1.10.4. Lehmann alternatives

$G(x) = F^{\delta}(x)$, where $x \in \Re$ and $\delta \in \Re^{+}$, $\psi(u) = u^{\delta}$. For example, let $F(x) = u$ and $G(x) = (F(x))^{\delta}$. But $\psi(u) = G(F^{-1}(u))$ (by definition) and therefore $\psi(u) = (F(F^{-1}(u)))^{\delta} = u^{\delta}$.

- For $\delta = 1$: $\psi(u) = u = F(x)$.
- For $\delta = 2$: $\psi(u) = u^2 = F^2(x)$.

4.1.10.5. Proportional hazards alternatives

$G(x) = 1-(1-F(x))^{\gamma}$, where $x \in \Re$ and $\gamma \in \Re^{+}$, $\psi(u) = 1-(1-u)^{\gamma}$. For example, let $F(x) = u$ and $G(x) = 1-(1-F(x))^{\gamma}$. But $\psi(u) = G(F^{-1}(u))$ (by definition) and therefore $\psi(u) =$ $1-(1-F(F^{-1}(u)))^{\gamma} = 1-(1-u)^{\gamma}$.

4.1.10.6. Summary

Although a lot of research has been done in the last few years regarding Lehmann and proportional hazard alternatives (see for example Van der Laan and Chakraborti (1999)), more remains to be done. Van der Laan and Chakraborti (1999) showed that the power of a precedence test can be determined for both the Lehmann and proportional hazards alternatives. The body of literature on Lehmann and proportional hazards alternatives is growing. However, in our opinion, a discussion on this topic is better postponed for the future.

4.2. The Shewhart-type control chart with runs-type signalling rules

4.2.1. Introduction

Chakraborti, Eryilmaz and Human (2006) considered enhancing the precedence charts with *2-of-2* type signalling rules. The *2-of-2* DR and *2-of-2* KL rules were defined previously (see Section 3.2). Recall that the *2-of-2* KL chart signals when two of the most recent charting statistics both fall either on or above or on or below the control limits, whereas the *2-of-2* DR chart signals when the charting statistics fall either both on or above or both on or below or one on or above (below) and the next one on or below (above) the control limits. We illustrate these procedures using the Montgomery (2001) piston ring data.

4.2.2. Example

Example 4.1

A sign-like control chart based on the Montgomery (2001) piston ring data

We illustrate the sign-like control charts using a set of data from Montgomery (2001, Tables 5.1 and 5.2) on inside diameters of piston rings manufactured by a forging process. Table 5.1 of Montgomery (2001) contains 25 retrospective or Phase I samples, each of size five, that were collected when the process was thought to be in-control. When working with individual observations, we have $25 \times 5 = 125$, i.e. $m = 125$, individual observations. Table 5.2 of Montgomery (2001) contains 15 prospective or Phase II samples, each of five observations.

In order to implement the control charts, the charting constants are needed. Generally, one finds the chart constants so that a specified ARL_0, such as 500 or 370, is obtained. For the precedence type charts, symmetric control limits are used so that $b = m - a + 1$ and only one charting constant a (≥ 1) needs to be found. Possible control limits for the three charts are shown in Table 4.2 for $m = 125$, $n = 5$ and $j = 3$, along with the corresponding *FAR* and ARL_0 values. The basic Shewhart-type precedence chart is referred to as the *1-of-1* chart.

Table 4.2. In-control average run length (ARL_0), false alarm rate (FAR) and chart constant (a) for the *1-of-1*, *2-of-2* DR and *2-of-2* KL precedence charts when $m = 125$, $n = 5$ and $j = 3$ *.

	1-of-1		***2-of-2* DR**			***2-of-2* KL**		
a	ARL_0	***FAR***	***a***	ARL_0	***FAR***	***a***	ARL_0	***FAR***
5	1315.98	0.001865	19	464.38	0.004020	19	819.47	0.002355
6	695.09	0.002948	20	344.73	0.005195	20	608.81	0.003019
7	413.80	0.004358	21	260.69	0.006627	21	460.54	0.003823
8	267.40	0.006164	22	200.46	0.008356	22	354.09	0.004788

Thus, for an ARL_0 of 500, one can take $a = 7$ and $b = 119$ so that the control limits for the *1-of-1* precedence chart are the 7th and the 119th ordered values of the reference sample. Thus $L\hat{C}L = X_{(7:125)} = 73.984$ and $U\hat{C}L = X_{(119:125)} = 74.017$, which yield an in-control average run length of 413.80 and a FAR of 0.0044. A plot of the sample medians for the *1-of-1* chart is shown in Figure 4.3. It is seen that the *1-of-1* precedence chart signals on the 12th sample in the prospective phase.

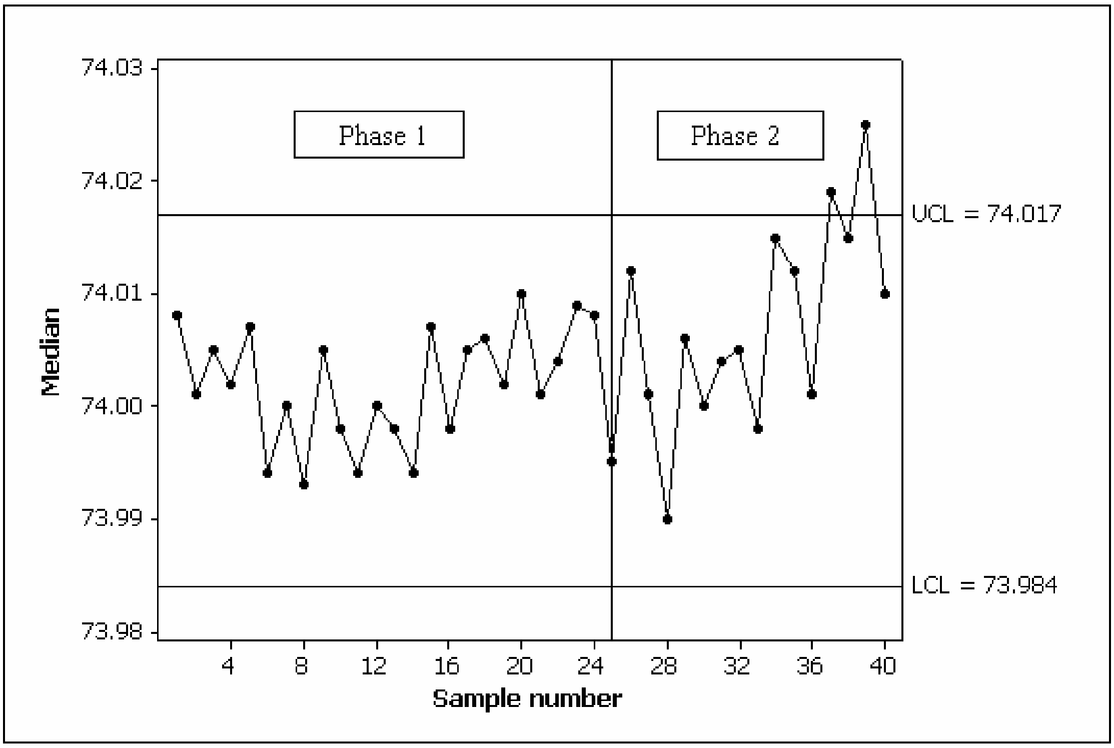

Figure 4.3. *1-of-1* Precedence chart for the Montgomery (2001) piston ring data.

* Table 4.2 appears in Chakraborti, Eryilmaz and Human (2006), Table 3.

For the *2-of-2* DR chart, take $a = 19$ so that $b = 125 - 19 + 1 = 107$ and the resulting limits, $L\hat{C}L = X_{(19:125)} = 73.992$ and $U\hat{C}L = X_{(107:125)} = 74.012$, yield an ARL_0 and FAR of 464.38 and 0.0040, respectively. Note however that if one chooses $a = 20$ so that $b = 106$, the control limits are $L\hat{C}L = X_{(20:125)}$ and $U\hat{C}L = X_{(106:125)}$ and the ARL_0 decreases to 344.73, whereas the *FAR* slightly increases to 0.0052. The *2-of-2* DR chart is shown in Figure 4.4.

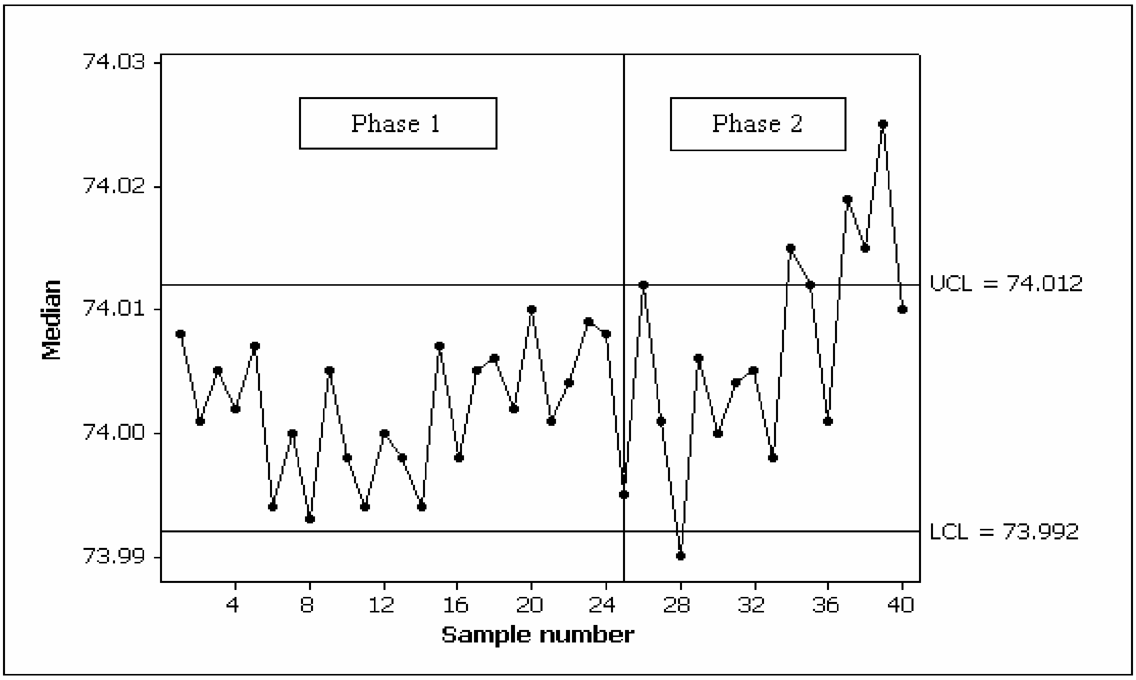

Figure 4.4. 2-*of*-*2* DR precedence chart for the Montgomery (2001) piston ring data.

For the *2-of-2* KL chart take $a = 21$ so that $b = 125 - 21 + 1 = 105$ so that $L\hat{C}L = X_{(21:125)} = 73.992$ and $U\hat{C}L = X_{(105:125)} = 74.011$, and this yields an ARL_0 of 460.54 and a *FAR* of 0.0038, respectively. This *2-of-2* KL chart is almost identical to the DR chart in Figure 4.4.

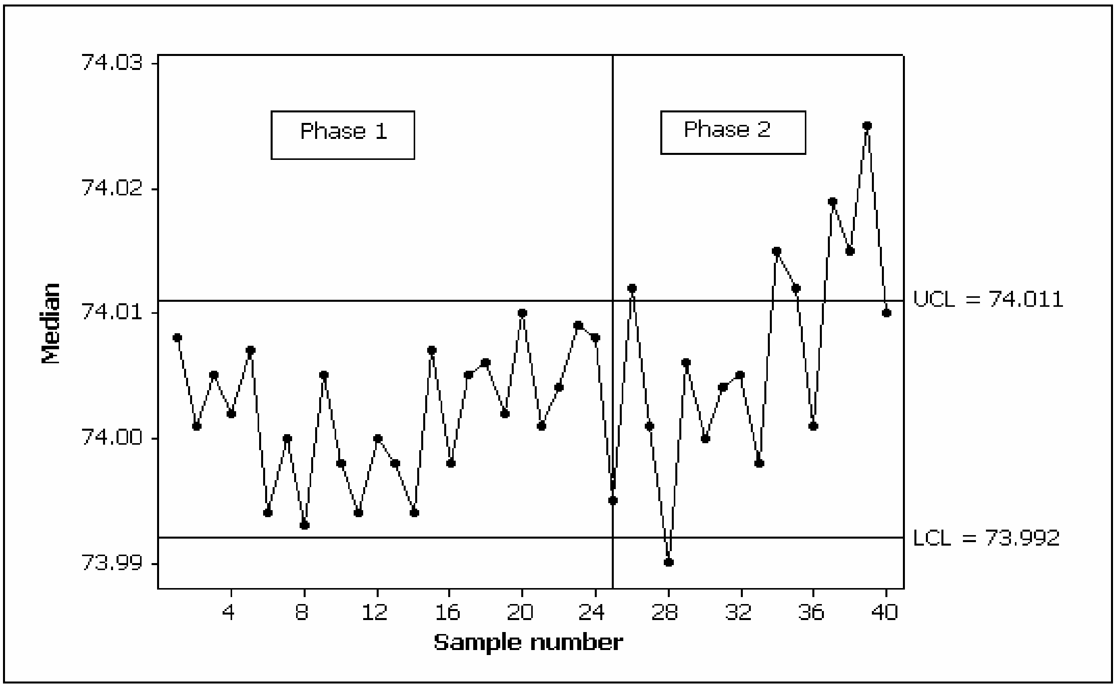

Figure 4.5. 2-*of*-2 KL precedence chart for the Montgomery (2001) piston ring data.

Both the *2-of-2* DR and KL charts signal on the 10th sample in the prospective phase. Note, however, that the achieved *FAR* values for all three charts are much larger than the nominal *FAR* of 0.0027.

4.2.3. Summary

In this chapter we examined sign-like control charts with runs-type signalling rules. We illustrated these procedures using the piston ring data from Montgomery (2001) to help the reader to understand the subject more thoroughly. There are many advantages to using these nonparametric charts (see Section 1.4). Chakraborti, Eryilmaz and Human (2006) draw attention to two advantages in particular, namely, that these charts can be applied as soon as the required order statistics are observed (recall that both the control limits and the charting statistic are based on order statistics), whereas for the Shewhart $\overline{X}$ charts one needs the full dataset to calculate the average. Moreover, these charts can be adapted to and applied in the case of ordinal data. As a result Chakraborti, Eryilmaz and Human (2006) recommend that these charts be used in practice.

Chapter 5: Signed-rank-like charts

5.1. The Shewhart-type control chart

5.1.1. Introduction

The statistics used in nonparametric control charts are mostly signs, ranks and signed-ranks and related to nonparametric procedures, such as the Wilcoxon signed-rank test and the Mann-Whitney-Wilcoxon rank-sum test. When considering nonparametric tests based on ranks, such tests deal with the ranking of independent, identically distributed (iid) random variables (under the assumption that the process is in-control). In Chapter 5 we consider nonparametric tests that involve ranking random variables that are exchangeable (again, this holds under the assumption that the process is in-control), meaning that each possible ranking is equally likely. Randles and Wolfe (1979) state that the term *rank-like* is used to describe a type of test procedure where the variables that are ranked are not the original observations, but are, instead, functions of them. The term *rank-like* was first introduced by Moses (1963). Moses's rank-like test is a nonparametric test for comparing differences in dispersion between two samples in which the medians are not equal. This requires randomly allocating the sample observations into two subgroups, ranking the subgroups according to their dispersion indexes and calculating the ranks sums for each subgroup. It should be noted that although Moses's rank-like test uses rankings of iid random variables (under the assumption that the process is in-control), these variables are not the original observations, but instead, functions of them. Bakir (2006) considered what are called *signed-rank-like* (SRL) statistics and used these to construct distribution-free charts. He uses the median of a reference sample (taken when the process was operating in-control) to estimate the unknown in-control process center.

5.1.2. Definition of the signed-rank-like test statistic

Assume that a reference sample of size $m > 1$, $X_1, X_2, ..., X_m$, is available from an in-control process with an unknown continuous cdf $F(x)$. Let $Y_{i1}, Y_{i2}, ..., Y_{in}$, $i = 1,2,...,$ denote the i^{th} test sample of size n. In case U the median of the in-control distribution (assumed to be symmetric) is unknown and can be estimated by the median of a reference sample, say *M*.

Let R_{ij}^{*} denote the rank of $\left|y_{ij}-M\right|$ within the subgroup $\left(\left|y_{i1}-M\right|,\ldots,\left|y_{in}-M\right|\right)$ for $i=1,2,3\ldots$. R_{ij}^{*} can be calculated using

$$R_{ij}^{*}=1+\sum_{k=1}^{n}(I\mid y_{ik}-M\mid<\mid y_{ij}-M\mid) \text{ for } j=1,2,\ldots,n \tag{5.1}$$

where I is the indicator function defined by $I(x)=1,0$ if x is true or false.

The charting statistic is given by

$$SRL_{i}=\sum_{j=1}^{n}sign\left(y_{ij}-M\right)R_{ij}^{*} \text{ for } i=1,2,3\ldots \tag{5.2}$$

where $sign(x)$ = -1, 0, 1 if $x<0$, $=0$, >0. The charting statistic, SRL_{i}, is a direct analog of the plotting statistic SR_{i} used in case K. If the charting statistic SRL_{i} falls between the two control limits, that is, $LCL<SRL_{i}<UCL$, the process is considered to be in-control. If the charting statistic SRL_{i} falls on or outside one of the control limits, that is $SRL_{i}\leq LCL$ or $SRL_{i}\geq UCL$, the process is considered to be out-of-control.

Example 5.1

A Shewhart-type signed-rank-like statistic for the Montgomery (2001) piston ring data

We illustrate the Shewhart-type signed-rank-like chart using a set of data from Montgomery (2001; Tables 5.1 and 5.2) on the inside diameters of piston rings manufactured by a forging process. Table 5.1 of Montgomery (2001) contains the reference sample of size $m=125$ (see example 4.1 for an explanation of why m is equal to 125 (and not 25 like some of the earlier examples) and the median of this reference sample equals 74.001, i.e. $M=74.001$.

Panel *a* of Table 5.1 exhibits the individual observations of 15 independent samples, each of size 5 i.e. $n=5$. The absolute deviations $\left|y_{ij}-M\right|$ and $sign\left(y_{ij}-M\right)$ are shown in panel *b* and panel *c* of Table 5.1, respectively. The rank R_{ij}^{*} and the $sign\left(y_{ij}-M\right)R_{ij}^{*}$ values are shown in panel *a* and panel *b* of Table 5.2, respectively. Panel *c* of Table 5.2 holds the SRL-values i.e. SRL_{i} for $i=1,2,3,\ldots,15$.

Table 5.1. Data and calculations for the signed-rank-like chart.*

Sample number	Panel *a*					Panel *b*					Panel *c*				
	y_{i1}	y_{i2}	y_{i3}	y_{i4}	y_{i5}	$\lvert y_{i1}-M\rvert$	$\lvert y_{i2}-M\rvert$	$\lvert y_{i3}-M\rvert$	$\lvert y_{i4}-M\rvert$	$\lvert y_{i5}-M\rvert$	$sign(y_{i1}-M)$	$sign(y_{i2}-M)$	$sign(y_{i3}-M)$	$sign(y_{i4}-M)$	$sign(y_{i5}-M)$
1	74.012	74.015	74.030	73.986	74.000	0.011	0.014	0.029	0.015	0.001	1	1	1	-1	-1
2	73.995	74.010	73.990	74.015	74.001	0.006	0.009	0.011	0.014	0.000	-1	1	-1	1	0
3	73.987	73.999	73.985	74.000	73.990	0.014	0.002	0.016	0.001	0.011	-1	-1	-1	-1	-1
4	74.008	74.010	74.003	73.991	74.006	0.007	0.009	0.002	0.01	0.005	1	1	1	-1	1
5	74.003	74.000	74.001	73.986	73.997	0.002	0.001	0.000	0.015	0.004	1	-1	0	-1	-1
6	73.994	74.003	74.015	74.020	74.004	0.007	0.002	0.014	0.019	0.003	-1	1	1	1	1
7	74.008	74.002	74.018	73.995	74.005	0.007	0.001	0.017	0.006	0.004	1	1	1	-1	1
8	74.001	74.004	73.990	73.996	73.998	0.000	0.003	0.011	0.005	0.003	0	1	-1	-1	-1
9	74.015	74.000	74.016	74.025	74.000	0.014	0.001	0.015	0.024	0.001	1	-1	1	1	-1
10	74.030	74.005	74.000	74.016	74.012	0.029	0.004	0.001	0.015	0.011	1	1	-1	1	1
11	74.001	73.990	73.995	74.010	74.024	0.000	0.011	0.006	0.009	0.023	0	-1	-1	1	1
12	74.015	74.020	74.024	74.005	74.019	0.014	0.019	0.023	0.004	0.018	1	1	1	1	1
13	74.035	74.010	74.012	74.015	74.026	0.034	0.009	0.011	0.014	0.025	1	1	1	1	1
14	74.017	74.013	74.036	74.025	74.026	0.016	0.012	0.035	0.024	0.025	1	1	1	1	1
15	74.010	74.005	74.029	74.000	74.020	0.009	0.004	0.028	0.001	0.019	1	1	1	-1	1

* See SAS Program 10 in Appendix B for the calculation of the values in Table 5.1.

Table 5.2. Calculations for the signed-rank-like chart*.

Sample number	**Panel *a***					**Panel *b***					**Panel *c***
	R^*_{i1}	R^*_{i2}	R^*_{i3}	R^*_{i4}	R^*_{i5}	$sign(y_{i1}-M)R^*_{i1}$	$sign(y_{i2}-M)R^*_{i2}$	$sign(y_{i3}-M)R^*_{i3}$	$sign(y_{i4}-M)R^*_{i4}$	$sign(y_{i5}-M)R^*_{i5}$	SRL_i
1	2	3	5	4	1	2	3	5	-4	-1	5
2	2	3	4	5	1	-2	3	-4	5	0	2
3	4	2	5	1	3	-4	-2	-5	-1	-3	-15
4	3	4	1	5	2	3	4	1	-5	2	5
5	3	2	1	5	4	3	-2	0	-5	-4	-8
6	3	1	4	5	2	-3	1	4	5	2	9
7	4	1	5	3	2	4	1	5	-3	2	9
8	1	2.5	5	4	2.5	0	2.5	-5	-4	-2.5	-9
9	3	1.5	4	5	1.5	3	-1.5	4	5	-1.5	9
10	5	2	1	4	3	5	2	-1	4	3	13
11	1	4	2	3	5	0	-4	-2	3	5	2
12	2	4	5	1	3	2	4	5	1	3	15
13	5	1	2	3	4	5	1	2	3	4	15
14	2	1	5	3	4	2	1	5	3	4	15
15	3	2	5	1	4	3	2	5	-1	4	13

The control limits are chosen to give a certain false alarm rate or in-control ARL. A symmetric two-sided chart is obtained by choosing $LCL = -UCL$. For $n = 5$, the control limits for the signed-rank-like chart are set at ± 15. These control limits yield an in-control ARL of 16 and a FAR of 0.0626 (these values were obtained by the use of a simulation study (see SAS Program 6 in Appendix B) where $m = 500$ and $n = 5$). With such a small in-control average run length, many false alarms will be signalled by this chart leading to possible loss of time and resources. The chart is shown in Figure 5.1 with control limits at ± 15.

* See SAS Program 10 Appendix B for the calculation of the values in Table 5.2.

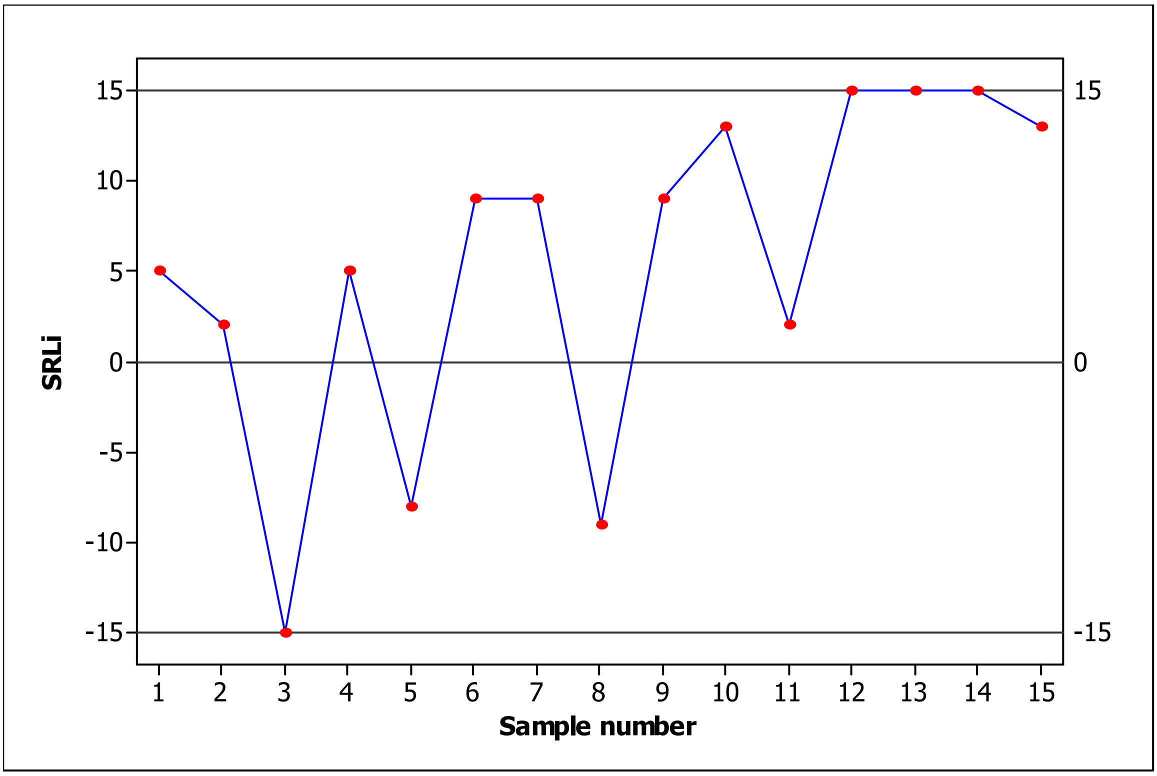

Figure 5.1. Shewhart-type signed-rank-like control chart for Montgomery (2001) piton ring data.

Observations 3, 12, 13 and 14 lie on the upper control limit which indicates that the process is out-of-control starting at sample 3. It appears most likely that the process median has shifted upwards from the target value of 74*mm*. Corrective action and a search for assignable causes is necessary.

5.1.3. Distribution-free properties

We want to establish that the charting statistic SRL_i is distribution-free. If the latter is true, then the signed-rank-like chart based on the SRL_i statistic will be distribution-free. To establish that SRL_i is distribution-free, we first have to look at some properties. Randles and Wolfe (1979) provided various definitions and theorems that are useful in this text.

Definition 1

(See Definition 1.3.1. of Randles and Wolfe (1979), pg. 13)

Two random variables S and T are said to be *equal in distribution* if they have the same cdf. To denote 'equal in distribution' we use the notation $S \overset{d}{=} T$.

Definition 2

(See Definition 1.3.6 of Randles and Wolfe (1979), pg. 15)

A collection of random variables $X_1, X_2, ..., X_n$ is said to be *exchangeable* if for every permutation $(\alpha_1, \alpha_2, ..., \alpha_n)$ of the integers $(1,2,...,n)$, $(X_1, X_2, ..., X_n) \overset{d}{=} (X_{\alpha_1}, X_{\alpha_2}, ..., X_{\alpha_n})$.

Theorem 1

(See Theorem 1.3.7 of Randles and Wolfe (1979), pg. 16)

If $\underline{X} \overset{d}{=} \underline{Y}$ and $U(\cdot)$ is a (measurable) function (possibly vector valued) defined on the common support of these random variables, then $U(\underline{X}) \overset{d}{=} U(\underline{Y})$.

Theorem 2

(See Theorem 11.2.3 of Randles and Wolfe (1979), pg. 356)

Let $\underline{X}_i = (X_{i1}, X_{i2}, ..., X_{ip})$, $i = 1,2,...,n$ be a random sample from some p-variate continuous distribution. Let $g(\cdot)$ be any function of n p-vectors that is symmetric in its arguments. Let $h(\cdot,\cdot)$ be any real-valued function of a p-tuple and the function values of $g(\cdot)$ and define the random variables $W_i = h(\underline{X}_i, g(\underline{X}_1, ..., \underline{X}_n))$, $i = 1,2,...,n$. Then $W_1, W_2, ..., W_n$ are exchangeable random variables, i.e. $(W_1, W_2, ..., W_n) \overset{d}{=} (W_{\alpha_1}, W_{\alpha_2}, ..., W_{\alpha_n})$ where $(\alpha_1, \alpha_2, ..., \alpha_n)$ is any permutation of $(1,2,...,n)$.

Theorem 2 can be generalized to complement our problem. Suppose $\underline{X}_1 = \underline{X} = (X_1, X_2, ..., X_m) \sim F$ and $\underline{X}_2 = \underline{Y} = (Y_1, Y_2, ..., Y_n) \sim G$ are independent random samples and F and G are continuous distributions. Let $g(\cdot)$ be a function of $\underline{X}$ that is symmetric in its arguments and let $h(\cdot,\cdot)$ be any real-valued function of $\underline{Y}$ and the function values of $g(\cdot)$. Then define $W_j = h(Y_j, g(\underline{X})) = h(Y_j, g(X_1, ..., X_m))$ for $j = 1,...,n$. Then, from Theorem 2, we have that $W_1, W_2, ..., W_n$ are exchangeable random variables when $F = G$.

Corollary 1

(See Corollary 2.4.5 of Randles and Wolfe (1979), pg. 50)

Let $S(\underline{\Psi}, \underline{R}^*)$ be a statistic that depends on the observations $X_1, X_2, ..., X_n$ only through $\Psi_1, \Psi_2, ..., \Psi_n$ and $\underline{R}^*$. Then the statistic $S(\cdot)$ is distribution-free over Θ, the collection of joint distributions of n iid continuous random variables, each symmetrically distributed about zero.

Corollary 2

(See Corollary 11.2.5 of Randles and Wolfe (1979), pg. 357)

Let $W_1, W_2, ..., W_n$ be defined as in Theorem 2 and let R_i^* denote the rank of W_i among $W_1, W_2, ..., W_n$. If $P(W_i = W_j) = 0$ for every $i \neq j$, then $P(\underline{R}^* = \underline{r}) = \dfrac{1}{n!}$ for every $\underline{r}$, a permutation of the integers $(1, ..., n)$. Thus any statistic that is a function of the sample observations $\underline{X}_1, ..., \underline{X}_n$ only through the ranks $R_i^*, ..., R_n^*$ is nonparametric distribution-free over the class of all p-variate continuous distribution.

Lemma 1

(See Lemma 2.4.2 of Randles and Wolfe (1979), pg. 49)

Let Z be a continuous random variable with a distribution that is symmetric about 0. Then the random variables $|Z|$ and $\Psi = \Psi(Z)$ are stochastically independent.

Establishing that the charting statistic SRL_i is distribution-free for an in-control process

The first step in establishing that the charting statistic SRL_i is distribution-free, is by proving that when the process is in-control, i.e. $F = G$, $V_1, V_2, ..., V_n$ are exchangeable random variables, where V_j is defined as $V_j = |Y_j - M|$, $j = 1,2,...,n$, and M is the median of $X_1, ..., X_m$. The proof to this follows from Theorem 2 by setting $g(\underline{X}) = g(X_1, ..., X_m) = M$ and $W_j = h(Y_j, g(X_1, ..., X_m)) = |Y_j - M|$.

The second step in establishing that the charting statistic SRL_i is distribution-free, is by proving that when $F = G$, $U_1, U_2, ..., U_n$ are exchangeable random variables, where U_j is defined as $U_j = sign(Y_j - M)$, $j = 1,2,...,n$. The proof to this follows from Theorem 2 by setting $g(\underline{X}) = g(X_1,..., X_m) = M$ and $W_j = h(Y_j, g(X_1,..., X_m)) = sign(Y_j - M)$.

The next step is to prove that when $F = G$, the joint distribution of $U_1, U_2, ..., U_n$ is distribution-free. To prove this we need to keep two things in mind. The first being that $P(U_j = 1 \text{ or } 0) = \frac{1}{2}$ since $(Y_j - M)$ is symmetric about zero when $F = G$. The second fact to recall is that $U_1, U_2, ..., U_n$ are exchangeable when $F = G$. The proof follows straightforwardly by combining these two facts.

In addition, when $F = G$, $U_j = sign(Y_j - M)$ and $V_j = |Y_j - M|$ for $j = 1,2,...,n$, are independent random variables. The proof follows from Lemma 1, since the distribution of $(Y_j - M)$ is symmetric about zero when $F = G$.

Next, we define $\underline{R}^* = (R_1^*, R_2^*, ..., R_n^*)$ where $R_j^* = 1 + \sum_{k=1}^{n} I(V_k < V_j)$ $= 1 + \sum_{k=1}^{n} I(|Y_k - M| < |Y_j - M|)$ for $j = 1,2,...,n$ (note that R_j^* is directly comparable to R_{ij}^* in equation (5.1)). Therefore, $\underline{R}^*$ is the vector of ranks of $V_1, V_2, ..., V_n$, i.e. $\underline{R}^*$ is the vector of ranks of $|Y_1 - M|, |Y_2 - M|, ..., |Y_n - M|$. We can prove, using Corollaries 1 and 2, that when $F = G$, any statistic that depends on the observations only through $U_1, U_2, ..., U_n$, i.e. $sign(Y_1 - M), sign(Y_2 - M), ..., sign(Y_n - M)$, and $\underline{R}^*$ is distribution-free over the class of continuous symmetric distributions. Consequently, the statistic $SRL_i = \sum_{j=1}^{n} sign(y_{ij} - M) R_{ij}^*$ is distribution-free. Since SRL_i is now known to be distribution-free, so is the signed-rank-like chart.

5.1.4. Simulation study

Bakir (2006) performed a simulation study where the robustness of the standard Shewhart $\overline{X}$ chart and that of the proposed Shewhart signed-rank-like chart are compared using the contaminated normal distribution. The contaminated normal distribution has been considered by various authors in an SPC context (see, for example, Wu, Zhao and Wang (2002) and Sheu and Yang (2006)). The cdf of the contaminated normal distribution is given by

$$\Phi_p(\theta,\sigma^2) = (1-p)\Phi(\theta,1) + p\Phi(\theta,\sigma^2) \tag{5.3}$$

where $(0 \le)$ p (≤ 1) denotes the percentage of contamination, σ^2 (>0) denotes the severity of contamination and Φ denotes the cdf of the normal distribution, respectively. It should be noted that if $p = 0$ and $\theta = 0$ equation (5.3) reduces to the standard normal distribution. Bakir (2006) proved, through simulation, that if a process is contaminated by outliers it is ill-advised to use the standard Shewhart $\overline{X}$ chart, especially if the percentage of contamination (p) and/or the severity of contamination (σ^2) is high, i.e. $p > 0.01$ and/or $\sigma^2 > 4$. Bakir concludes that the Shewhart $\overline{X}$ chart is not robust against outliers, whereas the proposed Shewhart signed-rank-like chart is robust against outliers for all possible combinations of (p,σ^2). This is what we expected to find: the Shewhart signed-rank-like chart wouldn't be affected by outliers, since the median from the reference sample, the signs from the test sample and the ranks from the test sample aren't affected by outliers (recall that $SRL_i = \sum_{j=1}^{n} sign(y_{ij} - M)R_{ij}^*$).

Table 5.3 shows the simulated values of the ARL_0's of the two-sided Shewhart $\overline{X}$ chart for all possible combinations of (p,σ^2) with p =0.01, 0.05, 0.10, 0.15, 0.20, 1 and σ^2 = 4, 9, 16. These values are graphically illustrated in Figure 5.2. These simulated values are for a stable process with the presense of sporadic outliers. The case where the process is operational with no outliers, i.e. $p = 0$, is also given for reference. In these simulation studies 500 reference samples, each of size $m = 39$, were generated from the standard normal distribution. In addition, 500 test samples, each of size $n = 10$, were generated from the contaminated normal distribution.

Table 5.3. Simulated values of the ARL_0's for the two-sided Shewhart $\overline{X}$ chart[*].

$LCL/UCL = \pm 2.8$			Severity of contamination: $\sigma^2 = 4$	$\sigma^2 = 9$	$\sigma^2 = 16$
			Level of severity		
Percentage of contamination		**Level of p**	**Low**	**Moderately high**	**High**
	$p = 0$ (0%)	**None**	163	163	163
	$p = 0.01$ (1%)	**Low**	159	115	70
	$p = 0.05$ (5%)	**Moderately high**	90	39	21
	$p = 0.10$ (10%)	**Moderately high**	61	22	12
	$p = 0.15$ (15%)	**High**	41	15	8
	$p = 0.20$ (20%)	**High**	33	11	6

Intuitively, we would expect the ARL to decrease (which would lead to an increase in the number of false alarms) as the percentage and/or severity of contamination increases. This is evident by looking at the lowest (p,σ^2) combination, i.e. $(p,\sigma^2) =$ (0, 4), opposed to the highest (p,σ^2) combination, i.e. $(p,\sigma^2) =$ (0.20, 16). The former shows that the ARL equals 163 when the process is operational with no outliers, whereas the latter shows that the ARL equals 6 when both the percentage and severity of contamination are high. These numbers indicate that there should be about 27 times as many false alarms when $(p,\sigma^2) =$ (0.20, 16) as opposed to $(p,\sigma^2) = (0,4)$.

Next, we look at what happens when both p and σ^2 are low. This is done by looking at the $(p,\sigma^2) =$ (0.01, 4) combination compared to the $(p,\sigma^2) =$ (0, 4) combination. The latter shows that the ARL equals 163 when the process is operational with no outliers, whereas the former shows that the ARL equals 159 when both the percentage and severity of contamination are low. These numbers indicate that there should be about the same number of false alarms when $(p,\sigma^2) =$ (0.01, 4) as opposed to $(p,\sigma^2) = (0,4)$.

[*] Table 5.3 appears in Bakir (2006), page 751, Table 1.

Next, we look at what happens when p is low, but σ^2 is moderately high. This is done by looking at the $(p,\sigma^2) =$ (0.01, 9) combination where the *ARL* has dropped to 115. This indicates that there should be about 1.42 times as many false alarms as the expected *ARL* of 163. Subsequently, we look at what happens when p is moderately high, but σ^2 is low. This is done by looking at the $(p,\sigma^2) =$ (0.05, 4) combination where the *ARL* has dropped to 90. This indicates that there should be about 1.81 as many false alarms as the expected *ARL* of 163. The rest of table can be interpreted similarly. The main conclusion that can be drawn from Table 5.3 is that it is ill-advised to use the Shewhart $\overline{X}$ chart when a process is contaminated by outliers, especially if the percentage of contamination (p) and/or the severity of contamination (σ^2) is high, i.e. $p > 0.01$ and/or $\sigma^2 > 4$.

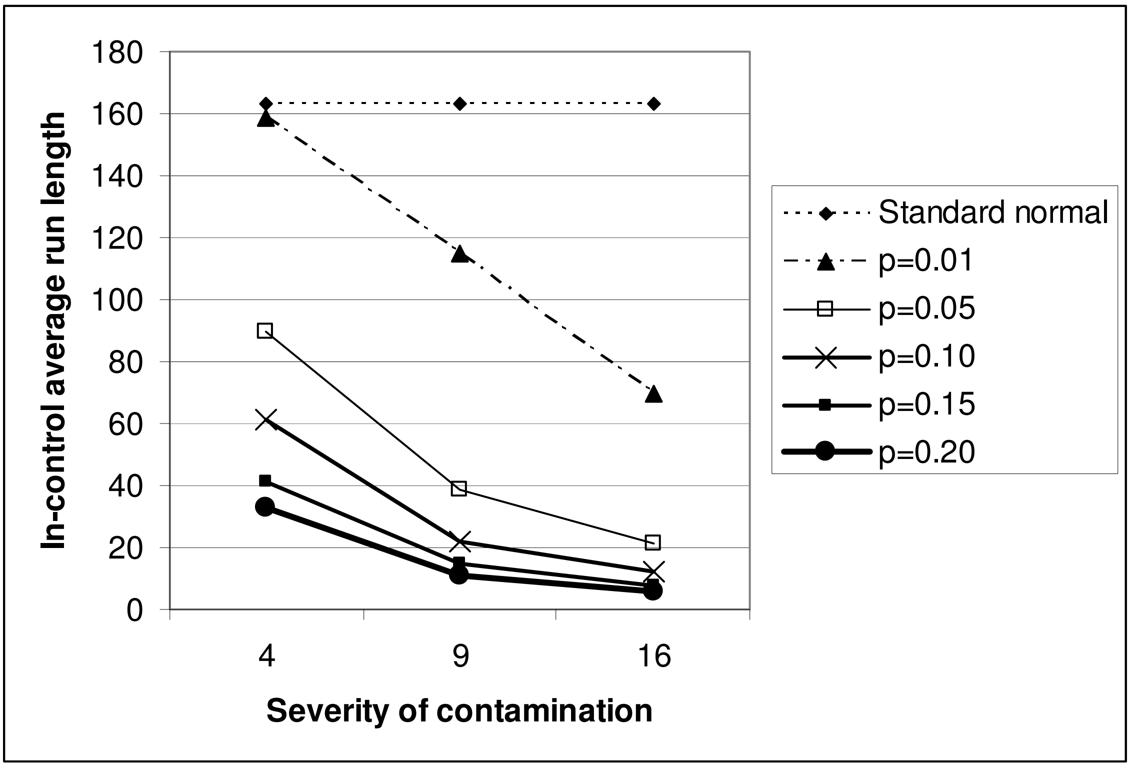

Figure 5.2. Simulated ARL_0 values for the two-sided Shewhart $\overline{X}$ chart for various values of p and σ^2.

5.1.5. Comparisons

The first comparison between the standard Shewhart $\overline{X}$ chart and the proposed Shewhart signed-rank-like chart.

Table 5.4. Simulated values of the ARL for the two-sided Shewhart $\overline{X}$ control chart ($\overline{X}_{CC}$) and the Shewhart signed-rank-like control chart (SRL_{CC})*.

	Severity of contamination											
	$\sigma^2 = 4$				$\sigma^2 = 9$				$\sigma^2 = 16$			
	Level of severity											
	Low				**Moderately high**				**High**			
	$p = 0.01$ (1%)		$p = 0.10$ (10%)		$p = 0.01$ (1%)		$p = 0.10$ (10%)		$p = 0.01$ (1%)		$p = 0.10$ (10%)	
Control chart	$\overline{X}_{CC}$	SRL_{CC}	$\overline{X}_{CC}$	SRL_{CC}	$\overline{X}_{CC}$	SRL_{CC}	$\overline{X}_{CC}$	SRL_{CC}	$\overline{X}_{CC}$	SRL_{CC}	$\overline{X}_{CC}$	SRL_{CC}
LCL/UCL	± 2.85	± 53	± 3.25	± 53	± 2.96	± 53	± 3.98	± 53	± 3.24	± 53	± 4.92	± 53
$\theta = 0.0$	170.0	166.0	165.0	166.0	167.4	166.0	166.7	166.0	164.6	166.0	167.0	166.0
$\theta = 0.2$	122.8	121.8	131.8	128.3	115.4	120.4	124.3	122.2	137.3	130.2	145.6	128.6
$\theta = 0.4$	43.9	50.9	59.8	54.4	56.4	60.7	65.4	66.2	60.6	61.7	88.3	65.0
$\theta = 0.6$	10.6	24.2	13.5	27.2	14.4	23.7	27.5	26.5	17.7	22.1	42.8	31.5
$\theta = 0.8$	3.5	11.1	5.4	10.6	4.2	9.3	11.9	10.3	6.7	8.4	22.6	11.3
$\theta = 1.0$	1.9	4.4	2.8	5.5	2.1	4.7	5.3	5.6	3.5	4.5	13.4	5.9

* Table 5.4 appears in Bakir (2006), page 754, Table 3.

Bakir (2006) compared the proposed Shewhart signed-rank-like chart to the Shewhart $\overline{X}$ chart using the contaminated normal distribution (the observations are normally distributed with occasional outliers). Both charts are designed to have approximately the same in-control average run length to ensure fair comparison between the charts. The out-of-control average run lengths were computed, using these chart constants, for various values of the median θ, the percentage of contamination (*p*) and the severity of contamination (σ^2). We typically want the ARL_δ to be small, i.e. the chart with the smallest ARL_δ will be the preferred chart.

From Table 5.4 we see that the median ranges from 0 (the in-control value) to 1 in increments of 0.2; the severity of contamination ranges from low to high, that is, $\sigma^2 = 4$ (low), $\sigma^2 = 9$ (moderately high) and $\sigma^2 = 16$ (high); and the percentage of contamination is taken to be 1% (low) and 10% (moderately high), respectively.

We start by investigating the lowest percentage and severity of contamination levels for the smallest process shift of 0.2. The ARL_δ of the Shewhart $\overline{X}$ chart (=122.8) is almost equivalent to the ARL_δ of the Shewhart signed-rank-like chart (=121.8). Therefore, for a low percentage and severity of contamination and a small process shift, both charts are performing equally well. More generally, for low to moderately high levels of p (= 0.01 or 0.1) and σ^2 (= 4 or 9) and small process shifts ($\theta =$0.2 or 0.4), the ARL_δ values of the Shewhart signed-rank-like chart are almost equivalent to the ARL_δ values of the Shewhart $\overline{X}$ chart.

In contrast, we investigate the highest percentage and severity of contamination for the largest process shift of 1. The ARL_δ of the Shewhart $\overline{X}$ chart (=13.4) is higher than the ARL_δ of the Shewhart signed-rank-like chart (=5.9). Consequently, we see that the Shewhart signed-rank-like chart performs better than the Shewhart $\overline{X}$ chart for a high percentage and severity of contamination and a large process shift. More generally, we find that for $p = 0.1$ and $\sigma^2 = 16$ the ARL_δ values of the Shewhart $\overline{X}$ chart are *all* higher than the ARL_δ values of the Shewhart signed-rank-like chart for *all* process shifts ($\theta =$0.2, 0.4, 0.6, 0.8 and 1). As a result we conclude that the Shewhart signed-rank-like chart performs better than the Shewhart $\overline{X}$ chart for high levels of p and σ^2 over all process shifts.

It should be noted that there are various cases where the Shewhart $\overline{X}$ chart performs better than the Shewhart signed-rank-like chart. An illustration of the latter is, for example, for low to moderately high levels of p (= 0.01 or 0.1), $\sigma^2 = 4$ and large process shifts ($\theta = 0.6$, 0.8 and 1) the ARL_δ values of the Shewhart $\overline{X}$ chart are lower than those of the Shewhart signed-rank-like chart. Hence, in some cases the Shewhart $\overline{X}$ chart outperforms the Shewhart signed-rank-like chart and vice versa. Table 5.5 indicates which chart, between the Shewhart $\overline{X}$ chart and the proposed signed-rank-like chart, is the preferred chart for various values of p, σ^2 and θ. The term 'comparable' in Table 5.5 implies that the proposed signed-rank-like chart is as efficient as the Shewhart $\overline{X}$ chart.

Table 5.5. Summary of the first comparison between the Shewhart $\overline{X}$ chart and the proposed signed-rank-like chart*.

	$\sigma^2 = 4$	$\sigma^2 = 9$	$\sigma^2 = 16$
$p = 0.01$	Small shifts: Comparable Large shifts: $\overline{X}$	Small shifts: Comparable Large shifts: $\overline{X}$	Small shifts: Comparable Large shifts: $\overline{X}$
$p = 0.10$	Small shifts: Comparable Large shifts: $\overline{X}$	All shifts: Comparable	All shifts: Signed-rank-like chart

The second comparison between the standard Shewhart $\overline{X}$ chart and the proposed Shewhart signed-rank-like chart.

The out-of-control ARL is examined for three distributions, namely, the Normal, Laplace and Cauchy distributions, respectively. Recall that we want the ARL_δ to be small in all cases.

* Small shifts refer to $\theta = 0.2$ or 0.4, whereas large shifts refer to $\theta = 0.6, 0.8$ or 1.

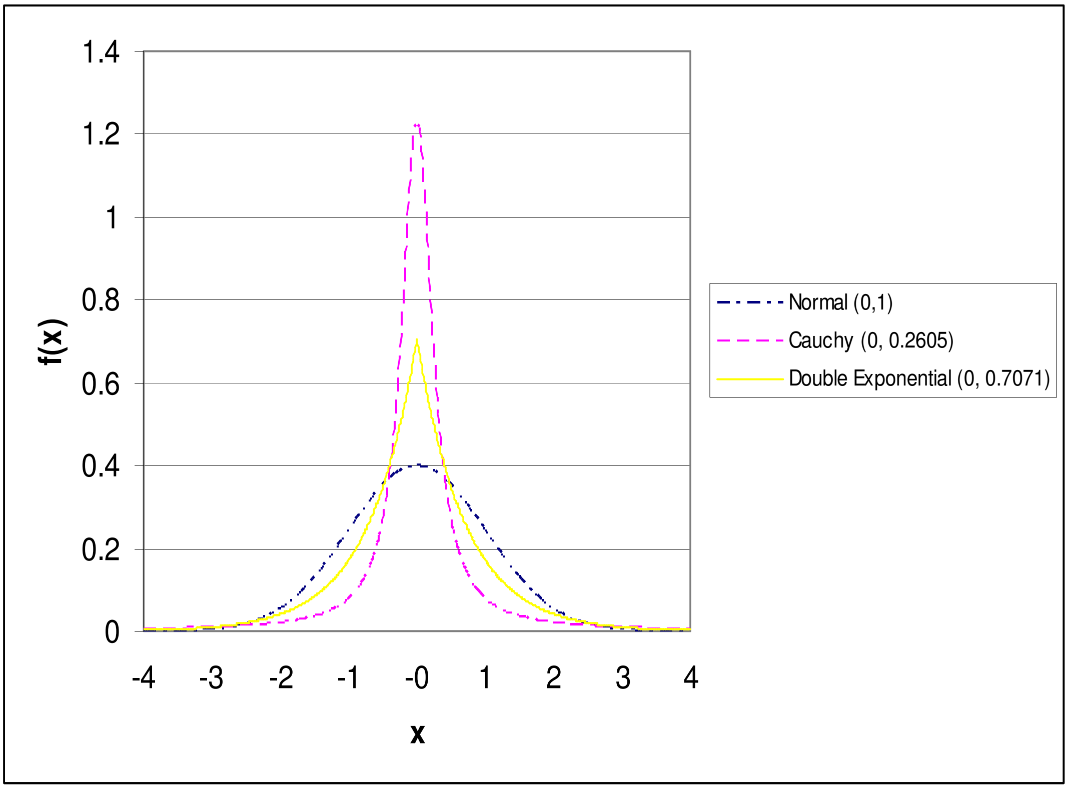

Figure 5.3. The shapes of the three distributions under consideration.

(i) The Normal distribution

For the Normal distribution we would expect the out-of-control performance of the Shewhart $\overline{X}$ chart to be better than that of the Shewhart signed-rank-like chart. The chart constants for both the Shewhart signed-rank-like and Shewhart $\overline{X}$ charts are chosen such that the in-control average run length is approximately equal ($ARL_0 \approx 164$) for both charts: $LCL/UCL_{\overline{X}} = \pm 2.80$ and $LCL/UCL_{SRL} = \pm 53$. The out-of-control average run length values were computed, using these chart constants, for various values of the median θ. The median ranges from 0 (the in-control value) to 1 in increments of 0.2. The results are shown below in Figure 5.4.

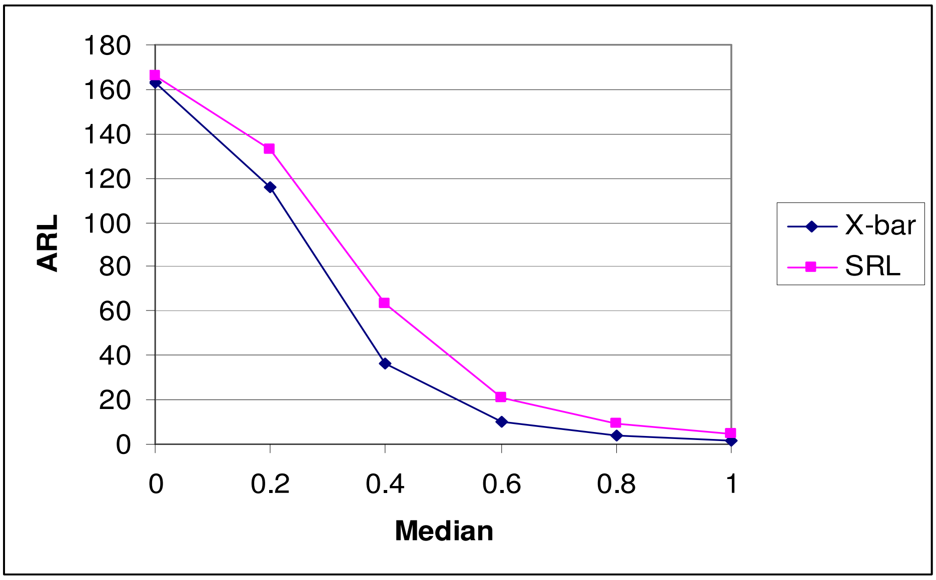

Figure 5.4. Comparison of the Shewhart signed-rank-like chart with the Shewhart $\overline{X}$ chart under Normal shift alternatives.

When comparing the Shewhart signed-rank-like chart with the Shewhart $\overline{X}$ chart under Normal shift alternatives we find that the Shewhart $\overline{X}$ chart is performing better than the Shewhart signed-rank-like chart, since the out-of-control average run length values for the Shewhart $\overline{X}$ chart are smaller than the out-of-control average run length values for the Shewhart signed-rank-like chart. However, it should be noted that the differences are small and it appears to fade away when the process is shifted from its in-control value of 0 to values greater than 0.8.

(ii) The Double Exponential distribution

The Double Exponential distribution, also called the Laplace distribution, is comparable to the Normal distribution (since they are both symmetric around 0), but it has heavier tails (see Figure 5.3). As a result, there are higher probabilities associated with extreme values when working with the Double Exponential distribution as opposed to using the Normal distribution. The scale parameter λ of the Double Exponential distribution is set equal to $1/\sqrt{2}$ so that the Double Exponential distribution has a standard deviation of 1. For the Double Exponential distribution we would expect the out-of-control performance of the Shewhart signed-rank-like chart to be better than that of the Shewhart $\overline{X}$ chart. The chart constants for both the Shewhart signed-rank-like and Shewhart $\overline{X}$ chart are chosen such that the in-control average run length is approximately equal ($ARL_0 \approx 150$) for both charts: $LCL/UCL_{\overline{X}} = \pm 2.85$ and $LCL/UCL_{SRL} = \pm 53$. The results are shown below in Figure 5.5.

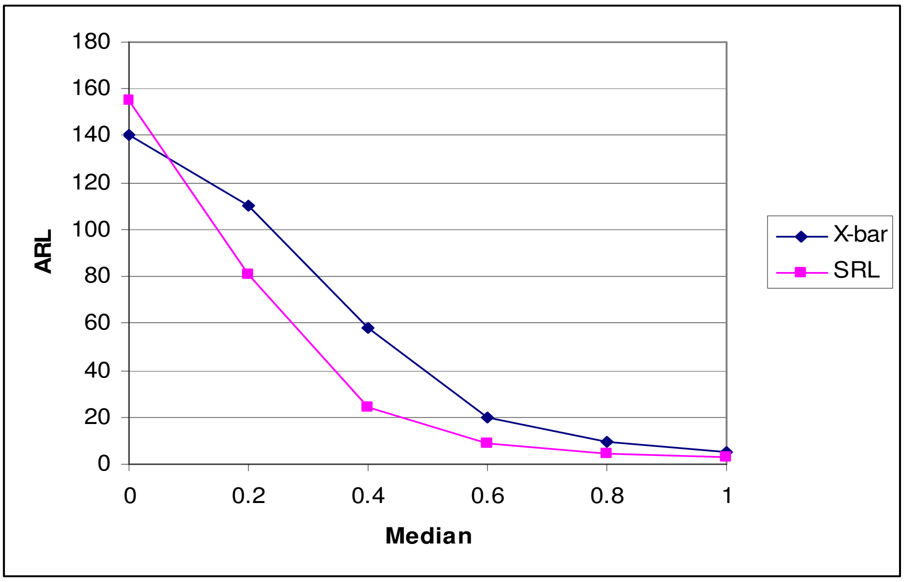

Figure 5.5. Comparison of the Shewhart signed-rank-like chart with the Shewhart $\overline{X}$ chart under Double Exponential shift alternatives.

When comparing the Shewhart signed-rank-like chart with the Shewhart $\overline{X}$ chart under Double Exponential shift alternatives we find that the Shewhart signed-rank-like chart is performing better than the Shewhart $\overline{X}$ chart, since the out-of-control average run length values for the Shewhart signed-rank-like chart are smaller than the out-of-control average run length values for the Shewhart $\overline{X}$ chart. However, it should be noted that the differences are small and it appears to fade away when the process is shifted from its in-control value of 0 to values greater than 0.8.

(iii) The Cauchy distribution

The scale parameter λ of the Cauchy distribution is set equal to 0.2605 so that the Cauchy distribution has a probability of 0.95 to the left of 1.645 (which is also the case for the standard normal distribution). For the Cauchy distribution we would expect the out-of-control performance of the Shewhart signed-rank-like chart to be better than that of the Shewhart $\overline{X}$ chart. The chart constants for both the Shewhart signed-rank-like and Shewhart $\overline{X}$ chart are chosen such that the in-control average run length is approximately equal ($ARL_0 \approx 164$) for both charts: $LCL/UCL_{\overline{X}} = \pm 22$ and $LCL/UCL_{SRL} = \pm 53$. The results are shown below in Figure 5.6.

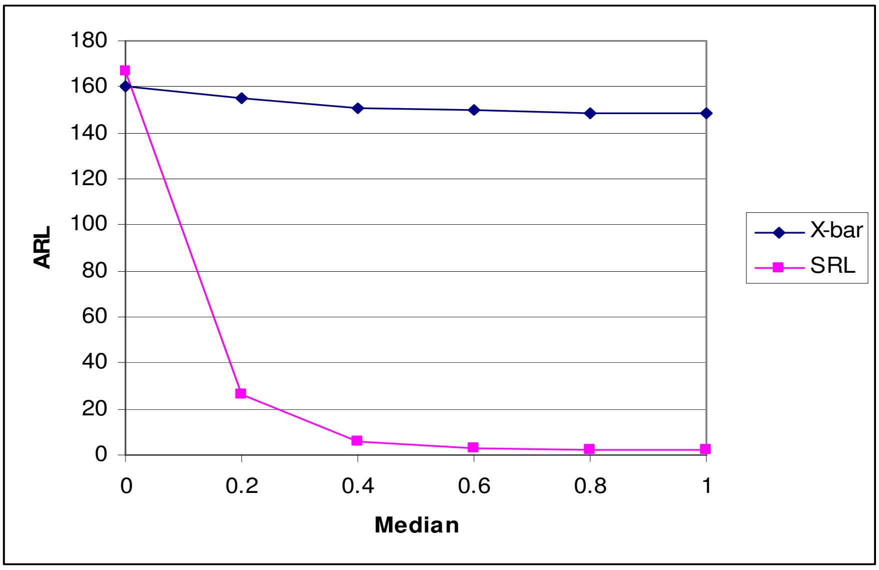

Figure 5.6. Comparison of the Shewhart signed-rank-like chart with the Shewhart $\overline{X}$ chart under Cauchy shift alternatives.

When comparing the Shewhart signed-rank-like chart with the Shewhart $\overline{X}$ chart under Cauchy shift alternatives we find that the Shewhart signed-rank-like chart is performing better than the Shewhart $\overline{X}$ chart, since the out-of-control average run length values for the Shewhart signed-rank-like chart are smaller than the out-of-control average run length values for the Shewhart $\overline{X}$ chart. It should be noted that these differences are large for all values of the median θ.

In conclusion we found that the Shewhart signed-rank-like chart performs better than the Shewhart $\overline{X}$ chart under heavy tailed distributions. In addition, recall that the Shewhart $\overline{X}$ chart is not robust against outliers, whereas the proposed Shewhart signed-rank-like chart is, for the most part, robust against outliers. These are two key motivations to why the user should rather use the Shewhart signed-rank-like chart as opposed to using the Shewhart $\overline{X}$ chart.

Table 5.6. Summary of the second comparison between the Shewhart $\overline{X}$ chart and the proposed signed-rank-like chart.

Distribution	Preferred control chart
Normal	Shewhart $\overline{X}$ chart
Double Exponential	Shewhart signed-rank-like chart
Cauchy	Shewhart signed-rank-like chart

5.1.6. The tabular CUSUM control chart

Bakir (2006) proposed a tabular CUSUM signed-rank-like chart. Generally, the standardized upper one-sided CUSUM is given by

$$S_i^+ = \max[0, S_{i-1}^+ + y_i - k] \quad \text{for } i = 1,2,3,... \tag{5.4}$$

while the resulting standardized lower one-sided CUSUM is given by

$$S_i^- = \min[0, S_{i-1}^- + y_i + k] \quad \text{for } i = 1,2,3,... \tag{5.5}$$

or

$$S_i^{-*} = \max[0, S_{i-1}^{-*} - y_i - k] \quad \text{for } i = 1,2,3,... \tag{5.6}$$

The two-sided standardized CUSUM is constructed by running the upper and lower one-sided standardized CUSUM charts simultaneously and signals at the first i such that $S_i^+ \geq h$ or $S_i^- \leq -h$.

The chart proposed by Bakir (2006) instead uses the cumulative sum of the statistic SRL_i (defined in (5.2)) with a stopping rule. A CUSUM signed-rank-like chart can be obtained by replacing y_i in expressions (5.4), (5.5) and (5.6) with SRL_i. In other words, for the upper one-sided CUSUM signed-rank-like chart we use

$$S_i^+ = \max[0, S_{i-1}^+ + SRL_i - k] \quad \text{for } i = 1,2,3,... \tag{5.7}$$

to detect positive deviations from zero. A signalling event occurs for the first i such that $S_i^+ \geq h$.

For a lower one-sided CUSUM signed-rank-like chart we use

$$S_i^- = \min[0, S_{i-1}^- + SRL_i + k] \quad \text{for } i = 1,2,3,... \tag{5.8}$$

or

$$S_i^{-*} = \max[0, S_{i-1}^{-*} - SRL_i - k] \quad \text{for } i = 1,2,3,... \tag{5.9}$$

to detect negative deviations from zero. A signalling event occurs for the first i such that $S_i^- \leq -h$ (if expression (5.8) is used) or $S_i^{-*} \geq h$ (if expression (5.9) is used).

The corresponding two-sided CUSUM chart signals for the first i at which either one of the two inequalities is satisfied, that is, either $S_i^+ \geq h$ or $S_i^- \leq -h$. Starting values are typically chosen to equal zero, that is, $S_0^+ = S_0^- = 0$.

A CUSUM signed-rank-like chart can also be constructed by replacing y_i in expressions (5.4), (5.5) and (5.6) with the standardized signed-rank-like statistic.

Although Bakir (2006) provided the general idea of how to construct a CUSUM signed-rank-like control chart, he failed to do any simulation studies or to give any tables that can be used for the implementation of the chart. More research is necessary on CUSUM signed-rank-like control charts, for example, one could look at the implementation of the CUSUM signed-rank-like chart and study its performance.

5.1.7. The EWMA control chart

Bakir (2006) proposed an EWMA signed-rank-like chart. Generally, an EWMA control chart scheme accumulates statistics $X_1, X_2, X_3, \ldots$ with the plotting statistics defined as

$$Z_i = \lambda X_i + (1-\lambda)Z_{i-1} \tag{5.10}$$

where $0 < \lambda \leq 1$ is a constant called the weighting constant. The starting value Z_0 is often taken to be zero.

A nonparametric EWMA-type of control chart based on the signed-rank-like statistic can be obtained by replacing X_i in expression (5.10) with SRL_i. Therefore, the EWMA signed-rank-like chart accumulates the statistics $SRL_1, SRL_2, SRL_3, \ldots$ with the plotting statistics defined as

$$Z_i = \lambda SRL_i + (1-\lambda)Z_{i-1} \tag{5.11}$$

where $0 < \lambda \leq 1$ and the starting value Z_0 could be taken to equal zero, i.e. $Z_0 = 0$.

An EWMA signed-rank-like chart can also be constructed by replacing X_i in expression (5.10) with the standardized signed-rank-like statistic.

Although Bakir (2006) provided the general idea of how to construct an EWMA signed-rank-like control chart, he failed to do any simulation studies or to give any tables that can be used for the implementation of the chart. More research is necessary on EWMA signed-rank-like control charts, for example, one could look at the implementation of the EWMA signed-rank-like chart and study its performance.

5.1.8. Summary

In this chapter we examined the Shewhart-type signed-rank-like chart proposed by Bakir (2006). We illustrated these procedures using the piston ring data from Montgomery (2001) to help the reader to understand the subject more thoroughly. The proposed chart is recommended when the process distribution is known to be heavy-tailed or to be contaminated by occasional outliers. We also briefly looked at CUSUM- and EWMA-type signed-rank-like charts. Although Bakir (2006) provided general ideas on how to construct CUSUM- and EWMA-type signed-rank-like control charts, he failed to do any simulation studies or to give any tables that can be used for the implementation of these charts. More research is necessary on CUSUM- and EWMA-type signed-rank-like control charts, for example, one could look at the implementation of these charts and study their performance.