Chapter 6: Mann-Whitney-Wilcoxon control charts

6.1. The Shewhart-type control chart

6.1.1. Introduction

While the precedence chart is a step in the right direction, the precedence test is not the most popular or the most powerful of nonparametric two-sample tests. That honour goes to the Mann-Whitney test (see for example, Gibbons and Chakraborti, 2003). The Mann-Whitney (hereafter MW) test (equivalent to the popular Wilcoxon rank-sum test) is a well-known nonparametric competitor of the two-independent-sample *t*-test. The test is known to be more powerful than the precedence test for light tailed distributions and hence MW charts are expected to be more efficient for such cases. Of course, the MW chart is also distribution-free and therefore has the same in-control robustness advantage as the precedence chart, namely that its in-control distribution is completely known. Park and Reynolds (1987) considered Shewhart-type control charts for monitoring the location parameter of a continuous process in case U. One of the special cases of their charts is the MW chart based on the Mann-Whitney-Wilcoxon (hereafter MWW) statistic. The control limits of these charts are established using Phase I reference data. However, they only considered properties of this chart when the reference sample size approaches infinity. Chakraborti and Van de Wiel (2003) considered the Shewhart-type MW chart for finite reference sample size, studied its properties, and provided tables for its implementation. These authors show that in some cases the MW chart is more efficient than the precedence chart.

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 $F(x)$. Let $Y_1^h, Y_2^h, ..., Y_{n_h}^h$, $h = 1,2,...,$ denote the h^{th} test sample of size n_h. Let $G^h(y)$ denote the cdf of the distribution of the h^{th} Phase II sample. $G^h(y) = G(y) \quad \forall h$, since the Phase II samples are all assumed to be identically distributed. Accordingly, the superscript h can be suppressed from this point forward. For convenience, assume that the Phase II samples are all of the same size, n. The Mann-Whitney test is based on the total number of (X, Y) pairs where the Y-observation (Phase II sample) is strictly greater than the X-observation (Phase I sample).

6.1.2. Plotting statistic

The Mann-Whitney statistic is defined to be

$$M_{XY} = \text{the number of pairs } (X_i, Y_j) \text{ with } Y_j > X_i \tag{6.1}$$

for $i = 1,2,...,m$ and $j = 1,2,...,n$. Expression (6.1) can be written as

$$M_{XY} = \sum_{i=1}^{m}\sum_{j=1}^{n} I(Y_j > X_i) \tag{6.2}$$

where $I(Y_j > X_i)$ is the indicator function, i.e.

$$I(Y_j > X_i) = \begin{cases} 1 & \text{if} \quad Y_j > X_i \\ 0 & \text{if} \quad Y_j \le X_i \end{cases}.$$

There are a total of mn (X_i, Y_j) pairs for each Phase II sample. Therefore, if all the Y-observations are greater than the X-observations, M_{XY} would be equal to mn. On the other hand, if all the Y-observations are smaller than the X-observations, M_{XY} would be equal to 0. Therefore, we have that $0 \le M_{XY} \le mn$. For large values of M_{XY}, that is, if a large number of the Y-observations are greater than the X-observations, this would be indicative of a positive shift from the X to the Y distribution. On the other hand, for small values of M_{XY}, that is, if a large number of the Y-observations are smaller than the X-observations, this would be indicative of a negative shift from the X to the Y distribution.

The proposed MW chart plots the M_{XY} statistics, that is, $M_{XY}^1, M_{XY}^2, \ldots$, versus the test sample number. M_{XY} is referred to as the plotting statistic. The chart signals if the plotting statistic falls on or above the upper control limit (UCL) or if the plotting statistic falls on or below the lower control limit (LCL). Since the in-control distribution of the plotting statistic, M_{XY}, is symmetric about the mean $\frac{mn}{2}$ (see Gibbons and Chakraborti (2003)), the control limits are taken to be symmetric. Because of symmetry, we have that $P(M_{XY} = a) = P(M_{XY} = mn - a)$ for the constant a with $0 \le a \le mn$, so it is reasonable to take $L_{mn} = mn - U_{mn}$ where U_{mn} and L_{mn} denote the upper and lower control limits, respectively. If the plotting statistic M_{XY} falls between the control limits, that is, $L_{mn} < M_{XY} < U_{mn}$, the process is declared to be in-control, whereas if the plotting statistic

M_{XY} falls on or outside one of the control limits, that is, if $M_{XY} \leq L_{mn}$ or $M_{XY} \geq U_{mn}$, the process is declared to be out-of-control.

6.1.3. Properties of the run-length distribution

Result 6.1: Probability of a signal - conditional

Let $p_G(\underline{x})$ denote the probability of a signal with any test (Phase II) sample, given the reference sample $(X_1, X_2, ..., X_m) = (x_1, x_2, ..., x_m)$ (in short, $\underline{X} = \underline{x}$).

$$p_G(\underline{x}) = 2P_G\left(M_{\underline{x}Y} \geq U_{mn}\right)$$

$$\begin{aligned} p_G(\underline{x}) &= P_G(\text{Signal} \mid \underline{X} = \underline{x}) \\ &= P_G\left(M_{\underline{x}Y} \leq L_{mn}\right) + P_G\left(M_{\underline{x}Y} \geq U_{mn}\right) \\ &= P_G\left(M_{\underline{x}Y} \leq mn - U_{mn}\right) + P_G\left(M_{\underline{x}Y} \geq U_{mn}\right) \\ &= 2P_G\left(M_{\underline{x}Y} \geq U_{mn}\right). \end{aligned}$$

The last equality follows on account of symmetry (see Section 6.1.2). From Result 6.1 it can be seen that the calculation of $p_G(\underline{x})$ essentially requires the calculation of the upper-tailed probability $P_G\left(M_{\underline{x}Y} \geq U_{mn}\right)$. More detail on this point appears in Section 6.1.4.

Result 6.2: Probability of no signal - conditional

$$P_G(\text{No Signal} \mid \underline{X} = \underline{x}) = 1 - p_G(\underline{x}) = 1 - 2P_G\left(M_{\underline{x}Y} \geq U_{mn}\right)$$

Result 6.3: Run-length distribution - conditional

$$P(N = k \mid \underline{X} = \underline{x}) = (1 - p_G(\underline{x}))^{k-1}(p_G(\underline{x})) \quad \text{for} \quad k = 1,2,3,...$$

The conditional run length, denoted by $N \mid \underline{X} = \underline{x}$, will have a geometric distribution with parameter $p_G(\underline{x})$, because all the Phase II samples are independent if we condition on the reference sample. A detailed motivation for using the method of conditioning is given by

Chakraborti (2000), but, in brief, the signalling events are dependent and by means of conditioning on the reference sample we don't have to be concerned about the dependence. Consequently we have that

$$N \mid \underline{X} = \underline{x} \sim GEO\left(p_G(\underline{x})\right)$$

$$P\left(N = k \mid \underline{X} = \underline{x}\right) = \left(1 - p_G(\underline{x})\right)^{k-1}\left(p_G(\underline{x})\right) \quad \text{for} \quad k = 1,2,3,...$$

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

$$P\left(N \le k \mid \underline{X} = \underline{x}\right) = \sum_{i=1}^{k}\left(1 - p_G(\underline{x})\right)^{i-1}\left(p_G(\underline{x})\right) \quad \text{for} \quad k = 1,2,3,...$$

Result 6.4: Average run-length – conditional

$$CARL = E_G\left(N \mid \underline{X} = \underline{x}\right) = \frac{1}{p_G(\underline{x})}$$

Since the conditional run length, denoted by $N \mid \underline{X} = \underline{x}$, has a geometric distribution with parameter $p_G(\underline{x})$, the conditional average run length is given by

$$CARL = E_G\left(N \mid \underline{X} = \underline{x}\right) = \frac{1}{p_G(\underline{x})}.$$

Result 6.5: Average run-length – unconditional

$$UARL = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \upsilon\left(G(x_1), G(x_2), ..., G(x_m)\right) dF(x_1) \ldots dF(x_m)$$

where

υ is some function of G and $x_1, x_2, ..., x_m$.

$$\begin{aligned} & UARL \\ &= E(N) \\ &= E_F\left(E_G(N \mid \underline{X} = \underline{x})\right) \\ &= E_F\left(\frac{1}{p_G(\underline{x})}\right) \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{1}{p_G(\underline{x})} dF(x_1) \ldots dF(x_m). \end{aligned} \tag{6.3}$$

The second equality in (6.3) follows from extending the notion of expectation to the conditional framework. The third equality in (6.3) follows from Result 6.4. The fourth equality in (6.3) follows from the definition of expected values (see, for example, Bain and Engelhardt (1992)).

From (6.3) it can be seen that the unconditional ARL is an m-dimensional integral, since the reference sample is of size m. Equation (6.3) can be expressed differently by writing $\frac{1}{p_G(\underline{x})}$ as $\upsilon\left(P(Y_j < x_1), P(Y_j < x_2),..., P(Y_j < x_m)\right) = \upsilon\left(G(x_1), G(x_2),..., G(x_m)\right)$, where υ is some function of G and $x_1, x_2,..., x_m$. By substituting $\frac{1}{p_G(\underline{x})}$ in (6.3) with $\upsilon\left(G(x_1), G(x_2),..., G(x_m)\right)$ we obtain

$$UARL = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \upsilon\left(G(x_1), G(x_2),..., G(x_m)\right) dF(x_1) \ldots dF(x_m). \tag{6.4}$$

Recall that a process is said to be in-control when $G = F$. Therefore, the in-control (unconditional) ARL is obtained by substituting $G = F$ into the equation for the unconditional ARL given in (6.3) and we obtain the m-dimensional integral

$$UARL_0 = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{1}{p_F(\underline{x})} dF(x_1) \ldots dF(x_m), \tag{6.5}$$

where the subscript 0 refers to the in-control state.

In the out-of-control case the unconditional ARL is given by the m-dimensional integral

$$UARL_\delta = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{1}{p_G(\underline{x})} dF(x_1) \ldots dF(x_m), \tag{6.6}$$

where δ signifies a shift between F and G.

Recall that $\frac{1}{p_G(\underline{x})}$ was re-written as $\upsilon\left(G(x_1), G(x_2), ..., G(x_m)\right)$ where υ is some function of G and $x_1, x_2, ..., x_m$. Similarly, $\frac{1}{p_F(\underline{x})}$ can be re-written as $\upsilon\left(F(x_1), F(x_2), ..., F(x_m)\right)$ and we obtain

$$
\begin{aligned}
&UARL_0 \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \upsilon\left(F(x_1), F(x_2), ..., F(x_m)\right) dF(x_1) \ldots dF(x_m) \\
&= \int_{0}^{1} \cdots \int_{0}^{1} \upsilon(u_1, u_2, ..., u_m) du_1 \ldots du_m \\
&= \int_{0}^{1} \cdots \int_{0}^{1} \frac{1}{p_U(\underline{u})} du_1 \ldots du_m, \qquad (6.7)
\end{aligned}
$$

by the probability integral transformation (see, for example, Gibbons and Chakraborti (2003)). The subscript U refers to the uniform(0,1) distribution and $p_U(\underline{u})$ is the conditional probability of a signal at any test sample, given the reference sample, when the process is in-control.

Recall that for the in-control case, the distributions of both the reference and test samples can be assumed to be uniformly(0,1) distributed*, which shows that the unconditional ARL, for the in-control situation, of the MW chart does not depend on the underlying process distributions F and G. The same argument can be used to show that the in-control run length distribution does not depend on the underlying process distributions F and G, thus establishing that the proposed MW chart is distribution-free.

We have to calculate the unconditional ARL, for the in-control situation, using (6.7) to implement the chart. Following this, we have to calculate the unconditional ARL, for the out-of-control situation, using (6.3) to evaluate chart performance. We run into two problems in doing so, that is, (i) we don't have exact formulas for the signal probabilities $p_G(\underline{x})$ and $p_U(\underline{u})$; and (ii) it could be difficult and time-consuming estimating (6.3) and (6.7), since both

* For the in-control case, the distributions of both the reference and test samples can be assumed to be uniformly(0,1) distributed. This is due to the well-known probabitliy integral transformation (see, for example, Gibbons and Chakraborti (2003)).

unconditional average run length formulas (for the in-control and out-of-control situations, respectively) are m-dimensional integrals.

Chakraborti and Van de Wiel (2003) proposed a possible solution to both of these problems. Their proposed solution proceeds in two steps. Firstly, fast computations (or approximations) of the signal probabilities are done. It should be noted that although the computation of $p_G(\underline{x})$ will be discussed in detail (in the following section), the computation of $p_U(\underline{u})$ is omitted, since it follows similarly to the computation of $p_G(\underline{x})$. Secondly, Monte Carlo simulation is applied to approximate the unconditional ARL's (for the in-control and out-of-control situations, respectively). The Monte Carlo estimates are given by

$$A\hat{R}L \approx \frac{1}{K}\sum_{i=1}^{K}\frac{1}{p_G(\underline{x}_i)} \tag{6.8}$$

and

$$A\hat{R}L_0 \approx \frac{1}{K}\sum_{i=1}^{K}\frac{1}{p_U(\underline{u}_i)} \tag{6.9}$$

where K denotes the number of Monte Carlo samples, $\underline{x}_i = (x_{i1}, x_{i2}, ..., x_{im})$ and $\underline{u}_i = (u_{i1}, u_{i2}, ..., u_{im})$ denote the i^{th} Monte Carlo sample, $i = 1,2,...,K$, of which each element is taken from some specified F for the $A\hat{R}L$ (for the out-of-control situations) and from the uniform(0,1) distribution for the $A\hat{R}L_0$ (for the in-control situation).

One concern is the size of K, that is, how may Monte Carlo samples should be used? Although larger sizes of K can result in more accurate approximations and smaller Monte Carlo errors, using larger Monte Carlo samples may be more time-consuming. This concern will be addressed in Section 6.1.6.

6.1.4. The computation of the signal probability

The Mann-Whitney statistic, given in (6.2), can be written in a simpler (more straightforward) form given by $M_{\underline{x}\underline{Y}} = \sum_{j=1}^{n} C_j$, where C_j denotes the number of x-observations that precede Y_j, $j = 1,2,...,n$. Also recall that since $p_G(\underline{x}) = 2P_G\left(M_{\underline{x}\underline{Y}} \geq U_{mn}\right)$,

the calculation of $p_G(\underline{x})$ essentially requires the calculation of the upper-tailed probability $P_G\left(M_{\underline{x}Y} \geq U_{mn}\right)$. The computation of the latter proceeds in two steps, namely: (i) listing of all n-tuples $(C_1, C_2, ..., C_n)$ for which the sum is greater than or equal to U_{mn}; and (ii) the summation of the probabilities for these tuples.

The Central Limit Theorem states that if S_n is the sum of n variables, then the distribution of S_n approaches the normal distribution as n approaches infinity, i.e. $S_n \rightarrow$ Normal distribution as $n \rightarrow \infty$. Using this result, we can find a normal approximation to the upper-tailed probability $P_G\left(M_{\underline{x}Y} \geq U_{mn}\right)$, since $M_{\underline{x}Y} = \sum_{j=1}^{n} C_j$ approaches the normal distribution as $n \rightarrow \infty$. Although using a normal approximation to the upper-tailed probability is a possible solution, it is not ideal. The reason being that although normal approximations work well when n is large (and improve as sample size increases), normal approximations do not work well when n is considered small. In our applications we typically use sample sizes that may be considered small and as a result using normal approximations would be somewhat unattractive. Clearly, a better approach is needed.

6.1.5. Saddlepoint approximations

Saddlepoint approximations (or saddlepoint expansions) provide good approximations (with a small relative error) to very small tail probabilities. Consequently, saddlepoint approximations can be applied to the problem of finding $p_G(\underline{x})$, which is usually set to be rather small (typically 0.0027). Jensen (1995) provides ample justifications for the application of saddlepoint expansions when approximating small probabilities. In Chapter 2 of Jensen (1995) the classical saddlepoint approximations for tail probabilities for sums of independent random variables are given. For our problem (the calculation of the upper-tailed probability $P_G(M_{\underline{x}Y} \geq U_{mn})$), we make use of the Lugannani-Rice formula (hereafter LR-formula) which is a saddlepoint expansion formula.

Prior to defining the LR-formula, a few concepts will be explained. To begin with, let a_l denote the probability that l x-observations (given $\underline{X} = \underline{x}$) precede Y_j for $j = 1,2,...,n$

and $l = 0,1,...,m$, respectively. Therefore, $a_l = P(C_j = l \mid \underline{X} = \underline{x}) = P(x_{(l)} < Y_j \le x_{(l+1)})$ where $x_{(1)} \le x_{(2)} \le \ldots \le x_{(m)}$ are the order statistics.

Since the pgf provides a very useful tool for studying the sum of independent random variables we turn to the conditional probability generating function (pgf) of C_j, and subsequently to the conditional pgf of $M_{\underline{x}Y}$. In view of the fact that C_j, $j = 1,2,...,n$, is a random variable whose possible values are restricted to the nonnegative integers $\{0,1,...,m\}$, the conditional pgf of C_j is given by

$$\Pi_1(z) = \sum_{l=0}^{m} P(C_j = l \mid \underline{X} = \underline{x}) z^l = \sum_{l=0}^{m} a_l z^l \ . \tag{6.10}$$

It's a well-known fact that if, for example, X and Y are independent random variables with probability generating functions $\Pi_X(z)$ and $\Pi_Y(z)$, respectively, we have that

$$\Pi_{X+Y}(z) = \Pi_X(z)\Pi_Y(z) \tag{6.11}$$

(see, for example, Bain and Engelhardt (1992)).

$M_{\underline{x}Y}$ is the sum of n independent identical variables (recall that $M_{\underline{x}Y} = \sum_{j=1}^{n} C_j$) and therefore, by using (6.10) and (6.11), the conditional pgf of $M_{\underline{x}Y}$ is given by

$$\Pi_2(z) = \sum_{j=0}^{mn} P(M_{\underline{x}Y} = j) z^j = \left(\sum_{j=0}^{m} a_j z^j \right)^n \ . \tag{6.12}$$

By implication C_j, for $j = 1,2,...,n$, are independent identically distributed, conditionally.

Next we examine the cumulant generating function (cgf) of C_j. The cgf is just the logarithm of the moment generating function (mgf). Mathematically, the mgf and the cgf are equivalent. The cgf generates the mean and variance, instead of the uncentered moments. We can think of $\kappa'(t)$ and $\kappa''(t)$ as the mean and variance, respectively, where $\kappa(t)$ denotes the cgf. Hence, the cgf of C_j can be obtained by taking the logarithm of the pgf in (6.12) at the point $z = e^t$. As a result, the cgf of C_j is given by

$$\kappa(t) = \log\left(\left(\sum_{l=0}^{m} a_l z^l\right)_{z=e^t}\right) = \log\left(\sum_{l=0}^{m} a_l e^{tl}\right). \tag{6.13}$$

The first and second order derivatives of the cgf is simply $\kappa'(t)$ and $\kappa''(t)$ so that $\kappa'(t) = m(t)$ and $\kappa''(t) = \sigma^2(t)$. Also, let $\mu = \frac{U_{mn}}{n}$ and $\overline{M}_{\underline{x}Y} = \frac{M_{\underline{x}Y}}{n}$. The saddlepoint, γ, is the solution to the equation $m(t) = \mu$. In other words, we solve $m(t) = \mu$ for t (see Theorem 3 in Appendix A for a detailed discussion on saddlepoint techniques).

Finally, we want to use a saddlepoint expansion to approximate the upper-tailed probability $P_G(M_{\underline{x}Y} \geq U_{mn})$. Jensen (1995) defined an upper-tailed probability, denoted by $P(\overline{X} \geq x)$, in equation (3.3.17) on page 79 by

$$P(\overline{X} \geq x) = (1 - \Phi(r))(1 + O(n^{-2})) + \phi(r)\left(\frac{1}{\lambda} - \frac{1}{r} + O(n^{-3/2})\right) \tag{6.14}$$

with

$$\lambda = \sqrt{n}\left(1 - e^{(-\hat{\theta}(x))}\right)\sigma(\hat{\theta}(x)) \text{ and } r = (\text{sgn}(\hat{\theta}(x)))\left(2n(\hat{\theta}(x)x - \kappa(\hat{\theta}(x)))\right)^{1/2} \tag{6.15}$$

where $\hat{\theta}(x)$ denotes the saddlepoint, $\text{sgn}(\hat{\theta}(x)) = 1, -1$ or 0 depending on whether $\hat{\theta}(x)$ is positive, negative or zero and $O(\cdot)$ is the big O function. In general, the notation $f(n) = O(g(n))$ means there is a real constant $c > 0$ and an integer n_0 such that $|f(n)| \leq c|g(n)|$ for all $n \geq n_0$ and where $f(n)$ and $g(n)$ are functions of the variable n. In other words, the notation $f(n) = O(g(n))$ states that the function $|f(n)|$ is bounded above by a constant multiple of the function $|g(n)|$ for all sufficiently large values of n indicated by $n \geq n_0$. Getting back to equations (6.14) and (6.15) it should be noted that the derivation of r was done separately on page 75 of Jensen (1995) using equations (3.3.2) and (3.3.3). The Lugannani and Rice (1980) paper was the first to give formula (6.14). Although they were the first to give formula (6.14), their paper is perhaps not easy to read. However, Daniels (1987) has given a very readable account where formula (6.14) is also given. In this thesis we mostly refer to Jensen (1995), because Jensen's textbook gives a rigorous account of the underlying mathematical theory of saddlepoint methods.

Using (6.14) and (6.15) we obtain

$$P\left(M_{\underline{x}Y} \geq U_{mn}\right) = P\left(\overline{M}_{\underline{x}Y} \geq \frac{U_{mn}}{n}\right) = P\left(\overline{M}_{\underline{x}Y} \geq \mu\right) \approx 1 - \Phi(r) + \phi(r)\left(\frac{1}{\lambda} - \frac{1}{r}\right) \quad (6.16)$$

with

$$\lambda = \sqrt{n}(1 - e^{-\gamma})\sigma(\gamma) \text{ and } r = \left(\text{sgn}(\gamma)\right)\left(2n(\gamma\mu - \kappa(\gamma))\right)^{1/2} \quad (6.17)$$

where γ denotes the saddlepoint.

Using (6.16) we can approximate the signal probability $p_G(\underline{x})$ given in Result 6.1.

6.1.6. Monte Carlo simulation

We run into a problem when computing the out-of-control and in-control unconditional average run lengths, since both formulas (see equations (6.3) and (6.7)) are m-dimensional integrals. A solution to this problem is using Monte Carlo simulation. Monte Carlo methods are based on the use of random numbers and probability statistics to investigate problems. It consists of a collection of ways for generating random samples on a computer and then using them to solve problems by providing approximate solutions to those problems. Moreover, Monte Carlo methods are useful for obtaining numerical solutions to problems which are too complicated to solve analytically and are, in this thesis, used to evaluate multiple integrals. Monte Carlo simulation is applied here to approximate the unconditional *ARL*'s for the in-control and out-of-control situations, respectively. It should be noted that these are approximations to *m*-dimensional integrals (see equations (6.3) and (6.7) for the out-of-control and in-control unconditional average run length formulas, respectively). The Monte Carlo estimates are given by (6.8) and (6.9), respectively, and by studying these formulas we see that the computations of $p_G(\underline{x})$ (for the out-of-control situation) and $p_U(u)$ (for the in-control situation) are repeated K times to obtain the Monte Carlo estimates given in (6.8) and (6.9).

Monte Carlo simulation used to approximate the unconditional ARL for the in-control situation

Chakraborti and Van de Wiel (2003) proposed five methods for computing (or approximating) ARL_0. The first three methods are similar in the sense that they all make use of Monte Carlo simulation using (6.9), but they differ in the way that $p_U(\underline{u})$ is computed or approximated. The five methods are as follows:

(i) Exact

Monte Carlo simulation is applied to approximate the ARL_0 using (6.9), with $p_U(\underline{u})$ computed exactly using (6.12).

(ii) Lugannani-Rice formula

Monte Carlo simulation is applied to approximate the ARL_0 using (6.9), with $p_U(\underline{u})$ computed approximately using (6.16).

(iii) Normal Approximation

Monte Carlo simulation is applied to approximate the ARL_0 using (6.9), with $p_U(\underline{u})$ computed using a normal approximation.

The first three methods have the same problem, namely, that we need to compute $p_U(\underline{u})$ K times for K Monte Carlo reference samples, where a reference sample is drawn from the uniform(0,1) distribution. Each element is taken from the uniform(0,1) distribution, since we're approximating the *in-control* average run length. The number of Monte Carlo reference samples K should be taken large enough so that the Monte Carlo error is acceptably small and, consequently, using methods (i), (ii) or (iii) may be time-consuming. By fixing the reference sample we would only need to compute $p_U(\underline{u})$ once. This is done in the fourth method by using the empirical cdf of $X_1, X_2, ..., X_m$.

(iv) Fixed reference sample

Recall that $F(x)$ denotes the unknown continuous cdf of each of $X_1, X_2, ..., X_m$. Let $F_m(\underline{x})$ denote the empirical cdf of $X_1, X_2, ..., X_m$. By the law of strong numbers (see, for example, Bain and Engelhardt (1992)), when m is large, the empirical cdf $F_m(\underline{x})$ converges to $F(\underline{x})$ (which is the cdf of the uniform(0,1) distribution), i.e. $F_m(\underline{x}) \to F(\underline{x})$ as $m \to \infty$, almost surely for fixed $\underline{x}$. Using this, we can replace the i^{th} reference sample observation by the $\left(i/(m+1)\right)^{th}$ quantile, $i = 1,2,...,m$, of the uniform(0,1) distribution. Since this quantile is equal to $i/(m+1)$ (say, $q_i = i/(m+1)$), we can approximate ARL_0 by $1/p_U(\underline{q})$ where $\underline{q} = (q_1, q_2, ..., q_m) = \left(1/(m+1), ..., m/(m+1)\right)$. It should be noted that one should only use the empirical cdf (and as a result fix the reference sample to $\underline{x} = \underline{u} = \underline{q}$) when m is large. Using this method we only require one reference sample and we only compute $p_U(\underline{u})$ once.

(v) Reciprocal of the false alarm rate

A quick way to approximate the ARL_0 is by using the fact that if the charting statistics, $M_{XY}^1, M_{XY}^2, ...$, were independent, the ARL_0 would be equal to the reciprocal of the false alarm rate, i.e. $ARL_0 = \frac{1}{FAR} = \frac{1}{2P\left(M_{XY} \geq U_{mn}\right)}$. When implementing this method, the FAR is estimated using the Fix-Hodges approximation formula (see Fix and Hodges (1955)). This approximation improves the normal approximation by including moments of order three and higher. Since the charting statistics are in fact dependent, we can only use the reciprocal of the false alarm rate as a quick approximation to the ARL_0. Further motivation for using the reciprocal of the false alarm rate as a quick approximation to the ARL_0 is given by Chakraborti (2000). In that paper the author showed that for the Shewhart $\overline{X}$ chart, $\frac{1}{FAR}$ can be used as a

lower bound to the ARL_0. Following this, we use the reciprocal of the false alarm rate as a quick approximation to the ARL_0.

Methods (iv) and (v) have the advantage that we don't have to draw K reference samples, since we approximate ARL_0 by $1/p_U(\underline{q})$ (using method (iv)) and by $1/2P(M_{XY} \geq U_{mn})$ (using method (v)).

The five abovementioned methods show one how to calculate (or approximate) the unconditional ARL_0 corresponding to a given value of the UCL. A table containing values of the unconditional ARL_0, for various values of m and n, is provided (see Table 1, Chakraborti and Van de Wiel (2003)). The table is given on the next page for reference. K is kept constant ($K = 1000$) to obtain a fair comparison regarding the computing times. The values in Table 6.1 were computed using all five abovementioned methods. The table shows two computing times. The first computing time is the time it took a 3.2GHz Pentium PC with 512MB of internal RAM to compute the values using Mathematica 6.0. The second computing time (given in brackets) is the computing time found by Chakraborti and Van de Wiel (2003) using a 1.7GHz Pentium PC with 128MB of internal RAM.

Certain in-control average run length values (indicated by ** in Table 6.1) could not be computed within a practical time. Chakraborti and Van de Wiel (2003) determined these computing times by multiplying the computing time for K = 1 by 1 000 and, consequently, getting a computing time for K = 1 000. In this paper the same course of action was taken to estimate the computing times for K = 1 000. From Table 6.1 we see that the 3.2GHz Pentium PC with 512MB of internal RAM is at least three times faster than the 1.7GHz Pentium PC with 128MB of internal RAM. Interpreting the times in Table 6.1 we find that the exact method is exceptionally time-consuming, particularly so as m increases. Similarly, using the LR-formula is also very time-consuming, again, particularly as m increases, but it's not as severely time-consuming as the exact method. Although fast approximations are given by the normal approximation, they are inaccurate. Fast approximations are also given by the fixed-reference-sample method and the reciprocal-of-the-false-alarm-rate method.

Table 6.1[*]**.** ARL_0 approximations and computing times for various m and n values and $U_{mn} = 4344$ [†].

m	***n***	**Exact**		**Lugannani-Rice formula**		**Normal Approximation**		**Fixed reference sample**		**Reciprocal of the false alarm rate**	
		$\hat{ARL}_0$	Time (sec.)	$\hat{ARL}_0$	Time (sec.)	$\hat{ARL}_0$	Time (sec.)	$\hat{ARL}_0$	Time (sec.)	$\hat{ARL}_0$	Time (sec.)
50	5	486	18 (54)	506	12 (36)	307	0.33 (1.00)	403	0.02 (0.05)	247	0.003 (0.01)
	10	504	132 (395)	505	12 (34)	327	0.33 (1.00)	524	0.02 (0.05)	226	0.003 (0.01)
	25	488	1618 (4850)	491	10 (31)	425	0.40 (1.20)	694	0.02 (0.05)	119	0.003 (0.01)
100	5	496	75 (220)	505	18 (48)	219	0.40 (1.20)	478	0.02 (0.05)	353	0.003 (0.01)
	10	505	640 (1920)	506	16 (47)	339	0.42 (1.30)	531	0.02 (0.05)	332	0.003 (0.01)
	25	**	8900 (26168)	503	18 (48)	422	0.42 (1.30)	683	0.03 (0.06)	233	0.003 (0.01)
500	5	491	3544 (10633)	496	70 (207)	226	0.40 (1.20)	492	0.07 (0.20)	445	0.003 (0.01)
	10	**	24500 (73516)	513	60 (179)	367	0.60 (1.70)	537	0.07 (0.21)	484	0.003 (0.01)
	25	**	$2.53*10^5$ ($7.59*10^5$)	494	60 (176)	445	0.55 (1.60)	578	0.10 (0.29)	450	0.003 (0.01)
1000	5	**	10601 (31766)	500	120 (356)	235	0.70 (2.10)	513	0.16 (0.48)	471	0.003 (0.01)
	10	**	$1.15*10^5$ ($3.42*10^5$)	499	126 (373)	355	0.80 (2.40)	516	0.18 (0.49)	488	0.003 (0.01)
	25	**	$1.05*10^6$ ($3.15*10^6$)	500	117 (348)	442	0.61 (1.70)	548	0.20 (0.63)	482	0.003 (0.01)
2000	5	**	$0.57*10^5$ ($1.71*10^5$)	503	240 (713)	234	0.70 (2.10)	506	0.22 (0.67)	474	0.003 (0.01)
	10	**	$0.48*10^6$ ($1.44*10^6$)	504	221 (659)	354	0.64 (1.90)	513	0.24 (0.71)	499	0.003 (0.01)
	25	**	$0.43*10^7$ ($1.29*10^7$)	509	229 (676)	446	0.71 (2.10)	531	0.48 (1.41)	497	0.003 (0.01)

[*] Chakraborti and Van de Wiel (2003) wrote a Mathmatica program to approximate the ARL_0 for a given m, n and value of the UCL. This Mathematica program can be downloaded using the website www.win.tue.nl/~markvdw. For more details on this Mathematica program see Mathematica Program 1 in Appendix B.

[†] Table 6.1 appears in Chakraborti and Van de Wiel (2003), Table 1. It should be noted that Chakraborti and Van de Wiel (2003) failed to say what the value of the UCL was set equal to when constructing this table. Turning to their Mathematica program we see that the user specific parameters are set equal to m=1000, n=5 and UCL=4344. Recall that only the UCL needs to be specified, since the LCL can be calculated using $L_{mn}=mn-U_{mn}$.

In Table 6.1 we have the exact formula and various approximations for the unconditional ARL_0. We see that although the exact formula gives us ARL_0 values close to 500 (which is desirable), the exact computations are very time-consuming for most values of m and n. Focusing on the approximations, the closer an ARL_0 value is to 500, the better the approximation. Using this criteria we see that the normal approximation is inaccurate for all values of m and n. The fixed-reference-sample and the reciprocal-of-the-false-alarm-rate approximations are relatively good for $m \geq 1000$ and, in particular, the fixed-reference-sample approximation performs better for 'small' values of n ($=5$ or 10) than for n 'large' ($n = 25$). It seems that the best approximation is the LR-formula, since all the corresponding ARL_0 values are close to 500. In summary, the best method of calculating the unconditional ARL_0 is by using the exact formula, if it's not too time-consuming, otherwise the LR-formula is the best approximation.

Monte Carlo simulation used to approximate the unconditional ARL for the out-of-control situation

Monte Carlo simulation is used to approximate the unconditional ARL for the out-of-control situation. There are concerns about the number of Monte Carlo samples used, namely, that although larger sizes of K will result in more accurate approximations and smaller Monte Carlo errors, using larger Monte Carlo samples may be time-consuming or computationally expensive or both.

Since the unconditional ARL is the average of the conditional $ARL_G(\underline{X})$ over all possible $\underline{X}'s$ and the K Monte Carlo reference samples are independent, the Monte Carlo standard error of the estimate $\hat{ARL}$ is given by

$$\sigma_{mc} = \frac{\sigma\left(ARL_G(\underline{X})\right)}{\sqrt{K}} \tag{6.18}$$

where $\sigma\left(ARL_G(\underline{X})\right)$ denotes the unknown standard deviation of $ARL_G(\underline{X})$. From (6.18) we see that the standard error decreases with the square root of the number of Monte Carlo samples used. If we, for example, quadruple the number of Monte Carlo samples used, we will half the standard error. While increasing K is one technique for reducing the standard

error, doing so can be time-consuming or computationally expensive or both. Clearly, a better approach is needed.

Let D denote some specified value such that

$$s_{mc} = \frac{s\left(ARL_G(\underline{X})\right)}{\sqrt{K}} \leq D. \tag{6.19}$$

where the sample standard deviation $s\left(ARL_G(\underline{X})\right)$ is used to estimate $\sigma\left(ARL_G(\underline{X})\right)$ and, subsequently, $s_{mc} = \frac{s\left(ARL_G(\underline{X})\right)}{\sqrt{K}}$ is used to estimate $\sigma_{mc} = \frac{\sigma\left(ARL_G(\underline{X})\right)}{\sqrt{K}}$.

We want to find the smallest K such that (6.19) is satisfied. We start by taking K 'small', say $K = 100$, for example, and then we compute the corresponding standard error s_{mc}. If (6.19) is not satisfied we increase K and the process is repeated until the standard error is smaller than or equal to some specified value D. It should be noted that D could also be taken to be some percentage of the estimate $A\hat{R}L$. By implementing (6.19), we find an accurate approximation (with a small Monte Carlo error) of the unconditional ARL for the out-of-control situation.

6.1.7. Determination of chart constants

Up to this point we've addressed the problem where one has to calculate the (unknown) unconditional ARL_0 for a given (known) upper control limit. In this section we address the opposite problem where one has to calculate the (unknown) upper control limit for a specified (known) ARL_0. In order to solve the latter problem, we use an iterative procedure based on linear interpolation. An initial value for the *UCL*, say $UCL^{(1)}$, is needed to start the iteration. We can limit our search of $UCL^{(1)}$ (and ultimately of *UCL*) to integer values between 0 and *mn*, since the MW charting statistic only takes on integer values between 0 and *mn* (recall that $0 \leq M_{XY} \leq mn$). In addition, we use the fact that ARL_0 is strictly increasing in *UCL* (and subsequently, $p_U(\underline{u})$ is strictly decreasing in *UCL*). Let the desired unconditional $ARL_0 = 500$.

To obtain $UCL^{(1)}$, the fixed-reference-sample approximation or the reciprocal-of-the-FAR approximation can be used, since they are both very fast and relatively accurate approximations. For the latter approach, we equate the reciprocal of the false alarm rate to 500, meaning that we have to solve for $\frac{1}{2(FH(u))} = 500$, where $FH(u)$ denotes the Fix-Hodges approximation for the upper tail probability $P_0(M_{\underline{xY}} \geq u)$. We estimate (through Monte Carlo simulation) ARL_0 at $UCL^{(1)}$ using (6.9), where the LR-approximation is used to calculate $p_U(\underline{u})$. In doing so, we obtain a new ARL_0, say $ARL_0^{(1)}$. If $ARL_0^{(1)}$ is smaller than 500, we increase the value of $UCL^{(1)}$ by a specified amount, say s, to obtain $UCL^{(2)} = UCL^{(1)} + s$. On the contrary, if $ARL_0^{(1)}$ is greater than 500, we decrease the value of $UCL^{(1)}$ to obtain $UCL^{(2)} = UCL^{(1)} - s$. Using $UCL^{(2)}$, the search procedure is repeated until ARL_0 is 'satisfactorily close' to the target value of 500.

A question arises: How close is 'satisfactorily close'? To answer this Chakraborti and Van de Wiel (2003) suggest using a target interval, say $500 \pm \lambda 500$, where λ denotes the percentage deviation from the target value that is acceptable. Suppose we allow a deviation of 3%, i.e. $\lambda = 0.03$, the search procedure stops at the l^{th} step if $485 \leq ARL_0^{(l)} \leq 515$. The larger this margin, the faster the algorithm, and as a result, the faster a solution is found. If the specifications can't be met, the algorithm returns one or more solutions for which ARL_0 is close to the target value. If $\lambda = 0$, the search procedure stops at the l^{th} step if $UCL^{(l)}$ has a corresponding $ARL_0 < 500$ and $UCL^{(l)} + 1$ has a corresponding $ARL_0 \geq 500$, and as a result, the practitioner has to decide whether to use $UCL^{(l)}$ or $UCL^{(l)} + 1$.

We illustrate this search procedure with an example. Suppose the reference sample size is 50 (m = 50), the test sample size is 5 (n = 5) and we want to find the chart constants (U_{mn} and $L_{mn} = mn - U_{mn}$) such that the specified target $ARL_0 = 500$. Suppose we specify that a 3% deviation from the target value is acceptable, i.e. $\lambda = 0.03$. In doing so, the search procedure stops when $485 \leq \hat{ARL}_0 \leq 515$ and yields the corresponding chart constants. In addition, we specify that the Monte Carlo standard error, s_{mc}, be smaller than or equal to 2.5% of the estimate of ARL_0. Then $D = 0.025 \times 500 = 12.5$ is the maximum value of the

standard error of the estimate that we allow. The output from the program is given in Table 6.2.

Table 6.2. Finding chart constants for m=50, n=5, target ARL_0=500, $\lambda = 0.03$ and $D = 12.5$*.

1/(false alarm rate approximation)
(1) ucl= 222 lcl= 28 ARL0= 500
Fixed reference sample approximation
(2) ucl= 222 lcl= 28 ARL0= 874.22
(3) ucl= 212 lcl= 38 ARL0= 206.763
(4) ucl= 216 lcl= 34 ARL0= 351.068
(5) ucl= 218 lcl= 32 ARL0= 467.529
LR-approximation
(6) ucl= 218 lcl= 32 ARL0= 571.016 smc= 15.4659 5% perc= 122.654 K= 2000
(7) ucl= 208 lcl= 42 ARL0= 138.825 smc= 3.46778 5% perc= 44.3363 K= 1061
(8) ucl= 216 lcl= 34 ARL0= 426.735 smc= 11.7639 5% perc= 95.169 K= 2000
(9) ucl= 217 lcl= 33 ARL0= 494.728 smc= 13.6655 5% perc= 105.761 K= 2000
{231.156,Null}

From Table 6.2 it can be seen that one iteration has been carried out under the reciprocal-of-the-*FAR* approximation and that four iterations have been carried out under the fixed-reference-sample approximation. These five iterations didn't take long, since both the reciprocal-of-the-*FAR* and the fixed-reference-sample approximations are fast approximations. For each of these five iterations, the values of U_{mn} (denoted ucl in the output), L_{mn} (denoted lcl in the output) and the corresponding unconditional ARL_0 (denoted ARL0 in the output) are given.

From Table 6.2 it can also be seen that four iterations have been carried out under the Lugannani-Rice approximation. For each of these four iterations, the values of U_{mn}, L_{mn}, the corresponding unconditional ARL_0, the standard error of the estimated ARL_0 (denoted smc in the output), the estimated 5th percentile of the conditional in-control *ARL* distribution (denoted 5% perc in the output) and the number of Monte Carlo samples used to obtain the estimates (denoted by K in the output) are given. In total there were nine iterations that have been carried out in approximately 231 seconds. The final chart constants are found at iteration number 9. They are $U_{mn} = 217$ and $L_{mn} = 33$ with a corresponding unconditional

* The values in Table 6.2 were obtained by running the Mathematica program provided by Chakraborti and Van de Wiel (2003). See Mathematica Program 1 in Appendix B for more information on this Mathematica program.

$ARL_0 = 494.728$. The 5th percentile of the conditional in-control *ARL* distribution is equal to 105.761, meaning that 95% of all reference samples (that could possibly have been taken from the in-control process) will generate a conditional ARL_0 of *at least* 106. Previously stated is the fact that K is chosen such that the standard error of the estimate is smaller than or equal to 2.5% of the estimate. Stated differently, K is chosen such that $s_{mc} \leq D$. When studying iteration number 9, we see that this condition is not satisfied, since $s_{mc} = 13.6655 > 12.5$. The reason for this is that the maximum number of Monte Carlo samples is set to 2 000 in the Mathematica program. If there were no restriction put on the number of Monte Carlo samples, the iterative procedure would have increased K and repeated the process until the standard error is smaller than or equal to $D = 12.5$. Table 6.3 contains the chart constants for various values of *m* and *n* with $\lambda = 0.03$ and where s_{mc} must be smaller than or equal to 2.5% of the estimate of ARL_0.

Table 6.3. Control limits for various values of *m* and *n*[*].

m	***n***	$ARL_0 = 370$		$ARL_0 = 500$	
		L_{mn}	U_{mn}	L_{mn}	U_{mn}
50	5	35	215	33	217
	10	115	385	111	389
	25	400	850	393	857
100	5	69	431	65	435
	10	231	769	224	776
	25	805	1695	793	1707
500	5	348	2152	328	2172
	10	1170	3830	1128	3872
	25	4081	8419	4016	8484
1000	5	698	4302	653	4347
	10	2344	7656	2268	7732
	25	8169	16831	8058	16942
2000	5	1397	8603	1309	8691
	10	4682	15318	4540	15460
	25	16392	33608	16145	33855

[*] The values in Table 6.3 were obtained by running the Mathematica program provided by Chakraborti and Van de Wiel (2003). See Mathematica Program 1 in Appendix B for more information on this Mathematica program. Table 6.3 also appears in Chakraborti and Van de Wiel (2003), Table 3.

Example 6.1

A Mann-Whitney control chart based on the Montgomery (2001) piston ring data

For the piston-ring data with $m = 125$ (see example 4.1 for an explanation of why m is equal to 125 (and not 25 like some of the earlier examples)) and $n = 5$, Chakraborti and Van de Wiel (2003) found the upper and lower control limits of the Shewhart-type MW chart to be 540 and 85, respectively. The control limits are obtained by setting the user-specific parameters equal to $m = 125$, $n = 5$, target $ARL_0 = 400$, $\lambda = 0.02$ and $D = 6$ in the Mathematica program provided by Chakraborti and Van de Wiel (2003). By setting $D = 6$ we require that $s_{mc} \leq 6$. By setting $\lambda = 0.02$ we specify that a 2% deviation from the target value is acceptable. In doing so, the search procedure stops when $392 \leq \hat{ARL}_0 \leq 408$ and yields the corresponding chart constants, which (in this case) equal $L_{mn} = 85$ and $U_{mn} = 540$. The fifteen Phase II samples and the reference sample lead to fifteen MW statistics shown in Table 6.4 (read from left to right and to left) and the MW control chart is shown in Figure 6.1.

Table 6.4. Phase II MW statistics for the Piston-ring data in Montgomery (2001)*.

429.0	333.0	142.5	370.5	241.5	410.5	393.0	240.5
471.0	486.0	340.5	561.0	575.5	601.5	484.5	

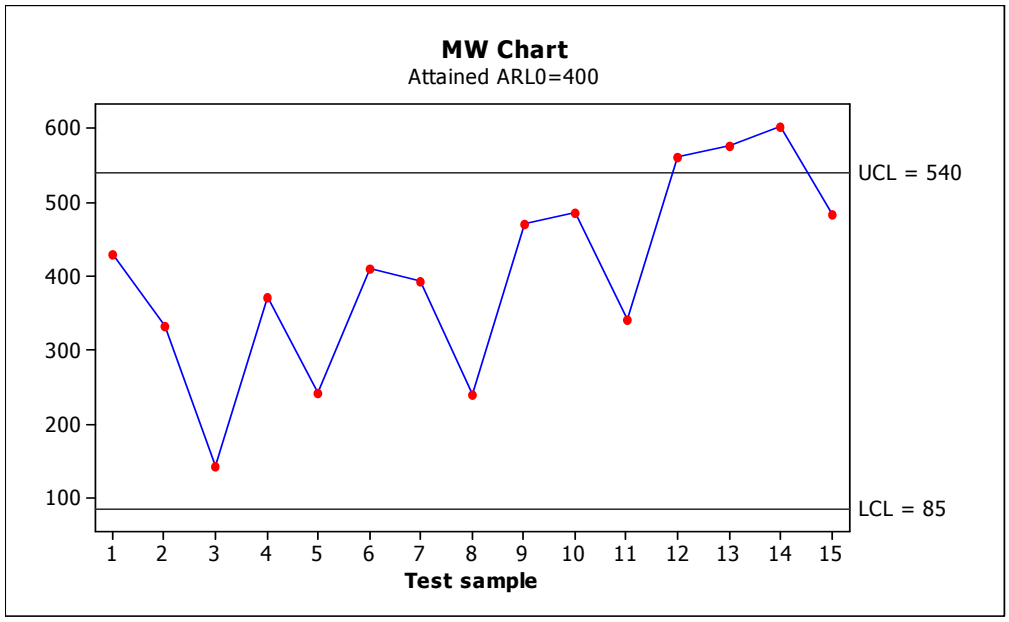

Figure 6.1 MW Chart for the Montgomery (2001) piston ring data.

*The values in Table 6.4 were calculated using Minitab.

It is seen that all but three of the test groups, 12, 13 and 14 are in-control. The conclusion from the MW chart is that the medians of test groups 12, 13 and 14 have shifted to the right in comparison with the median of the in-control distribution, assuming that G is a location shift of F. It may be noted that the Shewhart $\overline{X}$ chart shown in Montgomery (2001) led to the same conclusion with respect to the means. Of course, the advantage with the MW chart (and with any nonparametric chart) is that it is distribution-free, so that regardless of the underlying distribution, the in-control ARL of the chart is roughly equal to 400 and there is no need to worry about (non-) normality, as one must for the $\overline{X}$ chart. For comparison purposes, the distribution-free *1-of-1* precedence chart for this data for an unconditional $ARL_0 = 400$ is found to be $LCL = 73.982$ and $UCL = 74.017$ for an attained $ARL_0 \approx 414.0$. Consequently, the precedence chart declares the 12^{th} and the 14^{th} groups to be out of control but not the 13^{th} group, unlike the MW and the Shewhart chart. This is not entirely surprising since the MW test is generally more powerful than the precedence test.

6.1.8. Control chart performance

The performance of a control chart is usually judged in terms of certain characteristics associated with its run-length distribution. For the most part the ARL is used to evaluate chart performance, since it indicates, on average, how long one has to wait before the chart signals. Some researchers have advocated using other characteristics than the ARL, such as percentiles of the run length distribution (see Section 2.1.5 for a detailed discussion on this issue). Chakraborti and Van de Wiel (2003) examined the ARL, the 5^{th} and the 95^{th} percentiles (denoted ρ_5 and ρ_{95}, respectively) of the *conditional* distribution for the MW chart. The question may be raised about why the authors decided to use (only) the *conditional* distribution when both the conditional *and* unconditional distributions provide key information concerning the performance of a chart. Recall that the unconditional distribution results from averaging over all possible reference samples and in practice researchers would (almost certainly) not have the benefit of averaging.

Chakraborti and Van de Wiel (2003) compared the MW chart to the Shewhart $\overline{X}$ chart. For the latter we assume case UU, when both the mean and variance are unknown, and consequently both parameters need to be estimated from the reference sample. Therefore, the MW chart is compared to the Shewhart $\overline{X}$ chart *with estimated parameters*. Additionally,

both charts are designed to have the same in-control average run length ($ARL_0 \approx 500$). The latter two conditions are necessary to ensure a fair comparison between the MW and Shewhart $\overline{X}$ chart. The in-control case is considered first.

In-control performance

For the in-control case a lower order percentile, specifically the 5th percentile, should be examined (the 95th percentile is also examined for completeness). Recall that we want the in-control ARL to be large and, by the same token, large values of the 5th percentile are desirable. The test sample size, n, was taken to equal 5 for all cases, whereas the reference sample size, m, varies from 50 to 2000. Both the reference and test samples were drawn from a normal distribution, specifically a $N(0,1)$ distribution. The results were obtained using $K = 1000$ simulations and are shown in Table 6.5.

Table 6.5. The 5th and 95th percentiles and standard deviations of the conditional in-control distribution with $n = 5$ and $ARL_0 \approx 500$ *.

m	MW chart				Shewhart $\overline{X}$ Chart			
	Upper Control Limit	ρ_5	ρ_{95}	Standard Deviation	Upper Control Limit	ρ_5	ρ_{95}	Standard Deviation
50	217	97	1292	553	3.01996	49	1619	854
75	326	146	1219	461	3.05156	87	1379	645
100	435	182	1146	358	3.06535	112	1290	463
150	654	251	1090	315	3.07715	154	1197	377
300	1304	284	845	197	3.08607	232	927	235
500	2172	322	700	140	3.08848	270	828	174
750	3258	360	677	107	3.08935	314	765	140
1000	4347	379	674	83	3.08969	338	721	121
2000	8691	420	629	55	3.09007	376	651	84

From Table 6.5 we find that for the MW chart with $m = 100$ and a control chart constant of 435 ($U_{mn} = 435$), $\rho_5 = 182$, meaning that 95% of the in-control average run lengths are at least 182, whereas for the Shewhart $\overline{X}$ chart with $m = 100$ and a control chart

* The values in Table 6.5 were obtained by running the Mathematica program provided by Chakraborti and Van de Wiel (2003). See Mathematica Program 1 in Appendix B for more information on this Mathematica program. Table 6.5 also appears in Chakraborti and Van de Wiel (2003), Table 4.

constant of 3.06535, $\rho_5 = 112$, meaning that 95% of the in-control average run lengths are at least 112. Since 182 > 112 it can be concluded that the in-control performance of the MW chart is better than that of the Shewhart $\overline{X}$ chart with estimated parameters. Moreover, all the 5th percentiles of the MW chart are larger than those of the Shewhart $\overline{X}$ chart with estimated parameters, particularly for $m \leq 150$, further supporting the statement that the in-control performance of the MW chart is better than that of the Shewhart $\overline{X}$ chart with estimated parameters. The estimated standard deviations are given in Table 6.5 to give some information about the variability of ARL_0. All the estimated standard deviations for the MW chart are smaller than those of the Shewhart $\overline{X}$ chart with estimated parameters, further supporting the statement that the in-control performance of the MW chart is better.

Out-of-control performance

For the out-of-control case a higher order percentile, specifically the 95th percentile, is examined. Recall that we want the out-of-control ARL to be small and, by the same token, small values of the 95th percentile are desirable. These two performance measures, the ARL_δ and ρ_{95}, are examined for three distributions, namely, the Normal, Laplace and Gamma(2,2) distributions, respectively. The motivation for examining these three distributions is that we would like to examine a symmetric (Normal), asymmetric (Gamma(2,2)) and heavy-tailed (Laplace) distribution, respectively. The Laplace distribution is comparable to the Normal distribution, but it has heavier tails, while the Gamma(2,2) distribution is positively skewed.

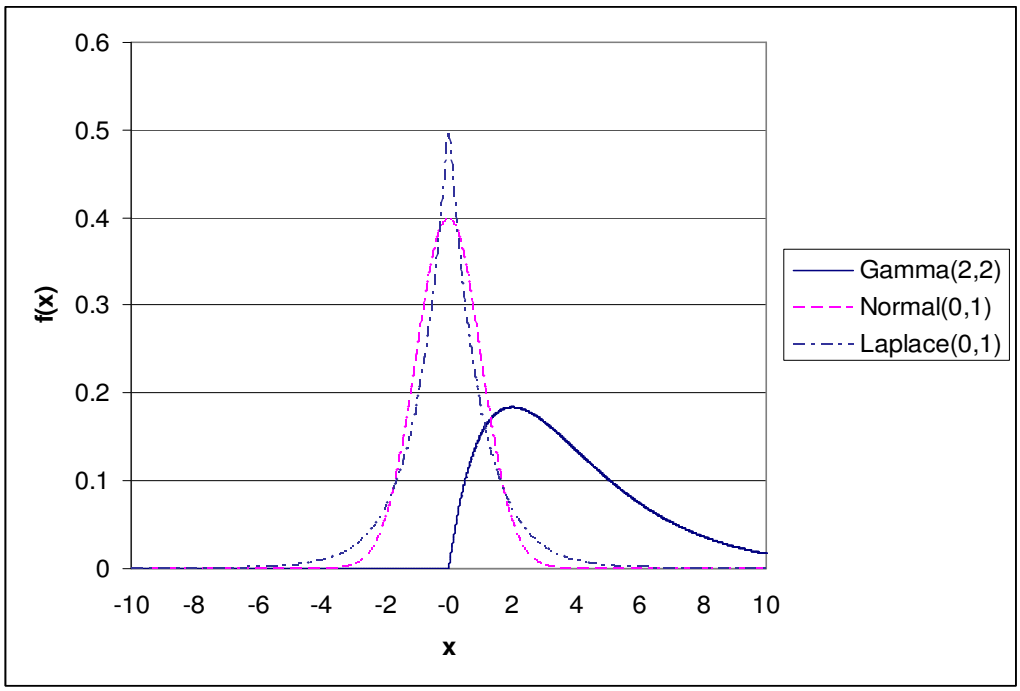

Figure 6.2. The shapes of the three distributions under consideration.

(i) The normal distribution

For the Normal distribution we would expect the out-of-control performance of the Shewhart $\overline{X}$ chart with estimated parameters to be better than that of the MW chart. The reason for this being that it's typical for normal theory methods to outperform nonparametric methods when the normality assumption is met. A two-sided chart was applied in the case of the Normal distribution. The test sample size, n, was taken to equal 5, whereas the reference sample size, m, was taken to equal 100. The chart constants for both the MW and Shewhart $\overline{X}$ chart are chosen such that the in-control average run length is approximately equal ($ARL_0 \approx 500$) for both charts. ARL_δ and the 95th percentiles of the distribution of ARL_δ were computed, using these chart constants, for various values of δ, where δ is the unknown shift parameter (recall that shift alternatives are denoted as $G(x) = F(x-\delta)$). The results are shown below in Figure 6.3.

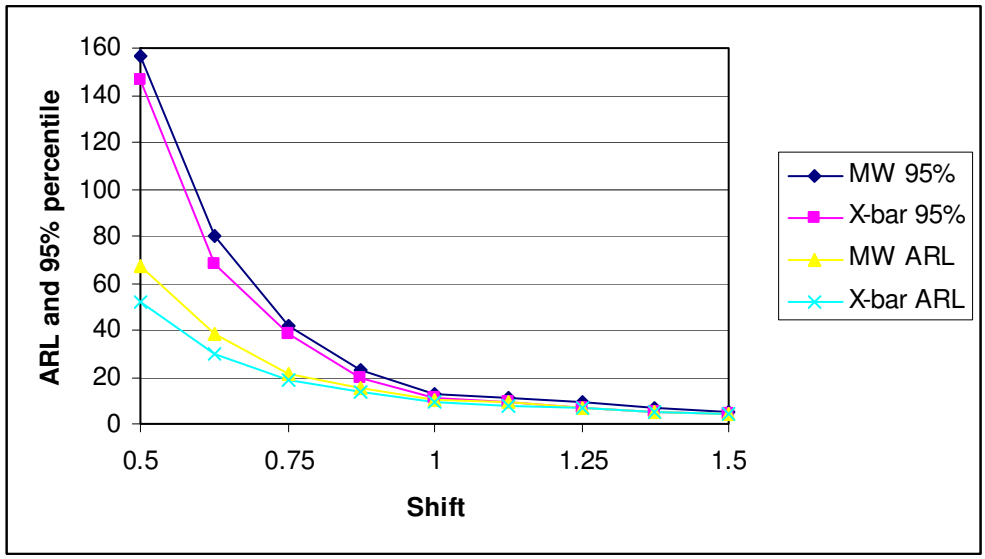

Figure 6.3. Comparison of the MW chart with the Shewhart $\overline{X}$ chart for the Normal distribution.

When comparing the MW chart with the Shewhart $\overline{X}$ chart under Normal shift alternatives we find that the Shewhart $\overline{X}$ chart is performing only slightly better than the MW chart, since the 95th percentiles for the Shewhart $\overline{X}$ chart are smaller than the 95th percentiles for the MW chart. However, it should be noted that the differences are small and it appears to fade away when the shift is greater than one. A similar pattern holds for the ARL_δ's.

(ii) The Laplace distribution

The Laplace distribution, also called the Double-Exponential distribution, is comparable to the Normal distribution, but it has heavier tails (see Figure 6.2). As a result, there are higher probabilities associated with extreme values when working with the Laplace distribution as opposed to using the Normal distribution. For the Laplace distribution we would expect the out-of-control performance of the MW chart to be better than that of the Shewhart $\overline{X}$ chart. The reason for this being that it's typical for nonparametric methods to outperform normal theory methods when the distribution in question is heavy-tailed (see, for example, Gibbons and Chakraborti (2003)). A two-sided chart was applied to the Laplace distribution. For consistency, $n = 5$ and $m = 100$ (the same values were used under Normal shift alternatives) and the chart constants for both the MW and Shewhart $\overline{X}$ chart are chosen

such that the in-control average run length is approximately equal ($ARL_0 \approx 500$) for both charts.

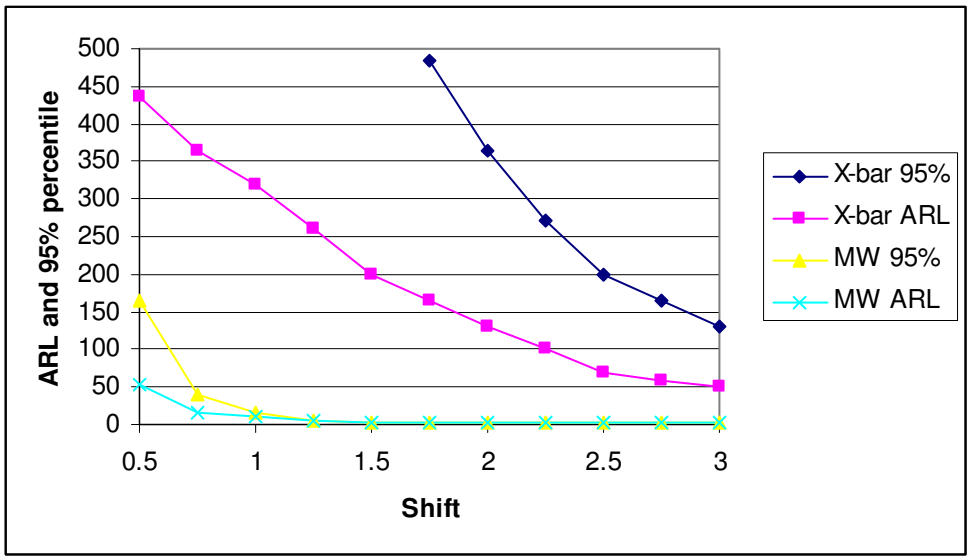

Figure 6.4. Comparison of the MW chart with the Shewhart $\overline{X}$ chart for the Laplace distribution.

When comparing the MW chart with the Shewhart $\overline{X}$ chart under Laplace shift alternatives we find that the MW chart is performing much better than the Shewhart $\overline{X}$ chart, since the 95th percentiles for the MW chart are smaller than the 95th percentiles for the Shewhart $\overline{X}$ chart. It should be noted that these differences are reasonably large for all shifts, indicating that the MW chart is performing a great deal better than the Shewhart $\overline{X}$ chart.

(iii) The Gamma distribution

From Figure 6.2 it can be seen that the Gamma(2,2) distribution is positively skewed. For the Gamma(2,2) distribution we would expect the out-of-control performance of the MW chart to be better than that of the Shewhart $\overline{X}$ chart. An upper one-sided chart was applied to the Gamma(2,2) distribution. For consistency, $n = 5$ and $m = 100$ (the same values were used under Normal and Laplace shift alternatives) and the chart constants for both the MW and Shewhart $\overline{X}$ chart are chosen such that the in-control average run length is approximately equal ($ARL_0 \approx 500$) for both charts.

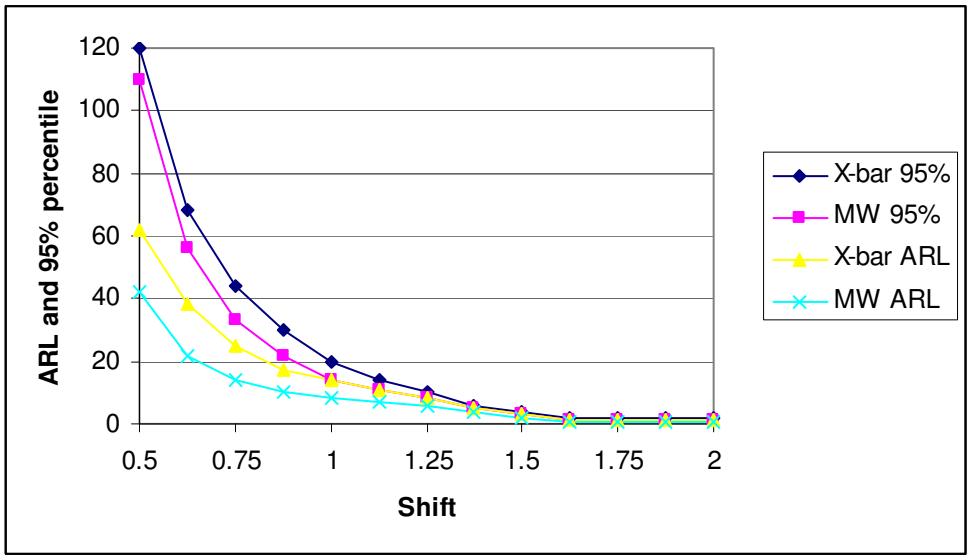

Figure 6.5. Comparison of the MW chart with the Shewhart $\overline{X}$ chart for the Gamma distribution.

When comparing the MW chart with the Shewhart $\overline{X}$ chart under Gamma(2,2) shift alternatives we find that the MW chart is performing better than the Shewhart $\overline{X}$ chart, since the 95th percentiles for the MW chart are smaller than the 95th percentiles for the Shewhart $\overline{X}$ chart. It should be noted that these differences are not as large as the differences obtained using the Laplace distribution.

The graphs were also constructed for larger values of m, but since the graphs were very similar to the given figures, they we omitted. In conclusion we found that the MW chart performs better than the Shewhart $\overline{X}$ chart with estimated parameters under heavy tailed and skewed distributions.

6.1.9. Summary

In Section 6.1 we examined a Shewhart-type chart based on the Mann-Whitney-Wilcoxon statistic. We illustrated these procedures using the piston ring data from Montgomery (2001) to help the reader to understand the subject more thoroughly. One practical advantage of the nonparametric Shewhart-type Mann-Whitney control chart is that there is no need to assume a particular parametric distribution for the underlying process (see Section 1.4 for other advantages of nonparametric charts).

6.2. The tabular Phase I CUSUM control chart

6.2.1. Introduction

Zhou, Zou and Wang (2007) (hereafter ZZW) proposed a Phase I CUSUM control chart for individual observations based on the Mann-Whitney-Wilcoxon statistic. They compared their proposed control chart to the likelihood ratio test (LRT) chart of Sullivan and Woodall (1996) and the CUSUM LT chart of Koning and Does (2000).

Suppose that a sample of size n, $X_1, X_2, ..., X_n$, is available with an unknown continuous cdf, $F(x, \mu_i)$ $i = 1,2,...,n$, where μ_i denotes the location parameter. In this set up, let μ_i denote the population mean of X_i. An out-of-control condition is a shift in the location parameter to some different value. The problem of detecting a shift in a parameter of the process is similar to sequential change-point detection (see, for example, Hawkins and Zamba (2005)). Various authors have studied the change-point problem; see for example Hawkins (1977), Sullivan and Woodall (1996), Hawkins, Qiu and Kang (2003) and Hawkins and Zamba (2005). In a change-point model, all the observations up to the change-point have the same distribution, say $F(x, \mu_a)$, while the remaining observations have the same distribution, say $F(x, \mu_b)$, i.e.

$$X_i = \begin{cases} F(x, \mu_a) & \text{for} \quad i = 1,2,...,t \\ F(x, \mu_b) & \text{for} \quad i = t+1, t+2,...,n \end{cases}$$

where t, with $1 \le t < n$, is the change-point. If $\mu_a = \mu_b$ the process is said to be in-control, whereas if $\mu_a \neq \mu_b$ the process is declared to be out-of-control. ZZW give an estimate for the position of the change-point, $\hat{\tau}$, as $\hat{\tau} = \underset{1<t<n}{\arg\max}\{|SMW_t|\}$ [*] (see Pettitt (1979)), where SMW_t is defined in (6.21). One can also look for multiple shifts, especially if the dataset is large. This could be done by partitioning the data at the location of the change-point then repeating the process on each subset of observations. This continues until no evidence of additional change-points is given. For example, if there are two shifts we have

[*] *Arg max* stands for the argument of the maximum, that is, the value of the given argument for which the value of the given expression attains its maximum value. For example, arg max $\{f(x)\}$ is the value of x for which $f(x)$ has the largest value.

$$X_i = \begin{cases} F(x,\mu_a) & \text{for} \quad i = 1,2,...,\tau_1 \\ F(x,\mu_b) & \text{for} \quad i = \tau_1+1,\tau_1+2,...,\tau_2 \\ F(x,\mu_c) & \text{for} \quad i = \tau_2+1,\tau_2+2,...,n \end{cases}$$

where τ_1 and τ_2 are the two change-points, respectively. ZZW lay emphasis on the fact that their proposed CUSUM Mann-Whitney chart is not intended to be used for detecting multiple shifts, but they still expect the chart to have good detecting performance if the mean shifts (μ_a, μ_b and μ_c) are all in the same direction, i.e. μ_a, μ_b and μ_c form either a decreasing or an increasing sequence.

The Mann-Whitney statistic* is defined to be the number of (X_i, X_j) pairs with $X_i > X_j$ where $i = 1,2,...,t$ and $j = t+1, t+2,...,n$. This can be written as

$$MW_t = \sum_{i=1}^{t}\sum_{j=t+1}^{n} I(X_j < X_i) \text{ for } t = 1,2,...,n-1 \qquad (6.20)$$

where $I(X_j < X_i)$ is the indicator function, i.e.

$$I(X_j < X_i) = \begin{cases} 1 & \text{if} \quad X_j < X_i \\ 0 & \text{if} \quad X_j \geq X_i \end{cases}.$$

The expected value, variance and standard deviation of the Mann-Whitney statistic is easy to find by using the relationship

$$MW_t = W_t - \frac{t(t+1)}{2}$$

where W_t is the well-known Wilcoxon rank-sum test statistic, that is, $W_t = \sum_{i=1}^{t} R_i$ and $R_1, R_2,..., R_t$ are the ranks of the t observations $x_1, x_2,..., x_t$ in the complete sample of n observations. The expected value and variance of W_t is given by (see Gibbons and

* The MW_t statistic is directly related to the well-known Mann-Whitney U test statistic (see Gibbons and Chakraborti (2003)) where the Mann-Whitney U test statistic is defined as the number of times Y precedes X in the combined ordered arrangement of the two samples, $X_1, X_2,..., X_m$ and $Y_1, Y_2,..., Y_n$, into a single sequence of $N = m+n$ variables. Then the U test statistic is defined as $U = \sum_{i=1}^{m}\sum_{j=1}^{n} D_{ij}$ where $D_{ij} = 1, 0$ if $Y_j < X_i$, $Y_j > X_i$ $\forall i, j$.

Chakraborti (2003)) $E(W_t)=\frac{t(n+1)}{2}$ and $\text{var}(W_t)=\frac{t(n-t)(n+1)}{12}$. As a result, the expected value, variance and standard deviation of MW_t is given by

$$E(MW_t)=E\left(W_t-\frac{t(t+1)}{2}\right)=E(W_t)-\frac{t(t+1)}{2}=\frac{t(n-t)}{2},$$

$$\text{var}(MW_t)=\text{var}\left(W_t-\frac{t(t+1)}{2}\right)=\frac{t(n-t)(n+1)}{12}, \text{ and}$$

$$stdev(MW_t)=\sqrt{\text{var}(MW_t)}=\sqrt{\frac{t(n-t)(n+1)}{12}}.$$

It follows that the standardized value of MW_t is given by

$$SMW_t=\frac{MW_t-E(MW_t)}{stdev(MW_t)}=\frac{MW_t-\frac{t(n-t)}{2}}{\sqrt{\frac{t(n-t)(n+1)}{12}}}. \qquad (6.21)$$

If all X_i-observations $(i=1,2,...,t)$ are smaller than the X_j-observations $(j=t+1, t+2,...,n)$, MW_t would be equal to zero. On the other hand, if all X_i-observations are greater then the X_j-observations, MW_t would equal $t\times(n-t)$. Therefore we have that $0\le MW_t\le t(n-t)$.

If the process is in-control, the distribution of MW_t is symmetric about its mean, $\frac{t(n-t)}{2}$ for each *t*, and large values of MW_t, that is, if a large number of X_i-observations are greater than the X_j-observations, would be indicative of a negative shift, whereas small values of MW_t would be indicative of a positive shift. If there are ties present, i.e. if any $X_i=X_j$, then recall that for a continuous random variable the probability of any particular value is zero; thus, $P(X=a)=0$ for any a. Since the distribution of the observations is assumed to be continuous, $P(X_i-X_j=0)=0$. Theoretically, ties should occur with zero probability, but in practice ties do occur. In case of the occurrence of ties, a correction to the variance of MW_t can be made by multiplying the variance by the factor $1-\frac{\sum_{i=1}^{r}g_i(g_i^2-1)}{n(n^2-1)}$ where g_i denotes the frequency of the i^{th} value and *r* denotes the distinct number of values in

the total of n observations, respectively. Since the sum over all frequencies equal n we have that $\sum_{i=1}^{r} g_i = n$. Consequently, the variance of MW_t (which is also the variance of W_t) is given by

$$\text{var}(MW_t) = \text{var}(W_t) = \left(\frac{t(n-t)(n+1)}{12}\right)\left(1 - \frac{\sum_{i=1}^{r} g_i(g_i^2 - 1)}{n(n^2-1)}\right).$$

The charting statistic for the proposed CUSUM Mann-Whitney chart is obtained by replacing y_i by SMW_t in the equations for the classic standardized CUSUM chart (these equations are given in Section 2.3 numbers (2.35), (2.36) and (2.37)).

The resulting upper one-sided CUSUM is given by

$$S_i^+ = \max[0, S_{i-1}^+ + SMW_i - k] \quad \text{for } i = 1,2,3,... \tag{6.22}$$

while the resulting lower one-sided CUSUM is given by

$$S_i^- = \min[0, S_{i-1}^- + SMW_i + k] \quad \text{for } i = 1,2,3,... \tag{6.23}$$

or

$$S_i^{-*} = \max[0, S_{i-1}^{-*} - SMW_i - k] \quad \text{for } i = 1,2,3,... \tag{6.24}$$

The two-sided CUSUM is constructed by running the upper and lower one-sided CUSUM charts simultaneously and signals at the first i such that $S_i^+ \geq h$ or $S_i^- \leq -h$. The starting values, S_0^- and S_0^+, are typically set equal to zero, that is, $S_0^- = 0$ and $S_0^+ = 0$.

6.2.2. Determination of chart constants

ZZW take the reference value, k, to equal 2. They motivate their choice of k by stating that smaller values of k lead to quicker detection of smaller shifts. Their simulation studies also support the decision of setting $k = 2$, since the simulation results show that the corresponding control chart has good performance. The decision interval, h, is chosen such that a desired *FAP*, denoted by α, is attained. ZZW considered h for various combinations of α and n. The table is given below for reference.

Table 6.6. Simulated h values for the CUSUM Mann-Whitney chart[*].

	α							
n	**0.07**	**0.06**	**0.05**	**0.04**	**0.03**	**0.02**	**0.01**	**0.0075**
20	0.753	0.885	1.053	1.306	1.656	2.194	3.247	3.671
30	1.267	1.490	1.788	2.187	2.719	3.615	5.531	6.371
40	1.774	2.111	2.525	3.081	3.882	5.258	7.612	9.134
50	2.362	2.779	3.329	4.102	5.194	6.988	10.236	11.993
60	2.940	3.454	4.124	4.989	6.328	8.401	12.480	14.147

From Table 6.6 we observe that h increases as n increases and α decreases. As pointed out by Sullivan and Woodall (1996), it is not important for the preliminary application to find exact control limits that correspond to a specific *FAP*. Instead it is sufficient to use computationally convenient limits having approximately the desired performance. Consequently, ZZW derived a formula to approximate the decision interval h:

$$h = (-0.0905n + 0.5221)\log\alpha - 0.1923n + 1.0248. \qquad (6.25)$$

Using equation (6.25) to approximate the decision interval we obtain the following values for h.

Table 6.7. Approximated h values for the CUSUM Mann-Whitney chart[†].

	α							
n	**0.07**	**0.06**	**0.05**	**0.04**	**0.03**	**0.02**	**0.01**	**0.0075**
20	0.604	0.802	1.037	1.324	1.695	2.217	3.110	3.480
30	1.087	1.425	1.825	2.314	2.945	3.834	5.354	5.985
40	1.571	2.048	2.613	3.305	4.196	5.452	7.599	8.490
50	2.055	2.672	3.401	4.295	5.446	7.069	9.844	10.995
60	2.538	3.295	4.190	5.285	6.697	8.687	12.089	13.500

Comparing the approximated h values with the simulated results in Table 6.6 it is clear that the approximated decision interval using equation (6.25) performs very well as they agree well with the values of Table 6.6.

6.2.3. Performance comparison

The performance of a control chart is usually judged in terms of certain characteristics associated with its run-length distribution. In a Phase I setting, the *FAP*, which is the probability of at least one false alarm out of many comparisons, is used for performance comparison as opposed to using the *FAR*, which is the probability of a single false alarm

[*] Table 6.6 appears in ZZW, page 5, Table 1.
[†] The values in Table 6.7 were generated using Microsoft Excel.

involving only a single comparison. By using the *FAP* we take into account that the signaling events are dependent. ZZW looked at the *FAP*, the true signal probability (*TSP*) and the average true signal probability (*ATSP*) for performance comparison*. The *TSP* is the probability of a signal when a shift has occurred. The *ATSP* is defined as $ATSP = \sum_{k=1}^{n-1} F(k)TSP_k$ where TSP_k denotes the *TSP* of a control chart when the shift occurs after the k^{th} observation and $F(k)$ denotes the distribution of the position of the shifts. To ensure fair comparison between two charts, charts with the same *FAP* are considered and the chart with the larger *TSP* (or *ATSP*) is the preferred chart. In their paper, ZZW assumes that the position of the shift is uniformly distributed so that the position of the shift is equally likely at any point and, under this assumption, the CUSUM Mann-Whitney chart, the CUSUM chart for detecting the linear trend (CUSUM LT) chart (see Koning and Does (2000)) and the likelihood ratio test (LRT) chart (see Sullivan and Woodall (1996)) are compared. For the LRT and CUSUM LT charts the assumption of normality is necessary, whereas with the CUSUM Mann-Whitney chart no assumption about the underlying process distribution needs to be made. The performances of these charts are compared for five distributions, namely the Normal, Chi-square, Student *t*, Weibull and Lognormal, respectively. We would expect to find that the CUSUM Mann-Whitney chart performs better compared to the CUSUM LT and LRT charts when the distribution is skewed or heavy-tailed.

(i) The standard normal distribution

For the standard normal distribution both a step shift and a linear trend shift are used to evaluate chart performance. Recall that charts with the same *FAP* ($= 0.05$) are considered to ensure fair comparison and that the chart with the larger *ATSP* is the preferred chart.

* The terms *TSP* and *ATSP* are fairly new and are introduced by Sullivan and Woodall (1996).

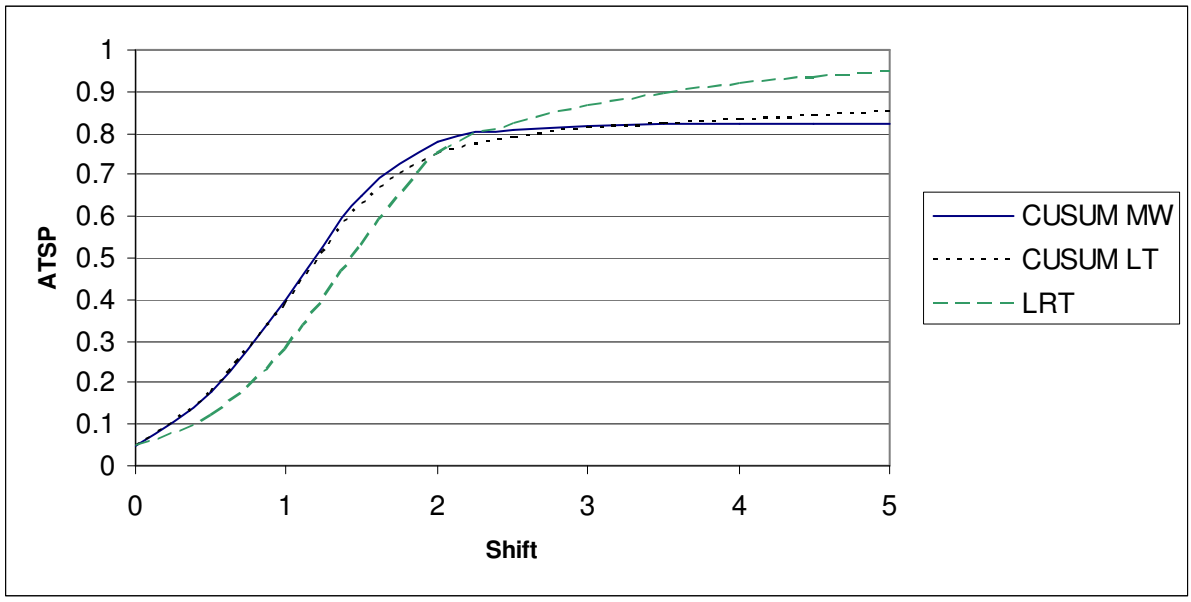

Figure 6.6. The *ATSP* values for a single step shift when the data is from a $N(0,1)$ distribution.

When comparing the CUSUM Mann-Whitney chart with the CUSUM LT chart we find that these charts have comparable performance. When comparing all three charts we find that the CUSUM Mann-Whitney chart has a slight disadvantage in detecting large shifts.

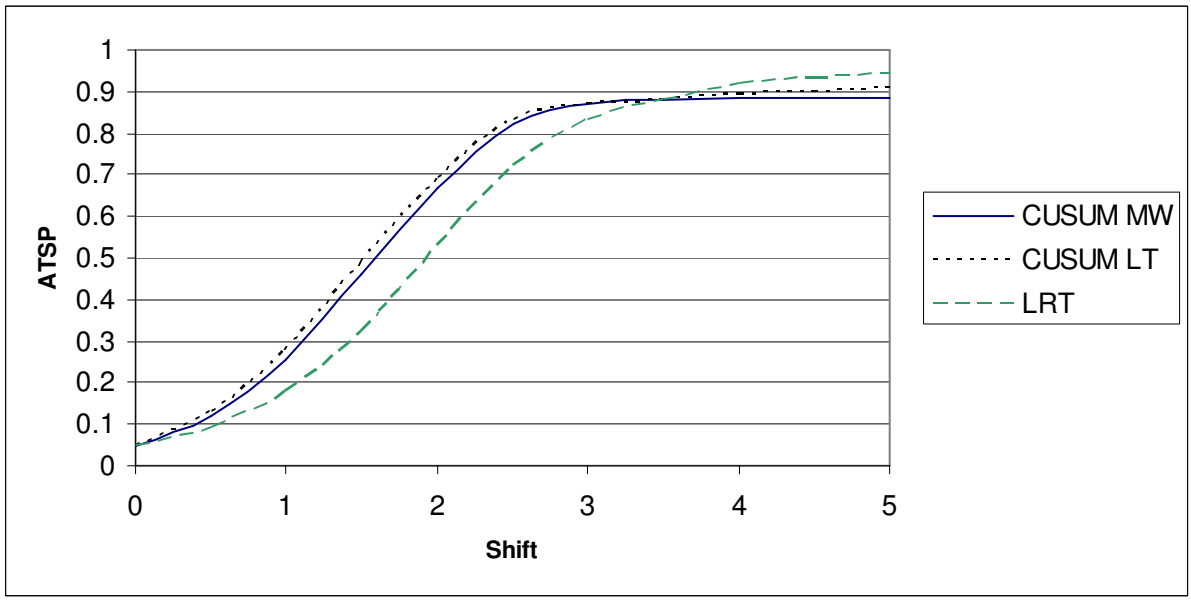

Figure 6.7. The *ATSP* values for a linear trend shift when the data is from a $N(0,1)$ distribution.

When comparing the CUSUM Mann-Whitney chart with the CUSUM LT chart we find that the CUSUM LT chart is performing slightly better than the CUSUM Mann-Whitney chart. When comparing all three charts we find that the CUSUM Mann-Whitney chart has a slight disadvantage in detecting large shifts. It is worth mentioning that even for normally distributed data the CUSUM Mann-Whitney chart is performing very well. The performance of the CUSUM Mann-Whitney chart could be improved by changing the reference value to some other value (recall that ZZW set the reference value equal to 2).

(ii) The *t*-distribution

The *t*-distribution with degrees of freedom 2 is symmetric around zero and as the number of degrees of freedom increases, the difference between the *t*-distribution and the standard normal distribution becomes smaller and smaller.

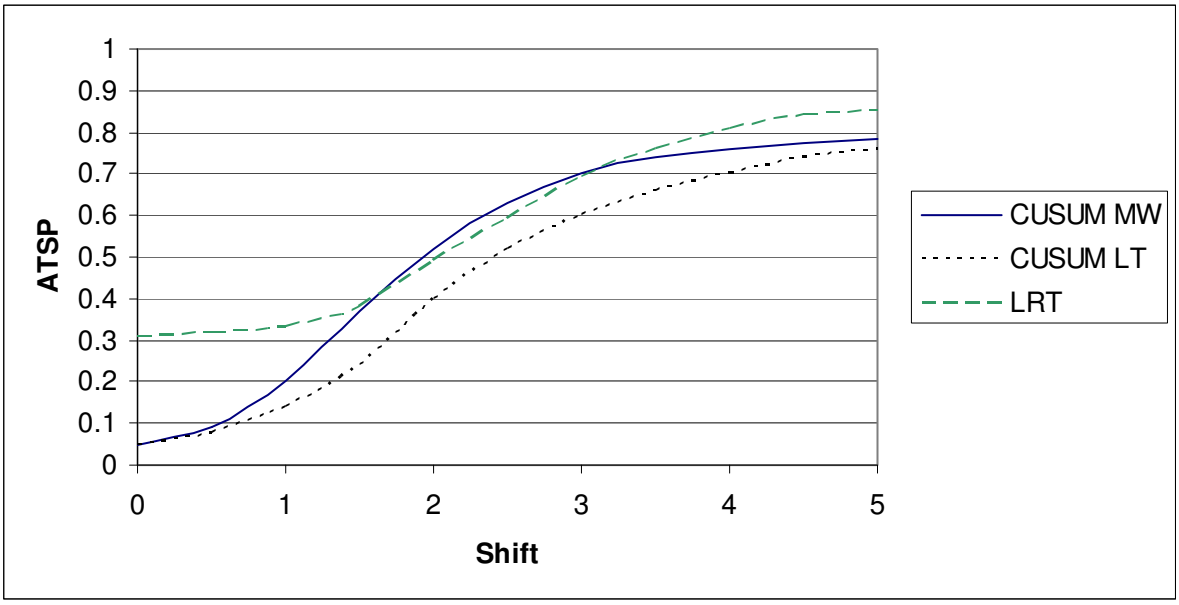

Figure 6.8. The *ATSP* values for a single step shift when the data is from a $t(2)$ distribution.

Recall that charts with the same *FAP* $(=0.05)$ are considered to ensure fair comparison and that the chart with the larger *ATSP* is the preferred chart. From Figure 6.8 we can see that the LRT chart can not obtain the specified *FAP* of 0.05. Consequently, the LRT chart is not compared to the other charts under $t(2)$ shift alternatives. When comparing the CUSUM Mann-Whitney chart with the CUSUM LT chart we find that the CUSUM Mann-Whitney chart is performing better than the CUSUM LT chart, since the *ATSP* values for the

CUSUM Mann-Whitney chart are larger than that of the CUSUM LT chart. It should be noted that the differences are relatively small over all values of the shift.

(iii) The Chi-square distribution

The Chi-square distribution is highly skewed to the right and as a result we would expect the performance of the CUSUM Mann-Whitney chart to be better than that of the CUSUM LT and LRT charts (since they have the additional assumption of normality).

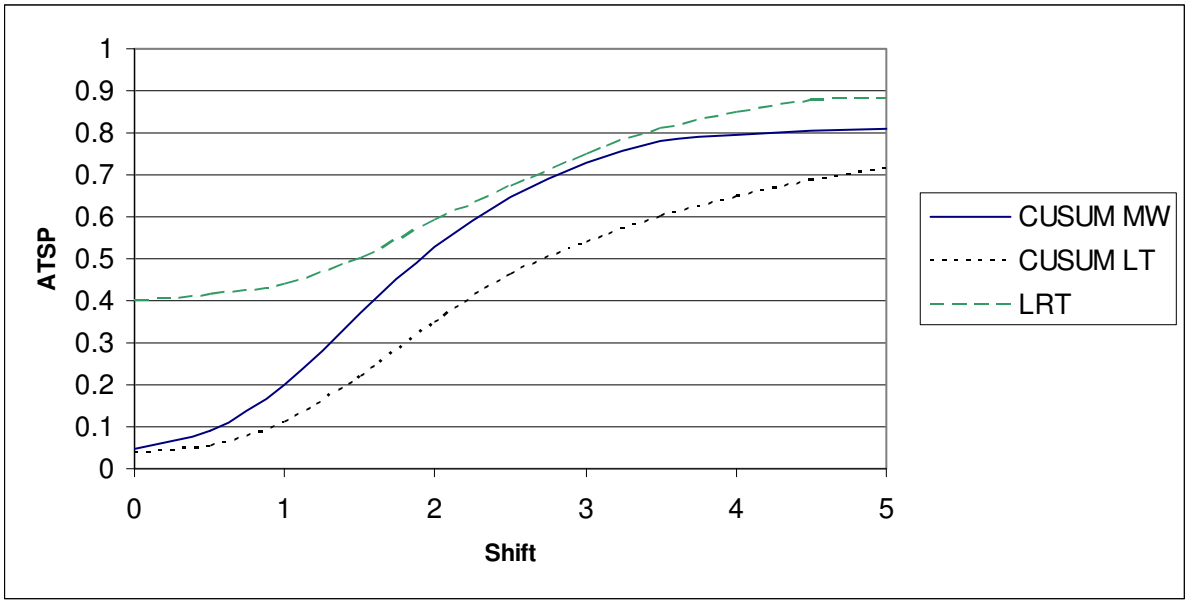

Figure 6.9. The *ATSP* values for a single step shift when the data is from a $\chi^2(2)$ distribution.

Similar to the previous comparison, we see that the LRT chart can not obtain the specified *FAP* of 0.05. Consequently, the LRT chart is not compared to the other charts under $\chi^2(2)$ shift alternatives. When comparing the CUSUM Mann-Whitney chart with the CUSUM LT chart we find that the CUSUM Mann-Whitney chart is performing better than the CUSUM LT chart, since the *ATSP* values for the CUSUM Mann-Whitney chart are larger than that of the CUSUM LT chart. It should be noted that the differences are larger than those under $t(2)$ shift alternatives.

(iv) The Weibull distribution

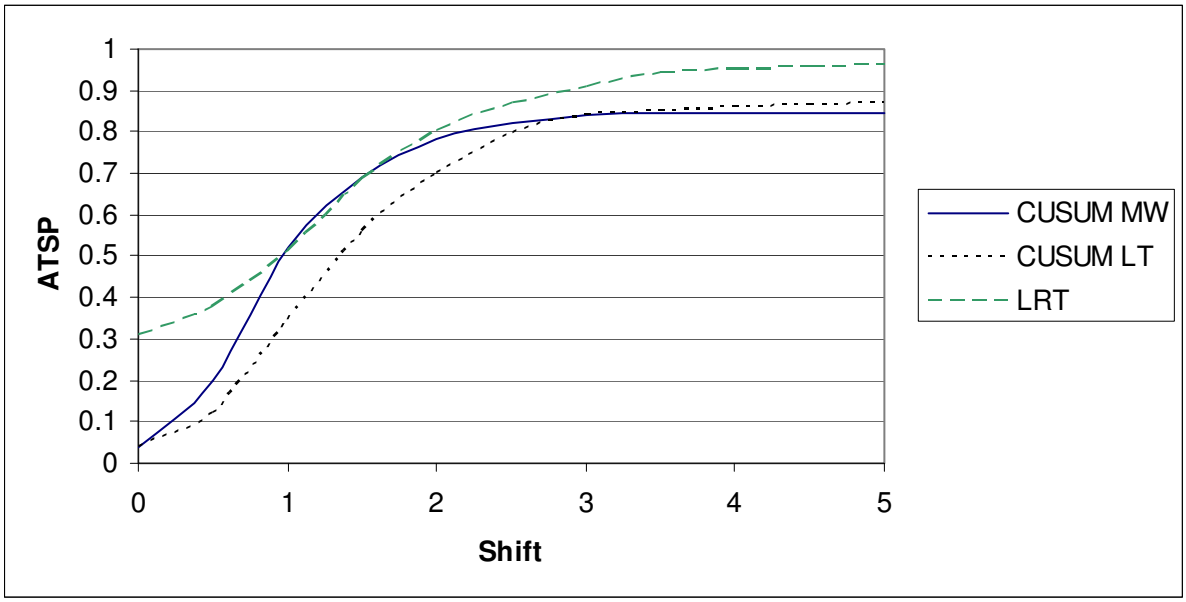

Figure 6.10. The *ATSP* values for a single step shift when the data is from a Weibull(1,1) distribution.

Similar to the previous two comparisons, we see that the LRT chart can not obtain the specified *FAP* of 0.05. Consequently, the LRT chart is not compared to the other charts under Weibull(1,1) shift alternatives. When comparing the CUSUM Mann-Whitney chart with the CUSUM LT chart we find that the CUSUM Mann-Whitney chart is performing better than the CUSUM LT chart for small shift sizes, whereas, for large shift sizes the opposite is true, although to a very small extent.

(v) The lognormal distribution

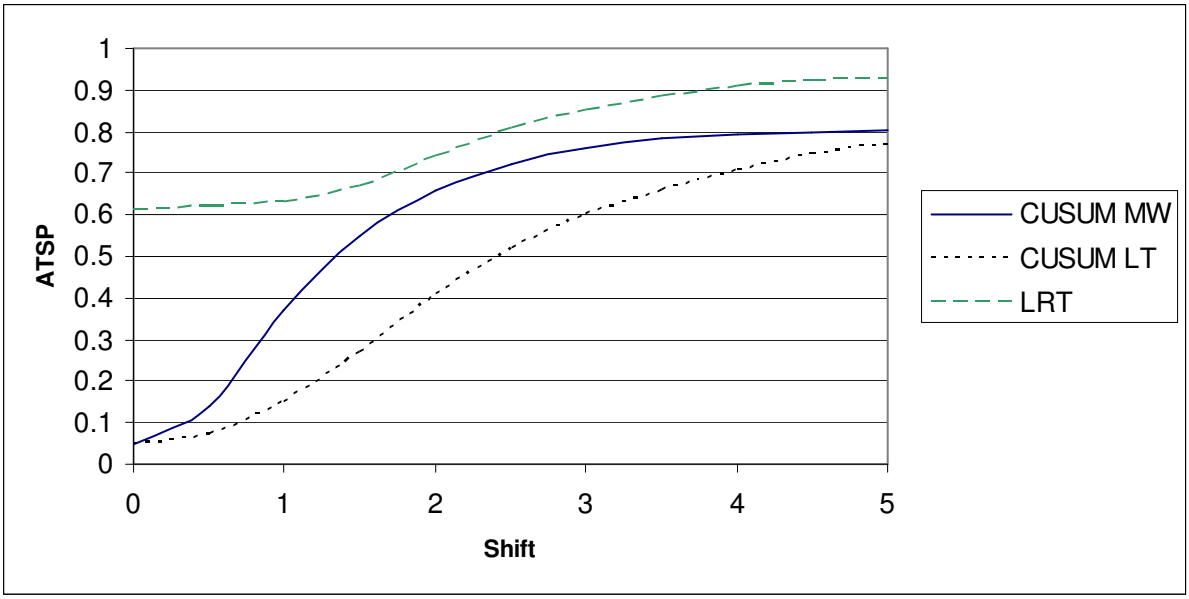

Figure 6.11. The *ATSP* values for a single step shift when the data is from a lognormal(0,1) distribution.

Similar to the previous three comparisons, we see that the LRT chart can not obtain the specified *FAP* of 0.05. Consequently, the LRT chart is not compared to the other charts under lognormal(0,1) shift alternatives. When comparing the CUSUM Mann-Whitney chart with the CUSUM LT chart we find that the CUSUM Mann-Whitney chart is performing better than the CUSUM LT chart, since the *ATSP* values for the CUSUM Mann-Whitney chart are larger than that of the CUSUM LT chart.

Table 6.8. A summary of the performances of the CUSUM Mann-Whitney, CUSUM LT and LRT charts for five different distributions.

Distribution	**Type of shift**	**Preferred control chart***
Normal(0,1)	Linear trend shift	For shifts < 3.5: 1) CUSUM LT 2) CUSUM MW 3) LRT For shifts > 3.5: 1) LRT 2) CUSUM LT 3) CUSUM MW
Normal(0,1)	Single step shift	For shifts < 2.5: 1) CUSUM MW or CUSUM LT (comparable performance) 2) LRT For shifts > 2.5: 1) LRT 2) CUSUM LT or CUSUM MW (comparable performance)
$t(2)$	Single step shift	1) CUSUM MW 2) CUSUM LT
$\chi^2(2)$	Single step shift	1) CUSUM MW 2) CUSUM LT
Weibull(1,1)	Single step shift	For shifts < 3.0: 1) CUSUM MW 2) CUSUM LT For shifts > 3.0: 1) CUSUM LT 2) CUSUM MW
Lognormal(0,1)	Single step shift	1) CUSUM MW 2) CUSUM LT

Example 6.2

A CUSUM Mann-Whitney control chart

We illustrate the CUSUM Mann-Whitney control chart using a set of simulated data used by Sullivan and Woodall (1996; Table 2) and ZZW (2007; Table 2). This data set is ideal for use in this Phase I problem, since it is known to have a single step shift in the mean. There are 30 observations, i.e. $n = 30$, which are distributed as follows: $X_i \sim N(0,1)$ for $i \leq 15$ and $X_i \sim N(1,1)$ for $i > 15$ (a value of 1 was added to the last 15 observations causing the data

* The control charts are ranked from the most preferred to the least preferred. The LRT chart could not be compared to the CUSUM MW and CUSUM LT charts under $t(2)$, $\chi^2(2)$, Weibull(1,1) and Lognormal(0,1) shift alternatives, since the LRT chart could not obtain the specified *FAP* of 0.05.

to exhibit a step shift in the middle of the sample). Clearly, there is a known change-point at $t = 15$ where the mean has shifted from 0 to 1. The Mann-Whitney statistics (MW_t), the corresponding expected values ($E(MW_t)$), standard deviations ($stdev(MW_t)$) and standardized values (SMW_t), respectively, are given in Table 6.9. The CUSUM S_i^+ and S_i^- values are also given in Table 6.9 and illustrated in Figure 6.12. The starting values are set equal to zero, that is, $S_0^+ = S_0^- = 0$ (as recommended by Page (1954)).

Table 6.9. Data and calculations for the CUSUM Mann-Whitney chart when $k = 2$.*

i	X_i	MW_t	$E(MW_t)$	$stdev(MW_t)$	SMW_t	S_i^+	S_i^-
1	-0.69	6	14.5	8.655	-0.982	0.000	0.000
2	0.56	20	28.0	12.028	-0.665	0.000	0.000
3	-0.96	23	40.5	14.465	-1.210	0.000	0.000
4	-0.11	29	52.0	16.391	-1.403	0.000	0.000
5	-0.25	33	62.5	17.970	-1.642	0.000	0.000
6	0.45	41	72.0	19.287	-1.607	0.000	0.000
7	-0.26	42	80.5	20.394	-1.888	0.000	0.000
8	0.68	54	88.0	21.323	-1.595	0.000	0.000
9	0.22	57	94.5	22.096	-1.697	0.000	0.000
10	-2.10	49	100.0	22.730	-2.244	0.000	-0.244
11	0.65	56	104.5	23.236	-2.087	0.000	-0.331
12	-1.49	47	108.0	23.622	-2.582	0.000	-0.913
13	-2.49	35	110.5	23.894	-3.160	0.000	-2.073
14	-1.11	25	112.0	24.055	-3.617	0.000	-3.690
15	0.23	23	112.5	24.109	-3.712	0.000	-5.402
16	2.16	35	112.0	24.055	-3.201	0.000	-6.603
17	1.95	45	110.5	23.894	-2.741	0.000	-7.344
18	1.54	52	108.0	23.622	-2.371	0.000	-7.715
19	0.67	52	104.5	23.236	-2.259	0.000	-7.974
20	1.09	54	100.0	22.730	-2.024	0.000	-7.998
21	1.37	56	94.5	22.096	-1.742	0.000	-7.740
22	0.69	55	88.0	21.323	-1.548	0.000	-7.288
23	2.26	61	80.5	20.394	-0.956	0.000	-6.244
24	1.86	63	72.0	19.287	-0.467	0.000	-4.711
25	0.62	55	62.5	17.970	-0.417	0.000	-3.128
26	-1.04	34	52.0	16.391	-1.098	0.000	-2.226
27	2.30	37	40.5	14.465	-0.242	0.000	-0.468
28	0.07	20	28.0	12.028	-0.665	0.000	0.000
29	1.49	15	14.5	8.655	0.058	0.000	0.000
30	0.52						

* See SAS Program 9 in Appendix B for the calculation of the values in Table 6.9. This table also appears in ZZW, page 7, Table 2.

Table 6.9 also appears in ZZW, page 7, Table 2. It should be noted that the CUSUM S_i^- values that we obtained in Table 6.9 are different from those in ZZW, since they used equation (6.24) to calculate S_i^-, whereas we used equation (6.23).

As illustration, the expected value ($E(MW_t)$), standard deviation ($stdev(MW_t)$), standardized value (SMW_t), CUSUM S_i^+ and S_i^- values will be calculated for $t = 1$.

$$E(MW_1) = \frac{1(30-1)}{2} = 14.5, \quad stdev(MW_t) == \sqrt{\frac{1(30-1)(30+1)}{12}} = 8.655,$$

$$SMW_t = \frac{MW_t - E(MW_t)}{stdev(MW_t)} = \frac{6-14.5}{8.655} = -0.982,$$

$$S_1^+ = \max[0, S_0^+ + SMW_1 - k] = \max[0, 0 + (-0.982) - 2] = 0,$$

$$S_1^- = \min[0, S_0^- + SMW_1 + k] = \min[0, 0 + (-0.982) + 2] = 0.$$

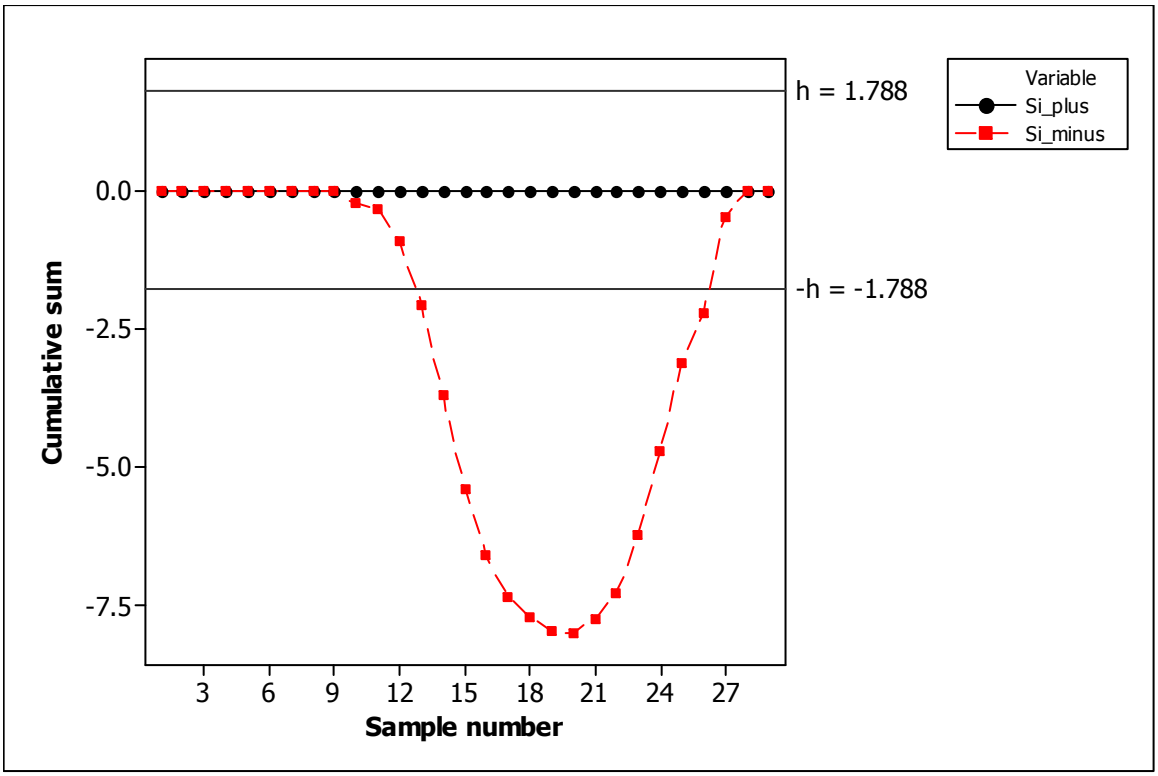

Figure 6.12. The CUSUM Mann-Whitney chart with $n = 30$, $k = 2$ and $h = 1.788$.

For a sample size of 30 and a desired *FAP* of 0.05, the decision interval is taken to be 1.788 (see Table 6.6). From Figure 6.12 we see that the process is out-of-control starting at sample number 13 using the CUSUM Mann-Whitney chart, whereas the LRT chart of

Sullivan and Woodall (1996) indicated that observation 15 is the most likely location of the shift. Hence, the CUSUM Mann-Whitney chart detects that the mean has shifted upwards.

6.2.4. Summary

ZZW found that the Phase I CUSUM Mann-Whitney chart has good performance compared to the CUSUM LT chart for all distributions, except for the Weibull distribution for large shifts. Their proposed nonparametric chart for preliminary analysis can be useful for quality practitioners in applications where not much is known or can be assumed about the process distribution. Although a lot has been accomplished in the last few years regarding the development of control charts based on the Mann-Whitney statistic, more remains to be done. In terms of research, work needs to be done on a Phase II CUSUM-type chart based on the Mann-Whitney statistic for individual observations and subgroups (recall that the control chart proposed by ZZW is a Phase I CUSUM-type chart for preliminary analysis of individual observations). Also recall that ZZW assumes that the position of the shift is uniformly distributed. One could, for future research, consider other distributions for the position of the shift. Furthermore, work needs to be done on Phase I and Phase II EWMA-type charts based on the Mann-Whitney statistic. Clearly, there are lots of opportunities for future research.

Chapter 7: Concluding remarks

In this thesis, we mentioned some of the key contributions and ideas and a few of the more recent developments in the area of univariate nonparametric control charts. We considered the three main classes of control charts: the Shewhart, CUSUM and EWMA control charts and their refinements. 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. We described the sign and signed-rank control charts under each of the three classes in Chapters 2 and 3, respectively. In Chapter 4 we only considered the Shewhart-type sign-like control chart, since the CUSUM- and EWMA-type control charts have not been developed for the sign-like case. In Chapter 5 we only considered the Shewhart-type signed-rank-like control chart, since the CUSUM- and EWMA-type control charts have not been developed for the signed-rank-like case. Finally, in Chapter 6 we only considered the Shewhart- and CUSUM-type Mann-Whitney-Wilcoxon control charts, since the EWMA-type control chart has not been developed for the Mann-Whitney-Wilcoxon statistic. Clearly, there are lots of opportunities for future research.

We considered decision problems under both Phase I and Phase II (see Section 1.5 for a distinction between the two phases). In all the sections of this thesis we considered Phase II process monitoring, except in Section 6.2 where a CUSUM-type control chart for the preliminary Phase I analysis of individual observations based on the Mann-Whitney two-sample test is proposed. Although the field of preliminary Phase I analysis is interesting and the body of literature on Phase I control charts is growing, more research is necessary on Phase I nonparametric control charts in general.

We only discussed univariate nonparametric control charts designed to track the location of a continuous process, since very few charts are available for scale. Therefore, future research needs to be done on monitoring the scale and simultaneously monitoring the location and the scale of a process.

There has been other work on nonparametric control charts. Among these, for example, Albers and Kallenberg (2004) studied conditions under which the nonparametric

charts become viable alternatives to their parametric counterparts. They consider Phase II charts for individual observations in case U based on empirical quantiles or order statistics. The basic problem is that for the very small *FAR* typically used in the industry, a very large reference sample size is usually necessary to set up the chart. They discuss various remedies for this problem.

Another area that has received some attention is control charts for variable sampling intervals (VSI). In a typical control charting environment, the time interval between two successive samples is fixed, and this called a fixed sampling interval (FSI) scheme. VSI schemes allow the user to vary the sampling interval between taking samples. This idea has intuitive appeal since when one or more charting statistics fall close to one of the control limits but not quite outside, it seems reasonable to sample more frequently, whereas when charting statistics plot closer to the centerline, no action is necessary and only a few samples might be sufficient. On the point of VSI control schemes see for example, Amin (1987), Reynolds et al (1990), Rendtel (1990), Saccucci et al (1992) and Amin and Hemasinha (1993). These researchers examined combining the VSI approach with the Shewhart, CUSUM and the EWMA control schemes, respectively. They demonstrated that the VSI control schemes are more efficient than the corresponding FSI control schemes. VSI control schemes use a long sampling interval between successive samples when the plotting statistic is close to target and a shorter sampling interval otherwise. Initially, the short sampling interval could be used for the first few samples to offer protection at start-up. Amin and Widmaier (1999) compared the Shewhart $\overline{X}$ charts with sign control charts, under the FSI and VSI schemes, on the basis of *ARL* for various shift sizes and several underlying distributions like the normal distribution and distributions that are heavy-tailed and/or asymmetric like the double exponential and the gamma. It is seen that the nonparametric VSI sign charts are more efficient than the corresponding FSI sign charts.

We hope this thesis leads to a wider acceptance of nonparametric control charts among practitioners and promotes further interest in the development of nonparametric control charts.