Chapter 2

Univariate and Multivariate Control Charts

2.1 Introduction

Control charts are one of the main tools of statistical process control. The literature on control charts is huge. In this chapter we try to present the main univariate and multivariate control charts along with their basic properties. We have to emphasize that the control charts presented in this chapter and their properties are by no means a detailed review of all charts.

In Section 2.2, we present the main characteristics of a control chart and a discussion of its evaluation using the most known measures. Univariate Shewhart Control Charts for data in subgroups and individual data for the mean and the variance are given in Section 2.3 for both variables and attributes. Cumulative Sum (CUSUM) charts for the mean and variance and their properties are described in Section 2.4. Exponentially Weighted Moving Average (EWMA) Control Charts are summarized in Section 2.5. Section 2.6 presents the multivariate Shewhart control charts for the process mean and dispersion for variables and attributes. Finally, Section 2.7 gives the multivariate CUSUM and EWMA control charts.

2.2 The fundamental characteristics of a control chart

When we have a production process there is usually a target value. We want our process to achieve this target for every product. However, in every process there is an inherent random variability. Therefore, no matter how good we design the whole procedure or how accurate our engines are, we expect to be close to the target value but not always on this value. The existence of this variability affects our process.

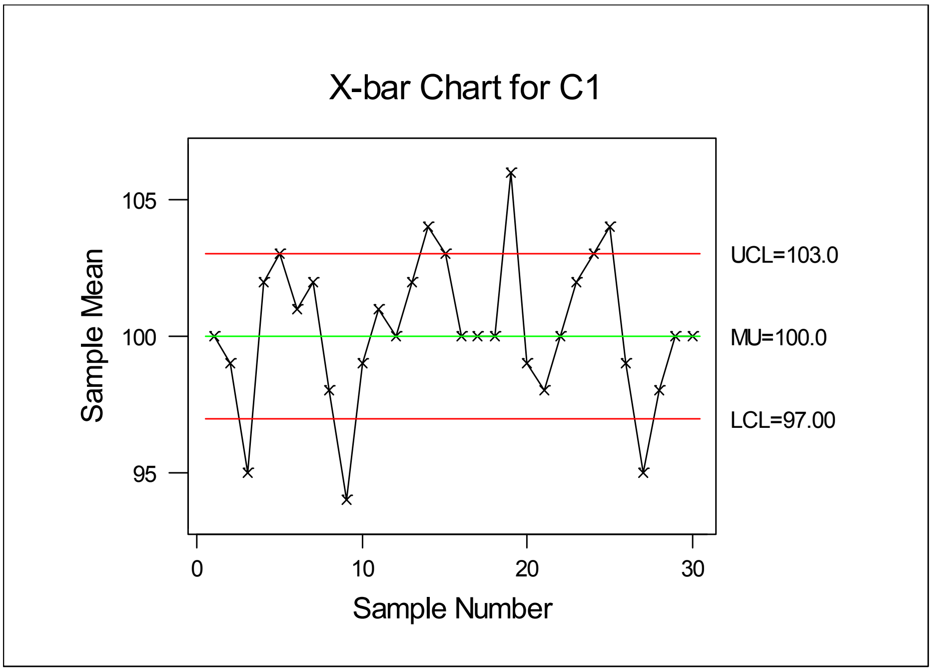
There are two different "versions" of this variability. The common cause (chance cause) variability is the natural variability every process experiences. Its existence is due to randomness as we can find purely random variability from one product to another. A process that operates with only common cause variability is said to be in-control. The special cause (assignable cause) variability is a result of factors that are not purely random. These factors cause heterogeneity in the process and as a result they affect it, leading to low quality product. A process that operates in the presence of special causes of variability is said to be out-of-control. This type of variability can be detected with control charts giving us the ability to remove its effect and therefore reduce the overall variability. As a result, removing special causes leads to an improvement of the quality of the product.

Common cause variability is the remainder of the variability after every component of special cause has been removed. In order to remove common cause variability we have to alter the process itself. However, a goal in today's industrial and technological world must be the continuous quality improvement. Under this perspective we have to stress that today's common cause can be a tomorrow's special cause. As the inspection process improves and the target for quality is constant we may be able to identify as special cause, a up to now identified common cause.

Special causes of variability can be divided in two different groups; transient special causes and persistent special causes. Transient special causes are those causes that affect a process for a short time until their reappearance in a future point in time. Persistent special causes are those causes that when they occur they stay in the process until they are detected and removed.

A control chart is a graphical representation of a characteristic of the process under investigation. It is used as the main tool to identify special causes of variability in a process. On the horizontal axis we have the number of the sample drawn from the process or the time that the sample was inspected. On the vertical axis we have the value of the characteristic measured for each sample or for the time of the horizontal axis. A straight line connects the successive points indicating the level of the characteristic in time or in successive samples. There are also three usually straight lines that stand for the upper control limit (UCL) the center line (CL) and the lower control limit (LCL). An example of a control chart is given in Figure 2.1.

Figure 2.1. A typical control chart



We assume that a process operates under control when the line connecting the sequence of points does not cross UCL or LCL. When a point is plotted outside these limits we assume that the process is in an out-of-control state and corrective actions must be taken in order to remove the assignable cause that led to this problem. The values of UCL and LCL are chosen usually in such a way that when the process is in-control the probability of a point plotting outside these limits is very small. However, there are some cases that even when all the points plot inside the control limits we characterize the process as being in an out-of-control state. Such cases are for example when we see a series of nine successive points plotting all above (below) the center line or when we see six successive points in a row steadily increasing or decreasing. We have to state here that the removal of any cause is not the objective of a control chart. A control chart simply indicates that an assignable cause may exist. It is the management's or the operator's job to act in order to get rid of the problem, if it exists.

In the literature, two distinct phases of control charting practice have been discussed (see, e.g. Woodall (2000)). In Phase I, charts are used for retrospectively testing whether the process was in-control when the first subgroups were being drawn. In this phase, the charts are used as aids to the practitioner, in bringing a process into a state of statistical control. Once this is accomplished, the control chart is used to define what is meant by statistical control. This is referred as the retrospective use of control charts. In general, there is a lot more going on in this phase than just charting some data. During this phase the practitioner is studying the process very intensively. The data collected are then analyzed in an attempt to answer the question "were the data collected from an in-control process?".

In Phase II, control charts are used for testing whether the process remains in-control when future subgroups are drawn. In this phase, the charts are used for monitoring the process for any change from an in-control state. At each sampling stage, the practitioner asks the question "has the state of process changed?". The meaning of in-control, in this phase, is usually determined by the values of the process parameters e.g., the mean and standard deviation for univariate continuously distributed variables. The values of the parameters are either given to the practitioner or they are estimated from the historical data known to be under control from Phase I. Note that in this phase the data is not taken as being from an in-control process unless the data provide evidence against no change in the process. Using these data to define what is meant by the process being in-control might lead to use an out-of-control process to define a state of statistical in-control. Woodall (2000) states that much work, process understanding and process improvement is often required in the transition from Phase I to Phase II.

In a control chart we have two objectives. Firstly, when a process is in-control, we want our chart to signal (false alarm) infrequently. In statistical terms we want the chart to operate with the planned probability of the statistic computed to plot outside the control limits if we are in-control. Secondly, when a process is out-of-control, we want the chart to signal as soon as possible. In statistical terms we want the probability of the statistic computed to plot in-control if we are out-of-control to be as small as possible. Different measures for evaluating the performance of a chart, concerning the previous two objectives, have been proposed. The most known measure is the average run length (ARL), which is based on the run length (RL) distribution. The number of observations (individual data), or samples (data in subgroups), needed for a control chart to signal is a run length or alternatively one observation of the RL distribution. The mean of the RL distribution is the ARL, which is actually the average number of observations needed for a control chart to signal. Page (1954) defined the average run length as follows: When the quality remains constant the average run length of a process inspection scheme is the expected number of articles sampled before action is taken. Ewan and Kemp (1960) gave a somewhat different definition; When the quality remains constant the average run length of an inspection scheme is the expected number of samples obtained before action is taken. Usually, along with the ARL, the standard deviation of the run length (SDRL) is computed. Alternatively, the ARL is expressed as the average number of observations to signal (ANOS). A measure similar to the ARL is the average time to signal (ATS), which is the average time needed for a control chart to signal and it is actually a product of the ARL and the sampling interval used in the case of fixed sampling.

From the preceding discussion we see that all these measures are related to the ARL. However the sole use of the ARL has been criticized (see, e.g. Barnard (1959), Bissell (1969), Woodall (1983) and Gan (1993b, 1994)).The disadvantage of the ARL is the skewness of the run length distribution in the out-of-control case and in non-normality and as a result the misleading conclusions one can draw based on the ARL. An alternative measure is the median run length (MRL), which is more credible since it is less affected by the skewness (see, e.g. Gan (1993b, 1994)).

A typical method of comparing control charts is based on the calculation of their average run length (ARL) (Woodall (1985)). Assume that independent random samples of size n are drawn successively from a process that measure the quality of a characteristic. Assume also that the sample means $\overline{x}_1, \overline{x}_2, ...$ are normally distributed with known variance σ^2/n. Consider as the objective of the control chart to keep the in-control mean equal to the target value μ_0. If $E(\overline{x}_i) = \mu, i = 1, 2, ...$ the parameter $\theta = \sqrt{n}\left|\mu - \mu_0\right|/\sigma$ denotes the shift in the mean measured in units of the standard error of the sample mean. We assume that any shift in the mean away from the target value occurs prior to the implementation of the control chart.

Let M_0, M_1 denote the in-control and out-of-control regions respectively. The in-control region M_0 contains all values of θ that correspond to acceptable shifts. Although "acceptable shifts" is an oxymoron, there is a meaningful explanation. When the shift in a process is very slight the attempt to adjust the process can lead to over-correction and introduce extra variability into the process. Duncan (1974) and Wetherill (1977) observe that low ARL values for small deviations from the target value is a drawback when some slack in the process is acceptable. The out-of-control region M_1 contains all values of θ for which a control procedure should give an out-of-control signal.

A control chart has an ARL value of at least L_0, when $\theta \in M_0$, and at most L_1, when $\theta \in M_1$. Consider two procedures A, B that are to be compared. If the ARL profile of A is above that of B for $\theta \in M_0$ and below that of B for $\theta \in M_1$, then procedure A is considered to be uniformly better than procedure B.

2.3 Univariate Shewhart Control Charts

The most known control charts are the Shewhart type control charts. They owe their name to Walter Shewhart who established them in his pioneering work in 1931. They are used to detect transient special causes in a process. This property is the result of the fact that Shewhart Control Charts are memoryless. In the following we present the Shewhart control charts for variables and attributes.

2.3.1 Control charts for the mean for data in subgroups

Assume that we have a variable that is normally distributed with mean μ and standard deviation σ. We assume that μ and σ are both known. Let $x_1, x_2, ..., x_n$ be a sample of n independent and identically distributed observations drawn from our production process. Then the average of this sample $\overline{x}$ is distributed as a normal variable with mean μ and standard deviation $\sigma/\sqrt{n}$. Therefore, we can use as control limits for each sample

$$\begin{aligned} UCL &= \mu + Z_{\alpha/2}\sigma/\sqrt{n} \\ LCL &= \mu - Z_{\alpha/2}\sigma/\sqrt{n} \end{aligned} \tag{2.1}$$

where UCL and LCL are the upper and lower control limits respectively, $Z_{\alpha/2}$ is the inverse of the normal cumulative distribution function for probability $\alpha/2$ and α is the probability that an in-control sample will plot outside these limits. If all the points (samples) plot inside the control limits we claim that we have an in-control process. This plot is a Phase II Shewhart chart for the mean.

However, in real world we usually do not know the values of μ and σ. Consequently, we have to estimate them. Therefore, the control limits in such a case will not be fixed numbers, but rather random variables. The control limits in this case for Phase I

Shewhart chart for the mean are

$$\begin{aligned}\widehat{UCL} &= \widehat{\mu} + k\widehat{\sigma}/\sqrt{n} \\ \widehat{LCL} &= \widehat{\mu} - k\widehat{\sigma}/\sqrt{n}\end{aligned}$$

where $\widehat{\mu}$ and $\widehat{\sigma}$ are the estimates for the mean and the standard deviation, respectively and k is a constant used to specify the width of the control limits usually taken to be equal to 3. If a point plots above $\widehat{UCL}$ or below $\widehat{LCL}$ we have an indication that this point (sample) is from an out-of-control process. Let $\overline{x}_1$, $\overline{x}_2$, ..., $\overline{x}_m$ be the sample means from samples each with n observations. Then, an estimate for the mean is $\widehat{\mu} = \overline{\overline{x}}$, the average of all the sample means. If the process is in-control this estimator is normally distributed with mean μ and variance $\sigma^2/(mn)$. For the standard deviation three different estimators have been proposed. The first one is based on the range. Let $R_1, R_2, ..., R_m$ denote the range for each of the m samples and $\overline{R}$ the average of these ranges. Then, a control charts' unbiased estimator is given by $\overline{R}/d_2$. The estimated control limits for the $\overline{X}$ chart are given by

$$\begin{aligned}\widehat{UCL} &= \overline{\overline{x}} + k\overline{R}/(d_2\sqrt{n}) \\ \widehat{LCL} &= \overline{\overline{x}} - k\overline{R}/(d_2\sqrt{n})\end{aligned} \tag{2.2}$$

where d_2 is the mean of the random variable R/σ and is a function of the sample size n. Details on the derivation of d_2 along with its values for different sample sizes can be found in textbooks, see e.g. Montgomery (2001).

A second version of the estimated control limits for the mean is based on a different unbiased estimator for the standard deviation. Let $S_1, S_2, ..., S_m$ denote the standard deviation for each of the m samples and $\bar{S} = \frac{1}{m}\sum_{i=1}^{m} S_i$ their average. An unbiased estimator for σ is $\overline{S}/c_4$ (see e.g. Ryan(2000)) where

$$c_4 = \left(\frac{2}{n-1}\right)^{1/2} \frac{\Gamma\left(n/2\right)}{\Gamma\left((n-1)/2\right)}$$

and $\Gamma(\cdot)$ stands for the gamma function, where the gamma function is defined as

$$\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx,\ z > 0.$$

The control limits will be

$$\begin{aligned} \widehat{UCL} &= \overline{\overline{x}} + k\overline{S}/(c_4\sqrt{n}) \\ \widehat{LCL} &= \overline{\overline{x}} - k\overline{S}/(c_4\sqrt{n}) \end{aligned} \tag{2.3}$$

A third version for such type of limits is based on $\overline{U} = \sqrt{\frac{1}{m}\sum_{i=1}^m S_i^2}$, where $\overline{U}/c_{4,m}$ is an unbiased estimator of σ and

$$c_{4,m} = \frac{\sqrt{2}\Gamma((m(n-1)+1)/2)}{\sqrt{m(n-1)}\Gamma(m(n-1)/2}.$$

The control limits using this estimator will be

$$\begin{aligned} \widehat{UCL} &= \overline{\overline{x}} + k\overline{U}/(c_{4,m}\sqrt{n}) \\ \widehat{LCL} &= \overline{\overline{x}} - k\overline{U}/(c_{4,m}\sqrt{n}) \end{aligned}$$

It can be proved for the three different unbiased estimators that $Var\left(\overline{U}/c_{4,m}\right) \leq Var\left(\overline{S}/c_4\right) \leq Var\left(\overline{R}/d_2\right)$ (Derman and Ross (1995)). Therefore, a preferable estimator for σ is $\overline{U}/c_{4,m}$.

Yang and Hillier (1970) proposed a somewhat different Phase I control charts. The control limits for the statistic plotted at time i are not functions of the ith sample. One or more of the other $m-1$ samples are used to estimate μ_0 and σ_0. If $\widehat{\mu}_{(i)}$ and $\widehat{\sigma}_{(i)}$ are the estimators for the mean and standard deviation respectively when only the i sample is removed the control limits will be

$$\begin{aligned}\widehat{UCL} &= \widehat{\mu}_{(i)} + k\widehat{\sigma}_{(i)}/\sqrt{n} \\ \widehat{LCL} &= \widehat{\mu}_{(i)} - k\widehat{\sigma}_{(i)}/\sqrt{n}\end{aligned}$$

Champ and Chou (2003) compared the performance of the two Phase I control charts the one using the m samples and the other using $m-1$ samples and concluded that the one using all samples gives better results.

The ARL for Shewhart charts is given by

$$ARL = 1/\Pr(\text{a point plots outside the control limits}). \tag{2.4}$$

It has to be stressed though that this relationship holds in the case of known parameters. If the parameters are unknown and as a result they have to be estimated a different relationship holds. This matter is studied in detail in chapter 3.

Several authors have dealt with the Shewhart chart for the mean and have proposed improvements or modifications. For instance see Champ and Woodall (1987), Reynolds et al. (1988) and Quesenberry (1995a).

2.3.2 Control charts for the variability for data in subgroups

Assume that we have again a variable from a stable process that is normally distributed with mean μ and standard deviation σ comprising m samples of size n each. We assume that σ is known. In this process we want to keep the variability in-control. Then, it can be proved that the mean and standard deviation of the range of a sample from this process are $E(R) = d_2\sigma$ and $SD(R) = d_3\sigma$ where d_2 and d_3 are functions of the sample size n. Computation of d_2 and d_3 and values for different sample sizes can be found in textbooks, see e.g. Montgomery (2001). Then the Phase II control limits for the variability using the range will be

$$
\begin{aligned}
UCL &= d_2\sigma + kd_3\sigma \\
LCL &= d_2\sigma - kd_3\sigma
\end{aligned}
\tag{2.5}
$$

The value of k is selected in the same way as in the chart for the mean. The most common value is 3. However, the selection in this case is actually approximating the 0.9973 probability limits for the mean when $\alpha = 0.0027$.

The usual design of the R-chart involves control limits that have equal tail probabilities (see, e.g. (2.5)). However, in such a case it is possible to have an interval (σ_1, σ_2) with $\sigma_1 < \sigma_2$ and for each σ in this interval $ARL(\sigma) > ARL(\sigma_0)$, where $ARL(\sigma_0)$ is the in control ARL. Such a chart is called a biased R chart. Champ (2001) showed how to design an ARL unbiased R control chart.

Another way to compute Phase II control limits for the variability is through the standard deviation. It can be proved that $E(S) = c_4\sigma$ and $SD(S) = \sigma\sqrt{1-c_4^2}$. Then the control limits will be

$$
\begin{aligned}
UCL &= \left(c_4 + 3\sqrt{1-c_4^2}\right)\sigma \\
LCL &= \left(c_4 - 3\sqrt{1-c_4^2}\right)\sigma
\end{aligned}
\tag{2.6}
$$

Usually, we do not know the value of σ and therefore we have to estimate it from past data. As in the case of the mean let $R_1, R_2, ..., R_m$ denote the range for each of the m samples and $\overline{R}$ the average of these ranges. An estimate based on the range as already mentioned is

$$\widehat{\sigma}_R = \frac{\overline{R}}{d_2}.$$

Then, the Phase I control limits are

$$
\begin{aligned}
\widehat{UCL} &= \left(1 + \frac{3d_3}{d_2}\right) \bar{R} \\
\widehat{LCL} &= \left(1 - \frac{3d_3}{d_2}\right) \bar{R}.
\end{aligned}
\tag{2.7}
$$

A different estimate used is

$$
\bar{S} = \frac{1}{m} \sum_{i=1}^{m} S_i
$$

where m is the number of past samples used, $S_i^2 = \frac{1}{n-1} \sum_{j=1}^{n} \left(X_j - \bar{X}_i\right)^2$ is the unbiased estimator of σ^2 and n is the sample size. However, we know that S is not an unbiased estimator of σ. It has been proved, as already mentioned, that an unbiased estimate of σ is $\bar{S}/c_4$ and that the standard deviation of S equals $\sigma\sqrt{1 - c_4^2}$. The upper and lower control limits of the chart known as the Phase I S chart are

$$
\begin{aligned}
\widehat{UCL} &= \left(1 + \frac{3}{c_4}\sqrt{1 - c_4^2}\right) \bar{S} \\
\widehat{LCL} &= \left(1 - \frac{3}{c_4}\sqrt{1 - c_4^2}\right) \bar{S}
\end{aligned}
\tag{2.8}
$$

Approaches making use of these limits are known as the three sigma approaches based on the normal approximation proposed by Shewhart in the early thirties. However, it is easy to prove that this approximation is not satisfactory since as is known

$$
\frac{(n-1)S^2}{\sigma^2} \sim X_{n-1}^2 \tag{2.9}
$$

Although this approximation is not accurate, it is usually used as a first check (see e.g. Ryan (2000), Klein (2000), Lowry, Champ and Woodall (1995)).

A modification of the control limits (2.6) and (2.8) based on property (2.9) uses probability limits in place of the three sigma limits (see e.g. Ryan(2000)). If the value of the standard deviation σ is known the Phase II control limits are

$$\begin{aligned} UCL &= \sigma\sqrt{\frac{\chi^2_{0.999}}{n-1}} \\ LCL &= \sigma\sqrt{\frac{\chi^2_{0.001}}{n-1}} \end{aligned} \tag{2.10}$$

In these limits, if the process variability operates in-control, the probability that the standard deviation of future subgroups will fall between them is 0.998, which is approximately equal to the 0.9973, the probability assumed when using the 3 sigma ones. If the true standard deviation is not known we use its unbiased estimate $\bar{S}/c_4$. The Phase I limits then become

$$\begin{aligned} \widehat{UCL} &= \frac{\bar{S}}{c_4}\sqrt{\frac{\chi^2_{0.999}}{n-1}} \\ \widehat{LCL} &= \frac{\bar{S}}{c_4}\sqrt{\frac{\chi^2_{0.001}}{n-1}}. \end{aligned} \tag{2.11}$$

Yang and Hillier (1970) proposed different Phase I control limits using the same way of thinking as in the case of the mean by excluding sample i from the calculation. Champ and Chou (2003) compared the performance of these different Phase I limits and concluded that the standard limits and the ones proposed by Yang and Hillier can be designed to be equivalent.

The ARL of the control charts for the variability of data in subgroups is given by the relationship (2.4) as this relationship is valid for all Shewhart charts with known parameters. More details on Shewhart charts for variability and related work can be found in Lowry, Champ and Woodall (1995), Klein (2000) and Sim(2000).

2.3.3 Control charts for individual data

Let $X_i, i = 1, .., n$ represent independent and identically distributed observations from a $N(\mu, \sigma^2)$ process. If the parameters μ and σ^2 are known, the Phase II X chart control limits are

$$
\begin{aligned}
UCL &= \mu + 3\sigma \\
LCL &= \mu - 3\sigma
\end{aligned}
$$

Usually, these parameters are not known and they have to be estimated. In this case, the variability is usually controlled using moving ranges. Nevertheless, Nelson (1982), Roes et al. (1993) and Rigdon et al. (1994) have recommended either against the use of the moving range chart or its use together with the classical X chart. Moreover, Sullivan and Woodall (1996a) showed that a moving range control chart does not contribute significantly to the identification of out-of-control situations. For these reasons we do not present it here. Therefore, the use of the X control chart for monitoring both the process mean and standard deviation is recommended. The Phase I control limits of the X control chart are

$$
\begin{aligned}
\widehat{UCL} &= \bar{X} + 3\widehat{\sigma} \\
\widehat{LCL} &= \bar{X} - 3\widehat{\sigma}
\end{aligned}
$$

where $\bar{X}$ is an unbiased estimate of the mean of the process and $\widehat{\sigma}$ is an estimate of the standard deviation σ of the process. Usually, the estimate of the standard deviation used is $\overline{MR}/d_2$ where $\overline{MR}$ denotes the average of the moving ranges and d_2 is the usual function of the sample size n used to make the estimator unbiased. However, Cryer and Ryan (1990) showed that a preferable estimate of σ is s/c_4 where c_4 is defined the same way as in the case of rational subgroups and s is the standard deviation of the observations. Sullivan and Woodall (1996a) proposed a Phase I control chart for independent observations that uses the log-likelihood function and is used to detect shifts in both the mean and the variance. This chart is shown to have better performance in comparison to the X chart or the combined X and MR chart. Moreover, it performs well for detecting sustained shifts in the distribution but not that well for outliers.

2.3.4 Control Charts for Attributes

When an item is produced or purchased it is inspected in order to identify if it satisfies a number of specifications. An item that does not satisfies those specifications is called a defective or a non-conforming item. These defectives lead to rework or they are characterized as scrap or second quality product. In any case we have a loss of money or working time or both. In order to avoid such products, control charts for the characteristics (attributes) have been developed (see, e.g. Woodall (1997), Ryan (2000) and Montgomery (2001)).

Assume that we have a random sample of n units and we inspect them for possible nonconforming items. The fraction nonconforming is defined as the ratio of the number of nonconforming items in a population to the total number of items in that population. Suppose the production is operating in a stable manner, such that the probability that any unit will not conform to specifications is p, and that successive units are produced independently. If d is the number of units of products that are nonconforming, then d has a binomial distribution with parameters n and p, that is

$$P(d = x) = \binom{n}{x} p^x (1-p)^{n-x},\ x = 0, 1, 2, ..., n$$

where $E(d) = np$ and $V(d) = np(1-p)$.

The sample fraction nonconforming is defined as the ratio of the number of nonconforming items in a sample to the total number of items in that sample that is

$$\widehat{p} = \frac{d}{n}$$

where $E(\widehat{p}) = p$ and $V(\widehat{p}) = p(1-p)/n$.

If the true fraction nonconforming p in the production process is known or is a standard value specified by management, then the Phase II control limits for the p chart are defined as

$$\begin{aligned} UCL &= p + 3\sqrt{\frac{p(1-p)}{n}} \\ LCL &= p - 3\sqrt{\frac{p(1-p)}{n}} \end{aligned}$$

where the charting statistic is $\widehat{p}_i$, for sample i.

If the true fraction nonconforming p is not known, then it must be estimated from observed data. The usual procedure is to select m preliminary samples, each of size n. Then the average of these m individual sample fractions nonconforming is

$$\overline{p} = \frac{\sum_{i=1}^{m} d_i}{mn} = \frac{\sum_{i=1}^{m} \widehat{p}_i}{m}$$

and the Phase I control limits are defined as

$$\begin{aligned} \widehat{UCL} &= \overline{p} + 3\sqrt{\overline{p}(1-\overline{p})/n} \\ \widehat{LCL} &= \overline{p} - 3\sqrt{\overline{p}(1-\overline{p})/n} \end{aligned}$$

where the charting statistic is again $\widehat{p}_i$, for sample i.

For a constant sample size it is also possible to plot on a control chart the number of nonconforming units, rather than the fraction nonconforming. This chart is called the np control chart. If the true fraction nonconforming p in the production process is known or is a standard value specified by management, then the Phase II control limits are defined as

$$\begin{aligned} UCL &= np + 3\sqrt{np(1-p)} \\ LCL &= np - 3\sqrt{np(1-p)} \end{aligned}$$

where the charting statistic is $n\widehat{p}_i$, for each sample.

If the true fraction nonconforming p in the production process is not known, then the average of the m preliminary individual sample fractions nonconforming $\overline{p}$ is used and the Phase I control limits are defined as

$$\begin{aligned}\widehat{UCL} &= n\overline{p} + 3\sqrt{n\overline{p}(1-\overline{p})}\\ \widehat{LCL} &= n\overline{p} - 3\sqrt{n\overline{p}(1-\overline{p})}\end{aligned}$$

where the charting statistic is $n\widehat{p_i}$, for each sample. If we have a signal on a p chart we will have also one in an np chart because of the relation between the two charted statistics. Therefore, we may say that these charts are equivalent for a constant sample size.

Borror and Champ (2001) proposed Phase I charts for p and np charts based on the recommendation of Yang and Hillier (1970). Borror and Champ (2001) compared the false alarm rate performance of the standard and the new charts and concluded that the new chart has a higher false alarm rate. Additionally, the performance of the standard Phase I charts is not satisfactory. Therefore the practitioner should use such charts carefully, keeping in mind the possibility of larger number of false alarms than what should be expected from the design of the charts.

If we count the number of defects or nonconformities in a sampling unit then we can plot them in a control chart. This chart is used to control the total number of non-conformities in a unit. In such a chart we usually assume that the number of non-conformities in sample of constant size follows a Poisson distribution. If x is the number of nonconformities and $c>0$ is the parameter of the Poisson distribution, then

$$P(x) = \frac{e^{-x}c^{x}}{x!}, x = 0, 1, 2, ...$$

If the true value of c in the production process is known or is a standard value specified by management, then the Phase II control limits are defined as

$$\begin{aligned} UCL &= c + 3\sqrt{c} \\ LCL &= c - 3\sqrt{c} \end{aligned}$$

where the charting statistic is $\widehat{c}_i$ the number of nonconformities in sample i.

If the true value of c in the production process is not known, then the Phase I control limits are defined as

$$\begin{aligned} \widehat{UCL} &= \overline{c} + 3\sqrt{\overline{c}} \\ \widehat{LCL} &= \overline{c} - 3\sqrt{\overline{c}} \end{aligned}$$

where $\overline{c}$ is the average number of nonconformities in a preliminary sample of inspection units and it is used as an estimate of c. The charting statistic in this case is $\widehat{c}_i$, the number of nonconformities in sample i, again.

If we want to develop a control chart for a sample of n sampling units or for a sampling unit that is n times larger than the standard sampling unit, we may set up a control chart based on the average number of nonconformities per inspection unit. Specifically, let $u = c/n$, then since c is distributed as a Poisson random variable the Phase II control limits for this chart are

$$\begin{aligned} UCL &= u + 3\sqrt{\frac{u}{n}} \\ LCL &= u - 3\sqrt{\frac{u}{n}} \end{aligned}$$

in the case that u is known or is a standard value specified by management. If the true value of u is not known then the Phase I control limits will be

$$\begin{aligned}\widehat{UCL} &= \overline{u} + 3\sqrt{\frac{\overline{u}}{n}} \\ \widehat{LCL} &= \overline{u} - 3\sqrt{\frac{\overline{u}}{n}}\end{aligned}$$

where $\overline{u}$ is the average number of nonconformities per inspection unit from a preliminary sample and it is used as an estimate of u.

If Y_i is the number of conforming items between the $(i-1)th$ and the ith nonconforming item from a stable process with the in-control probability of a nonconforming item be p_0 then this process is a sequence of independent Bernoulli trials with the same probability p_0. Therefore, Y_i+1 is a geometric random variable with parameter p_0. Then, the Phase II control limits for this chart are

$$\begin{aligned}UCL &= \frac{\ln(\alpha/2)}{\ln(1-p_0)} - 1 \\ LCL &= \frac{\ln(1-\alpha/2)}{\ln(1-p_0)}\end{aligned}$$

If p_0 is unknown the Phase I control limits are

$$\begin{aligned}\widehat{UCL} &= \frac{\ln(\alpha/2)}{\ln(1-N/m)} - 1 \\ \widehat{LCL} &= \frac{\ln(1-\alpha/2)}{\ln(1-N/m)}\end{aligned}$$

where $\widehat{p_0} = N/m$, N is the number of nonconforming items in a total of m items sampled.

In all the above control limits we can not accept a negative value. For this reason if the lower control limit is negative we set it equal to zero.

2.4 Cumulative Sum (CUSUM) Control Chart

CUSUM control charts were introduced by Page in 1954. They are used to identify persistent causes in a variable instead of Shewhart charts. This ability is attributed to the fact that they have a memory as they are based on successive sums of the observations minus a constant. Generally, we can say that CUSUM charts are able to detect small to moderate shifts whereas Shewhart charts are able to detect large shifts.

Let $x_1, x_2, ...x_n$ be n independent and identically distributed observations drawn from our production process and μ is the process mean target. Then, we define as the Phase II CUSUM control chart, the function

$$S_i = \sum_{i=1}^{n} (X_i - \mu)$$

plotted against the observation number. In the case of subgrouped data instead of each observation we have the corresponding sample mean.

A more usual way of calculating the CUSUM for an upward shift in mean is by the formulas

$$\begin{aligned} S_0^+ &= 0 \\ S_i^+ &= max(0, S_{i-1}^+ + (X_i - \mu) - k) \end{aligned}$$

where k is a constant called reference value. The CUSUM chart gives a signal if $S_i^+ > h$ where h is a value we choose to give the desired in-control ARL called decision interval. The corresponding CUSUM scheme for detecting downward shifts is

$$\begin{aligned} S_0^- &= 0 \\ S_i^- &= \min(0, S_{i-1}^- + (X_i - \mu) + k) \end{aligned}$$

and it signals if $S_i^- < -h$. There is a certain way to compute the values of k and h, which is related to the distribution of X_i's. Hawkins and Olwell's (1998) textbook is an excellent reference on this subject. We have to state here that in the case of standard normal data with $k = 3$ and $h = 0$ we end up with the classic Shewhart $\overline{X}$ chart for the mean. Moreover, in the case of subgrouped data we modify the preceding schemes and in place of each observation X_i we have the sample mean.

Koning and Does (2000) presented Phase I CUSUM control charts using recursive residuals. They showed that their chart has a better performance than the Likelihood Ratio chart of Sullivan and Woodall (1996a) and Q chart of Quesenberry (1995a).

In the following subsections for CUSUM charts we focus in the case of continuous distributed variables. The case of discrete distributed observations has been also examined see e.g., Lucas (1985), Gan (1993a) and Hawkins and Olwell (1998).

2.4.1 Optimality of the CUSUM

Assume that $x_1, x_2, ..$ are independent and identically distributed random variables that are observed sequentially. Let $x_1, x_2, .., x_{m-1}$ have (in-control) distribution function F_0 and $x_m, x_{m+1}, ..$ have (out-of-control) distribution function $F_1 \neq F_0$. The two distributions are known but the time of change m is assumed unknown.

Many schemes can detect such a change (e.g. Shewhart charts). These schemes are classified by the expected time until the process signals while it remains in-control (false alarm rate). Among all procedures with the same false alarm rates, the optimal procedure is the one that detects changes quicker. Or we could say that among all procedures with the same in-control expected number of samples until signal, the optimal procedure has the smallest expected time until it signals a change when the process shifts to the out-of-control state.

Moustakides (1986) proved that the CUSUM scheme was optimal in the above sense. Specifically, among all tests with the same in-control expected number of samples until signal, the CUSUM had the smallest out-of-control expected number of samples.

The optimality of the CUSUM is for detecting a shift to a single specific out-of-control distribution. The CUSUM that is optimal for detecting one particular shift is not optimal for detecting a different shift. For a different shift a different CUSUM will be optimal. However, while a CUSUM for detecting a shift of one standard deviation is optimal only for this shift, it performs nearly as well as the optimal CUSUM for all shifts that are not too far from one standard deviation.

2.4.2 Average Run Length (ARL)

Two different methods for the computation of the ARL have been developed; the integral equation method and the Markov chain approach. Page (1954) used integral equations for the computation of the ARL. Let the distribution function of a single score x be $F(x)$ and let $L(z)$ be the ARL of the one sided case. $L(0)$ stands for the ARL with an initial value of zero. Then, for $0 \leq z < h$

$$L(z) = 1 + L(0)F(-z) + \int_0^h f(x-z)L(x)dx$$

We may explain the above integral equation with the following description: the expected run length of a test which is now at z equals 1 (the next observation) plus the probability that the next observation will return the CUSUM to zero multiplied by the expected run length from $z = 0$ plus the integral over the probabilities that the CUSUM lands somewhere between zero and h multiplied by the respective expected run lengths from the new value of the CUSUM. Van Dobben de Bruyn (1968) gives a discussion on the derivation of this equation. Additionally, Wetherill (1977) gave an almost identical relationship but from a somewhat different way of thinking. Others that have dealt with the same problem are Ewan and Kemp (1960) and recently Champ, Rigdon and Scharnagl (2001) that give a general method for obtaining integral equations used in the evaluation of many control charts.

The Markov Chain approach begins by approximating the problem of obtaining the average run length (ARL) and then obtains an exact solution to the approximate problem. The integral equation approach begins with the exact problem and finds an approximate solution to it. Champ and Rigdon (1991) compared integral and Markov chain approaches. They propose the integral equation as a preferable method when an integral equation can be found. On the other hand, there are situations, where only the Markov chain approach seems appropriate.

Brook and Evans (1972) were the first to propose the new method for computing the ARL based on a Markov chain. This method applies to both discrete and continuous variables. In the case of continuous variables let Z be the quality characteristic we want in-control, which is continuously distributed. Consider the one-sided case where we accumulate the deviations of Z from a reference value k and this procedure stops if we reach the upper decision boundary h or if the cumulative sum equals zero. A Markov process with continuous state space can represent this scheme.

Suppose that the Markov chain has $t+1$ states labeled $E_0, E_1, ...E_t$ where E_t is the absorbing state. The probability that the chain remains in the same state at the next step should correspond to the case where the cumulative sum does not change in value by more than a small amount say $0.5w$, meaning that the next value of Z does not differ from the reference value k by more than $0.5w$. The value of w determines the width of the grouping interval that is used to discretize the probability distribution of Z. This value must be carefully chosen because properties like average run length and percentage points are highly affected by the width of the decision interval. In order to avoid unwilling behavior a further restriction is the following; the probability of a jump from E_i to the absorbing state E_t should be equal to the probability that the cumulative sum for $(Z-k)$ jumps beyond the point h from a position in $(0, h)$ which corresponds approximately to the state E_i. Therefore

$$w = 2h/(2t-1) \tag{2.12}$$

The transition probabilities for the Markov chain are for $i = 0, 1, \ldots, t-1$ as follows

$$\begin{aligned} P_{i0} &= pr(E_i \to E_0) = pr(Z - k \leq -iw + 0.5w) \\ P_{ij} &= pr(E_i \to E_j) = pr((j-i)w - 0.5w \leq Z - k \leq (j-i)w + 0.5w), 1 \leq j \leq t-1 \\ P_{it} &= pr(E_i \to E_t) = pr((t-i)w - 0.5w < Z - k) \end{aligned}$$

Also, $pr(E_0 \to E_t) = pr(Z - k > h)$ for any w that satisfies relation (2.12). Let $p_r = pr(rw - 0.5w < Z - k \leq rw + 0.5w)$ and $F_r = pr(Z - k \leq rw + 0.5w)$ then the transition probability matrix $\mathbf{P}$ has the following form

$$\mathbf{P} = \begin{bmatrix} F_0 & p_1 & p_2 & \ldots & p_j & \ldots & p_{t-1} & 1 - F_{t-1} \\ F_{-1} & p_0 & p_1 & \ldots & p_{j-1} & \ldots & p_{t-2} & 1 - F_{t-2} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots \\ F_{-i} & p_{1-i} & p_{2-i} & \ldots & p_{j-i} & \ldots & p_{t-1-i} & 1 - F_{t-1-i} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots \\ F_{1-t} & p_{2-t} & p_{3-t} & \ldots & p_{j-(t-1)} & \ldots & p_0 & 1 - F_0 \\ 0 & 0 & 0 & \ldots & 0 & \ldots & 0 & 1 \end{bmatrix}.$$

A relation that holds is $(\mathbf{I} - \mathbf{R})\,\boldsymbol{\mu}^{(s)} = s\mathbf{R}\boldsymbol{\mu}^{(s)}$, s=2,3,... where R is the matrix obtained from the transition probability matrix $\mathbf{P}$ by deleting the last row and column (those referring to the absorbing state E_t), $\mathbf{I}$ is the identity matrix and $\boldsymbol{\mu}^{(s)}$ is the vector of the sth factorial moments for the random variables $X_0, X_1, ..., X_{t-1}$. For s=1 the equation becomes $(\mathbf{I} - \mathbf{R})\,\boldsymbol{\mu} = \mathbf{1}$, where the vector $\mathbf{1}$ has each of its t elements equal to unity. The first element of the vector $\boldsymbol{\mu}$ gives the average run length for a CUSUM chart starting from zero and in general the ith element gives the mean of the run-length distribution when starting from state E_i, i=0,1,...,t-1. We have to state here that the above procedure, suitably modified (Brook and Evans (1972)), can be used for the computation of the ARL of discrete distributed observations also.

The two different computations of the ARL presented are for a one-sided scheme. In the case that we have a two-sided scheme Van Dobben de Bruyn (1968) showed that

$$\frac{1}{ARL} = \frac{1}{ARL^{+}} + \frac{1}{ARL^{-}},$$

where ARL^{+} is the ARL of an upward scheme and ARL^{-} is the ARL of a downward scheme.

The evaluation of CUSUM charts is usually done by the computation of the ARL. Two reasons for this are, firstly, that the computation of the run length distribution is difficult in most situations (see Page (1954), Ewan and Kemp (1960), Brook and Evans (1972), Woodall (1983,1984), Waldman (1986)) and secondly that the in-control run length distribution is approximately geometric, therefore it can be characterized by the ARL. On the contrary, the in-control run length distribution of a CUSUM chart is highly skewed and accordingly conjectures on the ARL can be misleading because the form of the run length distribution changes with a shift in the mean. Therefore the ARL is not a sufficient measure for the performance of the chart. Barnard (1959) and Bissell (1969) have criticized the use of the ARL only and they have proposed instead the simultaneous use of percentage points.

On the other hand, the median run length (MRL) is a quantity that we can rely on because it is more meaningful and more readily understood (see Gan (1994)). For example when the out-of-control MRL is 50, this means that half of all the run lengths are less than 50.

2.4.3 Fast Initial Response (FIR)

Lucas and Crosier (1982) extended the calculation of the average run length by using a head start value S_0 different than zero. The calculation of the ARL for several head start values showed that for a moderate value the in-control ARL has a small percentage decrease while the out-of-control ARL has a large percentage decrease. Therefore we may design a FIR CUSUM, with an almost equivalent in-control ARL and a smaller out-of-control ARL than a standard CUSUM, by increasing h (decision interval) slightly to compensate for the small decrease in ARL caused by the head start.

Lucas and Crosier (1982) recommended a head-start of $S_0 = h/2$. This recommendation is a result of the fact that a CUSUM scheme is a sequence of Wald tests (Page (1954)) with null hypothesis that the mean is zero and alternative hypothesis that the mean is $2k$.

The ARL computation of one-sided schemes is easily done using the proposed procedures of section 2.4.2. However, in the case of a two-sided FIR CUSUM scheme the computation is modified (Yashchin (1985)). Let H^+ and $-H^-$ denote the head starts of an upward and a downward CUSUM scheme, respectively. Also, let $A^+(s)$ and $A^-(s)$ denote the ARL of an upward and a downward CUSUM scheme with head starts s and $-s$, respectively. Then, the ARL of a two-sided FIR CUSUM is

$$ARL = \frac{A^+(H^+)A^-(0) + A^-(H^-)A^+(0) - A^+(0)A^-(0)}{A^+(0) + A^-(0)}.$$

This result holds if the following condition is satisfied

$$k^+ + k^- \geq \max\left(H^+ + H^- - \min(h^+, h^-), \left|h^+ - h^-\right|\right)$$

where k^+, k^- and h^+, h^- are the upward and downward reference values and decision intervals, respectively. This condition ensures that if the upward CUSUM signals the downward CUSUM will be at zero and vice-versa.

2.4.4 CUSUM control chart for process variability

A method for keeping in-control the process dispersion was developed by Chang and Gan (1995). Assume that the process mean is in-control and let $s_1^2, s_2^2, \ldots$ be successive sample variances observed from a process based on a sample of size n. The upper and lower CUSUM charts are obtained through the plotting of

$$C_i^+ = max(0, C_{i-1}^+ + y_i - k_C^+)$$

and

$$C_i^- = max(0, C_{i-1}^- + y_i + k_C^-)$$

against i respectively, for $i = 1, 2, \ldots$, where k_C^+, k_C^- are constants, $y_i = \log(s_i^2)$, $C_0^+ = u$ for $0 \leq u < h^+$ and $C_0^- = v$ for $-h^- \leq v < 0$. The upper CUSUM chart is used to detect increases in the variance and there is an out-of-control signal at the first i for which $C_i^+ > h^+$. The lower CUSUM chart is used to detect decreases in the variance and there is an out-of-control signal at the first i for which $C_i^- < -h^-$. In practice, we usually have to estimate the in-control variance because it is not known and this is done by taking samples from a process, which is assumed to be in-control.

The probability density function of $\log(s_i^2)$, when the measures of the quality characteristic are independent, identically and normally distributed, is

$$f(y) = \frac{\exp\left[ay - \exp\left(y\right)/\beta\right]}{\Gamma\left(a\right)\beta^a}, -\infty < y < \infty$$

where a=$(n-1)/2$, $\beta = 2\sigma^2/(n-1)$ and $\Gamma(a)$ is the gamma function. Let $H(u)$ be the ARL function of an upper CUSUM chart given that $C_0^+ = u$ where $0 \leq u \leq h^+$. Then, an approach similar to Page (1954) for the computation of the ARL is the following

$$H(u) = 1 + H(0)\Pr\left(\log\left(s^2\right) \leq k_C^+ - u\right) + \int_0^{h^+} f(x + k_C^+ - u)H(x)dx.$$

In practice, most of the processes are out-of-control at the beginning and a FIR CUSUM is recommended for a faster detection of this situation. As we have already said Lucas and Crosier (1982) have recommended using $h/2$ as the head start value for monitoring normal means. The distribution of $\log(s_i^2)$ is approximately normal, so $h^+/2$ and $-h^-/2$ are recommended here as head start values for the upper and lower CUSUM charts.

In order to design a CUSUM chart we have to determine the values of h and k. Chang and Gan (1995) provided tables for various values of sample sizes and for the out-of-control standard deviation which we want to detect quickly.

A CUSUM chart based on a larger sample size will be more sensitive than a CUSUM chart based on a smaller sample size for detecting changes in σ. For the two-sided case, the two one-sided CUSUM charts that are optimal for detecting a specific shift in either direction can be run simultaneously so as to detect changes in standard deviation in both directions.

The CUSUM chart described here is based on the assumption that the measures of the quality characteristic are independent, identically and normally distributed. For non-normally distributed observations the ARL values are different and especially for distributions with a tail larger than the normal one, the ARL tends to be small, so the false alarm rate is higher. When observations are positively serially correlated, then the CUSUM is less effective in detecting increases in σ, because the sample variance decreases as the serial correlation increases.

For other CUSUM charts developed for monitoring process variance see Yashchin (1994) and Srivastava (1997).

2.5 Exponentially Weighted Moving Average (EWMA) Control Chart

The EWMA chart was introduced by Roberts (1959) and it is used as the CUSUM chart to detect persistent shifts in a variable. Its ability is to signal faster than the Shewhart charts for small and moderate shifts but not that fast for large shifts. Generally, we can say that its performance is similar to the performance of the CUSUM chart.

In the following subsections we present the EWMA chart for continuous variables. The case of discrete variables has been studied by Gan (1990), Borror et al. (1998), Quesenberry (1995b) and Quesenberry (1995c).

2.5.1 EWMA Control Chart for the mean

Let the mean μ and standard deviation σ of a process to be known. The EWMA chart for individual observations is defined as

$$Z_i = \lambda x_i + (1-\lambda) Z_{i-1},\ Z_0 = \mu$$

where x_i is the observation at time i = 1, 2, ..., λ is a smoothing parameter that takes values between 0 and 1 and Z_0 is the initial value. When the value of λ is close to 0, the EWMA chart can detect small to moderate shifts in the process mean, when λ is close to unity the EWMA can detect large shifts in the process mean and when $\lambda = 1$ it is actually the $\overline{X}$ chart. As a starting value, instead of the in-control process mean, we can use the target value. The control limits of this chart are

$$\begin{aligned} UCL &= \mu + L\frac{\sigma}{\sqrt{n}}\sqrt{\left(\frac{\lambda}{2-\lambda}\right)\left[1-(1-\lambda)^{2i}\right]} \\ LCL &= \mu - L\frac{\sigma}{\sqrt{n}}\sqrt{\left(\frac{\lambda}{2-\lambda}\right)\left[1-(1-\lambda)^{2i}\right]}, \end{aligned} \tag{2.13}$$

where L is a constant used to specify the width of the control limits, μ is the mean of the process and $\frac{\sigma}{\sqrt{n}}\sqrt{\left(\frac{\lambda}{2-\lambda}\right)\left[1-(1-\lambda)^{2i}\right]}$ the standard deviation of Z_i when the process is in-control. In case the EWMA chart is used for some time, instead of control limits (2.13), we may use their limiting values

$$\begin{aligned} UCL &= \mu + L\frac{\sigma}{\sqrt{n}}\sqrt{\left(\frac{\lambda}{2-\lambda}\right)} \\ LCL &= \mu - L\frac{\sigma}{\sqrt{n}}\sqrt{\left(\frac{\lambda}{2-\lambda}\right)} \end{aligned} \tag{2.14}$$

since $\lim\limits_{i\to\infty}\left(\frac{\sigma^2}{n}\left(\frac{\lambda}{2-\lambda}\right)\left[1-(1-\lambda)^{2i}\right]\right) = \frac{\lambda\sigma^2}{(2-\lambda)n}$ (see e.g., Lucas and Saccucci (1990)). In this case, $\frac{\sigma}{\sqrt{n}}\sqrt{\lambda/(2-\lambda)}$ is the asymptotic standard deviation of Z_i. In the case of subgrouped data instead of a single observation we have the sample mean of the observations at time i. The control limits are correspondingly modified.

The main features of the EWMA chart are the same as the ones for the CUSUM except of the optimality. The computation of its run length distribution and the ARL can be done by the exact way using integral equations (Crowder(1987)). The ARL $L(u)$ of a two-sided EWMA chart for the mean given that the EWMA starts at u is computed through the relation

$$L(u) = 1 + \frac{1}{\lambda} \int_{-h}^{h} f\left(\frac{y-(1-\lambda)u}{\lambda}\right) L(y)dy$$

where y_i's are assumed to be independent, identically distributed observations with probability density function $f(\cdot)$, h is the upper control limit and $-h$ the lower control limit. This can be explained as follows; if for the first observation y_1, we have that $|(1-\lambda u)+\lambda y_1| > h$ then we have a signal. On the other hand, if this relation does not hold, the run length continues to move from $(1-\lambda u)+\lambda y_1$ and $L((1-\lambda u)+\lambda y_1)$ stands for the additional run length.

The approximation method of the Markov chain is the other alternative (Lucas and Saccucci (1990)). The ARL in this case is computed by

$$ARL = (\mathbf{I}-\mathbf{R})^{-1}\mathbf{1},$$

where $\mathbf{I}$ is the identity matrix, $\mathbf{1}$ is a vector of unities and $\mathbf{R}$ is a submatrix of the transition probability matrix $\mathbf{P}$, where

$$\mathbf{P} = \begin{bmatrix} \mathbf{R} & (\mathbf{I}-\mathbf{R})\,\mathbf{1} \\ \mathbf{0}^T & 1 \end{bmatrix}.$$

If p_{jk} is the probability that the control statistic goes from state j to state k then

$$p_{jk} = \Pr\left[\lambda^{-1}\left\{(S_k-\delta)-(1-\lambda)S_j\right\} < Y_i \le \lambda^{-1}\left\{(S_k+\delta)-(1-\lambda)S_j\right\}\right]$$

where S_k is the midpoint of the kth interval, $j = -m, ..., m$ and m is the number of states we will use. The larger the number of states the more accurate the computation will be.

The median run length and the fast initial response are properties that have been implemented and in the context of EWMA charts (see Lucas and Saccucci (1990) and Gan (1993b)). For further discussion of the EWMA charts and other modifications see Robinson and Ho (1978), Hunter (1986), Saccucci and Lucas (1990), Domangue and Patch (1991), Ingolfson and Sachs (1993), Steiner (1999).

2.5.2 EWMA control chart for process variability

Several publications dealing with the subject of keeping in-control the process variance using an EWMA chart have appeared in the literature like Wortham and Ringer (1971), Wortham (1972), Sweet (1986), Ng and Case (1989), Domangue and Patch (1991), Crowder and Hamilton (1992), Hamilton and Crowder (1992), MacGregor and Harris (1993), Acosta-Mejia and Pignatiello (2000). In this subsection we present schemes for sample size larger than unity. The schemes for $n = 1$ are investigated in Chapter 4.

The EWMA chart of squared deviations from target (EWMA$_\mathrm{S}$) was proposed by Wortham and Ringer (1971) for detecting a shift in the process standard deviation. The statistic of this chart is given by

$$S_i = \lambda(x_i - \mu_0)^2 + (1 - \lambda)S_{i-1},\ S_0 = \sigma_0^2,$$

where λ is a smoothing parameter that takes values between 0 and 1 and S_0 is the initial estimated value of the mean squared error. It can be proved (MacGregor and Harris (1993)) that under normality the quantity S_i/σ^2 is approximately distributed as $\chi^2(\nu)/\nu$ where the degrees of freedom ν depend on the parameter λ, the correlation of the x_i's and the degrees of freedom associated with the initial value. If we assume that the process mean is on target and the variance is σ_0^2 then the control limits of S_i are the $\alpha/2$ and $1 - \alpha/2$ percentiles of $\sigma_0^2\chi^2(\nu)/\nu$ distribution. In case of independent and normally distributed observations we may plot $\sqrt{S_i}$ and the corresponding control limits are

$$\begin{aligned} UCL &= \sigma_0 \sqrt{\frac{\chi^2_{1-a/2}(\nu)}{\nu}} \\ LCL &= \sigma_0 \sqrt{\frac{\chi^2_{a/2}(\nu)}{\nu}} \end{aligned}$$

However, the above statistic has the property to respond both to changes in mean and in variance. Therefore, a statistic that would plot out of the control limits only in the case of variance shifts is desirable. Sweet (1986) proposed the use of an estimate of the process mean in each step in time. Specifically, let μ_i denote an estimate of the process mean at time i. Then

$$S_i = \lambda(x_i - \mu_i)^2 + (1-\lambda)S_{i-1},\ S_0 = \sigma_0^2.$$

A usually used estimate for the mean is the EWMA statistic for the mean (Z_i). MacGregor and Harris (1993) computed control limits for this statistic which is usually addressed as the Exponentially Weighted Moving Variance (EWMV).

Crowder and Hamilton (1992) proposed a different control chart based on $\ln\left(\widehat{\sigma}_i^2\right)$. The scheme is

$$C_i = \max\left\{(1-\lambda)\,C_{i-1} + \lambda y_i, \ln\left(\sigma_0^2\right)\right\},$$

where $C_0 = \ln\left(\sigma_0^2\right)$, λ is the usual constant taking values between 0 and 1 and $y_i = \ln\left(\widehat{\sigma}_i^2\right)$. This statistic can be used to identify only upward shifts in the variance. The UCL of this chart in case of independent observations is given by

$$UCL = K\sqrt{\left(\frac{\lambda}{2-\lambda}\right)\left\{\frac{2}{n-1} + \frac{2}{(n-1)^2} + \frac{4}{3\,(n-1)^3} - \frac{16}{15\,(n-1)^5}\right\}},$$

where K is a constant chosen together with λ so as to achieve the desired ARL. If $L(u)$

is the ARL of this chart with u the starting value then

$$L(u) = 1 + L(0)F\left(\frac{-(1-\lambda)u}{\lambda}\right) + \frac{1}{\lambda}\int_0^{UCL} f\left(\frac{y-(1-\lambda)u}{\lambda}\right) L(y)dy,$$

where $F(x)$ and $f(x)$ are the cumulative distribution function and the probability distribution function of the log-gamma distribution respectively.

2.6 Multivariate Shewhart Control Charts

Multivariate Shewhart Control Charts are analogous to the univariate ones but they involve in the computations several variables instead of one. The Phase I and II charts discussion does not change in this case. Sparks (1992), Wierda (1994), Lowry and Montgomery (1995), Fuchs and Kenett (1998), Ryan (2000) and other statisticians and engineers agree with the definition of the Phases given in Section 2.2. However, Alt (1985) gives a somewhat different definition for the two distinct phases of control charting practice. In the following the definition of Section 2.2 is followed.

A crucial matter in Multivariate Shewhart Control Charts is the sample size n of each rational subgroup. As Lowry and Montgomery (1995) suggest, the appropriate use of a test statistic (X^2 or T^2) can be broken into four categories: 1) Phase I and $n = 1$, working with individual observations; 2) Phase I and $n > 1$, working with rational subgroups; 3) Phase II and $n = 1$, working with individual observations; 4) Phase II and $n > 1$, working with rational subgroups.

Mason and Young (2002) recently published a textbook for the implementation of multivariate statistical process control in the case of Shewhart charts that discusses in detail several subjects.

2.6.1 Control Charts for the Process Mean $(n > 1)$

Assume that the vector $\mathbf{x}$ follows a p-dimensional normal distribution, denoted as $N_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, and that there are m samples of size $n > 1$ available from the process. A control chart can be based on the sequence of the following statistic

$$D_i^2 = n(\overline{\mathbf{x}}_i - \boldsymbol{\vartheta}_0)^t \mathbf{Z}_0^{-1} (\overline{\mathbf{x}}_i - \boldsymbol{\vartheta}_0),$$

where $\overline{\mathbf{x}}_i$ is the vector of the sample means of the ith rational subgroup, $\boldsymbol{\vartheta}_0$ and $\mathbf{Z}_0$ are the appropriate vector of means and the appropriate variance-covariance matrix in either Phase I or Phase II, respectively. The superscript t is used to define the transpose of a matrix. The D_i^2 statistic represents the Mahalanobis distance of any point from the target $\boldsymbol{\vartheta}_0$. Thus, if the value of the test statistic D_i^2 plots above the control limit (L_u), the chart signals a potential out-of-control process. Generally, control charts have both upper (L_u) and lower control limits (L_l). However, in this case only an upper control limit is meaningful, because extreme values of the D_i^2 statistic correspond to a point far away from the target $\boldsymbol{\vartheta}_0$, whereas small or zero values of the D_i^2 statistic correspond to points close to the target $\boldsymbol{\vartheta}_0$.

If $\boldsymbol{\vartheta}_0 = \overline{\overline{\mathbf{x}}}_0$, $\mathbf{Z}_0 = \overline{\mathbf{S}}$, $n > 1$ and $\overline{\mathbf{x}}_i$ is the mean of the ith observation then the $D_i^2/c_0(p,m,n)$ statistic follows an F distribution with p and $(mn - m - p + 1)$ degrees of freedom. Here $c_0(p,m,n) = [p(m-1)(n-1)]\,(mn - m - p + 1)^{-1}$, the parameter $\overline{\overline{\mathbf{x}}}$ is the overall sample mean vector and $\overline{\mathbf{S}}$ is the pooled sample variance-covariance matrix. Consequently, a multivariate Shewhart control chart for the process mean, with unknown parameters, has the following control limit

$$L_u = c_0(p,m,n) F_{1-a,p,mn-m-p+1}.$$

This control chart is called a Phase I T^2-chart**.** We must note that, for a Phase I T^2-chart the statement "if the process is in-control the probability of at least one of the D_i^2's being outside the control limits is a" does not hold. It does not hold because in this Phase the

D_i^2's are not independent (this is valid only for $i = 1$). In practical problems T^2-chart is typically recommended for the preliminary analysis of multivariate observations in process monitoring applications. Nedumaran and Pignatiello (2000) consider the issue of constructing retrospective T^2 control chart limits so as to control the overall probability of a false alarm at a specified value. Furthermore, Mason et. al. (2001) use the T^2-chart for monitoring batch processes in both Phase I and Phase II operations.

If $\boldsymbol{\vartheta}_0 = \overline{\overline{\mathbf{x}}}_0$, $\mathbf{Z}_0 = \overline{\mathbf{S}}$, $n > 1$ and $\overline{\mathbf{x}}_i$ is the mean of a future observation then the $D_i^2/c_1(p,m,n)$ statistic follows an F Distribution with p and $(mn - m - p + 1)$ degrees of freedom, where $c_1(p,m,n) = [p(m+1)(n-1)]\,(mn - m - p + 1)^{-1}$. Thus, a multivariate Shewhart control chart for the process mean, with unknown parameters, has the following control limit

$$L_u = c_1(p,m,n) F_{1-a,p,mn-m-p+1}.$$

This control chart is called a Phase II T^2-Chart.

If $\boldsymbol{\vartheta}_0 = \boldsymbol{\mu}_0$, $\mathbf{Z}_0 = \boldsymbol{\Sigma}_0$, $n > 1$ and $\overline{\mathbf{x}}_i$ is the mean of the ith observation then the D_i^2 statistic follows a X^2-distribution with p degrees of freedom. Therefore, a multivariate Shewhart control chart for the process mean, with known mean vector $\boldsymbol{\mu}_0$ and known variance-covariance matrix $\boldsymbol{\Sigma}_0$ has the upper control limit $L_u = X^2_{p,1-\alpha}$. This control chart is called a Phase II X^2-Chart.

The in-control ARL_0 of the multivariate Shewhart chart, when $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ are known, can be calculated as $ARL_0 = 1/\alpha$ where α is the probability that D_i^2 exceeds L_u. Furthermore, the out-of-control ARL_1 of the multivariate Shewhart chart depends on the mean vector and variance-covariance matrix only through the noncentrality parameter $\lambda^2(\boldsymbol{\mu}_1)$,

$$\lambda^2(\boldsymbol{\mu}_1) = n(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^t \boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) = n\boldsymbol{\delta}^t \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\delta},$$

where $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_0 + \boldsymbol{\delta}$ is a specific out-of-control mean vector. Hence, it is possible to consider the ARL_1 as a function of $\lambda(\boldsymbol{\mu}_1)$, the square root of $\lambda^2(\boldsymbol{\mu}_1)$, and construct an ARL_1 curve by using the equation $ARL_1 = 1/\,(1-\beta)$, where β is the probability of the event “Procedure fails to diagnose an out-of-control situation”. We have to note that the result that the ARL depends only on the noncentrality parameter is based on the assumptions that $\mathbf{\Sigma}_0$ is the known variance-covariance matrix and that random sampling is being done independently from a multivariate normal distribution.

The theory presented up to now considers the case of a pre-defined and fixed sample of size n. Jolayemi (1995) presented a power function model for determining sample sizes for the operation of a multivariate process control chart. Moreover, Aparisi (1996), gives a procedure for the construction of a control chart with adaptive sample sizes.

2.6.2 Control Charts for the Process Mean ($n = 1$)

For charts constructed using individual observations ($n = 1$), the test statistic for the ith individual observation has the form

$$D_i^2 = (\mathbf{x}_i - \boldsymbol{\vartheta}_0)^t \, \mathbf{Z}_0^{-1} \, (\mathbf{x}_i - \boldsymbol{\vartheta}_0) \, ,$$

where $\mathbf{x}_i$ is the ith observation, $i = 1, 2, ..., m$ following $N_p(\boldsymbol{\mu}_0, \mathbf{\Sigma}_0)$, $\boldsymbol{\vartheta}_0$ and $\mathbf{Z}_0$ are the appropriate vector of means and the variance-covariance matrix in either Phase I or Phase II, respectively.

If $\boldsymbol{\vartheta}_0 = \overline{\mathbf{x}}_m$, $\mathbf{Z}_0 = \mathbf{S}_m$ and $\mathbf{x}_i$ is the ith individual observation then the $D_i^2/d_0(m)$ statistic follows a Beta distribution with $p/2$ and $(m - p - 1)$ degrees of freedom, where $d_0(m) = (m-1)^2 m^{-1}$. Thus, a multivariate Shewhart control chart for the process mean, with unknown parameters, has the following control limit (Tracy et al. (1992))

$$L_u = d_0(m) B_{1-\alpha/2, p/2, (m-p-1)/2},$$

where $\overline{\mathbf{x}}_m$ is the overall sample mean and $\mathbf{S}_m$ is the sample variance-covariance matrix. This control chart is called a Phase I T^2-Chart. Alternative estimators of the variance-covariance matrix has been proposed by Sullivan and Woodall (1996b) and Chou et al. (1999).

If $\boldsymbol{\vartheta}_0 = \overline{\mathbf{x}}_m$, $\mathbf{Z}_0 = \mathbf{S}_m$ and $\mathbf{x}_i$ is a future individual observation then the $D_i^2/d_1(m,p)$ statistic follows an F distribution with p and $(m-p)$ degrees of freedom, where $d_1(m,p) = p(m+1)(m-1)\left[m(m-p)\right]^{-1}$. Therefore, a multivariate Shewhart control chart for the process mean, with unknown parameters, has the following control limits (Tracy et al. (1992))

$$L_u = d_1(m,p)F_{1-\alpha,p,m-p}.$$

This control chart is called a Phase II T^2-Chart.

If $\boldsymbol{\vartheta}_0 = \boldsymbol{\mu}_0$, $\mathbf{Z}_0 = \boldsymbol{\Sigma}_0$ and $\mathbf{x}_i$ is the *ith* observation then the D_i^2 statistic follows a X^2-distribution with p degrees of freedom (Seber (1984)). Consequently, a multivariate Shewhart control chart for the process mean, with known mean vector $\boldsymbol{\mu}_0$ and known variance-covariance matrix $\boldsymbol{\Sigma}_0$, has upper control limit $L_u = X^2_{p,1-\alpha}$. This control chart is called a Phase II X^2-Chart.

2.6.3 Control Charts for Process Dispersion

In the following, multivariate control charts for controlling process dispersion are presented. In the previous two subsections, it was assumed that process dispersion remained constant and equal to $\boldsymbol{\Sigma}$. This assumption, is generally not true, and must be validated in practice. Process variability is summarized in the $p \times p$ variance-covariance matrix $\boldsymbol{\Sigma}$ which contains $p \times (p+1)/2$ parameters. There are two single-number quantities for measuring the overall variability of a set of multivariate data. The first one is the determinant of the variance-covariance matrix, $|\mathbf{S}|$, which is called the generalized variance. The square root of this quantity is proportional to the area or volume generated by a set of data. The second one is the trace of the variance-covariance matrix, $tr\mathbf{S}$, the sum of the variances of the variables. In this subsection, two different control charts for the process dispersion are presented since different statistics can be used to describe variability.

Assume that the vector $\mathbf{x}$ follows a $N_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, and that there are m samples of size $n > 1$ available from the process. The first multivariate chart for the process dispersion can be based on the sequence of the following statistic

$$W_i = -pn + pn \ln n - n \ln \left[|\mathbf{A}_i| \, |\mathbf{\Sigma}_0|^{-1} \right] + trace \left(\mathbf{\Sigma}_0^{-1} \mathbf{A}_i \right)$$

for the ith sample, $i = 1, 2, ..., m$, where $\mathbf{A}_i = (n-1)\mathbf{S}_i$. The W_i statistic follows an asymptotic X^2-distribution with $p \times (p+1)/2$ degrees of freedom. Hence, a multivariate Shewhart control chart for process dispersion, with known mean vector $\boldsymbol{\mu}_0$ and known variance-covariance matrix $\mathbf{\Sigma}_0$ has the upper control limit $L_u = X^2_{p(p+1)/2, 1-\alpha}$. Therefore, if the value of the test statistic W_i plots above L_u, the chart signals a potential out-of-control process. This control chart is called a Phase II W chart.

The second chart is based on the sample generalized variance $|\mathbf{S}|$, where $\mathbf{S}$ is the $p \times p$ sample variance-covariance matrix. One approach in developing an $|\mathbf{S}|$-Chart is to utilize its distributional properties. Alt (1985) and Alt and Smith (1988) state that if there are two quality characteristics, then

$$\left[2(n-1) \, |\mathbf{S}|^{1/2} \right] |\mathbf{\Sigma}_0|^{-1/2} \text{ is distributed as a } X^2_{2n-4}.$$

Thus, the control limits for an $|\mathbf{S}|$-Chart are

$$\begin{aligned} L_u &= \left[|\mathbf{\Sigma}_0| \left(X^2_{2n-4, 1-\alpha/2} \right)^2 \right] [2(n-1)]^{-2} \\ L_l &= \left[|\mathbf{\Sigma}_0| \left(X^2_{2n-4, \alpha/2} \right)^2 \right] [2(n-1)]^{-2}, \end{aligned}$$

where L_u is the upper control limit and L_l is the lower control limit.

In a recent paper by Aparisi et al. (2001), the distribution of the $|\mathbf{S}|$-Chart is studied and suitable control limits are obtained for the case when there are more than two variables. Aparisi et al. (2001) propose the design of the $|\mathbf{S}|$ Chart with adaptive sample size to control process defined by two quality characteristics. Alt (1985) proposes a second approach in developing an $|\mathbf{S}|$-Chart by using only the first two moments of $|\mathbf{S}|$ and the property that most of the probability distribution of $|\mathbf{S}|$ is contained in the interval

$$E\left[|\mathbf{S}|\right] \pm 3\sqrt{V\left[|\mathbf{S}|\right]}.$$

Additionally, Alt and Smith (1988) propose a modification, the $|\mathbf{S}|^{1/2}$ Chart. Furthermore, Alt (1985) gives a proper unbiased estimator for $|\mathbf{\Sigma}_0|$, in order to define a Phase I control chart for controlling the process dispersion.

Although $|\mathbf{S}|$ is a widely used measure of multivariate variability, it is a relative simplistic scalar representation of a complex multivariate structure. Therefore, it can be misleading in some cases. Lowry and Montgomery (1995) present three sample covariance matrices for bivariate data that all have the same generalized variance and yet have distinctly different correlations. As a result, it is often desirable to provide more than the single number $|\mathbf{S}|$ as a summary of $\mathbf{S}$. The use of univariate dispersion charts as supplementary to a control chart for $|\mathbf{S}|$ is proposed by Alt (1985).

2.6.4 Multiattributes Control Charts

Patel (1973) was the first to deal with methods of quality control, when the p-dimensional observations are coming from a multivariate binomial or multivariate Poisson population. Specifically, Patel proposed a X^2 chart using an approximation to normality. Lu et al. (1998) developed a multivariate attribute control chart, called the MNP chart. The name of this chart stems from the fact that it is a straightforward extension of the univariate np chart. Let $\mathbf{p} = (p_1, p_2, ..., p_p)$ be the fraction nonconforming vector, $\mathbf{P}_0 = [\delta_{ij}]_{p\times p}$ the correlation matrix and $\mathbf{c} = (C_1, C_2, ..., C_p)$ the vector of counts of nonconforming units. Define

$$L = \sum_{i=1}^{p} \frac{C_i}{\sqrt{p_i}},$$

which is the weighted sum of the nonconforming units of all the quality characteristics in the sample. Since the nonconformance of a quality characteristic in one dimension may be more serious than in another dimension we want to take into account that information in the calculations. Montgomery (2001) suggested a statistic that uses this information.

Let $\mathbf{d} = (d_1, d_2, ..., d_p)$ denote the vector of the numbers of demerits, which indicates the severity of nonconformance in quality characteristics. Then the above statistic L can be extended as follows

$$L_D = \sum_{i=1}^{p} \frac{d_i C_i}{\sqrt{p_i}}.$$

For L_D Lu et al. (1998) proposed the following multivariate attribute chart

$$\begin{aligned} UCL &= n\sum_{j=1}^{p} d_j\sqrt{p_j} + 3\sqrt{n\left[\sum_{j=1}^{p} d_j^2(1-p_j) + 2\sum_{i<j}^{p}(d_i d_j \delta_{ij}\sqrt{(1-p_i)(1-p_j)})\right]} \\ LCL &= n\sum_{j=1}^{p} d_j\sqrt{p_j} - 3\sqrt{n\left[\sum_{j=1}^{p} d_j^2(1-p_j) + 2\sum_{i<j}^{p}(d_i d_j \delta_{ij}\sqrt{(1-p_i)(1-p_j)})\right]}. \end{aligned}$$

Given the values of the parameters, the control limits can be computed and the MNP chart can then be established using the above equation. If the real values are unknown, then they must be estimated. Furthermore, Lu et al. (1998) introduced a formula that can be used to calculate the appropriate sample size n of each rational subgroup and gave a procedure for the interpretation of an out-of-control signal.

Jolayemi (1999) proposed a multivariate attribute control chart (MACC), which is based on an approximation of the convolution of independent binomial variables and on an extension of the univariate np chart. When a process is monitored with respect to many independent attributes $X_1, X_2, ..., X_m$, each of which follows a binomial distribution, the distribution of the sum or the convolution of the number of defective items found in a sample of size n from the process, with respect to all m attributes, is well approximated by a binomial distribution with parameters mn and $\overline{p}_0$ (the mean of $p_1, p_2, ..., p_p$). Therefore, instead of plotting m different np charts we use a single one using the preceding approximation. The control limits of this chart are

$$\begin{aligned} UCL &= nm\overline{p}_0 + k\sqrt{nm\overline{p}_0(1-\overline{p}_0)} \\ LCL &= nm\overline{p}_0 - k\sqrt{nm\overline{p}_0(1-\overline{p}_0)}, \end{aligned}$$

where k (usually $k = 3$) is the constant that determines the width of the control limits and p_{0i}, i= 1, 2,..., m is the expected fraction defective produced with respect to attribute X_i when the process is in-control.

Therefore, a corrective action will be taken whenever the sum of the numbers of defective items found in a sample of size n, with respect to p attributes, exceeds an acceptance number c_p, where c_p is the largest integer less than or equal to the upper limit. The acceptance number c_p and the sample size n are given by the following equations

$$\begin{aligned} \sum_{i=0}^{c_p} \tbinom{nm}{i} \overline{p}_1^i (1-\overline{p}_1)^{nm-i} &= \beta \\ \sum_{i=0}^{c_p} \tbinom{nm}{i} \overline{p}_0^i (1-\overline{p}_0)^{nm-i} &= 1-\alpha, \end{aligned}$$

where $\overline{p}_1$ is the mean of p_{1i}, for $i = 1, 2, ..., p$, p_{1i} is the expected fraction defective produced with respect to attribute X_i when the process is out-of-control. The design of the MACC presupposes to solve the above equations with specified α and β values, to find the proper acceptance number c_p and the sample size n. Finally, Jolayemi (1999) gives a proper statistical procedure for the interpretation of an out-of-control signal.

2.7 Multivariate CUSUM and EWMA Control Charts

Multivariate Shewhart control charts use the information only from the current sample and they are relative insensitive to small and moderate shifts in the mean vector. Multivariate Cumulative Sum (CUSUM) and Exponentially Weighted Moving Average (EWMA) control charts are developed to overcome this problem.

The multivariate CUSUM and EWMA charts presented in the following subsections are Phase II control charts. Sullivan and Woodall (1998), recommend the use of multivariate CUSUM and EWMA charts for the preliminary analysis of multivariate observations.

2.7.1 CUSUM Type Control Charts

The multivariate CUSUM control charts are distinguished in two major categories. In the first case, the direction of the shift (or shifts) is considered to be known whereas in the second the direction of the shift is considered to be unknown (directionally invariant schemes).

We first present the CUSUM control charts for which we assume that the direction of the shift (or shifts) is known. Woodall and Ncube (1985) described how a p-dimensional multivariate normal process, can be monitored by using p two or one-sided univariate CUSUM charts for the p original variables or using p two or one sided univariate CUSUM charts for the p principal components. This multiple univariate CUSUM scheme is called the MCUSUM. The MCUSUM gives an out-of-control signal whenever any of the univariate CUSUM charts gives an out-of-control signal. The ARL performance in a multivariate process, is studied in the cases of independent and dependent quality characteristics.

Healy (1987) uses the fact that CUSUM charts can be viewed as a series of sequential probability ratio tests, in order to develop a multivariate CUSUM chart. Let $\mathbf{x}_i$ be the ith observation, which comes from a $N_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ with an in-control $p \times 1$ mean vector $\boldsymbol{\mu}_0$ and a known $p \times p$ common variance-covariance matrix $\boldsymbol{\Sigma}_0$. Denote $\boldsymbol{\mu}_1$ an out-of-control $p \times 1$ vector of means. The CUSUM for detecting a shift in $\boldsymbol{\mu}_0$ towards $\boldsymbol{\mu}_1$ may be written as

$$G_i^1 = \max\left[\left(G_{i-1}^1 + \mathbf{a}^t(\mathbf{x}_i - \boldsymbol{\mu}_0) - 0.5\lambda(\boldsymbol{\mu}_1)\right), 0\right],$$

where $\lambda(\boldsymbol{\mu}_1)$ is the square root of the noncentrality parameter and $\mathbf{a}^t = A/\lambda(\boldsymbol{\mu}_1)$ and $A = \left[(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^t \boldsymbol{\Sigma}_0^{-1}\right]$. This CUSUM scheme signals when $G_i^1 \geq H$. For detecting a shift in the mean of a multivariate normal random variable, the CUSUM procedure reduces to a univariate normal procedure. That is, all the available theory for calculating ARL, H, G_0^1 for a univariate normal CUSUM can also be used for multivariate normal CUSUM.

A similar procedure is proposed for controlling the process dispersion. The CUSUM for detecting a change in the variance-covariance matrix may be written as

$$G_i^2 = \max\left[\left(G_i^2 + D_i^2 - K\right), 0\right],$$

where $\mathbf{\Sigma}_1 = C\mathbf{\Sigma}_0$ (C is a real constant), $K = p\log C\,(C/(C-1))$ and

$$D_i^2 = (\mathbf{x}_i - \boldsymbol{\mu}_0)^t\, \mathbf{\Sigma}_0^{-1}\,(\mathbf{x}_i - \boldsymbol{\mu}_0)\,.$$

This CUSUM scheme signals when $G_i^2 \geq H$. We could not find in the literature any proposal for an analogous charting procedure in the case that the mean vector and the variance-covariance matrix have to be estimated.

Hawkins (1991) introduced CUSUMs for regression adjusted variables based on the idea that the most common situation in practice is that departures from control have some known structure. In particular, it is assumed that the mean shifts with magnitude δ in only one variable.

Consider the multiple regression of X_j, the jth variable of $\mathbf{x}$ on all other variables of $\mathbf{x}$. Let Z_j be the regression residual when the jth variable is regressed on all other variables, rescaled to unit variance. This may be used to test the hypothesis that there is not a shift in the μ_j against the alternative that there is. The regression residual Z_j is given by

$$\mathbf{Z} = [diag(\mathbf{\Sigma}^{-1})]^{-1/2}\mathbf{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_0)$$

whose null distribution is $N(0,1)$. Hawkins (1991, 1993) proposes to chart each Z_j using a CUSUM procedure because in general it is not known which of the p variables is out-of-control. For studying the p individual charts simultaneously, Hawkins (1991) proposed the following control charts

$$MCZ = \max(L_{i,j}^{+}, -L_{i,j}^{-}) \text{ and } ZNO = \sum_{j=1}^{p}(L_{i,j}^{+} + L_{i,j}^{-})^2,$$

where

$$L_{i,j}^{+} = \max\left(0, L_{i-1,j}^{+} + Z_{i,j} - k\right) \text{ and } L_{i,j}^{-} = \min\left(0, L_{i-1,j}^{-} + Z_{i,j} + k\right)$$

and $L_{0,j}^{+} = L_{0,j}^{-}$ $i = 1, 2, ..., m$. MCZ is the MCUSUM statistic introduced by Woodall and Ncube (1985) applied to the CUSUM for $\mathbf{Z}$. ZNO is the squared Euclidean norm of the resultant vectors of the CUSUM for upward and downward shifts in mean. The CUSUMs L^{+}, L^{-} test for shifts in location in the upward and downward directions, respectively. The plot of these CUSUMs on a common chart gives a powerful CUSUM control chart for location. An out-of-control signal occurs when any of these four CUSUMs exceeds the decision interval h. The values of h and k are selected as in any other CUSUM because this chart consists of separate random variables each following the $N(0,1)$ distribution. An out-of-control signal is indicated when MCZ and ZNO exceed a threshold value set to fix the in-control ARL. Hauck et al. (1999) applied multivariate statistical process monitoring and diagnosis with grouped regression-adjusted variables.

In the sequel, we present the directionally invariant CUSUM schemes. Crosier (1988) proposes two multivariate CUSUM schemes. The first CUSUM proposed by Crosier (1988) is a CUSUM of the scalars D_i, the square root of D_i^2 statistic, and is given by

$$G_i^3 = \max\left[\left(G_{i-1}^3 + D_i - K\right), 0\right]$$

where $G_0^3 \geq 0$ and $K \geq 0$. This scheme signals when $G_i^3 \geq H$, which is determined using the Markov chain approach. Crosier (1988) notes that a search for the optimal K produced a sequence that closely resemble the square root of the number of variables.

A similar CUSUM is proposed by Pignatiello and Runger (1990) defined as

$$G_i^4 = \max\left[0, G_{i-1}^4 + D_i^2 - k\right]$$

with $G_0^4 = 0$, and k chosen to be $0.5\lambda^2(\boldsymbol{\mu}_1) + p$, where p is the number of the variables. The process is out-of-control if G_i^4 exceeds an upper control limit H, which is determined using the Markov chain approach.

Crosier (1988) and Pignatiello and Runger (1990) found that ordinary one sided univariate CUSUMs based on successive values of D_i^2 or D_i statistic, respectively, do not have good ARL properties.

The second CUSUM proposed by Crosier (1988) is a CUSUM of vectors. A vector-valued scheme can be derived by replacing the scalar quantities of a univariate CUSUM scheme by vectors and is given by

$$G_i^5 = \left[\mathbf{g}_i^t \boldsymbol{\Sigma}_0^{-1} \mathbf{g}_i\right]^{1/2},$$

where

$$\mathbf{g}_i = \left\{ \begin{array}{cc} (\mathbf{g}_{i-1} + \mathbf{x}_i - \boldsymbol{\mu}_0)(1 - KC_i^{-1}) & \text{if } C_i > K \\ \mathbf{0} & \text{otherwise} \end{array} \right\}$$

and

$$C_i = \left[(\mathbf{g}_{i-1} + \mathbf{x}_i - \boldsymbol{\mu}_0)^t \, \boldsymbol{\Sigma}_0^{-1} \left(\mathbf{g}_{i-1} + \mathbf{x}_i - \boldsymbol{\mu}_0\right)\right]^{1/2}.$$

This scheme signals when $G_i^5 > H$, where H is chosen to provide a predefined in-control ARL, using simulation. Because of the fact that the ARL performance of this chart depends on the noncentrality parameter, Crosier (1988) recommends that $K = \lambda(\boldsymbol{\mu}_1)/2$ and $\mathbf{g}_0 = \mathbf{0}$. Both CUSUMs, as proposed by Crosier (1988) allow the use of recent enhancements in CUSUM schemes. Among the CUSUM schemes proposed by Crosier (1988) the vector-valued scheme has a better ARL performance than the scalar scheme.

The second CUSUM proposed by Pignatiello and Runger (1990) can be constructed by defining G_i^6 as

$$G_i^6 = \max\left\{\left[\mathbf{C}_i^t \boldsymbol{\Sigma}_0^{-1} \mathbf{C}_i\right]^{1/2} - kn_i, 0\right\},$$

where $G_0^6 = 0$, k is chosen to be $0.5\lambda(\boldsymbol{\mu}_1)$, $\boldsymbol{\mu}_1$ is a specified out-of-control mean, $\mathbf{C}_i$ equals

$$\mathbf{C}_i = \sum_{l=i-n_i+1}^{i} (\mathbf{x}_i - \boldsymbol{\mu}_0)$$

and n_i is the number of subgroups since the most recent renewal (i.e. zero value) of the CUSUM chart, formally defined as

$$n_i = \left\{ \begin{array}{c} n_{i-1} + 1, \text{ if } G_{i-1}^6 > 0 \\ 1, \text{otherwise} \end{array} \right\}.$$

This chart operates by plotting G_i^6 on a control chart with an upper control limit H which is again computed through simulation. If G_i^6 exceeds H then the process is out-of-control. Pignatiello and Runger (1990), proved that the ARL performance of the G_i^6 chart depends only on the square root of the noncentrality parameter and it is better in relation to G_i^5.

Ngai and Zhang (2001) gave a natural multivariate extension of the two-sided cumulative sum chart for controlling the process mean. Additionally, Chan and Zhang (2001) propose cumulative sum charts for controlling the covariance matrix.

2.7.2 Multivariate Exponentially Weighted Moving Average Charts

Let $\mathbf{x}_i^t$ be the ith p-dimensional observation. Also, assume that $\mathbf{x}_i$ follows a $N_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ with a known variance-covariance matrix $\boldsymbol{\Sigma}$ and a known p-dimensional mean vector $\boldsymbol{\mu}_0$. A multivariate EWMA control chart is proposed by Lowry et al. (1992) as follows

$$\mathbf{z}_i = \mathbf{R}(\mathbf{x}_i - \boldsymbol{\mu}_0) + (\mathbf{I} - \mathbf{R})\,\mathbf{z}_{i-1} = \sum_{j=1}^{i} \mathbf{R}\,(\mathbf{I} - \mathbf{R})^{i-j}\,(\mathbf{x}_j - \boldsymbol{\mu}_0),\ i = 1, 2, 3, ...,$$

where $\mathbf{R} = diag(r_1, r_2, ..., r_p)$, $0 \leq r_k \leq 1$ for $k = 1, 2, 3, ..., p$, and $\mathbf{I}$ is the identity matrix. If there is no a-priori reason to weight past observations differently for the p quality characteristics being monitored, then $r_1 = r_2 = ... = r_p = r$. The initial value $\mathbf{z}_0$ is usually obtained equal to the in-control mean vector of the process. It is obvious that if $\mathbf{R} = \mathbf{I}$ then the multivariate EWMA control chart is equivalent to the T^2-Chart. The multivariate EWMA chart gives an out-of-control signal if $\mathbf{z}_i^t \boldsymbol{\Sigma}_{\mathbf{z}_i}^{-1} \mathbf{z}_i > h$ where $\boldsymbol{\Sigma}_{\mathbf{z}_i}$ is the variance-covariance matrix of $\mathbf{z}_i$. The value h is calculated via simulation to achieve a specified in-control ARL. The ARL performance of the multivariate EWMA control chart depends only on the noncentrality parameter. This means that the multivariate EWMA has the property of directional invariance. The variance-covariance matrix of $\mathbf{z}_i$ is calculated by the following formula

$$\boldsymbol{\Sigma}_{\mathbf{z}_i} = \sum_{j=1}^{i} Var\left[\mathbf{R}\left(\mathbf{I} - \mathbf{R}\right)^{i-j}\left(\mathbf{x}_j - \boldsymbol{\mu}_0\right)\right] = \sum_{j=1}^{i} \mathbf{R}\left(\mathbf{I} - \mathbf{R}\right)^{i-j} \boldsymbol{\Sigma}\left(\mathbf{I} - \mathbf{R}\right)^{i-j} \mathbf{R}$$

or in case that $r_1 = r_2 = ... = r_p = r$

$$\boldsymbol{\Sigma}_{\mathbf{z}_i} = \left(1 - (1 - r)^{2i}\right) r / \left(2 - r\right) \boldsymbol{\Sigma}.$$

An approximation of the variance-covariance matrix $\boldsymbol{\Sigma}_{\mathbf{z}_i}$ when i approaches infinity, is the following

$$\boldsymbol{\Sigma}_{\mathbf{z}_i} = \frac{r}{2 - r} \boldsymbol{\Sigma}.$$

However, the use of the exact variance-covariance matrix of the multivariate EWMA, leads to a natural fast initial response for the multivariate EWMA chart.

In a univariate EWMA chart if the plotted statistic is on one side of the center line and a shift occurs on the other side the result is that the EWMA chart will need more observations until it signals. Such a problem is called inertia problem. Inertia problem may occur with the multivariate EWMA chart and the simultaneous use of a Shewhart type chart is proposed.

Lowry et al. (1992) studied the ARL of the multivariate EWMA. The ARL performance of the multivariate EWMA procedure depends only on $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ through the value of the noncentrality parameter. Since, the multivariate EWMA, the MCUSUM#1 and the vector CUSUM are all directionally invariant, these three charts can be compared to each other and to Hotelling's (1947) T^2-Chart. The comparison of these charts shows that the ARL performance of the multivariate EWMA is at least as good as those of vector-valued CUSUM and MCUSUM#1.

Rigdon (1995a, 1995b) gives an integral and a double integral equation for the calculation of in-control and out-of-control ARLs respectively. Moreover, Bodden and Rigdon (1999) developed a computer program for approximating the in-control ARL of the multivariate EWMA chart. Runger and Prabhu (1996) use a Markov chain approximation to determine the run length performance of the multivariate EWMA chart. This kind of analysis, can be applied whenever the multivariate control statistic can be modeled as a Markov chain and has the property of directional invariance. In addition, Prabhu and Runger (1997) provide recommendations for the selection of parameters for a multivariate EWMA chart. Molnau et. al. (2001) present a program that calculates the ARL for the multivariate EWMA when the values of the shift in the mean vector, the control limit and the smoothing parameter are known.

Kramer and Schmid (1997), proposed a generalization of the multivariate EWMA control scheme of Lowry et al. (1992) for multivariate time dependent observations. Sullivan and Woodall (1998) recommended the use of multivariate EWMA for the preliminary analysis of multivariate observations. Fasso (1999) developed a one-sided multivariate EWMA control chart, based on the restricted Maximum Likelihood Estimator.

Yumin (1996) proposed the construction of a multivariate EWMA using the principal components of the original variables. Runger et al. (1999) show how the shift detection capability of the multivariate EWMA can be significantly improved by transforming the original process variables to a lower-dimensional subspace through the use of the U transformation. The U transformation is similar to principal components transformation.

Margavio and Conerly (1995) developed two alternatives for the multivariate EWMA chart. The first of these alternatives is an arithmetic multivariate moving average while the second alternative is a truncated version of the multivariate EWMA.