THE UNIVERSITY OF MANITOBA

STATISTICAL QUALITY CONTROL:

UNIVARIATE AND MULTIVARIATE CASES

by

MARINA S. YOGENDRAN

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

Department of Statistics

WINNIPEG, MANITOBA

December, 1990

Canadä

STATISTICAL QUALITY CONTROL:
UNIVARIATE AND MULTIVARIATE CASES

BY

MARINA S. YOGENDRAN

A thesis submitted to the Faculty of Graduate Studies of
the University of Manitoba in partial fulfillment of the requirements
of the degree of

MASTER OF SCIENCE

© 1990

## TABLE OF CONTENTS

# ABSTRACT

Recently there has been a resurgence in using statistical techniques to improve the quality of the output of industrial production processes. Recognizing the random nature of the process, statistical procedures can be used to describe the process when it is in a state of "control". Using these baseline parameters, the process can then be judged as "in control" or "out of control".

Shewhart control charts are developed for the univariate and bivariate case. For most production processes, more than one related quality characteristics of the product has to be kept in control at the same time. Statistical process is then Multivariable. Although it is important to control each variable *individually,* it is necessary to consider the variables simultaneously. Since the variables are correlated with one another, individual control will not be sufficient to control the multivariable process. Hence the multivariate approach to quality control has been considered.

As an alternative to the Shewhart chart, the cumulative sum (CUSUM) control chart has been developed. The CUSUM chart tends to detect relatively small shifts in the process more effectively than the Shewhart charts. CUSUM control charts in the univariate and bivariate case are discussed here. The CUSUM control chart is developed by a sequential sampling procedure. The sequential probability ratio test (SPRT) is used to construct the V-mask Scheme to detect the shifts from in control to the out of control state.

To study the ARL of the CUSUM scheme the Markov chain approach has been considered. The transition probability matrix for this Markov chain is obtained and then the properties of this matrix are used to determine not only the average run lengths for the scheme, but also moments and percentage points of the run-length distribution and exact probabilities of run length. This method has been suggested for any discrete distribution

and is also as an accurate approximation with any continuous distribution for the random variable which is to be controled.

In dealing with bivariate data the role of correlation as measure of joint dispersion is very important. This area has received very little attention in the quality control literature. In the final chapter several different types of control charts are developed for the correlation coefficient. Both the exact and asymptotic distributions of the correlation coefficent are considered.

# ACKNOWLEDGEMENTS

I would like to express my most sincere thanks to my supervisor Professor Subrahmaniam Kocherlakota and to my advisor Professor Kathleen Kocherlakota for their advice, comments encouragement and criticisim. It is due to the constant stream of ideas and support recieved from them that made my stay at Winnipeg a beneficial part of my education.

My sincere thanks are due to Professors A. Thavaneswaran and S. Kocherlakota for their financial support from their grants given by the Natural Sciences and Engineering Council of Canada.

Most of all I am grateful to Professors K. Kocherlakota, S. Kocherlakota and A. Thavaneswaran for their guidance and patience during my stay here.

Finally, I thank my husband, my parents, my sister and brothers for their moral support given to me even though they themselves were living in the midst of a civil war.

# CHAPTER 0

# INTRODUCTION: SUMMARY

Recently there has been a resurgence in using statistical techniques to improve the quality of the output of industrial production processes. Recognizing the random nature of the process, statistical procedures can be used to describe the process when it is in a state of "control". Using these baseline parameters, the process can then be judged as "in control" or "out of control".

An excellent review of the history of statistical quality control and its impact on industry can be found in an article by A. J. Duncan(1986). Work in statistical quality began in North America in the 1920's with work at the Bell Telephone Laboratories under the leadership of W. E. Shewart, H. F. Dodge and H. G. Romig. During World War II the techniques were refined mainly by the Columbia University Research group under the leadership of A. Wald.

During the same period the British school was working on similar problems relating to the variation in quality of output and improving the quality using statistical procedures. Following World War II W. Edward Deming introduced statistical quality control procedures in Japan. This has led to development of products that consistantly maintain high quality. The demand in North America and Europe for products of quality equal to those produced in Japan has led to a renewed interest in the statistical community in the development of new and improved statistical techniques.

One of the main techniques developed for statistical quality control is the **control chart.** Control charts are used in deciding whether variations in the quality of the output of a process can be viewed as coming from "in control" process or whether the process has for some reason gone "out of control".

Generally a production process operates in an "in control" state producing acceptable products for a relatively long period of time. Occasionally a "shift" to an "out of control" state will occur and a larger proportion of nonacceptable items will be produced. A major objective of Statistical Quality Control is to detect this shift in the behaviour of the process and to return it to its "in control" state. Hence, the control chart is used as a monitoring, or on line, device to detect the shift in the process. The process is then adjusted to attain an acceptable level of quality.

In Chapter 1 univariate and bivariate control charts are discussed. Following the general model due to Shewhart (1931), univariate control charts are developed controlling for process means and process dispersion (variance and range).

The second section of this chapter deals with bivariate control charts. For most production processes, more than one related quality characteristics of the product have to be kept in control at the same time. Statistical process is then multivariable. Although it is important to control each variable *individually*, it is necessary to consider the variables simultaneously. Since the variables are correlated with one another, individual control will not be sufficient to control the multivariable process. The multivariate approach to quality control was originally considered by Hotelling(1947).

In Chapter 2 the control of two quality characteristics is considered both individually and simultaneously. When two related quality characteristics are considered, there are three parameters that have to be studied: the process mean, the process standard deviation and the correlation between the two characteristics. The control of the first two parameters mentioned are discussed in Chapter 2 and the latter is discussed in Chapter 3.

After a process has been judged to be in a state of statistical control, the effectiveness of a control chart in detecting departures from the "in control" state become important. As an alternative to the Shewhart chart, the cumulative sum (CUSUM) control chart was developed by Page (1954).

The CUSUM chart takes into account all the information in the sequence of observations, while the Shewhart Chart only uses the information about the process contained in the last plotted point. Thus the CUSUM chart tends to detect relatively small shifts in the process more effectively than the Shewhart charts. The CUSUM chart is used primarily to maintain the current control of the process. CUSUM control charts in the univariate and bivariate case are discussed in Chapter 2.

The CUSUM control chart is developed by a sequential sampling procedure. Suppose the observations are observed sequentially in time. Then the CUSUM control chart plots the cumulative sum of the sequence of observations against the number of observations. The change in trend on a CUSUM chart will indicate that a change has occurred in the process. It is not sufficient to detect a shift by visual inspection. The sequential probability ratio test (SPRT) is used to construct the so called V-mask Scheme.

The use of cumulated sums can be very effective in detecting shifts. The CUSUM will detect a shift in a process mean more quickly than a Shewhart chart. We say that its average run length (ARL) or the expected number of samples needed before the process mean plots "out of control" is less than for the Shewhart chart.

To study the ARL of the CUSUM scheme a different approach has been developed by Brook and Evans (1972). They consider the scheme as a Markov chain. The transition probability matrix for this Markov chain is obtained and then the properties of this matrix are used to determine not only the average run lengths for the scheme, but also moments and percentage points of the run-length distribution and exact probabilities of run length. This method has been suggested for any discrete distribution and is also as an accurate approximation with any continuous distribution for the random variable which is to be controlled.

As we have seen in dealing with bivariate data the role of correlation as measure of joint dispersion is very important. This area has received very little attention in the quality control literature. In Chapter 3 several different types of control charts are developed for

the correlation coefficient. Both the exact and the asymptotic distribution of the correlation coefficent are considered.

The Appendices contain the computer programs that were used. In some of the programs the statistical package SAS has been used. However, for the simulation studies the programs were written in FORTRAN incorporating some IMSL subroutines.

# CHAPTER 1

# SHEWHART CONTROL CHARTS

## 1.1  Univariate Quality Control Charts

In any production process, the quality of the product is determined by different characteristics of the product produced. The measure of the quality of the characteristic of interest is expected to be kept at a specified level throughout the production process. In any production process a certain amount of inherent or natural variability is present. The inherent variability is due to uncontrollable causes or "chance" causes. Another source of variability, which can be controlled may be present. This types of variability is said to be due to " assignable causes". A process that is operating in the presence of assignable causes is said to be "out of control".

Generally a production process operates in an "in control" state producing qualified products for a relatively long time. However, occasionally some assignable causes will occur making a " shift" to an out of control state, where a larger proportion of nonconforming units will be produced. A major objective of Statistical Quality Control is to quickly detect the shift in the process and to return it to its in control state after identifying an assignable cause and correcting it. Hence, the control chart is used as an on-line process control technique to detect the shift in the process.

### *1.1.1  General Model of a Shewhart Control Chart.*

Let w be a sample statistic that measures some quality characteristic of interest with mean $\mu_w$ and standard deviation $\sigma_w$. Then the control limits of a Shewhart Control Chart [Duncan (1974), Montgomery (1985)] are given by:

Upper Control Limit $\qquad$ $UCL = \mu_w + k\,\sigma_w$

Centre Line $\qquad$ $CL \;\;= \mu_w$

Lower Control Limit $\qquad$ $LCL \;= \mu_w - k\,\sigma_w$

where $k$ is the distance of the control limits from the centre line. Customarily $k$ is chosen to be 3, when the underlying distribution is normal, then the control limits are called " 3-sigma limits".

The above control limits are chosen so that if the process is in control, nearly all of the sample points will fall between the upper control limit and lower control limit. The process is assumed to be in control as long as the points plot within the control limits. However, a point that plots outside the control limits is interpreted as evidence that the process is out of control, and an assignable cause is sought.

The control chart is actually a test of the hypothesis that the process is in a state of control. In this context a Type I error is concluding the process is out of control when it is really in control. Similarly a Type II error is concluding the process is in control when it is really out of control.

W | UCL $= \mu_w + k\,\sigma_w$

CL $\;= \mu_w$

LCL $= \mu_w - k\,\sigma_w$

Samples or observations in the order of production
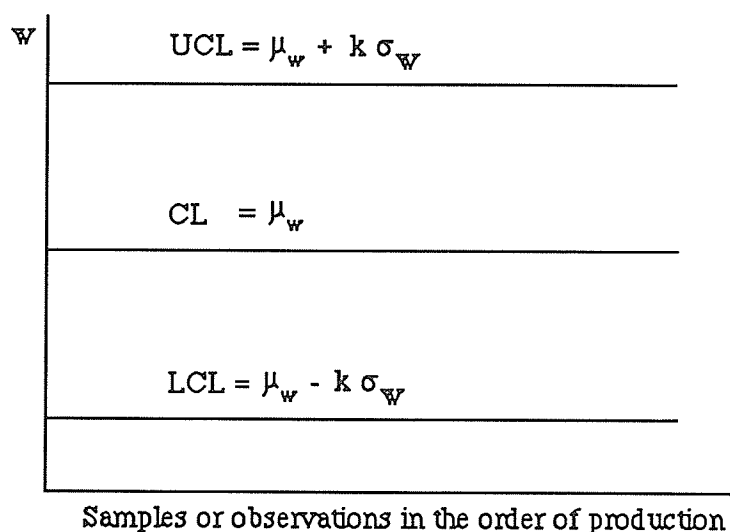
## 1.1.2 $\overline{\overline{X}}$ - Control Chart.

Suppose that the quality characteristic of interest is normally distributed with mean $\mu$ and standard deviation $\sigma$, where both $\mu$ and $\sigma$ are known. Let $x_1, x_2, \ldots, x_n$ be a sample of size n with average $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$. Then $\bar{x}$ is normally distributed with mean $\mu$ and standard deviation $\sigma/\sqrt{n}$. If the Type I error of testing the hypothesis that the process mean is at its specified value $\mu = \mu_0$ is $\alpha$, then it is expected that the $100(1-\alpha)\%$ of the sample mean $\bar{x}$'s should fall between $\mu_0 \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$. If $z_{\alpha/2} = 3$, then $\alpha=0.0027$ and the 3-sigma limits for the $\bar{x}$ - chart are given by

$$UCL = \mu_0 + 3\frac{\sigma}{\sqrt{n}}$$

$$CL \;\; = \mu_0$$

$$LCL = \mu_0 - 3\frac{\sigma}{\sqrt{n}}.$$

Generally, $\mu$ and $\sigma$ are unknown. Then they must be estimated from preliminary samples taken when the process is considered to be in control.


## 1.1.3 $\overline{\overline{X}}$ and R Control Charts.

Consider m samples, each containing n observations on the quality characteristic. Usually n is taken to be very small (n< 10). Let $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_m$ be the averages of the m samples. Then the best estimator of the process mean $\mu$ is the grand average

$$\bar{\bar{x}} = \frac{1}{m} \sum_{j=1}^{m} \bar{x}_j.$$

Thus $\bar{\bar{x}}$ would be the centre line on the $\bar{x}$ - chart. Now, to find an estimator for $\sigma$, two methods can be used, one is estimating from the sample standard deviations, and the other is estimating from the ranges of the m samples. In this section the method using ranges will be discussed.

Let $x_{i1}, x_{i2}, \ldots, x_{in}$ be the ith sample of size n and define

$$x_{i(max)} = \max(x_{i1}, x_{i2}, \ldots, x_{in})$$

$$x_{i(min)} = \min(x_{i1}, x_{i2}, \ldots, x_{in}).$$

Then the range $R_i$ of the ith sample is given by,

$$R_i = x_{i(max)} - x_{i(min)}$$

Then $W_i = R_i/\sigma$ is called the relative range of the ith sample. And the parameters of the distribution of W are functions of the sample size n.

To determine the control limits of the R-chart, we need the estimates of mean and standard deviation of R. The estimate of the mean of R can be taken as $\bar{R}$. Hence the centre line for this chart is $\bar{R}$ and the estimate of the standard deviation of R, $\hat{\sigma}_R$ can be found from the distribution of $W = R/\sigma$. The mean and standard deviation of W are $d_2$ and $d_3$ respectively, where $d_2$ and $d_3$ are known functions of the sample size n. The values of $d_2$ and $d_3$ for various sample sizes are given in [Duncan (1974), table M]. An estimator of $\sigma$ is given by

$$\hat{\sigma} = \frac{\bar{R}}{d_2}$$

where $\bar{R}$ is the average range;

$$\bar{R} = \frac{1}{m}\sum_{j=1}^{m} R_i.$$

Therefore

$$\hat{\sigma}_R = d_3 \hat{\sigma},$$

where $\hat{\sigma} = \frac{\bar{R}}{d_2}$.

That is,

$$\hat{\sigma}_R = d_3 \frac{\bar{R}}{d_2}.$$

Hence, the control limits for R - Chart with the usual 3-sigma limits are:

$$UCL = \bar{R} + 3 \; \hat{\sigma}_R = \bar{R} + 3 \, d_3 \, \frac{\bar{R}}{d_2}$$

$$CL = \bar{R}$$

$$LCL = \bar{R} - 3 \; \hat{\sigma}_R = \bar{R} - 3 \, d_3 \, \frac{\bar{R}}{d_2}$$

Let

$$D_3 = 1 - 3\frac{d_3}{d_2}, \quad D_4 = 1 + 3\frac{d_3}{d_2};$$

then the R chart control limits become:

$$UCL = \bar{R} \, D_4$$

$$CL = \bar{R}$$

$$LCL = \bar{R} \, D_3.$$

The constants $D_3$ and $D_4$ are tabulated for various values of n in [Duncan (1974), table M].

The m sample ranges should be plotted on a chart with these control limits, and if any points exceed the control limits, the assignable cause should be sought and corrected. A new set of control limits should be calculated for the remaining observations. The control limits of the $\bar{x}$ - Chart based on $\bar{R}$ as an estimator for $\sigma$ are given by:

$$UCL = \bar{\bar{x}} + \frac{3}{d_2\sqrt{n}} \, \bar{R}$$

$$CL = \bar{\bar{x}}$$

$$LCL = \bar{\bar{x}} - \frac{3}{d_2\sqrt{n}} \, \bar{R}.$$

Here the quantity

$$A_2 = \frac{3}{d_2\sqrt{n}}$$

is a constant that depends only on the sample size n; hence the control limits can be rewritten as:

$$\text{UCL} = \bar{\bar{x}} + A_2 \bar{R}$$

$$\text{CL} \;\; = \bar{\bar{x}}$$

$$\text{LCL} = \bar{\bar{x}} - A_2 \bar{R} \,.$$

The m sample averages should be plotted on a chart with the above control limits. If any of the points fall outside the limits, the process is considered to be "out of control" and an assignable cause is investigated.

The R-chart monitors process variability within each sample whereas the $\bar{X}$-chart measures variability between the samples. Hence when constructing $\bar{X}$ and R charts it is better to deal with the R-chart first and see whether the variability within samples is in control for all the samples. If the variability is in control, then the $\bar{X}$-chart can be constructed. If the R-chart indicates out of control, eliminate the assignable causes until it plots in control and then proceed with the $\bar{X}$-chart.

When preliminary samples are used to construct the $\bar{X}$ and R control charts, it is customary to treat the control limits as trial values.

### 1.1.4  $\bar{X}$ and  S  Control Charts.

When the sample size is moderately large, say n > 10, the Range method for estimating $\sigma$ loses statistical efficiency. In this case $\bar{X}$ and S charts are used, where the process standard deviation is estimated directly instead of indirectly through the use of R. Sometimes this S chart is refered to as the $\sigma$ chart.

From each preliminary samples, the sample mean $\bar{x}$ and the sample standard deviation s is calculated. The sample variance

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

is an unbiased estimate of $\sigma^2$; however, s is not an unbiased estimator of $\sigma$. An unbiased estimate is $c_4\sigma$, where $c_4$ is a constant depends only on n, that is:

$$c_4 = \sqrt{\frac{2}{n-1}} \; \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})} \; .$$

It can be shown that the variance of s:

$$\text{Var}(s) = \sigma^2 \, (1 - c_4^2) \; .$$

Using these results, $\bar{X}$ and S charts can be developed.

Consider the case where a standard value is given for $\sigma$. Since the mean of s is $c_4\sigma$, the centre line for the chart is $c_4\sigma$. Then the 3-sigma control limits for the s chart are:

$$\text{UCL} = c_4\sigma + 3\,\sigma\sqrt{1 - c_4^2}$$

$$\text{CL} \;\; = c_4\sigma$$

$$\text{LCL} = c_4\sigma - 3\,\sigma\sqrt{1 - c_4^2} \; .$$

Let

$$B_5 = c_4 - 3\sqrt{1 - c_4^2}, \quad B_6 = c_4 + 3\sqrt{1 - c_4^2} \; ;$$

where the constants $B_5$ and $B_6$ are tabulated in [Duncan (1974), table M] for each n.

Now the control limits become:

$$\text{UCL} = B_6\,\sigma$$

$$\text{CL} \;\; = c_4\sigma$$

$$\text{LCL} = B_5\,\sigma.$$

If no standard value is given for $\sigma$, then it must be estimated from the preliminary samples. Consider m preliminary samples each of size n, and let $s_i$ be the standard deviation of the $i^{th}$ sample. Then the average of the m sample standard deviations is,

$$\bar{s} = \frac{1}{m} \sum_{i=1}^{m} s_i \; .$$

The statistic $\dfrac{\bar{s}}{c_4}$ is an unbiased estimator of $\sigma$. Therefore the control limits of the S-chart would be:

$$UCL = \bar{s} + 3\frac{\bar{s}}{c_4}\sqrt{1-c_4^2}$$

$$CL = \bar{s}$$

$$LCL = \bar{s} + 3\frac{\bar{s}}{c_4}\sqrt{1-c_4^2}\ .$$

Let

$$B_3 = 1 - \frac{3}{c_4}\sqrt{1-c_4^2},\quad B_4 = 1 + \frac{3}{c_4}\sqrt{1-c_4^2}\ ;$$

where $B_3$ and $B_4$ are tabulated in [Duncan (1974), table M] for different values of n.

Now the control limits become,

$$UCL = B_4\,\bar{s}$$

$$CL = \bar{s}$$

$$LCL = B_3\,\bar{s}.$$

Note that $B_4 = \dfrac{B_6}{c_4}$ and $B_3 = \dfrac{B_5}{c_4}$.

When $\dfrac{\bar{s}}{c_4}$ is used to estimate $\sigma$, the control limits on the corresponding $\bar{X}$ chart would be:

$$UCL = \bar{\bar{x}} + \frac{3\,\bar{s}}{c_4\sqrt{n}}$$

$$CL = \bar{\bar{x}}$$

$$LCL = \bar{\bar{x}} - \frac{3\,\bar{s}}{c_4\sqrt{n}}\ .$$

Let the constant

$$A_3 = \frac{3}{c_4\sqrt{n}}$$

and the values of $A_3$ for each n is given in [Duncan (1974), table M], then the $\bar{X}$ chart control limits become:

$$UCL = \bar{\bar{x}} + A_3 \bar{s}$$

$$CL = \bar{\bar{x}}$$

$$LCL = \bar{\bar{x}} - A_3 \bar{s}.$$

### 1.1.6  $S^2$- chart.

The $S^2$ control chart is also sometimes used to detect process dispersion. Let $s^2$ be the sample variance. Since the underlying distribution of the production process is assumed to be normal the statistic $(n-1)s^2/\sigma^2$ is distributed as chi-square with (n-1) degrees of freedom. If the Type I error of testing the hypothesis that the process standard deviation is at its specified value $\sigma = \sigma_0$ is $\alpha$,

then

$$P \left\{ \chi^2_{n-1,1-\alpha/2} \leq (n-1) \frac{s^2}{\sigma_0^2} \leq \chi^2_{n-1,\alpha/2} \right\} = 1 - \alpha$$

where $\chi^2_{n-1,1-\alpha/2}$ and $\chi^2_{n-1,\alpha/2}$ denote the lower and upper $\alpha/2$ percentage points of the chi-square distribution on (n-1) degrees of freedom.

Then the $S^2$ chart control limits become,

$$UCL = \frac{\sigma_0^2}{n-1} \chi^2_{n-1,\alpha/2}$$

$$CL = \sigma_0^2$$

$$LCL = \frac{\sigma_0^2}{n-1} \chi^2_{n-1,1-\alpha/2}.$$

The above control limits are probability limits. If a standard value is not given for $\sigma^2$ then it should be estimated from the priliminary samples taken when the process is considered to be in control. Let $s_i^2$, $i = 1, 2, \ldots, m$ be the sample variance of the ith preliminary sample.

Then

$$\bar{s}^2 = \frac{1}{m} \sum_{i=1}^{m} s_i^2 \; .$$

Now the control limits are:

$$UCL = \frac{\bar{s}^2}{n\text{-}1} \; \chi^2_{n\text{-}1,\alpha/2}$$

$$CL \;\; = \bar{s}^2$$

$$LCL \;\; = \frac{\bar{s}^2}{n\text{-}1} \; \chi^2_{n\text{-}1,1-\alpha/2}.$$

Usually these $S^2$ control charts are plotted only with an upper control limit. Since a small variance is of advantageous in maintaining a control process, we can use only an upper control limit:

$$UCL = \frac{\sigma_0^2}{n\text{-}1} \; \chi^2_{n\text{-}1,\alpha}$$

when $\sigma = \sigma_0$ is given; otherwise,

$$UCL = \frac{\bar{s}^2}{n\text{-}1} \; \chi^2_{n\text{-}1,\alpha}.$$

These control limits are all valid only if the underlying distribution is normal.

## 1.2  Bivariate Control Charts.

For most production processes, more than one related quality characteristics of the product has to be kept in control at the same time. Statistical process is then Multivariable. Although it is important to control each variable *individually,* it is necessary to consider the variables simultaneously. Since the variables are correlated with one another, individual control will not be sufficient to control the multivariable process. The multivariate approach to quality control was originally considered by Hotelling (1947).

In this section the control of two quality characteristics is considered both individually and simultaneously. When two related quality characteristics are considered, there are three parameters that have to be studied: the process mean, the process standard deviation and the correlation between the two characteristics. The control of the first two parameters mentioned are discussed in the following sections and the later is discussed in the next chapter.

### *1.2.1  Control Charts for Process Means.*

Alt and Smith (1988) have given an excellent review of the multivariate process control technique which are currently available. Montgomery (1985) has illustrated the use of multivariate control charts using Hotelling's (1947) ideas. These techniques will be summarized for the bivariate case.

Consider two related quality characteristics. Let them be denoted by $X_1$ and $X_2$. Suppose that the underlying distribution of $X_1$ and $X_2$ is bivariate normal with mean vector $\underline{\mu}$ and variance- covariance matrix $\Sigma$; that is,

$$( X_1, X_2 ) \sim BVN ( \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho ).$$

The control of the process means of the characteristics $X_1$, $X_2$ can be monitored

individually via the usual $\overline{X}$ charts discussed in section 1.1. The $\overline{X}_1$ and $\overline{X}_2$ charts can be

plotted seperately to see whether the process mean is in control for $X_1$ and $X_2$ individually. This is equivalent to plotting them both on one chart as shown in figure 1.2.1, where $UCL_1$ and $LCL_1$ are the control limits for the $\overline{X}_1$ chart and $UCL_2$ and $LCL_2$ are the $\overline{X}_2$ chart control limits. The process is considered to be in control only if the sample means $\overline{x}_1$ and $\overline{x}_2$ fall within their respective control limits. When the two quality characteristics are considered simultaneously, the pair of means ($\overline{x}_1$, $\overline{x}_2$) should fall within the rectangular region. This can be very misleading, since the correlation is not taken into consideration and also the Type I error or the probability that a point plotting in control would not be equal to their claimed values for the individual $\overline{X}$ charts [Montgomery (1985), page 246]. When the correlation is taken into account, the control region will be elliptical in nature.
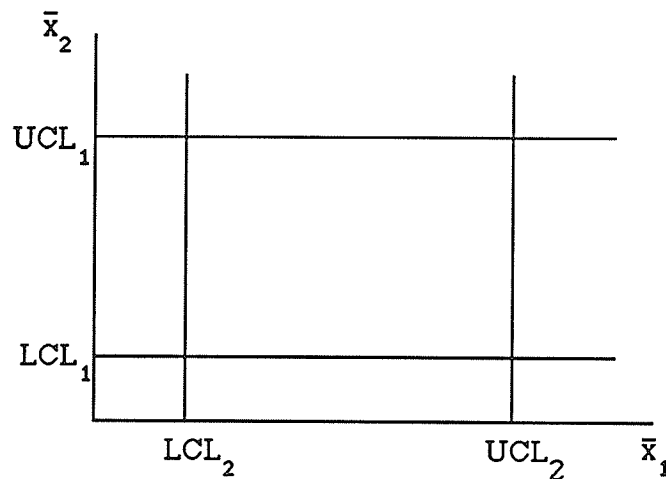


figure 1.2.1

First we will consider the case when the parameters are at their standard value. Let the standard values of $\underline{\mu}$ and $\Sigma$ be given by

$$\underline{\mu} = \underline{\mu}_0 = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

and

$$\Sigma = \Sigma_0 = \begin{bmatrix} \sigma_1^2 & \rho\,\sigma_1\sigma_2 \\ \rho\,\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

respectively. Suppose random samples of size n are taken and

$$\bar{x}_1 = \frac{1}{n}\sum_{i=1}^{n} x_{1i}, \quad \bar{x}_2 = \frac{1}{n}\sum_{i=1}^{n} x_{2i}$$

are determined, then a $\chi^2$ - chart is used to determine whether the process mean is at its specified value $\underline{\mu}_0$. Determining whether the process is in control is equivalent to testing the hypothesis:

$$H_0 : \underline{\mu} = \underline{\mu}_0 \quad vs \quad H_1 : \underline{\mu} \neq \underline{\mu}_0.$$

Here, instead of upper and lower control limits only an upper control limit is used. Since the quadratic form

$$n\,(\underline{\bar{x}} - \underline{\mu}_0)'\,\Sigma_0^{-1}\,(\underline{\bar{x}} - \underline{\mu}_0)$$

is distributed as $\chi^2$ with 2 degrees of freedom, the appropriate statistic is,

$$\chi_0^{\,2} = \frac{n}{(1-\rho^2)}\left\{ \left(\frac{\bar{x}_1 - \mu_1}{\sigma_1}\right)^2 + \left(\frac{\bar{x}_2 - \mu_2}{\sigma_2}\right)^2 \right.$$

$$\left. - 2\rho\left(\frac{\bar{x}_1 - \mu_1}{\sigma_1}\right)\left(\frac{\bar{x}_2 - \mu_2}{\sigma_2}\right) \right\}, \qquad (1.2.1)$$

where $\underline{\bar{x}} = (\bar{x}_1, \bar{x}_2)'$. Then for an $\alpha$ -level test $H_0$ is rejected if $\chi_0^{\,2} > \chi_{2,\alpha}^2$, where $\chi_{2,\alpha}^2$

is the upper $\alpha\%$ point of a chi square distribution with 2 degrees of freedom.

The equation 1.2.1 represents an ellipse centred at $(\mu_1, \mu_2)$ with the principal axes and orientation determined by $\rho$, $\sigma_1$ and $\sigma_2$. Therefore, a control region could be constructed such that the region is the interior and boundary of such an ellipse (figure 1.2.2). Hence using the control chart with UCL= $\chi_{2,\alpha}^2$ the control of the vector of process means could be monitored.

figure 1.2.2
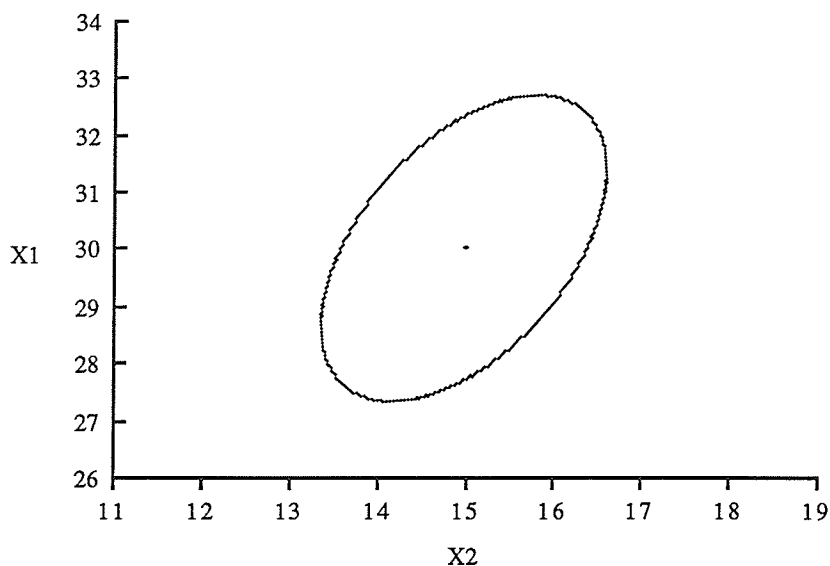


If a particular pair of sample means plots outside the elliptical region, the process is said to be out of control. We are now interested in which *one* (or *both*) of the characteristics may be causing the signal. Since the characteristics are correlated, it is not appropriate to test *each* of the means individually at the $\alpha$ level.

To control the overall type I error rate, Bonferroni tests or confidence intervals can be constructed [ Alt (1982), Morrison (1983) ]. Let $A_1, A_2, \ldots, A_p$ be p random events and $\bar{A}_1, \bar{A}_2, \ldots, \bar{A}_p$ their corresponding negations. Suppose that the $A_i$'s $(i = 1, 2, \ldots, p)$ and p hypotheses that we are interested in testing *simultaneously*. Then from the Bonferroni inequality

$$\Pr\left[\bigcap_{i=1}^{p} A_i\right] \geq 1 - \alpha$$

for a given family error rate $\alpha$ or confidence coefficient $1 - \alpha$. If we specify each individual test level at $\gamma = \alpha/p$, then

Pr ( one or more hypotheses will be rejected | all hypotheses are true )

$$= 1 - \Pr\left[\bigcap_{i=1}^{p} A_i\right] \geq p\left(\frac{\alpha}{p}\right) = \alpha.$$

The Bonferroni method of simultaneous tests and intervals merely requires that the error rate for the individual tests be divided by the number of tests p in the family. For this case, for p=2, individual $\bar{x}$ - charts for each characteristics with the probability of a type I error set at $\alpha/2$ will be constructed. The overall rate is then $\alpha$.

To illustrate this consider a set of bivariate data generated by simulation using SAS (Appendix A, program 1). Twenty samples of size n = 10 are generated from a normal population with parameters $\mu_1 = 30$, $\mu_2 = 15$, $\sigma_1^2 = 8$, $\sigma_2^2 = 4$ and $\rho = 0.5$. Then the sample averages for each characteristic $X_1$, $X_2$ and the $\chi_0^2$ statistic are calculated. The statistics $\bar{x}_1$, $\bar{x}_2$ and $\chi_0^2$ are tabulated in Table 1.1 for each sample.

Table 1.1

| Sample | $\bar{x}_1$ | $\bar{x}_2$ | $\chi_0^2$ |
|--------|-------------|-------------|------------|
| 1 | 30.5152 | 15.0318 | 0.40712 |
| 2 | 30.3481 | 15.3424 | 0.31185 |
| 3 | 31.0430 | 15.8084 | 2.00400 |
| 4 | 31.5144 | 15.9029 | 3.31713 |
| 5 | 31.1773 | 15.0891 | 2.08928 |
| 6 | 29.2462 | 14.6426 | 0.73775 |
| 7 | 31.6637 | 15.9283 | 3.84556 |
| 8 | 31.3846 | 15.3627 | 2.45006 |
| 9 | 31.8660 | 16.3256 | 5.83052 |
| 10 | 29.5721 | 15.5069 | 1.67306 |
| 11 | 31.1990 | 15.2104 | 1.94906 |
| 12 | 29.6743 | 14.8655 | 0.13382 |
| 13 | 29.8488 | 14.7799 | 0.12118 |
| 14 | 30.2235 | 15.2442 | 0.15337 |
| 15 | 30.4238 | 15.2223 | 0.24199 |
| 16 | 29.8453 | 16.0165 | 3.85466 |
| 17 | 29.9138 | 15.4282 | 0.71059 |

| 18 | 30.2293 | 15.1393 | 0.07701 |
| 19 | 29.9804 | 14.9192 | 0.01868 |
| 20 | 29.7927 | 16.0803 | 4.48931 |

The control limits for the $\bar{x}_1$ and $\bar{x}_2$ charts are,

$$UCL_1 = 30 + 3 \times 2/\sqrt{5} = 32.6833$$

$$LCL_1 = 30 - 3 \times 2/\sqrt{5} = 27.3167$$

and,

$$UCL_2 = 15 + 3 \times 2/\sqrt{10} = 16.8974$$

$$LCL_2 = 15 - 3 \times 2/\sqrt{10} = 13.1026$$

The upper control limit for the $\chi^2$ - chart is,

$$UCL = \chi^2_{2,0.0054} = 10.44.$$

All these sample statistics plots within the control limits; hence we conclude that the process mean for both characteristics is in control individually and simultaneously. Now, suppose we simulate (Appendix A, program 2) another set of 20 samples each of size n = 10, and we disturb the process mean for the last 15 samples as follows:

for samples from 6 to 10 , $\mu_1 = 30 + 0.65\, i$,

for samples from 11 to 15, $\mu_2 = 15 + 0.3\, i$,

and for samples from 16 to 20, $\mu_1 = 30 + 0.65\, i$;

$$\mu_2 = 15 + 0.3\, i, \qquad i = 1, \dots, 5.$$

Then the sample statistics are given in Table 1.2.

Table 1.2

| Sample | $\bar{x}_1$ | $\bar{x}_2$ | $\chi^2_0$ |
|---|---|---|---|
| 1 | 30.6513 | 16.0915 | 3.0025 |
| 2 | 29.6978 | 15.3289 | 0.7471 |
| 3 | 30.6084 | 14.8924 | 0.8099 |

| | | | |
|---|---|---|---|
| 4 | 29.7696 | 14.8658 | 0.0757 |
| 5 | 29.8697 | 14.2574 | 1.6383 |
| ------ | --------------- | --------------- | ------------ |
| 6 | 32.2180 | 14.8044 | 9.3489 |
| 7 | 32.1267 | 15.3518 | 6.1874 |
| 8 | 32.2015 | 15.8108 | 6.0617 |
| 9 | 31.6575 | 14.3945 | 8.1669 |
| 10 | 32.5314 | 14.7230 | 12.5888 * |
| ------ | --------------- | --------------- | ------------ |
| 11 | 30.0127 | 16.0900 | 3.9281 |
| 12 | 30.2065 | 15.2940 | 0.2161 |
| 13 | 30.0089 | 15.2117 | 0.1450 |
| 14 | 29.6574 | 16.6930 | 11.1168 * |
| 15 | 29.5477 | 16.0142 | 4.8507 |
| ------ | --------------- | --------------- | ------------ |
| 16 | 30.1623 | 15.1689 | 0.0744 |
| 17 | 30.3221 | 16.1253 | 3.5396 |
| 18 | 31.4625 | 15.5394 | 2.6754 |
| 19 | 33.0432 | 16.5076 | 12.1976 * |
| 20 | 32.5287 | 15.7286 | 8.0839 |

Though the sample process mean for characteristic $X_1$ has been increased by 0.65i, i = 1, ... , 5; from the 6th to the 10th samples, the $\bar{x}_1$ chart has not detected this shift but the $\chi^2$ - chart has detected it at the 10th sample. Similarly from the 11th to the 15th sample the process mean of the characteristic $X_2$ has been shifted by 0.3i, i = 1, ... , 5; and here again the $\bar{x}_2$ chart did not detect the shift but the $\chi^2$ - chart has detected the shift at the 14th sample. When both characteristics $X_1$ and $X_2$ are increased by 0.65i and 0.3i, i = 1, ... , 5; respectively through samples 16 to 20, the $\bar{x}_1$ chart was sensitive enough to detect the shift at the 19th sample like the $\chi^2$ - chart, though the $\bar{x}_2$ chart failed to detect the shift.

Hence from this example we see that, where two quality characteristics are considered the $\chi^2$ - chart is more sensitive in detecting the shift than the individual $\bar{x}$ charts.

Let us consider the case when the parameters are unknown. Suppose m preliminary samples, each of size n are taken. From these samples, the sample means, $\bar{x}_{1,i}$, $\bar{x}_{2,i}$ and the sample variances, $s_{1,i}^2$, $s_{2,i}^2$ of the characteristics $X_1$, $X_2$ and the sample covariance, $s_{12,i}$ of the two variables are calculated. Let

$$\bar{\bar{x}}_1 = \frac{1}{m} \sum_{i=1}^{m} \bar{x}_{1,i} \; , \quad \bar{\bar{x}}_2 = \frac{1}{m} \sum_{i=1}^{m} \bar{x}_{2,i}$$

and

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

where

$$S_i = \begin{bmatrix} s_{1,i}^2 & s_{12,i} \\ s_{12,i} & s_{2,i}^2 \end{bmatrix},$$

and also let

$$\underline{\bar{x}}_i = (\bar{x}_{1,i}, \bar{x}_{2,i})', \; i = 1,2, \ldots,m.$$

Then

$$\underline{\bar{\bar{x}}} = \begin{pmatrix} \bar{\bar{x}}_1 \\ \bar{\bar{x}}_2 \end{pmatrix} = \frac{1}{m} \sum_{i=1}^{m} \underline{\bar{x}}_i$$

is an unbiased estimator of $\underline{\mu}$, and $\bar{S}$ is an unbiased estimator of $\Sigma_0$. Hence we replace $\underline{\mu}$ and $\Sigma_0$ by $\underline{\bar{\bar{x}}}$ and $\bar{S}$ respectively. Then the resulting statistic is

$$T_0^2 = n (\underline{\bar{x}}_i - \underline{\bar{\bar{x}}})' \bar{S}^{-1} (\underline{\bar{x}}_i - \underline{\bar{\bar{x}}}) \tag{1.2.2}$$

$$= \frac{n s_1^2 s_2^2}{\det(\bar{S})} \left[ \left( \frac{\bar{x}_{1,i} - \bar{\bar{x}}_1}{\bar{s}_1} \right)^2 + \left( \frac{\bar{x}_{2,i} - \bar{\bar{x}}_2}{\bar{s}_2} \right)^2 \right.$$

$$\left. - 2 \bar{s}_{12} \left( \frac{\bar{x}_{1,i} - \bar{\bar{x}}_1}{\bar{s}_1} \right) \left( \frac{\bar{x}_{2,i} - \bar{\bar{x}}_2}{\bar{s}_2} \right) \right],$$

where

$$\bar{s}_1^2 = \frac{1}{m} \sum_{i=1}^{m} s_{1,i}^2, \quad \bar{s}_2^2 = \frac{1}{m} \sum_{i=1}^{m} s_{2,i}^2, \quad \bar{s}_{12} = \frac{1}{m} \sum_{i=1}^{m} s_{12,i}$$

and

$$\det(\bar{S}) = \bar{s}_1^2 \, \bar{s}_2^2 - \bar{s}_{12}^2.$$

$T_0^2$ is distributed as $\dfrac{2(m-1)(n-1)}{mn-m-1} F_{2,mn-m-1}$. Hence the $T_0^2$ statistics for the m preliminary

samples are plotted on a chart with

$$UCL = \frac{2(m-1)(n-1)}{mn-m-1} F_{2,mn-m-1,\alpha} \quad \text{and} \quad LCL = 0.$$

If one or more of the preliminary samples $T_0^2$ statistic plots out of control then they are

discarded and based on the rest of the preliminary samples, $\bar{\bar{x}}$ and $\bar{S}$ are recalculated and

again tested for control. This process is repeated until all the samples plot in control. Then

these in control values of $\bar{\bar{x}}$ and $\bar{S}$ are used as estimates of $\underline{\mu}$ and $\Sigma_0$. Now suppose a

new set of m samples each of size n are taken from the process. Let $\bar{x}_f$ denote the 2x1

vector of sample means, then by substituting $\bar{x}_f$ instead of $\bar{x}_i$ in equation(1.2.2), the

test statistic becomes

$$T_{0f}^2 = n \, (\bar{x}_f - \bar{\bar{x}})' \bar{S}^{-1} (\bar{x}_f - \bar{\bar{x}}) \tag{1.2.3}$$

And now $T_{0f}^2$ is distributed as $\dfrac{2(m+1)(n-1)}{mn-m-1} F_{2,mn-m-1}$. Hence $T_{0f}^2$ of each sample mean

is plotted on a control chart with

$$UCL = \frac{2(m+1)(n-1)}{mn-m-1} F_{2,mn-m-1,\alpha} \quad \text{and} \quad LCL = 0.$$

To illustrate this, consider the example where 10 preliminary samples each of size

10 are generated (Appendix A, program 3) with parameters

$$\mu_1 = 30, \quad \mu_2 = 15, \quad \sigma_1^2 = 8, \quad \sigma_2^2 = 4 \quad \text{and} \quad \rho_0 = 0.5,$$

except that the 3rd and 5th samples are generated as follows:

for the 3rd sample, $\mu_1 = 33$ and $\mu_2 = 17$,

for the 5th sample, $\mu_1 = 27$ and $\mu_2 = 13$.

Then the test statistic $T_0^2$ for each sample is given in Table 1.3.

Table 1.3

| Sample | $T_0^2$ | |
|---|---|---|
| 1 | 1.0433 | |
| 2 | 1.4317 | |
| 3 | 17.0695 | * |
| 4 | 4.2399 | |
| 5 | 9.9870 | * |
| 6 | 0.8851 | |
| 7 | 0.2356 | |
| 8 | 0.0054 | |
| 9 | 1.3498 | |
| 10 | 0.6277 | |

And the control limits for this chart are

UCL = 5.461 and LCL = 0.

The 3rd and 5th sample statistic falls outside the control limits. Hence we discard those two samples and recompute the statistic based on the remaining 8 samples (Appendix A, program 4), and the $T_0^2$ staistic for the 8 samples are listed in Table 1.4.

Table 1.4

| Sample | $T_0^2$ |
|---|---|
| 1 | 0.54296 |
| 2 | 1.04689 |
| 3 | 0.40840 |
| 4 | 2.29616 |
| 5 | 0.42761 |

|   |         |
|---|---------|
| 6 | 0.57674 |
| 7 | 0.11597 |
| 8 | 0.00328 |

The new control limits based on the 8 samples are

$$UCL = 5.324 \quad \text{and} \quad LCL = 0.$$

Now all the samples plot incontrol. Therefore the sample statistics based on these 8 samples are calculated:

$$\bar{\bar{x}} = \begin{pmatrix} 29.511 \\ 14.906 \end{pmatrix}$$

and

$$\bar{S} = \begin{pmatrix} 11.398 & 4.227 \\ 4.227 & 5.934 \end{pmatrix}$$

Now we generate (Appendix A, program 5) a new set of 10 samples each of size 10. Then the $T_{0f}^2$ statistic for these future samples are given in Table 1.5.

Table 1.5

| Sample | $T_{0f}^2$ |
|--------|------------|
| 1 | 0.80445 |
| 2 | 0.37780 |
| 3 | 2.32445 |
| 4 | 0.46154 |
| 5 | 0.06703 |
| 6 | 0.11760 |
| 7 | 1.08576 |
| 8 | 1.82544 |
| 9 | 7.25430 |
| 10 | 2.17544 |

The control limits for this chart are,

$$UCL = 6.674 \quad \text{and} \quad LCL = 0.$$

Here all the samples plot in control.

The sample statistics $\bar{\bar{x}}$ and $\bar{S}$ should be updated from time to time by taking preliminary samples when the control charts indicate in control state.

### 1.2.2 Control Charts for Process Dispersion.

As in the process mean case, the control of the process dispersion can be monitored by constructing S-charts and $S^2$-charts individually as discussed in section 1.1. Here again the correlation of the two variables is ignored. Hence we construct charts to monitor the control of process dispersion simultaneously. There are various kinds of control charts to detect process dispersion. That is basically all these charts check whether the covariance matrix of the process remain at the standard value $\Sigma_0$.

Suppose a random sample of size n is obtained and the sample variance-covariance matrix is computed from those samples. Let it be denoted by S.

$$S = \begin{bmatrix} s_1^2 & s_{12} \\ s_{12} & s_2^2 \end{bmatrix}$$

$$= \frac{1}{n-1} \begin{bmatrix} \sum_{k=1}^{n} (x_{1k} - \bar{x}_1)^2 & \sum_{k=1}^{n} (x_{1k} - \bar{x}_1)(x_{2k} - \bar{x}_2) \\ \sum_{k=1}^{n} (x_{1k} - \bar{x}_1)(x_{2k} - \bar{x}_2) & \sum_{k=1}^{n} (x_{2k} - \bar{x}_2)^2 \end{bmatrix}$$

The sample correlation coefficient for the two variables denoted by $r_{12}$ is given by

$$r_{12} = s_{12}/s_1 s_2.$$

Then the sample generalized variance is

$$|S| = s_1^2 s_2^2 - s_{12}^2$$

$$= s_1^2 s_2^2 (1 - r_{12}^2)$$

(i) $\left|S\right|^{1/2}$- chart : Probability limits.

The first chart we'll consider is the $\left|S\right|^{1/2}$- chart. Hoel (1937) has shown that the statistic

$$\frac{2(n-1)\left|S\right|^{1/2}}{\left|\Sigma_0\right|^{1/2}},$$

where

$$\left|\Sigma_0\right|^{1/2} = \sigma_1\,\sigma_2\,(1-\rho_0^2)^{1/2}$$

has a $\chi^2$ distribution with $2n-4$ degrees of freedom. Hence, the control limit of the $\left|S\right|^{1/2}$- chart becomes,

$$UCL = \frac{1}{2(n-1)}\left|\Sigma_0\right|^{1/2}\chi^2_{2n-4,\alpha/2}$$

$$LCL = \frac{1}{2(n-1)}\left|\Sigma_0\right|^{1/2}\chi^2_{2n-4,1-\alpha/2}.$$

If the sample statistic $\left|S\right|^{1/2}$ plots outside these limits, then the dispersion of the process is said to be out of control, and the assignable causes are sought. Note that these control limits are probability limits.

(ii) $\left|S\right|^{1/2}$- chart : 3 - sigma limits.

Another way of controlling the process dispersion is by constructing 3-sigma limits chart. For this the first two moments of $\left|S\right|^{1/2}$ are used. Since the statistic $\left|S\right|$ is distributed as

$$(n-1)^{-2}\left|\Sigma_0\right|\prod_{k=1}^{2}\chi^2_{n-k},$$

it follows that the rth moment of $\left|S\right|$ is given by

$$\left|S\right|^r = (n-1)^{-2r}\left|\Sigma_0\right|^r\prod_{k=1}^{2}Y_k^r,$$

where the chi-square variables $Y_k$'s are independent, and

$$E(Y_k^r) = 2^r \frac{\Gamma(\frac{n-k}{2} + r)}{\Gamma(\frac{n-k}{2})}, \quad \text{for } k = 1, 2.$$

Therefore,

$$E(|S|^r) = (n-1)^{-2r} |\Sigma_0|^r \; 2^{2r} \prod_{k=1}^{2} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-2}{2})}.$$

For $r = \frac{1}{2}$,

$$E(|S|^{1/2}) = \frac{2}{n-1} \frac{\Gamma(\frac{n-k}{2} + r)}{\Gamma(\frac{n-k}{2})} |\Sigma_0|^{1/2} = b_3 |\Sigma_0|^{1/2},$$

where

$$b_3 = \frac{2}{n-1} \frac{\Gamma(\frac{n-k}{2} + r)}{\Gamma(\frac{n-k}{2})}.$$

If $r = 1$,

$$E(|S|) = \frac{n-2}{n-1} |\Sigma_0| = b_1 |\Sigma_0|,$$

where

$$b_1 = \frac{n-2}{n-1}.$$

Therefore,

$$V(|S|^{1/2}) = E(|S|) - \left\{ E(|S|^{1/2}) \right\}^2 = (b_1 - b_3^2) |\Sigma_0|.$$

Hence, for the $|S|^{1/2}$- chart, three sigma control limits would be:

$$UCL = (b_3 + 3 \sqrt{b_1 - b_3^2}) |\Sigma_0|^{1/2},$$

$$CL = b_3 |\Sigma_0|^{1/2},$$

$$LCL = (b_3 - 3 \sqrt{b_1 - b_3^2}) |\Sigma_0|^{1/2}.$$

If the $\left|S\right|^{1/2}$ statistic calculated for each sample plots inside the control limits, the process is considered to be in control. If one or more of the sample statistic plots outside the limits the process is considered to be out of control and assignable causes are sought.

(iii) $W^*$ chart.

The final chart to monitor process dispersion is an analogue of the $S^2$ - chart. Anderson (1984) has shown that the Likelihood Ratio Test of

$$H_0 : \Sigma = \Sigma_0 \quad \text{vs} \quad H_1 : \Sigma \neq \Sigma_0$$

is modified to be unbiased, based on the following statistic

$$W^* = -2(n-1) - (n-1)\ln(\left|S\right|) + (n-1)\ln(\left|\Sigma_0\right|) + (n-1)\,\text{trace}(\Sigma_0^{-1} S),$$

$$(1.2.5)$$

where

$$\text{trace}(\Sigma_0^{-1} S) = \frac{1}{(1-\rho_0^2)}\left[\frac{s_1^2}{\sigma_1^2} + \frac{s_2^2}{\sigma_2^2} - 2\rho_0\frac{s_{12}}{\sigma_1\sigma_2}\right]$$

The upper 5% and 1% points of the distribution of $W^*$ are tabulated in [Anderson (1984), table7], under p=2, for various values of n-1.

To illustrate all these control charts discussed for process dispersion, we will consider the same example which we discussed earlier for process mean. For process dispersion we will construct the S-chart and $S^2$-chart for the two variables individually; for simultaneous control, the $\left|S\right|^{1/2}$- charts and the $W^*$ - chart are considered.

Again 20 samples each of size 10 are simulated (Appendix A, program 6), and the sample statistics necessary for these charts are calculated and tabulated in Table 1.6.

Table 1.6

| Sample | $s_1$ | $s_2$ | $s_1^2$ | $s_2^2$ | $\lvert S \rvert^{1/2}$ | $w^*$ |
|---|---|---|---|---|---|---|
| 1 | 3.37755 | 1.28291 | 11.4079 | 1.64586 | 4.15806 | 4.41484 |
| 2 | 2.17111 | 1.18312 | 4.7137 | 1.39978 | 1.77734 | 7.58624 |
| 3 | 2.31765 | 1.67198 | 5.3715 | 2.79552 | 3.10586 | 1.73137 |
| 4 | 2.82211 | 2.08354 | 7.9643 | 4.34113 | 5.25658 | 0.11231 |
| 5 | 2.51136 | 2.10146 | 6.3069 | 4.41614 | 5.27735 | 3.28196 |
| 6 | 1.95998 | 1.26131 | 3.8415 | 1.59091 | 2.39765 | 4.11898 |
| 7 | 3.59024 | 1.96691 | 12.8898 | 3.86873 | 7.01105 | 4.69817 |
| 8 | 2.67260 | 2.06409 | 7.1428 | 4.26045 | 5.18438 | 0.47737 |
| 9 | 2.66152 | 2.05126 | 7.0837 | 4.20768 | 4.97450 | 0.20130 |
| 10 | 1.85617 | 1.51116 | 3.4454 | 2.28362 | 2.65859 | 3.12394 |
| 11 | 2.91446 | 2.09679 | 8.4941 | 4.39651 | 4.32168 | 1.02222 |
| 12 | 2.11859 | 2.12041 | 4.4884 | 4.49613 | 3.66858 | 1.92709 |
| 13 | 2.96494 | 1.94756 | 8.7909 | 3.79299 | 5.28536 | 0.26536 |
| 14 | 2.95900 | 2.36012 | 8.7557 | 5.57015 | 4.86058 | 1.34825 |
| 15 | 2.98815 | 2.67485 | 8.9291 | 7.15483 | 5.79054 | 2.16091 |
| 16 | 3.85218 | 2.47527 | 14.8393 | 6.12695 | 7.11698 | 2.45625 |
| 17 | 3.11956 | 2.61704 | 9.7317 | 6.84888 | 5.30380 | 2.54901 |
| 18 | 3.24793 | 2.34258 | 10.5491 | 5.48768 | 4.16097 | 3.71294 |
| 19 | 3.09140 | 2.04355 | 9.5568 | 4.17609 | 3.80567 | 2.71231 |
| 20 | 2.11664 | 1.79249 | 4.4802 | 3.21302 | 3.14665 | 1.83096 |

The control limits for these charts are tabulated below

| | | UCL | LCL |
|---|---|---|---|
| $S_1$- chart | - | 4.721 | 0.781 |
| $S_2$- chart | - | 3.338 | 0.552 |
| $S_1^2$-chart | - | 22.36 | |
| $S_2^2$-chart | - | 11.18 | |
| $\lvert S \rvert^{1/2}$- chart  (i) probability limits | - | 7.838 | 1.881 |
| (ii) 3 - sigma limits | - | 8.97 | 0.00 |
| $W^*$ - chart ($\alpha = 0.05$) | - | 8.52 | |

An examination of the results for the twenty samples indicates that all the sample statistics are in control. Now we generate another set of 20 samples each of size 10 and disturb the process variance of each variable for the last 15 samples as follows:

for samples from 6 to 10, $\sigma_1^2 = 8*0.4*i$ , i=1,...,5

for samples from 11 to 15, $\sigma_2^2 = 4*0.2*i$ , i=1,...,5

and for samples from 16 to 20, $\sigma_1^2 = 8*0.4*i$ ;

$$\sigma_2^2 = 4*0.2*i , \quad i = 1,...,5.$$

A summary of the sample statistics are given in Table 1.7.

Table 1.7

| Sample | $s_1$ | $s_2$ | $s_1^2$ | $s_2^2$ | $\lvert s \rvert^{1/2}$ | $w^*$ | |
|--------|-------|-------|---------|---------|----------|-------|---|
| 1 | 2.31372 | 1.86484 | 5.3533 | 3.4776 | 4.0372 | 0.7159 | |
| 2 | 2.78896 | 1.95435 | 7.7783 | 3.8195 | 5.1747 | 0.5083 | |
| 3 | 2.83484 | 1.26325 | 8.0363 | 1.5958 | 2.9258 | 3.7397 | |
| 4 | 2.82750 | 1.46452 | 7.9948 | 2.1448 | 3.8003 | 1.5087 | |
| 5 | 2.54312 | 2.28576 | 6.4675 | 5.2247 | 4.4611 | 1.1550 | |
| 6 | 1.85864 | 2.77648 | 3.4545 | 7.7088 | 5.1072 | 11.1272 | * |
| 7 | 1.97094 | 2.21119 | 3.8846 | 4.8893 | 2.1995 | 8.9278 | * |
| 8 | 2.67969 | 1.32772 | 7.1807 | 1.7628 | 3.3009 | 2.3503 | |
| 9 | 4.42383 | 3.19566 | 19.5703 | 10.2123 | 11.8527 * | 9.7435 | * |
| 10 | 3.91406 | 2.11005 | 15.3199 | 4.4523 | 7.5983 | 3.5710 | |
| 11 | 1.55054 | 0.67928 | 2.4042 | 0.4614 | 1.0278 | 14.6112 | * |
| 12 | 2.63107 | 1.42597 | 6.9225 | 2.0334 | 3.2622 | 1.8721 | |
| 13 | 2.84769 | 1.45584 | 8.1093 | 2.1195 | 2.9763 | 3.3704 | |
| 14 | 3.38850 | 2.00686 | 11.4819 | 4.0275 | 5.0404 | 1.1096 | |

| 15 | 3.10072 | 1.87821 | 9.6145 | 3.5277 | 4.0554 | 1.5397 |
|----|---------|---------|--------|--------|--------|--------|
| 16 | 1.74575 | 1.12331 | 3.0476 | 1.2618 | 1.6536 | 7.6701 |
| 17 | 1.50127 | 1.33493 | 2.2538 | 1.7820 | 2.0003 | 7.1100 |
| 18 | 3.45474 | 1.49705 | 11.9352 | 2.2412 | 4.0734 | 3.1878 |
| 19 | 3.98015 | 1.80834 | 15.8416 | 3.2701 | 5.7561 | 3.5043 |
| 20 | 4.34182 | 1.48051 | 18.8514 | 2.1919 | 5.9633 | 8.2231 |

When the process variance $\sigma_1^2$ is shifted by 8*0.4*i (i=1,..,5), for the samples through 6 to 10, the $W^*$ - chart(at 5% level) detects the shift immediately at the 6th sample and the $|S|^{1/2}$- chart detects the shift at the 9th sample. None of the univariate charts detects the shift. For the 9th sample all the statistics are near to the UCL, although the second variable process variance is not shifted. When the process variance $\sigma_2^2$ is shifted by 4*0.2*i (i=1,..,5), for the samples through 11 to 15 only the $W^*$ - chart(at 5% level) detects the shift at the 11th sample. When both process standard deviations are shifted for samples 16 to 20, not one of the charts detectes the shift.

(iv) Charts using estimates.

When the parameters of the Bivariate normal distrbution are not specified, we use the sample estimates $\overline{\overline{x}}$ and $\overline{S}$ instead of $\underline{\mu}$ and $\Sigma_0$ respectively. For the $|S|^{1/2}$- chart , the unbiased estimate of $|\Sigma_0|^{1/2}$ has to be determined.

Let $|S^*|^{1/2}$ be the average of the m preliminary sample statistics $|S|^{1/2}$; that is,

$$|S^*|^{1/2} = \frac{1}{m} \sum_{i=1}^{m} |S_i|^{1/2}.$$

Then an unbiased estimate of $|\Sigma_0|^{1/2}$ is given by,

$$\frac{1}{b_3} |S^*|^{1/2}$$

Hence the probability limits of $|S|^{1/2}$- chart become:

$$UCL = \frac{\left|S^*\right|^{1/2}}{2b_3(n-1)} \; \chi^2_{2n-4,\alpha/2}$$

$$LCL = \frac{\left|S^*\right|^{1/2}}{2b_3(n-1)} \; \chi^2_{2n-4,1-\alpha/2}.$$

Similarly 3 - sigma limits of the $\left|S\right|^{1/2}$- chart become:

$$UCL = ( 1 + \frac{3}{b_3} \sqrt{b_1 - b_3^2} ) \left|S^*\right|^{1/2} ,$$

$$CL = \left|S^*\right|^{1/2} ,$$

$$LCL = ( 1 - \frac{3}{b_3} \sqrt{b_1 - b_3^2} ) \left|S^*\right|^{1/2} .$$

For the $W^*$ chart, the unbiased estimates of $\left|\Sigma_0\right|$ and $\Sigma_0^{-1}$ are needed. Let $\left|S_*\right|$ be the average of $\left|S\right|$ from the m preliminary samples. That is,

$$\left|S_*\right| = \frac{1}{m} \sum_{i=1}^{m} \left|S_i\right| ,$$

and

$$E\left|S_*\right| = b_1 \left|\Sigma_0\right| .$$

Hence $\dfrac{\left|S_*\right|}{b_1}$ is an unbiased estimate of $\left|\Sigma_0\right|$. Now, let $S_i^{-1}$ denote the inverse of the

sample variance-covariance matrix for subgroup i, i = 1, 2, ..., m. Kshirsagar (1971) shows that $\dfrac{n-4}{n-1} \; S_i^{-1}$ is an unbiased estimate of $\Sigma_0^{-1}$. And if,

$$S_*^{-1} = \frac{1}{m} \sum_{i=1}^{m} S_i^{-1},$$

then $\dfrac{n-4}{n-1} \; S_*^{-1}$ is an unbiased estimate of $\Sigma_0^{-1}$. Hence we substitute these unbiased

estimates in equation (1.2.5) the $W^*$ statistic becomes:

$$W^{**} = - 2(n-1) - (n-1) \ln(\left|S\right|) + (n-1) \ln(\frac{\left|S_*\right|}{b_1}) + (n-1) \operatorname{trace}(\frac{n-4}{n-1} \; S_*^{-1} \; S),$$

The values of the $W^{**}$ will be plotted on a control chart and the upper 5% and 1% points of the distribution of $W^{**}$ are tabulated in [Anderson (1984) table 7], under p=2, for various values of n-1.

Now consider the example done earlier, when standard values are not given , we use the same estimated values that were calculated when the samples were all in control. The estimates are,

$$\bar{\bar{x}} = \begin{pmatrix} 29.511 \\ 14.906 \end{pmatrix}$$

and

$$\bar{S} = \begin{pmatrix} 11.398 & 4.227 \\ 4.227 & 5.934 \end{pmatrix}$$

And the sample statistics for the S-charts, $S^2$-charts, $|S|^{1/2}$- chart and the $W^*$ - chart are computed (Appendix A, program 8) and are given in Table 1.8.

Table 1.8

| Sample | $s_1$ | $s_2$ | $s_1^2$ | $s_2^2$ | $|S|^{1/2}$ | $W^*$ |
|--------|-------|-------|---------|---------|-------------|-------|
| 1 | 4.03154 | 1.56393 | 16.2534 | 2.4459 | 5.9877 | 5.9155 |
| 2 | 2.59150 | 1.45283 | 6.7159 | 2.1107 | 2.5594 | 4.2348 |
| 3 | 2.76641 | 2.03969 | 7.6530 | 4.1603 | 4.4725 | 0.3019 |
| 4 | 3.36856 | 2.53436 | 11.3472 | 6.4230 | 7.5696 | 2.0835 |
| 5 | 2.99763 | 2.53590 | 8.9858 | 6.4308 | 7.5995 | 6.4842 |
| 6 | 2.33949 | 1.52969 | 5.4732 | 2.3399 | 3.4527 | 1.5306 |
| 7 | 4.28542 | 2.38054 | 18.3648 | 5.6670 | 10.0961 | 10.4273 |
| 8 | 3.19009 | 2.50522 | 10.1767 | 6.2761 | 7.4656 | 2.4601 |
| 9 | 3.17687 | 2.49269 | 10.0925 | 6.2135 | 7.1634 | 1.7791 |
| 10 | 2.21557 | 1.83287 | 4.9088 | 3.3594 | 3.8284 | 1.0071 |
| 11 | 3.47878 | 2.56306 | 12.1019 | 6.5693 | 6.2233 | 2.0086 |
| 12 | 2.52881 | 2.57799 | 6.3949 | 6.6460 | 5.2829 | 2.0691 |
| 13 | 3.53904 | 2.36925 | 12.5248 | 5.6134 | 7.6111 | 2.2336 |

| 14 | 3.53195 | 2.88176 | 12.4747 | 8.3046 | 6.9994 | 3.5278 |
| 15 | 3.56675 | 3.26042 | 12.7217 | 10.6303 | 8.3385 | 6.2049 |
| 16 | 4.59807 | 3.02793 | 21.1422 | 9.1683 | 10.2486 | 7.9417 |
| 17 | 3.72360 | 3.19568 | 13.8652 | 10.2123 | 7.6376 | 6.0850 |
| 18 | 3.87683 | 2.87015 | 15.0298 | 8.2377 | 5.9919 | 5.7436 |
| 19 | 3.68999 | 2.50536 | 13.6160 | 6.2769 | 5.4803 | 3.4425 |
| 20 | 2.52648 | 2.18176 | 6.3831 | 4.7601 | 4.5313 | 0.6011 |

The control limits for these charts are tabulated below

|  |  | UCL | LCL |
|---|---|---|---|
| $S_1$- chart | - | 5.793 | 0.959 |
| $S_2$- chart | - | 4.180 | 0.692 |
| $S_1^2$-chart | - | 31.864 |  |
| $S_2^2$-chart | - | 16.589 |  |
| $\lvert S \rvert^{1/2}$- chart (Probability limits) | - | 11.10 | 2.66 |
| $W^*$ - chart $(\alpha = 0.05)$ | - | 8.52 |  |

Except for the 8th sample, for which the $W^*$ - chart $(\alpha = 0.05)$ is out of control and $\lvert S \rvert^{1/2}$- chart also is almost near to the UCL, the other sample statistics are all in control. Therfore the 8th sample should be discarded.

# CHAPTER 2

# BASIC IDEAS IN CUSUM CONTROL CHARTS

## 2.1 The CUSUM Control Chart and the V-Mask Scheme.

The Cumulative Sum (CUSUM) Control Charts were proposed by Page (1954) as an alternative to the Shewhart Control Charts. The CUSUM chart takes into account all the information in the sequence of observations, while the Shewhart chart only uses the information about the process which is contained in the last plotted point. Thus the CUSUM control chart tends to detect relatively small shifts in the process more effectively than the Shewhart charts. The CUSUM chart is used primarily to maintain the current control of the process.

The CUSUM Control Chart is a sequential sampling procedure. Suppose the observations are observed sequentially in time. Then the CUSUM control chart plots the cumulative sum of the sequence of observations against the number of observations. The change in trend on a CUSUM chart will indicate that a change has occured in the process. It is not sufficient to detect a shift by visual inspection. Hence different decision rules have been proposed to detect the shift in the process quickly. The first decision rule considered here is called as the V-mask scheme. Johnson and Leone (1962, (a), (b), (c)) proposed the sequential probability ratio test to construct the so called V-mask Scheme.

### 2.1.1    The Sequential Probability Ratio Test:

The sequential probability ratio test (SPRT) [Ghosh, (1970, chapter 3)] is used to construct the V-mask scheme for the CUSUM control chart. Suppose a random sample of m values $x_1, x_2, \ldots, x_m$, are observed. The likelihood ratio $L_{1m}/L_{0m}$, is calculated, where $L_{1m}$ and $L_{0m}$ are the likelihoods of $x_1, x_2, \ldots, x_m$, under the alternative

hypothesis, $H_1$ and the null hypothesis, $H_0$ respectively. Based on the likelihood ratio three decisions are possible;

(i) $\dfrac{L_{1m}}{L_{0m}} \leq \dfrac{\alpha_1}{1 - \alpha_0}$ , $H_0$ is accepted

(ii) $\dfrac{L_{1m}}{L_{0m}} \geq \dfrac{1 - \alpha_1}{\alpha_0}$ , $H_1$ is accepted

(iii) $\dfrac{\alpha_1}{1 - \alpha_0} < \dfrac{L_{1m}}{L_{0m}} < \dfrac{1 - \alpha_1}{\alpha_0}$ , a future observation $x_{m+1}$ is taken ,

and the new ratio, $L_{1(m+1)} / L_{0(m+1)}$ is then calculated based on the m+1 observations. This process is repeated until a decision is reached. Here $\alpha_0$ , $\alpha_1$ are the approximate probabilities of Type I error and Type II error respectively.

## 2.1.2 Cumulative Sum Charts.

Consider the case when $x_1, x_2, \ldots, x_m$ are normal independent random variables with process mean $\mu$ and known variance $\sigma^2$. The hypothesis to be tested is:

$$H_0 : \mu = \mu_0 \text{ against } H_1 : \mu = \mu_0 + \delta\sigma , \delta > 0 .$$

that is, we are testing a positive shift in the target mean value. Then, in this case the likelihood ratio becomes

$$\frac{L_{1m}}{L_{0m}} = \frac{\left(\sqrt{2\pi}\,\sigma\right)^{-m} \exp\left\{ -\dfrac{1}{2\sigma^2} \sum_{i=1}^{m} (x_i - \mu_0 - \delta\sigma)^2 \right\}}{\left(\sqrt{2\pi}\,\sigma\right)^{-m} \exp\left\{ -\dfrac{1}{2\sigma^2} \sum_{i=1}^{m} (x_i - \mu_0)^2 \right\}}$$

$$= \exp\left\{ \frac{1}{2\sigma^2} \left( 2\delta\sigma \sum_{i=1}^{m} (x_i - \mu_0) - m\delta^2\sigma^2 \right) \right\}$$

$$= \exp\left\{ \frac{\delta}{\sigma} \sum_{i=1}^{m} (x_i - \mu_0) - \frac{1}{2}m\delta^2 \right\}$$

$$\text{if } \frac{1}{\sigma} \sum_{i=1}^{m} (x_i - \mu_0) \leq \frac{1}{\delta} \, ln \left( \frac{\alpha_1}{1 - \alpha_0} \right) + \frac{1}{2} m\delta \,, \quad \text{accept } \mu = \mu_0 \,.$$

$$\text{if } \frac{1}{\sigma} \sum_{i=1}^{m} (x_i - \mu_0) \geq \frac{1}{\delta} \, ln \left( \frac{1 - \alpha_1}{\alpha_0} \right) + \frac{1}{2} m\delta \,, \quad \text{accept } \mu = \mu_0 + \delta\sigma \,.$$

If the points $\left[ m \,, \frac{1}{\sigma} \sum_{i=1}^{m} (x_i - \mu_0) \right]$ are plotted on graph paper, the 'continuation region'

lies between two parrallel straight lines, each inclined at an angle $\tan^{-1} (\frac{1}{2} \delta)$ to the axis of

m, and with intercepts

$$\frac{1}{\delta} \, ln \left( \frac{\alpha_1}{1 - \alpha_0} \right) \,, \quad \frac{1}{\delta} \, ln \left( \frac{1 - \alpha_1}{\alpha_0} \right)$$

respectively on the axis of

$$\frac{1}{\sigma} \sum_{i=1}^{m} (x_i - \mu_0) \,.$$

Consider a third hypothesis

$$H_{-1} : \mu = \mu_0 - \delta\sigma$$

Armitage (1950) proposed the simultaneous application of two SPRT's,

(i)  $H_0$  vs  $H_1$

(ii)  $H_0$  vs  $H_{-1}$,

where  $\alpha_1 = \alpha_{-1}$. The graphical limits for the combined tests are shown in figure 2.1.

The cumulative sum up to the mth observation is denoted by $S_m$ , where

$$S_m = \frac{1}{\sigma} \sum_{i=1}^{m} (x_i - \mu_0) \,.$$

and

$$OP = d = \frac{2}{\delta^2} \, ln \left( \frac{1 - \alpha_1}{\alpha_0} \right).$$

If $\alpha_1$ is small, as we usually require, then d can be approximated as

$$d = -\frac{2}{\delta^2} \; ln \; \alpha_0.$$

The angle $B_1PO$ is denoted by

$$\theta = tan^{-1} \left(\frac{1}{2}\delta\right).$$
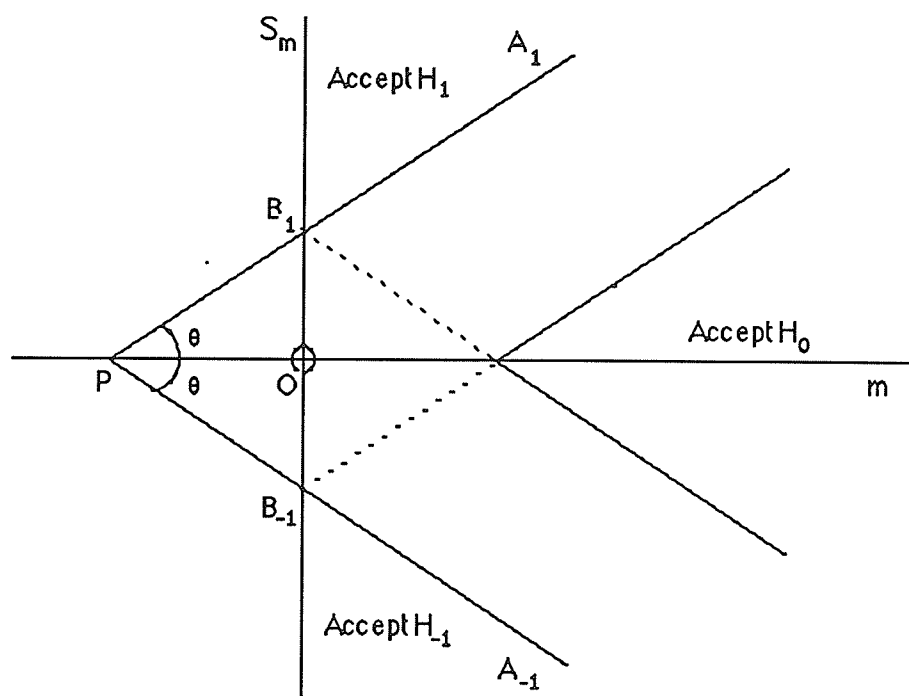
figure 2.1



The chart is interpreted by placing a mask (shaded area) over the chart as shown in figure 2.2 with the point 'O' over the last plotted point with the line OP horizontal. Any point lying below $A_1B_1$ is regarded as an indication of increase in the process quality characteristic we are measuring while points lying above $A_{-1}B_{-1}$ indicates a decrease.

In the CUSUM chart only two decisions are considered: one is rejecting $H_0$ ( or accept $H_1$) and the other is the combination of the other two decisions; that is, accept $H_0$ and the continue sampling. This is because the CUSUM procedure is constructed so that it detects only the sample points that are going out of control. Hence the acceptance of $H_0$ is considered as "continue sampling".

figure 2.2
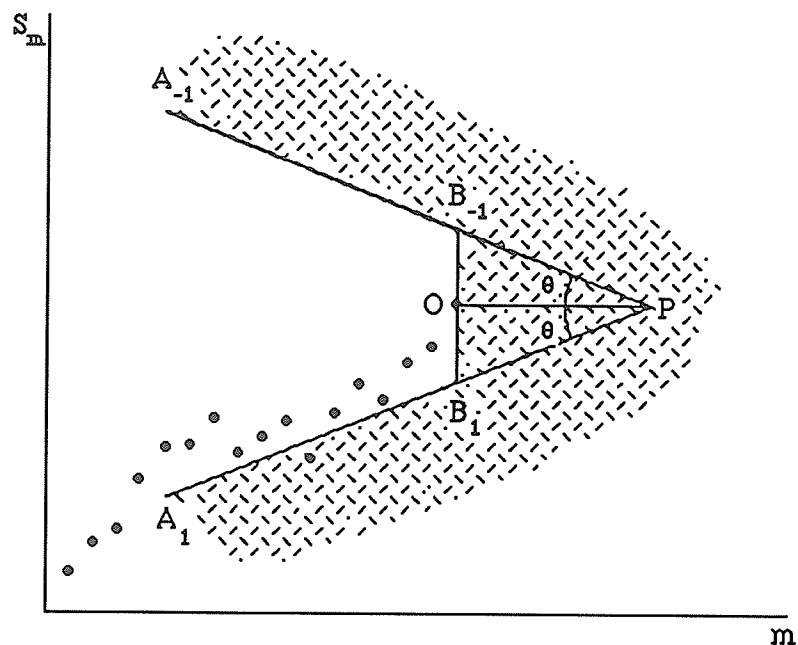


Consider the cumulative sum up to the (m+1)th observation, then

$$S_{m+1} = \frac{1}{\sigma} \sum_{i=1}^{m+1} (x_i - \mu_0)$$

$$= \frac{1}{\sigma} \sum_{i=1}^{m} (x_i - \mu_0) + \frac{x_{m+1} - \mu_0}{\sigma}$$

$$= S_m + \frac{x_{m+1} - \mu_0}{\sigma} .$$

Hence the ordinate of a plotted point equals the value of the immediately preceeding point plus the value of the statistic T; that is,

$$S_{m+1} = S_m + T_{m+1} \qquad \text{where } T_{m+1} = \frac{x_{m+1} - \mu_0}{\sigma}$$

$$= \sum_{j=1}^{m+1} T_j .$$

Thus the ordinate is the sum of the calculated values of the statistic T. Hence the name cumulative sum control charts.

### 2.1.3 Run Length and Average Run Length.

In an SPRT, the Average Sample Number(ASN) [Ghosh, B. K., (1970), section 3.5] is defined as the expected number of samples that falls inside the continuation region before it falls on either side for the first time. If it falls for the first time on the acceptance region, that is referred to as the ASN under $H_0$, and the other occurance is referred to as the ASN under $H_1$.

Run Length is the number of observations plotted before a point plots out-of-control for the first time. Average Run Length, ARL is the expected number of sample points that must be plotted before a point indicates an out-of-control condition. Hence ARL is the same as ASN under $H_1$ in an SPRT. For example if in a production process ARL=400, this means that an out-of-control signal goes after every 400 samples observed independent of whether the process is in control or not. Usually in a production process its appropriate to have a large ARL when the process is in control (i.e. ARL under $H_0$) and a small ARL when it is out of control (i.e. ARL under $H_1$).

Let

$$z_i = ln \frac{f_1(x_i)}{f_0(x_i)} \ , \ i = 1, 2, ..., m,$$

where $f_0, f_1$ are the density functions under $H_0$ and $H_1$, respectively. Then

$$z_i = -\frac{\delta^2}{2} + \frac{\delta}{\sigma}(x_i - \mu_0) \ , \ i = 1, 2, ..., m,$$

and

$$E_{H_0}(z) = -\frac{\delta^2}{2} \ .$$

Let

$$\lambda_m = \frac{L_{1m}}{L_{0m}} \ ,$$

then

$$E_{H_0}[\ ln\ \lambda_m\ ] = \alpha_0\ ln\left(\frac{1 - \alpha_1}{\alpha_0}\right) + (1 - \alpha_0)\ ln\left(\frac{\alpha_1}{1 - \alpha_0}\right).$$

Thus the ARL under $H_0$ is given by,

$$ARL_{H_0} = \frac{E_{H_0}[ln\ \lambda_m]}{E_{H_0}(z)}$$

$$= \frac{\alpha_0\ ln\left(\frac{1 - \alpha_1}{\alpha_0}\right) + (1 - \alpha_0)\ ln\left(\frac{\alpha_1}{1 - \alpha_0}\right)}{-\delta^2/2}.$$

To calculate the ARL under $H_1$,

$$E_{H_1}(z) = \frac{\delta^2}{2}\ .$$

$$E_{H_1}[\ ln\ \lambda_m\ ] = (1 - \alpha_1)\ ln\left(\frac{1 - \alpha_1}{\alpha_0}\right) + \alpha_1\ ln\left(\frac{\alpha_1}{1 - \alpha_0}\right).$$

Thus

$$ARL_{H_1} = \frac{E_{H_1}[ln\ \lambda_m]}{E_{H_1}(z)}$$

$$= \frac{(1 - \alpha_1)\ ln\left(\frac{1 - \alpha_1}{\alpha_0}\right) + \alpha_1\ ln\left(\frac{\alpha_1}{1 - \alpha_0}\right)}{\delta^2/2}.$$

If $\alpha_1$ is taken to be very small [Johnson and Leone, 1962(b),(c)] then,

$$ARL = \frac{-\ ln\ \alpha_0}{\delta^2/2} = \frac{-2ln\ \alpha_0}{\delta^2}\ .$$

As $\delta$ is varied, the values of ARL associated with different values of $\delta$ will trace out the ARL curve for the given control scheme.

## 2.1.4  CUSUM for Sample Means.

Here we use the sample mean instead of each observation in calculating the statistic T. Suppose each time a sample of size n is taken and the sample mean of the ith sample

$$\bar{x}_i = \frac{1}{n} \sum_{j=1}^{m} x_{ij} \qquad i=1, 2, \ldots, m$$

is calculated. The standard deviation of $\bar{x}$, $\sigma_{\bar{x}} = \sigma/\sqrt{n}$.

Then we plot the statistic

$$S_m = \frac{1}{\sigma_{\bar{x}}} \sum_{i=1}^{m} (\bar{x}_i - \mu_0) \qquad \text{against} \quad m.$$

Here the SPRT for the construction of CUSUM for sample means is based on the hypothesis:

$$H_0 : \mu = \mu_0$$

against

$$H_1 : \mu = \mu_1 = \mu_0 + \delta \sigma_{\bar{x}}, \quad \delta > 0 .$$

Then the dimensions of the V-mask for this chart are

$$\theta = \tan^{-1} (\tfrac{1}{2} \delta)$$

and

$$d = \frac{2}{\delta^2} \, ln \left( \frac{1 - \alpha_1}{\alpha_0} \right),$$

where

$$\delta = \frac{\mu_1 - \mu_0}{\sigma_{\bar{x}}} .$$

If $\alpha_1$ is small, as we usually require, then d can be approximated as

$$d = -\frac{2}{\delta^2} \, ln \, \alpha_0.$$

Hence to construct the V-mask, one procedure for determining the dimension's of the mask is:

(i)    Decide on the least size, D say, of absolute change in the mean which is desired to detect.  Calculate $\delta = D/\sigma_{\bar{x}}$ .

(ii)   Decide on the proportion of occassions ( that is samples ) on which you are willing to make a false statement that there has been a change in process average.  Call this proportion $2\alpha_0$ ( or $\alpha_0$, if only one-sided deviations from average are to be considered).

(iii)  Determine $\theta$ and d from the equations.

It appears that CUSUM chart will give more rapid indication of a change in mean if a small enough value of $\alpha_0$ is being used.  Since a value of about 0.001 is customory in most control charts,  CUSUM charts have a marked advantage.  The advantage decreases sharply as $\alpha_0$ increases.

## 2.1.5   *CUSUM for Sample Variances.*

The Cusum chart for variability in a process has been studied by Johnson and Leone, 1962(b),  based on sample variances and sample ranges.  But, in this section the variability of the process is discussed based on sample variances alone.

Suppose the variance is $\sigma_0^2$ when the process is in control.  We wish to detect an increase in variance from $\sigma_0^2$ to  $\sigma_1^2$ $(> \sigma_0^2)$.

We can construct an SPRT such that, the probability of failing to detect such a change is approximately  $\alpha_1$, and the probability of detecting a change when in reality, none has occured, is approximately $\alpha_0$.  The test is based on the hypothesis

$$H_0: \sigma^2 = \sigma_0^2 \quad \text{vs} \quad H_1: \sigma^2 = \sigma_1^2 \, (> \sigma_0^2).$$

Let $V_1, V_2, \ldots, V_m$ be the observed sample variances:

$$V_i = \sum_{j=1}^{N_i} \frac{(x_{ij} - \bar{x}_i)^2}{N_i - 1} \; ; \quad i = 1, 2, \ldots, m.$$

Consider the likelihood

$$L_m = \prod_{i=1}^{m} \prod_{j=1}^{N_i} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{ - \frac{1}{2\sigma^2} (x_{ij} - \bar{x}_i)^2 \right\}$$

$$= \prod_{i=1}^{m} \left( \frac{1}{\sqrt{2\pi}\,\sigma} \right)^{N_i} \exp\left\{ - \frac{1}{2\sigma^2} \sum_{j=1}^{N_i} (x_{ij} - \bar{x}_i)^2 \right\}$$

$$= \prod_{i=1}^{m} \left( \frac{1}{\sqrt{2\pi}\,\sigma} \right)^{N_i} \exp\left\{ - \frac{1}{2\sigma^2} (N_i - 1)\, V_i \right\}$$

$$= \left( \frac{1}{\sqrt{2\pi}\,\sigma} \right)^{\sum_{i=1}^{m} N_i} \exp\left\{ - \frac{1}{2\sigma^2} \sum_{i=1}^{m} (N_i - 1)\, V_i \right\}.$$

Hence, the likelihood ratio is,

$$\frac{L_{1m}}{L_{0m}} = \frac{\left( \dfrac{1}{\sqrt{2\pi}\,\sigma_1} \right)^{\sum_{i=1}^{m} N_i} \exp\left\{ - \dfrac{1}{2\sigma_1^2} \sum_{i=1}^{m} (N_i - 1)\, V_i \right\}}{\left( \dfrac{1}{\sqrt{2\pi}\,\sigma_0} \right)^{\sum_{i=1}^{m} N_i} \exp\left\{ - \dfrac{1}{2\sigma_0^2} \sum_{i=1}^{m} (N_i - 1)\, V_i \right\}}$$

$$= \left( \frac{\sigma_0}{\sigma_1} \right)^{\sum_{i=1}^{m} N_i} \exp\left\{ - \frac{1}{2} \left( \frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2} \right) \sum_{i=1}^{m} (N_i - 1)\, V_i \right\}$$

$$= \left( \frac{\sigma_0}{\sigma_1} \right)^{\sum_{i=1}^{m} N_i} \exp\left\{ - \frac{1}{2} \left( \frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2} \right) \sum_{i=1}^{m} \nu_i\, V_i \right\};$$

where $\nu_i = N_i - 1$. Then the continuation region is given by,

$$\frac{\alpha_1}{1 - \alpha_0} < \left(\frac{\sigma_0}{\sigma_1}\right)^{\sum\limits_{i=1}^{m} N_i} \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right)\sum\limits_{i=1}^{m} v_i V_i\right\} < \frac{1 - \alpha_1}{\alpha_0}.$$

The boundary between the continuation region and the acceptance region for $H_1$ has the

equation

$$\sum\limits_{i=1}^{m} v_i V_i = 2\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)^{-1} ln\left(\frac{1 - \alpha_1}{\alpha_0}\right)$$

$$+ 2\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)^{-1} ln\left(\frac{\sigma_1}{\sigma_0}\right)\sum\limits_{i=1}^{m} N_i$$

$$\sum\limits_{i=1}^{m} v_i V_i = \frac{2}{\frac{1}{\sigma_0^2}\left(1 - \frac{\sigma_0^2}{\sigma_1^2}\right)} ln\left(\frac{1 - \alpha_1}{\alpha_0}\right)$$

$$+ \frac{2}{\frac{1}{\sigma_0^2}\left(1 - \frac{\sigma_0^2}{\sigma_1^2}\right)} ln\left(\frac{\sigma_1}{\sigma_0}\right)\sum\limits_{i=1}^{m} N_i.$$

If $\alpha_1$ is small, this is approximately

$$\frac{1}{\sigma_0^2}\sum\limits_{i=1}^{m} v_i V_i = -\frac{2\, ln\, \alpha_0}{1 - \left(\frac{\sigma_0}{\sigma_1}\right)^2} + \frac{2\, ln\left(\frac{\sigma_1}{\sigma_0}\right)}{1 - \left(\frac{\sigma_0}{\sigma_1}\right)^2}\sum\limits_{i=1}^{m} N_i.$$

Hence, we construct a CUSUM control chart by plotting

$$\frac{1}{\sigma_0^2}\sum\limits_{i=1}^{m} v_i V_i \quad \text{against} \quad \sum\limits_{i=1}^{m} N_i, \quad \text{for each } m = 1, 2, \ldots,$$

that is we plot the points

$$\left( N_1, \frac{v_1 V_1}{\sigma_0^2} \right) \quad \left( N_1 + N_2, \frac{v_1 V_1 + v_2 V_2}{\sigma_0^2} \right) \cdots$$

As this CUSUM is constructed to detect the increase in the process variance, the out of control points of these plots can be detected using the lower limb, PQ of the V-mask as shown in figure 2.3, where the parameters of the chart are given by

$$d = - \frac{\ln \alpha_0}{\ln \left( \dfrac{\sigma_1}{\sigma_0} \right)},$$

and

$$\theta = \tan^{-1} \left\{ \frac{2 \ln \left( \dfrac{\sigma_1}{\sigma_0} \right)}{1 - \left( \dfrac{\sigma_0}{\sigma_1} \right)^2} \right\}.$$

figure 2.3
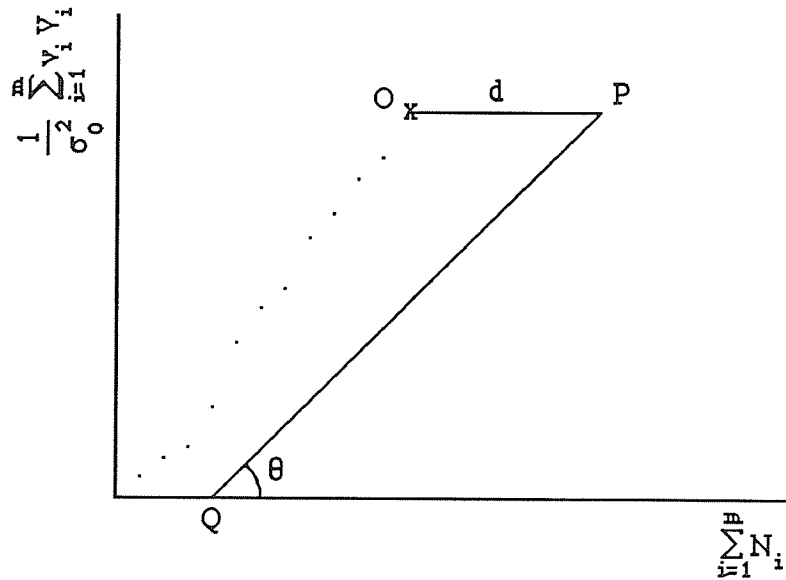
Sometimes it may also be desired to detect a reduction in process variance, from $\sigma_0^2$ to $\sigma_2^2$ ($< \sigma_0^2$), say. By a similar argument, we arrive at a control limit represented by the straight line P'Q' in figure 2.4.

In this case,

$$d' = - \frac{ln\ \alpha_0}{ln\left(\frac{\sigma_0}{\sigma_2}\right)} ,$$

and

$$\theta' = \tan^{-1}\left\{ \frac{2\ ln\left(\frac{\sigma_0}{\sigma_2}\right)}{\left(\frac{\sigma_0}{\sigma_2}\right)^2 - 1} \right\} .$$

figure 2.4
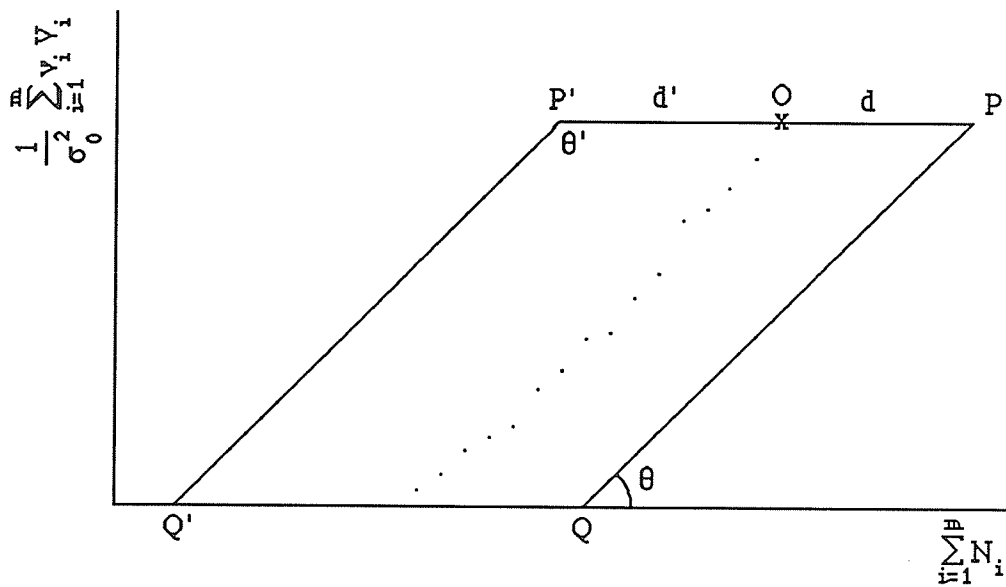
Note that in this case both control limits have positive slopes. This is to be expected, cause the ordinate $\dfrac{1}{\sigma_0^2} \displaystyle\sum_{i=1}^{m} v_i\, V_i$ cannot decrease as $m$ increases.

## 2.2    CUSUM Schemes.

### 2.2.1    Conversion of V- mask into two one-sided CUSUM Schemes.

The V-mask scheme can be converted into two one-sided schemes for the upward and downward shifts respectively. In the case of detecting an upward shift in the process average, the rule for deciding when such a shift has occurred is to compare the last point plotted with the lowest point previously plotted. If the difference exceeds a specified quantity h, the conclusion is that the shift has occured in the upward direction.

This decision rule picks up very small shifts which may not be of any importance. Hence the one-sided scheme on the upper side has been modified by the adoption of the reference value k. Under this modified procedure the quantities plotted on the chart are the cumulative sums of $\bar{x}$ - k , where k is a value greater than the target for $\bar{x}$. When the cumulative sum become less than zero a new series is began. When the cumulative sum of $\bar{x}$ - k exceeds the decision interval h, it is decided the process average has shifted above k. This procedure is much more easier to handle and it is sensitive only to the upward shifts in the process average that are considered to be of importance.

For the shifts on the negative side we choose k to be smaller than that of the target value and the test will be whether $\Sigma\,(\bar{x}$ - k) has a negative value less than that of -h. It is possible to have two one-sided schemes of the kind now discussed with upper and lower reference values, which is equivalent to using a V- shaped mask. Kemp (1961) demonstrates this equivalence as follows:

When the horizontal distance between the successive points on a CUSUM chart, measured in terms of unit distance on the vertical scale, is w and if the limbs of the V are inclined at an angle $\theta$ to the horizontal, the two one-sided schemes and the V-mask scheme will be equivalent if

$$k = w\,\sigma_{\bar{x}}\tan\theta = k_1 - \mu_0 = \mu_0 - k_2 \quad \text{and} \quad h = d\,\sigma_{\bar{x}}\tan\theta,$$

where $k_1$ and $k_2$ are the upper and lower reference values of the respective one-sided schemes. The target for $\bar{x}$ is the true mean $\mu_0$ and the variance is $\sigma_{\bar{x}}$.

Proof:

Let the last point plotted be denoted by O with cumulative sum equal to $S_m$ and let R be the point for which the cumulative sum is $S_{m-r}$, where

$$S_m = \frac{1}{\sigma_{\bar{x}}} \sum_{i=1}^{m} (\bar{x}_i - \mu_0).$$

The V-mask is placed such that the line OP lies parallel to the horizontal axis as shown in the diagram. The line OT is drawn parallel to the lower limb of the V-mask. The horizontal line through R and the perpendicular to the horizontal axis dropped from O meets at Q.

figure 2.5
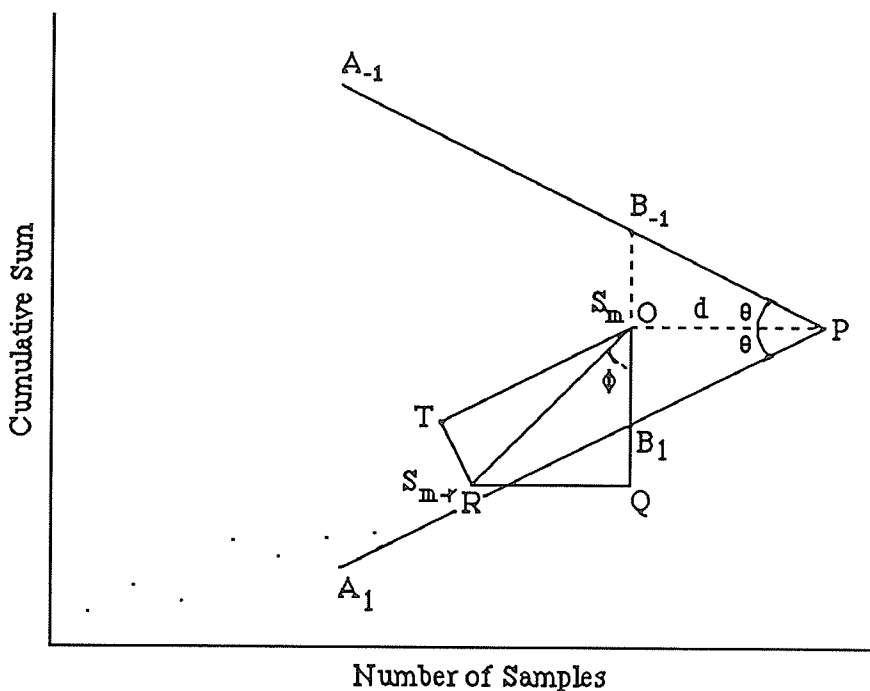


Then the angle TOQ is $\frac{\pi}{2} - \theta$, and if the angle ROQ is denoted by $\phi$, we have that

$$TR = OR \cos(\theta + \phi)$$

$$TR = OR \, ( \cos\theta \, \cos\phi - \sin\theta \, \sin\phi )$$

$$= OR \cos\phi \, \cos\theta - OR \sin\phi \, \sin\theta$$

$$= OQ \cos\theta - RQ \sin\theta$$

$$= (S_m - S_{m-r}) \cos\theta - rw \sin\theta$$

If

$$TR \geq d \, \sin\theta,$$

the path of the plotted points will cross the lower limb of the V-mask; that is,

$$S_m - S_{m-r} - rw \, \tan\theta \geq d \, \tan\theta$$

This implies that

$$\sum_{i=m-r+1}^{m} \left( \frac{\bar{x}_i - \mu_0}{\sigma_{\bar{x}}} - w \, \tan\theta \right) \geq d \, \tan\theta$$

$$\sum_{i=m-r+1}^{m} ( \bar{x}_i - \mu_0 - w \, \sigma_{\bar{x}} \tan\theta ) \geq d \, \sigma_{\bar{x}} \tan\theta \tag{2.2.1}$$

Then this is equivalent to cumulating the deviations $\bar{x} - k_1$, where $k_1 = \mu_0 + w \, \sigma_{\bar{x}} \tan\theta$, and using the decision interval $h = d \, \sigma_{\bar{x}} \tan\theta$. Similarly for the negative shifts it can be shown that with $k_2 = \mu_0 - w \, \sigma_{\bar{x}} \tan\theta$, the upper limb of the V-mask is crossed if

$$\sum_{i=m-r+1}^{m} ( \bar{x}_i - \mu_0 + w \, \sigma_{\bar{x}} \tan\theta ) \leq - d \, \sigma_{\bar{x}} \tan\theta . \tag{2.2.2}$$

For the first single-sided decision scheme the lower boundary is taken as zero and the upper boundary is used as a decision boundary; the value $\bar{x}_i - k_1$ is calculated for each sample and as soon as a positive value occurs, this and the subsequent values of $\bar{x}_i - k_1$ are accumulated. The cumulated sums are plotted until their path crosses either boundary. A decision that quality is off target is reached when the upper boundary is crossed; if the lower one is crossed plotting is terminated until another positive value of $\bar{x}_i - k_1$ is obtained. The second scheme is similar but here the upper boundary is taken at zero whilst

the lower one is the decision boundary, the values of $\bar{x}_i - k_2$ are cumulated when negative values of $\bar{x}_i - k_2$ are obtained.

Now to see what happens to the other scheme when one scheme crosses the decision boundary we will again refer to Kemp (1961). Let us define

$$k = w\, \sigma_{\bar{x}}\, \tan\theta \ , \quad h = d\, \sigma_{\bar{x}}\, \tan\theta \ \text{ and } \quad y_i = \bar{x}_i - \mu_0 \ .$$

Also let

$$C_j = \sum_{i=t}^{t+j} (y_i - k) \qquad \text{and} \qquad c_1 = \sum_{i=u}^{u+1} (y_i + k).$$

Consider the two sums $C_m$ and $c_n$, where

$$0 < C_j < h \quad (\, 0 \le j \le m),$$

$$0 > c_1 > -h \quad (\, 0 \le 1 \le n).$$

If the cumulative sums of the two schemes lie between their respective boundaries immediately after the rth sample is drawn, $C_m$ and $c_n$ will represent them if

$$m + t = u + n = r.$$

We now consider what happens when $C_{m+1} \ge h$, so that a decision that the process is off target is indicated by the first scheme. We then have

$$C_{m+1} = C_m + y_{m+1} - k \ \ge h,$$

so that

$$y_{m+1} \ge h + k - C_m \ .$$

The cumulative sum of the second scheme is now

$$c_{n+1} = c_n + y_{m+1} + k \ \ge \ c_n - C_m + h + 2k$$

To compute $c_n - C_m$, we consider three different cases: (i) $t > u$, (ii) $t = u$ and (iii) $t < u$.

case(i) : $t > u$.

From above

$$t = r - m \ \text{ and } \ u = r - n.$$

Therefore $t > u$ implies $n > m$.

Hence

$$C_m = \sum_{i=t}^{m+t} (y_i - k) = \sum_{i=t}^{m+t} y_i - (m + 1)k$$

$$c_n = \sum_{i=u}^{n+u} (y_i + k) = \sum_{i=u}^{n+u} y_i + (n + 1)k.$$

Thus

$$c_n - C_m = \sum_{i=u}^{n+u} y_i - \sum_{i=t}^{m+t} y_i + (n + m + 2)k$$

$$= \sum_{i=u}^{t-1} y_i + (n + m + 2)k$$

$$= \sum_{i=u}^{u+(t-u-1)} (y_i + k) - (t - u)k + (n + m + 2)k$$

$$= c_{t-u-1} - (n - m)k + (n + m + 2)k$$

$$= c_{t-u-1} - (2m + 2)k .$$

case(ii) :  $t = u$  (this implies $m = n$)

$$C_m = \sum_{i=t}^{m+t} y_i - (m + 1)k$$

$$c_n = \sum_{i=t}^{m+t} y_i + (m + 1)k$$

$$c_n - C_m = \sum_{i=t}^{m+t} y_i - \sum_{i=t}^{m+t} y_i + (2m + 2)k$$

$$= (2m + 2)k .$$

case(iii) :  $t < u$  (this implies $m > n$)

$$C_m = \sum_{i=t}^{m+t} y_i - (m + 1)k$$

$$c_n = \sum_{i=u}^{n+u} y_i + (n + 1)k .$$

Therefore,

$$c_n - C_m = \sum_{i=t}^{u-1} y_i + (n + m + 2)k$$

$$= - \sum_{i=t}^{t+(u-t-1)} (y_i - k) - (u - t)k + (n + m + 2)k$$

$$= - C_{u-t-1} - (m - n)k + (n + m + 2)k$$

$$= - C_{u-t-1} + (2n + 2)k$$

Now from all these three cases we have

$$c_{n+1} = \begin{cases} c_{t-u-1} + h + 2(m+2)k & (t > u) \\ h + 2(m+2)k & (t = u) \\ h + 2(n+2)k - C_{u-t-1} & (t < u). \end{cases}$$

Therefore $c_{n+1} > 0$.

A similar argument shows that $C_{m+1}$ must be less than zero when $c_{n+1} \leq$ -h.

When the cumulative sum immediately after the rth sample of one scheme lies between its boundaries and cumulation for the other has ceased, similar reasoning shows that only one decision boundary can be crossed at the next sample, and that the cumulative sum for the other cannot lie between its boundaries. It follows that a cumulation which ends on or above the upper boundary of the first scheme cannot cut short a cumulation which would otherwise have ended on or below the lower boundary of the second, and vice versa.

### 2.2.2   Two - sided CUSUM scheme.

The two-sided CUSUM scheme can be constructed for positive shifts and negative shifts in the mean with two boundaries for each as in Shewart control charts. That is for the positive shift the lower boundary as zero and the upper boundary as the decision interval h. And for the negative shifts the two-sided scheme is with the decision boundary -h as the lower boundary and the upper boundary as zero. Hence the procedure can be defined as

$$S_H(i) = \max [\, 0, \, \bar{x}_i - \mu_0 - k + S_H(i\text{-}1)\, ]$$

$$S_L(i) = \min [\, 0, \, \bar{x}_i - \mu_0 + k - S_L(i\text{-}1)\, ],$$

where $S_H(0) = S_L(0) = 0$, and k is the reference value. The values $S_H(0)$ and $S_L(0)$ are referred to as the head start values. If $S_H(i)$ exceed the decision interval h, or if $S_L(i)$ is less than -h, the process is considered to be out of control.

To make calculations easier we define,

$$z_i = \frac{\bar{x}_i - \mu_0}{\sigma_{\bar{x}}} \,; \qquad i = 1, 2, \ldots,$$

where $\bar{x}_i \sim N(\mu_0, \sigma_{\bar{x}})$; $i = 1, 2, \ldots$, and are identically independently distributed. As we derived in section 2.2.1 the values of k and h involves $\sigma_{\bar{x}}$. Now if we consider $z_i$'s in the scheme [ Fellner, (1990) ], the k and h values will be independent of $\sigma_{\bar{x}}$. Then the two schemes can be written as

$$S_H(i) = \max [\, 0, \, z_i - k + S_H(i\text{-}1)\, ]$$

$$S_L(i) = \max [\, 0, \, -z_i - k + S_L(i\text{-}1)\, ],$$

where $S_H(0) = S_L(0) = S_0$. For a prescribed value k, the sequence is considered to be out of control whenever either $S_H$ or $S_L$ exceeds a prespecified value h.

For the CUSUM scheme to be completly defined, the intial value $S_0$ of $S_H$ and $S_L$ must be defined. Customarily, the value of $S_0$ is set to zero. However, Lucas and Crosier

(1982) showed that by selecting a positive value for $S_0$ the scheme could be made more sensitive to the process mean at the start of the scheme, without greatly increasing the 'false alarm' rate. This gives the CUSUM scheme a 'fast initial response'.

The ARL of the scheme depends on the value of the process mean, which is unknown at the time of the process, as well as on the values of the specified parameters k, h and $S_0$. Ideally these parameters are set so that, when the process mean is at its target value $\mu_0$, the ARL is sufficiently large, while, when process mean is critically off target, the ARL is small.

Consider the following example which shows, how the ARL is behaving for different sets of values of k, h and $S_0$, when the process mean is at its target value $\mu_0$ and when it is off target. Using the IMSL subroutine GGNPM (Appendix B, program 1), samples of size 10 are generated by simulation, from $N(\mu, 1)$. For values of (k, h) as (0.25, 2.5), (0.50, 2.0), (0.75, 1.5) and (1.0, 1.0) the start up value $S_0$ is taken as h/2.

The ARL is found by repeating the procedure 1000 times and calculating the expected run length based on those 1000 run lengths. Here in this example we find the ARL based on the 1000 run lengths, under $\mu = 0.0, 0.5, 1.0, 2.0, 3.0$.

Table 2.1

| k | h | $\mu$ 0.0 | 0.5 | 1.0 | 2.0 | 3.0 |
|---|---|---|---|---|---|---|
| 0.25 | 2.5 | 11.012 | 5.597 | 2.819 | 1.362 | 1.065 |
| 0.50 | 2.0 | 17.062 | 7.870 | 3.492 | 1.442 | 1.068 |
| 0.75 | 1.5 | 21.058 | 10.105 | 4.297 | 1.545 | 1.086 |
| 1.00 | 1.0 | 21.652 | 11.484 | 4.904 | 1.680 | 1.090 |

The results in Fellner's paper almost tally with this results. These results are obtained by simulation, while his is analytical

## 2.3   Markov chain approach to find the ARL of a CUSUM scheme.

To study the ARL of the CUSUM scheme a different approach has been given by Brook and Evans (1972) in which the operation of the scheme is regarded as forming a Markov chain. We will briefly summarize their work here.

The transition probability matrix for this Markov chain is obtained and then the properties of this matrix are used to determine not only the average run lengths for the scheme, but also moments and percentage points of the run-length distribution and exact probabilities of run length. This method has been suggested for any discrete distribution and is also as an accurate approximation with any continuous distribution for the random variable which is to be controled.

### 2.3.1   The Markov chain approach.

Suppose samples of n items are taken at regular intervals and the number of defectives $D_i$ (i = 1, 2, 3, ...) are observed. For a reference value k, the process is considered to be in control if $D < k$. As soon as D exceeds k, we start plotting the cumulative sums

$$S_m = \frac{1}{\sigma} \sum_{i=1}^{m} (D_i - k) \quad \text{against} \quad m.$$

If $S_m$ becomes zero the process starts in control. When $S_m$ exceeds the decision interval, h an appropriate action is taken.

First we'll consider the discrete case. Let D be an integer random variable and k, h have positive values. Then the random variable $S_m$ can take on values 0, 1, 2, ..., h. Let us suppose if $S_m = i$, the scheme is in state $E_i$. Then each realization of this scheme can be considered as a Markov chain with states $E_0$, $E_1$, ......., $E_h$, where $E_h$ is an absorbing state. It is assumed that the process is initially in state $E_0$.

The transition probabilities from state $E_i$ ( $i = 0, 1, \ldots, h\text{-}1$ ) are determined by the probability distribution of D as follows:

$$p_{io} = \Pr ( E_i \longrightarrow E_0 ) = \Pr ( D \leq k - i ),$$

$$p_{ij} = \Pr ( E_i \longrightarrow E_j ) = \Pr ( D = k + j - i ), \quad j = 1, \ldots, h\text{-}1,$$

$$p_{ih} = \Pr ( E_i \longrightarrow E_h ) = \Pr ( D \geq k + h - i ).$$

The system forms a Markov chain whose transition probability matrix is constructed from the probability distribution of D, given h and k.

Let $P_r = \Pr ( D - k = r )$ and $F_r = \Pr ( D - k \leq r )$, then

$$p_{io} = \Pr ( D - k \leq - i ) = F_{-i}$$

$$p_{ij} = \Pr ( D - k = j - i ) = P_{j-i}, \quad j = 1, \ldots, h\text{-}1$$

$$p_{ih} = \Pr ( D - k \geq h - i ) = 1 - F_{h-i}.$$

Hence the transition probability matrix P, has the following form:

$$
P = \begin{bmatrix}
F_0 & P_1 & P_2 & \cdots & P_{h-1} & 1\text{-}F_{h-1} \\
F_{-1} & P_0 & P_1 & \cdots & P_{h-2} & 1\text{-}F_{h-2} \\
\vdots & \vdots & \vdots & & \vdots & \vdots \\
\vdots & \vdots & \vdots & & \vdots & \vdots \\
1\text{-}F_{1-h} & P_{2-h} & P_{3-h} & \cdots & P_0 & 1\text{-}F_0 \\
0 & 0 & 0 & \cdots & 0 & 1
\end{bmatrix}
$$

Most of the results required can be obtained by working with the matrix R, obtained from P by deleting the final row and column. In particular, all the eigenvalues of R are the same as P.

## 2.3.2  Sampling properties of CUSUM run length.

Let $X_i$ be the number of steps taken to reach the absorbing state $E_h$ for the first time, starting from $E_i$ and let $\mu_i^{(s)}$ be the sth factorial moment of $X_i$, where

$$\mu_i^{(s)} = E[X_i^{(s)}] = E\left[\prod_{j=0}^{s-1} (X_i - j)\right], \quad s = 2, 3, \ldots.$$

By considering the Markov chain one-step later, we have

$$\mu_i^{(s)} = \sum_{r=s}^{\infty} r^{(s)} \Pr(X_i = r)$$

$$= \sum_{r=s}^{\infty} r^{(s)} \sum_{j=0}^{h-1} p_{ij} \Pr(X_j = r-1)$$

$$= \sum_{j=0}^{h-1} p_{ij} \sum_{r=s}^{\infty} r^{(s)} \Pr(X_j = r-1).$$

From this it can be shown that

$$\mu_i^{(s)} = \sum_{j=0}^{h-1} p_{ij} (\mu_j^{(s)} + s\mu_j^{(s-1)}) \quad s = 2, 3, \ldots \; ; \quad j = 0, 1, \ldots, h-1.$$

In matrix form this becomes

$$\underline{\mu}^{(s)} = \begin{pmatrix} \mu_0^{(s)} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \mu_{h-1}^{(s)} \end{pmatrix} = R\,\underline{\mu}^{(s)} + s\,R\,\underline{\mu}^{(s-1)} \quad s = 2, 3, \ldots.$$

where R is the h x h matrix obtained from the transition probability matrix P by deleting the last column and row;  that is,

$$(I - R)\,\underline{\mu}^{(s)} = s\,R\,\underline{\mu}^{(s-1)} \quad s = 2, 3, \ldots, \tag{2.3.1}$$

where I is the hxh identity matrix and $\underline{\mu}^{(s)}$ is the vector of sth factorial moments for the random variables $X_0, X_1, \ldots, X_{h-1}$.

For the special case in which s=1, (2.3.1) becomes

$$(I - R)\,\underline{\mu} = \underline{j},$$

where the vector $\underline{j}$ has each of its h elements equal to unity; hence

$$\underline{\mu} = (I - R)^{-1} \underline{j}. \tag{2.3.2}$$

The first element of $\underline{\mu}$ is the ARL for a CUSUM chart starting from zero. In general, the ith element gives the mean of the run-length distribution when starting from state $E_i$ ($i = 0$, $1, \ldots, h\text{-}1$).

Let $\underline{X}$ be the random vector of runlengths with elements $X_0, X_1, \ldots, X_{h-1}$; then higher order factorial moments of $\underline{X}$ can be obtained recursively via (2.3.1):

$$\underline{\mu}^{(s)} = s (I - R)^{-1} R \, \underline{\mu}^{(s-1)}$$

$$= s \{(I - R)^{-1} - I\} \, \underline{\mu}^{(s-1)} \qquad s = 2, 3, \ldots. \tag{2.3.3}$$

The matrix multiplication in (2.3.3) can be avoided by premultiplying a scalar times the previous solution by the matrix $(I - R)^{-1} - I$. Since $(I - R)^{-1} R = R (I - R)^{-1}$, the solution to (2.3.3) can be written in either of the following forms:

$$\underline{\mu}^{(s)} = s! \, (I - R)^{-s} \, R^{s-1} \, \underline{j}$$

$$= s! \, R^{s-1} \, (I - R)^{-s} \, \underline{j} \qquad s = 1, 2, \ldots. \tag{2.3.4}$$

Hence the probability distribution of the run-length vector $\underline{X}$ can be regarded as a multidimensional generalization of a geometric distribution over the positive integers. For, let N be such that $\Pr ( N = r ) = p \, q^{r-1}$ ($r = 1, 2, \ldots$), where $p + q = 1$ ($0 < p < 1$); then its sth factorial moment is $s! \, q^{s-1} \, p^{-s}$ ($s = 1, 2, \ldots$). If R was diagonal, then the elements of $\underline{X}$ would each be geometric random variables. The scalar parameter p is replaced by the matrix $(I - R)$ and the elements of the vector $(I - R) \underline{j}$ are the probabilities of jumping to the absorbing state, the first occurence of which terminates the process.

Now we will consider an example with a numerical illustration of the moments of run length distribution. Suppose that D has a Poisson distribution with mean 4.1. An ARL of about 4 can be arranged by choosing $k = 3$ and $h = 4$. Using the subroutine

POPROB (Appendix B, program 2), the Poisson probabilities are obtained and the matrix R is calculated

$$R = \begin{pmatrix} 0.4142 & 0.1951 & 0.1600 & 0.1093 \\ 0.2238 & 0.1904 & 0.1951 & 0.1600 \\ 0.0845 & 0.1393 & 0.1904 & 0.1951 \\ 0.0166 & 0.0679 & 0.1393 & 0.1904 \end{pmatrix}$$

and from that the table 2.2 is obtained. Here again to calculate the $(I - R)^{-1}$ matrix the IMSL subroutine LINV1F is being used. From (2.3.2) and (2.3.3) we find $\mu^{(1)}$ by summing the rows of $(I - R)^{-1}$, $\mu^{(2)}$ by postmultiplying $\{(I - R)^{-1} - I\}$ by $2\mu^{(1)}$, and so on. And from these the central moments, coefficients of skewness and kurtosis which are shown in table 2.2 are determined.

Table 2.2

| State | $\mu$ | $\mu_2$ | $\mu_3$ | $\mu_4$ | $\sigma$ | $\sigma/\mu$ | $\mu_3/\sigma^3$ | $(\mu_4/\sigma^4)-3$ |
|---|---|---|---|---|---|---|---|---|
| 0 | 3.97 | 7.23 | 34.01 | 411.37 | 2.69 | 0.68 | 1.75 | 4.86 |
| 1 | 3.40 | 6.71 | 33.35 | 387.91 | 2.59 | 0.76 | 1.92 | 5.63 |
| 2 | 2.73 | 5.55 | 30.24 | 333.09 | 2.36 | 0.86 | 2.31 | 7.81 |
| 3 | 2.07 | 3.84 | 23.08 | 242.24 | 1.96 | 0.95 | 3.07 | 13.46 |

A basic comparison is with the geometric distribution over the positive integers with mean $1/p$. This has

$$\sigma^2 = q/p^2, \quad \mu_3/\sigma^3 = (1+q)/q^{1\backslash 2}, \quad (\mu_4/\sigma^4) - 3 = (q^2 + 4q + 1)/q.$$

In particular, if $1/p = 3.97$ then

$$\sigma^2 = 11.79, \quad \sigma = 3.43, \quad \mu_3/\sigma^3 = 2.02 \quad \text{and} \quad (\mu_4/\sigma^4) - 3 = 6.09.$$

Thus the run length from state $E_0$ has a smaller variance than a geometric random variable with the same mean and it also has smaller skewness and kurtosis.

## 2.3.3 The probability distribution of run length.

The last column of $P^r$; the rth power of the transition probability matrix P, gives $Pr(X_i \leq r)$; $i = 0,1,.....,h-1$, together with unity as the last element. If we partitioned the matrix P as:

$$P = \begin{bmatrix} R & | & \underline{p}_h \\ ----- & | & ----- \\ O^T & | & 1 \end{bmatrix},$$

where $\underline{p}_h$ is a vector of h elements and R is the hxh matrix discussed in the previous section, then

$$\underline{p}_h = (I - R)\, \underline{j}$$

Since P is a stochastic matrix, its row sums are equal to unity. Hence it follows that

$$P = \begin{bmatrix} R^r & | & (I - R^r)\underline{j} \\ ----- & | & ---------- \\ O^T & | & 1 \end{bmatrix} \qquad r = 1, 2, \ldots .$$

Let $\underline{B}_r$ be a vector of length h whose elements are the distribution functions of run length starting from states $E_0, E_1, ........, E_{h-1}$, that is

$$\underline{B}_r = \{Pr(X_0 \leq r), Pr(X_1 \leq r), \ldots ., Pr(X_{h-1} \leq r)\}^T \quad r = 1, 2, \ldots .$$

Then

$$\underline{B}_r = (I - R^r)\, \underline{j} \quad r = 1, 2, \ldots .$$

and the first element of $\underline{B}_r$ gives the cumulative probability for run length for a CUSUM scheme starting from zero.

Let $\underline{L}_r$ be a vector of length h whose elements are the values of the probability functions of run length starting from states $E_0, E_1, ........, E_{h-1}$, that is

$$\underline{L}_r = \{Pr(X_0 = r), Pr(X_1 = r), \ldots ., Pr(X_{h-1} = r)\}^T \quad r = 1, 2, \ldots .$$

Then

$$\underline{L}_1 = \underline{p}_h = (I - R)\, \underline{j},$$

$$\underline{L}_r = R \, \underline{L}_{r-1} = R^{r-1} \, \underline{L}_1, \quad r = 1, 2, \ldots.$$

which is closely related to the univariate geometric form

$$Pr(N = r) = p \, q^{r-1}, \quad r = 1, 2, \ldots.$$

For our example in (2.3.2), for which the average run length is 3.97, the probability distribution of run length starting from state $E_0$, and the corresponding probabilities for a geometric distribution with the same mean have been calculated (Appendix B, program 2) and are tabulated in table 2.3; and they are graphed in figure 2.6.

Table 2.3

| Run length | The actual run length probabilities | Geometric distribution probabilities |
|:---:|:---:|:---:|
| 1 | 0.1214 | 0.2516 |
| 2 | 0.2218 | 0.1883 |
| 3 | 0.1960 | 0.1409 |
| 4 | 0.1450 | 0.1055 |
| 5 | 0.1013 | 0.0789 |
| 6 | 0.0693 | 0.0591 |
| 7 | 0.0471 | 0.0442 |
| 8 | 0.0318 | 0.0331 |
| 9 | 0.0215 | 0.0248 |
| 10 | 0.0145 | 0.0185 |
| 11 | 0.0098 | 0.0139 |
| 12 | 0.0066 | 0.0104 |

figure 2.6
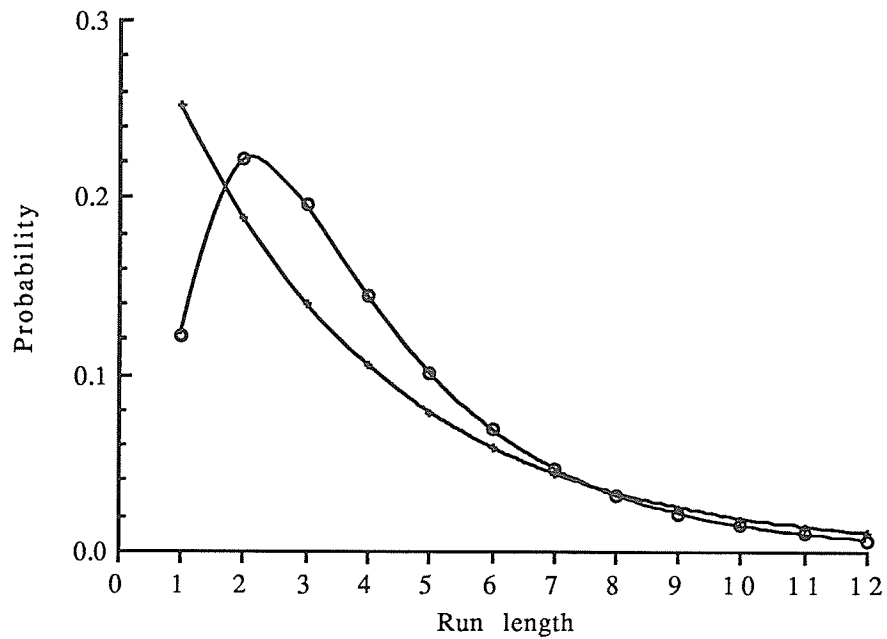


The geometric approximation to the run length distribution seems to be good in the tail probabilities only.

## 2.3.4   Applications to general continuous distributions.

Here in this section the case where the distribution of the random variable concerned is continuous is considered. Let Z be the continuous random variable that is considered to have had a shift in the mean. Here the one-sided type of decision interval scheme is considered. We accumulate the deviations of Z from a reference value k (while the cumulative sum is positive) until either we cross the decision boundary h or the cumulative sum reverts back to zero.

Suppose that we wish to represent the continuous scheme by a Markov chain having t+1 states labelled $E_0$, $E_1$, ......., $E_t$, where $E_t$ is absorbing. Then 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 $\frac{1}{2}$ w; that is, the next value of Z does not differ from the reference value k by more than $\frac{1}{2}$ w. The quantity w determines the width of the grouping interval involved in the discretization of the probability distribution of Z.

A further restriction to this is that 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 with the state $E_i$.

These requirements lead to

$$w = \frac{2h}{2t - 1} .$$  (2.3.5)

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

$$p_{io} = Pr\,(\,E_i \longrightarrow E_0\,) = \; Pr\,\{\; Z - k \leq - iw + \tfrac{1}{2}w \;\},$$

$$p_{ij} = Pr\,(\,E_i \longrightarrow E_j\,) = \; Pr\,\{\; (j - i)w - \tfrac{1}{2}w < Z - k \leq (j - i)w + \tfrac{1}{2}w \;\},$$

$$(1 \leq j \leq t\text{-}1)$$

$$p_{it} = Pr\,(\,E_i \longrightarrow E_t\,) = \; Pr\,\{\; (t - i)w - \tfrac{1}{2}w < Z - k \;\},$$

Note that,

$$Pr(\,E_0 \longrightarrow E_t\,) = \; Pr\,(\,Z - k > h\,)$$

for any choice of w satisfying (2.3.5).

If we write

$$P_r = Pr\,(\; rw - \tfrac{1}{2}w < Z - k \leq rw + \tfrac{1}{2}w\;)$$

and

$$F_r = Pr\,(\,Z - k \; \leq rw + \tfrac{1}{2}w\;),$$

then the transition probability matrix, P, for the Markov chain is constructed as before and has the general form shown in section 2.3.1, except that the states are labelled $E_0$, $E_1$, . . ., $E_t$. The number of states is now arbitrary and is not determined directly by h, as in the

discrete case. The matrix obtained from P by deleting the last row and column is dentoed by R, as before, and the sampling properties of the corresponding CUSUM scheme are determined in the manner described above for the discrete case.

Now we consider an example with a numerical illustration of the moments of run length distribution. Suppose that Z has a normal distribution with mean 1 and variance 1. We choose h = 3 and t = 5. Using the IMSL subroutine MDNOR the normal distribution probabilities are obtained (Appendix B, program 3), and from that the matrix R is calculated. Table 2.4 is obtained as in the case of discrete random variables. In this case the matrix R will be

$$R = \begin{pmatrix} 0.2525 & 0.2475 & 0.2475 & 0.1613 & 0.0685 \\ 0.0912 & 0.1613 & 0.2475 & 0.2475 & 0.1613 \\ 0.0228 & 0.0685 & 0.1613 & 0.2475 & 0.2475 \\ 0.0038 & 0.0189 & 0.0685 & 0.1613 & 0.2475 \\ 0.0004 & 0.0034 & 0.0189 & 0.0685 & 0.1613 \end{pmatrix}$$

Table 2.4

| State | $\mu$ | $\mu_2$ | $\mu_3$ | $\mu_4$ | $\sigma$ | $\sigma/\mu$ | $\mu_3/\sigma^3$ | $(\mu_4/\sigma^4)-3$ |
|-------|-------|---------|---------|---------|----------|--------------|-------------------|----------------------|
| 0 | 3.77 | 3.15 | 7.80 | 62.07 | 1.77 | 0.47 | 1.40 | 3.26 |
| 1 | 3.19 | 2.86 | 7.39 | 56.23 | 1.69 | 0.53 | 1.53 | 3.88 |
| 2 | 2.54 | 2.37 | 6.46 | 45.63 | 1.54 | 0.61 | 1.77 | 5.11 |
| 3 | 1.91 | 1.66 | 4.95 | 31.23 | 1.29 | 0.68 | 2.31 | 8.31 |
| 4 | 1.42 | 0.87 | 2.86 | 16.32 | 0.93 | 0.66 | 3.52 | 18.54 |

Then the ARL is the first element of the mean vector $\mu$, which is 3.77. If we use the geometric approximation $1/p = 3.77$, then the corresponding geometric distribuition

probabilities can be obtained.  Both probabilities for the run length distribution are tabulated

in Table 2.7, and are graphed and shown in figure 2.7.

Table 2.5

| Run length | The actual run length probabilities | Geometric distribution probabilities |
|---|---|---|
| 1 | 0.0228 | 0.2649 |
| 2 | 0.2226 | 0.1947 |
| 3 | 0.2814 | 0.1432 |
| 4 | 0.2053 | 0.1052 |
| 5 | 0.1235 | 0.0774 |
| 6 | 0.0685 | 0.0569 |
| 7 | 0.0365 | 0.0418 |
| 8 | 0.0191 | 0.0307 |
| 9 | 0.0099 | 0.0226 |
| 10 | 0.0051 | 0.0166 |
| 11 | 0.0026 | 0.0122 |
| 12 | 0.0013 | 0.0090 |
| 13 | 0.0007 | 0.0066 |
| 14 | 0.0004 | 0.0049 |
| 15 | 0.0002 | 0.0036 |

figure 2.7
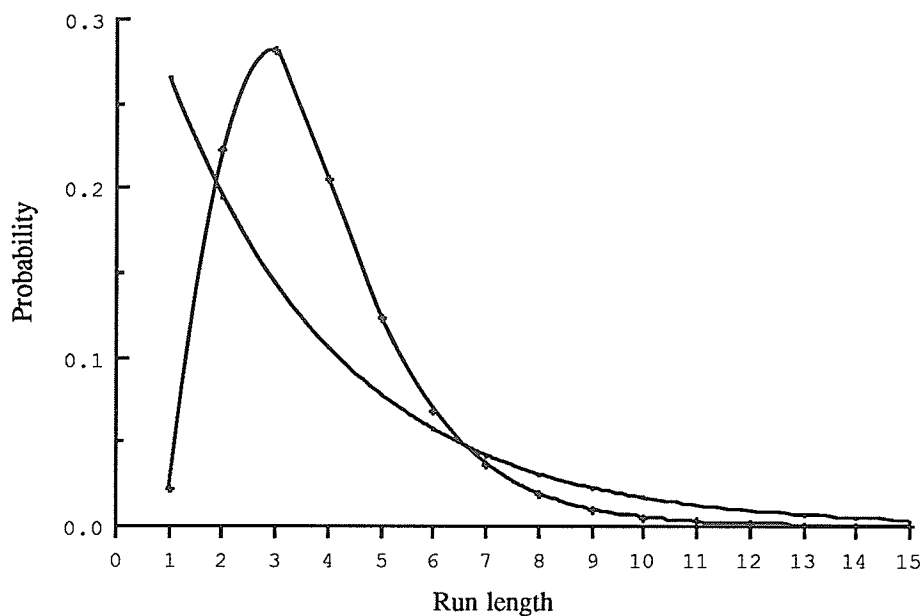


The geometric approximation to the actual run length distribution seems to be good between run lengths 6 and 7 only, although it seems close in the tail probabilities. Brook and Evans states that as r tends to infinity the limiting probability distribution of the run length is approximately geometric.

Consider another example for the continuous case. Suppose the random variable Z has an exponential distribution. Then to compute the R matrix, we need

$$F_r = Pr\left( Z - k \leq rw + \frac{1}{2}w \right),$$

$$= 1 - exp\left[ -\left( k + (r + \frac{1}{2})w \right) \right], \quad -(t-1) \leq r \leq (t-1),$$

where

$$w = \frac{2h}{2t - 1}.$$

and

$$P_r = F_r - F_{r-1}$$

If we consider the extreme probabilities,

$$F_{-(t-1)} = \Pr\left( Z - k \le -(t-1)w + \frac{1}{2}w \right)$$

$$= \Pr\left( Z \le k - (t-1)\frac{2h}{2t-1} + \frac{h}{2t-1} \right)$$

$$= \Pr\left( Z \le k - \frac{2t-3}{2t-1}h \right),$$

and

$$F_{t-1} = \Pr\left( Z \le k + (t-1)\frac{2h}{2t-1} + \frac{h}{2t-1} \right)$$

$$= \Pr\left( Z \le k + h \right).$$

Choose a large integer $I^*$, which is almost equal to $k + h$. From both extreme probabilities we choose

$$k = \frac{2t-3}{2t-1}h, \quad \text{and} \quad I^* = k + h.$$

This implies,

$$h = \frac{2t-1}{4(t-1)}I^*, \quad k = \frac{2t-3}{4(t-1)}I^*.$$

Hence, if the values of $I^*$ and t are given, the values of k and h can be obtained. In our example we took $I^*$ and t as, $I^* = 5$ and $t = 4$. This gives the R matrix as (Appendix B, program 4),

$$R = \begin{pmatrix} 0.9179 & 0.8111 & 0.5654 & 0.0000 \\ 0.0464 & 0.1068 & 0.2457 & 0.5654 \\ 0.0202 & 0.0464 & 0.1068 & 0.2457 \\ 0.0088 & 0.0202 & 0.0464 & 0.1068 \end{pmatrix}$$

and the ARL (the first element in the $\mu$ vector) is 112.05. Since the ARL is large the geometric mean p ($1/p = 112.05$), becomes really small. That makes the geometric approximation more appropriate.

Table 2.6

| Run length | The actual run length probabilities | Geometric distribution probabilities |
| --- | --- | --- |
| 1 | 0.0067 | 0.0089 |
| 2 | 0.0083 | 0.0088 |
| 3 | 0.0087 | 0.0088 |
| 4 | 0.0087 | 0.0087 |
| 5 | 0.0087 | 0.0086 |
| 6 | 0.0086 | 0.0085 |
| 7 | 0.0085 | 0.0085 |
| 8 | 0.0084 | 0.0084 |
| 9 | 0.0084 | 0.0083 |
| 10 | 0.0083 | 0.0082 |
| 11 | 0.0082 | 0.0082 |
| 12 | 0.0081 | 0.0081 |

As can be seen that almost all the run length probabilities are the same and constant up to 3 decimal values.

# CHAPTER 3

# STATISTICAL CONTROL OF THE CORRELATION COEFFICIENT

## 3.1 Distribution of the Sample Correlation Coefficient.

### *3.1.1 The distribution when $\rho = 0$.*

Suppose samples of size n are obtained from a bivariate Normal population with mean vector

$$\underline{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

and variance covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\,\sigma_1\sigma_2 \\ \rho\,\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}.$$

The sample mean vector

$$\underline{\bar{x}} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix}; \quad \text{where} \quad \bar{x}_j = \frac{1}{n}\sum_{k=1}^{n} x_{jk}, \quad \text{for} \quad j = 1, 2.$$

and the sample variance covariance matrix

$$S = \begin{bmatrix} s_1^2 & s_{12} \\ s_{12} & s_2^2 \end{bmatrix}$$

$$= \frac{1}{n-1} \begin{bmatrix} \sum_{k=1}^{n} (x_{1k} - \bar{x}_1)^2 & \sum_{k=1}^{n} (x_{1k} - \bar{x}_1)(x_{2k} - \bar{x}_2) \\ \sum_{k=1}^{n} (x_{1k} - \bar{x}_1)(x_{2k} - \bar{x}_2) & \sum_{k=1}^{n} (x_{2k} - \bar{x}_2)^2 \end{bmatrix}$$

are computed. The sample correlation coefficient, r, is given by

$$r = \frac{s_{12}}{s_1\,s_2}$$

$$= \frac{\sum\limits_{k=1}^{n} (x_{1k} - \bar{x}_1)(x_{2k} - \bar{x}_2)}{\sqrt{\sum\limits_{k=1}^{n} (x_{1k} - \bar{x}_1)^2 \sum\limits_{k=1}^{n} (x_{2k} - \bar{x}_2)^2}} .$$

For $\rho = 0$, the distribution of $r$ was given by Fisher (1915):

$$f(r; n, \rho=0) = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{\pi}\ \Gamma\left(\frac{n-2}{2}\right)} (1 - r^2)^{\frac{n-4}{2}}, \qquad -1 \leq r \leq 1 \qquad (3.1.1)$$

It can be shown that the transformed statistic

$$t = r \sqrt{\frac{n-2}{1-r^2}}$$

has Student's $t$ - distribution with $n - 2$ degrees of freedom.

Test of the hypothesis

$$H_o: \rho = 0 \quad \text{vs} \quad H_a: \rho \neq 0,$$

are then based on a $t$ - statistic with $n - 2$ degrees of freedom.

### 3.1.1 The distribution when $\rho \neq 0$.

The distribution of the sample correlation was first found by Fisher (1915):

$$f(r; n, \rho \neq 0) = \frac{(1 - \rho^2)^{\frac{n-1}{2}}}{\pi\ (n-3)!} (1 - r^2)^{\frac{n-4}{2}} \frac{d^{n-2}}{d(r\rho)^{n-2}} \left( \frac{\arccos(-\rho r)}{\sqrt{1 - \rho^2 r^2}} \right),$$

$$-1 \leq r \leq 1 \qquad (3.1.2)$$

And Anderson (1984) showed that (3.1.2) can be written as:

$$\frac{2^{n-3} (1 - \rho^2)^{\frac{n-1}{2}} (1 - r^2)^{\frac{n-4}{2}}}{\pi\ (n-3)!} \sum_{\alpha=0}^{\infty} \frac{(2\rho r)^\alpha}{\alpha!}\ \Gamma^2\left(\frac{n+\alpha-1}{2}\right),$$

$$-1 \leq r \leq 1 \qquad (3.1.3)$$

Hotelling (1953) suggested the following form, involving the hypergeometric function, which converges more rapidly than the infinite series in (3.1.3):

$$g\,(r\,;\,\rho)\;=\;\frac{(n-2)\,\Gamma(n-1)}{\sqrt{2\pi}\;\Gamma(n-\tfrac{1}{2})}\,(1-\rho^2)^{\frac{n-1}{2}}\,(1-r^2)^{\frac{n-4}{2}}\,(1-\rho r)^{-n+\frac{3}{2}}$$

$$F\left(\tfrac{1}{2},\tfrac{1}{2};\,n-\tfrac{1}{2};\,\frac{1+\rho r}{2}\right),\qquad -1\le r\le 1,\qquad(3.1.4)$$

where

$$F\,[a,\,b;\,c;\,z]\;=\;\sum_{j=0}^{\infty}\frac{\Gamma(a+j)\,\Gamma(b+j)\,\Gamma(c)}{\Gamma(a)\,\Gamma(b)\,\Gamma(c+j)}\,\frac{z^{j}}{j!}\,,$$

is the usual Gaussian hypergeometric function. David (1938) has tabulated the cumulative distribution of r, as

$$Pr(\,r\le r')\;=\;\int_{t=-1}^{r}f(t)\,dt\;=\;F(r';\,n,\rho)$$

for the selection of parameter values:

(i)  n = 3(1)25, 50, 100,200,400

(ii) $\rho$ = 0.0(0.1)0.9

(iii) r = -1.00(0.05)1.00

It is clear from the density (3.1.3) that

$$F(r';\,n,\,\rho)\;=\;1-F(-r';\,n,\,-\rho)$$

because the density for r, $\rho$ is equal to the density for -r, -$\rho$.

More recently Subrahmaniam and Subrahmaniam (1972) extended the David's tables to cover the values of n=26(1)49 for a variety of values of r. In addition these tables give the percentiles of the distribution of r for

$$p\;=\;\int_{t=-1}^{r}f(t)\,dt\;,$$

with  p = .01, .02, .025, .05, .10, .20, .80, .90, .95, .975, .98, .99.

When n is large, the asymptotic distribution of a sample correlation coefficient r is given by the statistic

$$z = \frac{1}{2} \ln\frac{1 + r}{1 - r}$$

called " Fisher's z". Here z has an asymptotic normal distribution with mean

$$\frac{1}{2} \ln\frac{1 + \rho}{1 - \rho},$$

and standard deviation

$$\frac{1}{\sqrt{n-3}}.$$

## 3.2   Control Charts for the Correlation Coefficient.

### 3.2.1   $\rho$ - control chart when $\rho = 0$.

To construct the control charts for the sample correlation we will use the distributions of r, that is discussed in the previous section 3.1.

When we consider the case of independence between the variables, that is testing for $\rho = 0$, we do not need to control the two variables simultaneously; it is sufficient to control them individually. In a particular production process to see whether the two quality characteristics that we are interested are independent, we will construct a $\rho$ control chart when $\rho = 0$. For this the statistic used is

$$t = r\sqrt{\frac{n - 2}{1 - r^2}}$$

and the limits of the control chart are given by

$$UCL = t_{n-2, \alpha/2}$$

$$LCL = -t_{n-2, \alpha/2}$$

where $t_{n-2, \alpha/2}$ denotes the upper $\alpha/2\%$ point of the student's t - distribution on n-2 degrees of freedom.

To see whether the process correlation is zero, take samples of size n and calculate the statistic  t  for each sample obtained and plot on  a chart with the above specified control limits and if the points fall outside the control limits then the variables considered in the process are not independent anymore.   If we want to keep the correlation zero, then the assignable causes have to be sought.

To illustrate the effectiveness of this chart to detect the shifts in $\rho$;  1000 samples each of size n=10 have been generated using the IMSL subroutine GGNPM (Appendix C: program 1), and $\rho$ has been changed from 0.1 through 0.9 in steps of 0.1.  The percentage of points plotted outside the limits in each case has been tabulated in Table 3.1.  Here the t-statistic is calculated and plotted on a control chart with

UCL=2.3060 and LCL=-2.3060.

Table 3.1

| $\rho$ | % of points plotted outside the limits |
|---|---|
| 0.1 | 5.6 |
| 0.2 | 9.3 |
| 0.3 | 12.8 |
| 0.4 | 21.6 |
| 0.5 | 32.5 |
| 0.6 | 46.7 |
| 0.7 | 66.7 |
| 0.8 | 87.2 |
| 0.9 | 98.2 |

From this we can see that as $\rho$ is shifted from zero the number of points that are falling out of control increases as the shift increases. When $\rho$=0.1 only 5.6% of the points fall outside the limits and when $\rho$=0.9 almost all the points are outside the limits.  Here the samples are of size n = 10.

When the sample size is increased, the shift is detected more quickly. To illustrate this 1000 samples were generated using the IMSL subroutine GGNPM (Appendix C, program 2) with $\rho=0.5$, for $n = 5, 15, 20, 30, 40$. For each n the percentage of points exceeding the limits is given in Table 3.2.

Table 3.2

| n | UCL | LCL | % of points exceeding the limits |
|---|-----|-----|-----------------------------------|
| 5 | 3.1824 | -3.1824 | 12.5 |
| 15 | 2.1604 | -2.1604 | 47.5 |
| 20 | 2.1009 | -2.1009 | 63.6 |
| 30 | 2.0484 | -2.0484 | 82.6 |
| 40 | 1.96 | -1.96 | 94.3 |

Here also as n increases the shift in $\rho$ is more obvious.

### 3.2.1 $\rho$ - control chart when $\rho \neq 0$.

When the two quality characteristics we take into consideration are correlated with the correlation coefficient $\rho_0$, we test whether the production process keeps the two quality characteristics correlation at a specified value $\rho_0$.

Then the statistic r is computed, and the control limits are given by

$$UCL = r_{n,\alpha/2}$$
$$LCL = r_{n,1-\alpha/2}.$$

where $r_{n,\gamma}$ is the value tabulated in table on Subramaniam's paper for the sample size n and $\gamma$% point. Also the test can be performed for specified values of $\rho = 0.1(0.1)0.9$. To illustrate this chart we generate 1000 samples each of size n=5, 10, 15, 20, 30, 50, 100, with $\rho = 0.5$ (Appendix C, program 3). And we test for each n, whether the shift is detected more rapidly when $\rho$ is varied from 0.1 through 0.9 in steps of 0.1.

The control limits for the charts for each n and $\rho$ are obtained from table [Subrahmaniam and Subrahmaniam (1972)] and are given in table 3.3. In the program the READ statement reads the values in table 3.3.

Table 3.3

| $\rho$ | Control limits | n 5 | 10 | 15 | 25 | 50 |
|--------|----------------|-----|-----|-----|-----|-----|
| 0.1 | UCL | 0.9026 | 0.6917 | 0.5865 | 0.4788 | 0.3693 |
|     | LCL | -0.8482 | -0.5633 | -0.4331 | -0.3063 | -0.1828 |
| 0.2 | UCL | 0.9223 | 0.7440 | 0.6517 | 0.5553 | 0.4551 |
|     | LCL | -0.8099 | -0.4839 | -0.3427 | -0.2086 | -0.0812 |
| 0.3 | UCL | 0.9386 | 0.7901 | 0.7107 | 0.6259 | 0.5362 |
|     | LCL | -0.7606 | -0.3916 | -0.2410 | -0.1021 | 0.0265 |
| 0.4 | UCL | 0.9521 | 0.8310 | 0.7641 | 0.6915 | 0.6131 |
|     | LCL | -0.6955 | -0.2835 | -0.1263 | 0.0142 | 0.1409 |
| 0.5 | UCL | 0.9635 | 0.8673 | 0.8127 | 0.7523 | 0.6861 |
|     | LCL | -0.6072 | -0.1556 | 0.0037 | 0.1416 | 0.2624 |
| 0.6 | UCL | 0.9732 | 0.8997 | 0.8570 | 0.8089 | 0.7553 |
|     | LCL | -0.4839 | -0.0034 | 0.1518 | 0.2814 | 0.3916 |
| 0.7 | UCL | 0.9814 | 0.9288 | 0.8974 | 0.8617 | 0.8211 |
|     | LCL | -0.3056 | 0.1793 | 0.3211 | 0.4352 | 0.5293 |
| 0.8 | UCL | 0.9885 | 0.9549 | 0.9345 | 0.9109 | 0.8837 |
|     | LCL | -0.0384 | 0.4005 | 0.5157 | 0.6048 | 0.6760 |
| 0.9 | UCL | 0.9947 | 0.9786 | 0.9686 | 0.9569 | 0.9433 |
|     | LCL | 0.3733 | 0.6700 | 0.7401 | 0.7923 | 0.8327 |

For each n the percentage of points exceeding the limits is given in table 3.4, for the values of n = 5, 10, 15, 25, 50; and for $\rho$ = 0.1(0.1)0.9.

Table 3.4

| percentage of points | | n | | | |
|---|---|---|---|---|---|
| exceeding the limits | 5 | 10 | 15 | 25 | 50 |
| 0.1 | 0.0 | 23.0 | 34.8 | 54.9 | 86.8 |
| 0.2 | 0.0 | 14.8 | 23.3 | 36.7 | 62.9 |
| 0.3 | 0.0 | 10.5 | 14.6 | 20.2 | 35.9 |
| 0.4 | 0.0 | 5.3 | 7.8 | 9.0 | 15.1 |
| $\rho$ 0.5 | 0.0 | 7.3 | 6.0 | 4.5 | 5.1 |
| 0.6 | 0.0 | 8.7 | 8.0 | 9.4 | 18.6 |
| 0.7 | 0.0 | 12.7 | 19.9 | 34.1 | 56.7 |
| 0.8 | 0.0 | 31.3 | 51.2 | 76.1 | 97.4 |
| 0.9 | 14.2 | 72.7 | 92.8 | 99.2 | 100.0 |

When n=5 the shift in $\rho$ is not detected until it shifts to 0.9. Since the actual value of $\rho$ is 0.5 the points exceeding the limits is around 5% only. For larger n the shift is detected more effectively and also the detection of the shift when $\rho$ is shifted further from 0.5 is quite obvious too.

## 3.3   The Sequential Testing procedures.

As we have seen in Chapter 2, the CUSUM schemes are actually sequential sampling procedures. In this section we consider some sequential procedures to test the change in the sample correlation.

The CUSUM control charts for the sample mean and variance have been discussed in Chapter 2. But unlike sample mean and sample variance the sample correlation coefficient takes into account the variation between the two variables and that makes it difficult to construct the CUSUM chart using the SPRT as we did earlier.

Essentially the CUSUM control charts have been constructed so that they take into consideration all the informations in the sequence of observations. In this section we consider three different sequential procedures so that all these procedures take into account all the information in the samples as one by one the samples enter the procedure. First we construct a CUSUM chart using the Fisher's-Z statistic.

### 3.3.1   CUSUM Chart for the correlation coefficient.

In section 2.1, when we constructed the CUSUM charts for the mean, the rejection region boundary line had the form of a linear equation

$$S_m = a\, m + b,$$

where

$$a = \frac{\delta}{2}$$

and

$$b = \frac{1}{\delta} \ln \left( \frac{1 - \alpha_1}{\alpha_0} \right).$$

We plotted $S_m$ against $m$ for each sample $m = 1, 2, \ldots$ .

If we proceed in the same way to construct the CUSUM chart for $\rho$ using the SPRT based on the hypothesis

$$H_0: \rho = \rho_0 \quad vs \quad H_1: \rho = \rho_1 \; (> \rho_0)$$

the likelihood ratio is quite complicated. It is not possible to seperate the partial sums and write them as a linear equation on the sample size. Instead we will use the Fisher's-Z statistic, which is asymptotically normally distributed, to construct the CUSUM chart for $\rho$.

Using the notation of section 3.1, it can be shown that the Fisher's-Z statistic

$$z = \frac{1}{2} \ln\frac{1 + r}{1 - r}$$

is approximately normally distributed with mean

$$\eta_0 = \frac{1}{2} \ln\frac{1 + \rho_0}{1 - \rho_0},$$

and standard deviation

$$\sigma = \frac{1}{\sqrt{n-3}} .$$

Now to test the hypothesis:

$$H_0: \rho = \rho_0 \quad vs \quad H_1: \rho = \rho_1 \; (> \rho_0),$$

let us consider the statistic

$$y_i = \frac{z_i - \eta_0}{\sigma}; \quad i = 1, 2, \ldots,$$

where $z_i$'s are the Fisher's-Z statistics.

Now using this statistic we could test for

$$H_0: \mu = \eta_0 \quad vs \quad H_1: \mu = \eta_1 ,$$

where

$$\eta_1 = \frac{1}{2} \ln\frac{1 + \rho_1}{1 - \rho_1} .$$

This is analogous to a test on the mean. Hence the CUSUM chart can be constructed based on $y_i$'s, where $y_i \sim N(0, 1)$; $(i = 1, 2, \ldots)$. As discussed in section 2.2.2, we can construct a two-sided CUSUM scheme for $y_i$'s. That is,

$$S_H (i) = \max [\, 0, \ y_i - k + S_H (i\text{-}1) \,]$$

$$S_L (i) = \max [\, 0, \ - y_i - k + S_L (i\text{-}1) \,],$$

where $S_H (0) = S_L (0) = S_0$. For a prescribed value k, the sequence is considered to be out of control whenever either $S_H$ or $S_L$ exceeds a prespecified value h.

To illustrate this, we consider an example to find the ARL under $H_0$ and $H_1$.

Samples of size 10 are generated each time, using the subroutine ZCAL (Appendix C, program 4), from a bivariate normal distribution with mean vector

$$\underline{\mu} = \begin{bmatrix} 30 \\ 15 \end{bmatrix},$$

variances $\sigma_1^2 = 8$, $\sigma_2^2 = 4$ and correlation coefficient $\rho$.

The target value for $\rho$ is taken as $\rho_0 = 0.1$. Then samples are generated for $\rho = 0.1, 0.5$ and $0.9$. For values of k, 0.5, 1.0 and 1.5; h , 2.25, 1.50 and 0.75 and the initial value $S_0 = h/2$.

Table 3.5

| k | h | $\rho$ ($\mu$) | | |
|---|---|---|---|---|
| | | 0.1(0.10) | 0.5(0.55) | 0.9(1.47) |
| 0.5 | 2.25 | 20.695 | 2.243 | 1.015 |
| 1.0 | 1.50 | 35.551 | 3.141 | 1.009 |
| 1.5 | 0.75 | 31.568 | 3.972 | 1.026 |

In Table 3.5, the ARL's are tabulated for $\rho$ values, with the corresponding $\mu$ values in paranthesies, where

$$\mu = \frac{1}{2} \ln \frac{1+\rho}{1-\rho} \,.$$

Hence it makes it easy to compare with the results in Fellner 's (1990), table 1, where the values of ARL's are tabulated for a two-sided CUSUM scheme where the process has a distribution with mean $\Delta$ and variance 1. Hence in our example $\Delta = \mu/\sigma$, where $\sigma = 0.38$.

### 3.3.2   Sequential procedure for the exact distribution of r.

In this procedure a different approach is considered to construct a control chart. Here we start with a sample size n and calculate the sample correlation coefficient $r_n$ based on those n samples and begin the SPRT with $n \geq 3$, using the exact distribution of the correlation coefficient given in (3.1.4) [ Kollerstrom and Wetherill (1979), Kocherlakota, Kocherlakota and Balakrishnan (1983)].

Consider ( $x_{1i}$, $x_{2i}$ ), $i = 1, 2, \ldots$ as a sequence of independent observations from the bivariate normal population with parameters as mentioned above. For a sample of size n, let $\bar{x}_{1(n)}$ and $\bar{x}_{2(n)}$ be the sample means and the sample correlation coefficient be

$$r_n = \frac{\sum\limits_{k=1}^{n} \left( x_{1k} - \bar{x}_{1(n)} \right) \left( x_{2k} - \bar{x}_{2(n)} \right)}{\sqrt{\sum\limits_{k=1}^{n} \left( x_{1k} - \bar{x}_{1(n)} \right)^2 \sum\limits_{k=1}^{n} \left( x_{2k} - \bar{x}_{2(n)} \right)^2}} \,.$$

For $n \geq 3$, the probability density function of $r_n$ is given by

$$g\left( r_n ; \rho \right) = \frac{(n-2)\,\Gamma(n-1)}{\sqrt{2}\,\Gamma(\tfrac{1}{2})\,\Gamma(n-\tfrac{1}{2})} \left( 1 - \rho^2 \right)^{\frac{n-1}{2}} \left( 1 - r_n^2 \right)^{\frac{n-4}{2}}$$

$$\left( 1 - \rho r_n \right)^{-n+\frac{3}{2}} F\left( \frac{1}{2}, \frac{1}{2}; n-\frac{1}{2}; \frac{1+\rho r_n}{2} \right),$$

where $F(a, b; c; z)$ is Gauss' hypergeometric function defined by the series

$$F\left( a, b; c; z \right) = \sum\limits_{j=0}^{\infty} \frac{\Gamma(a+j)\,\Gamma(b+j)\,\Gamma(c)}{\Gamma(a)\,\Gamma(b)\,\Gamma(c+j)} \frac{z^j}{j!} \,,$$

and $-1 < r_n < 1$.

For this case, the invariant sequential probability ratio test (SPRT) of

$$H_0: \rho \le \rho_0 \quad \text{vs} \quad H_1: \rho \ge \rho_1,$$

$-1 < \rho_0 < \rho_1 < 1$, is based on the log-likelihood ratio for $n \ge 3$,

$$z_n(r_n) = \frac{n-1}{2} \ln\left(\frac{1 - \rho_1^2}{1 - \rho_0^2}\right) - (n - \frac{3}{2}) \ln\left(\frac{1 - \rho_1 r_n}{1 - \rho_0 r_n}\right)$$

$$+ \ln\left\{\frac{F(\frac{1}{2}, \frac{1}{2}; n - \frac{1}{2}; \frac{1 + \rho_1 r_n}{2})}{F(\frac{1}{2}, \frac{1}{2}; n - \frac{1}{2}; \frac{1 + \rho_0 r_n}{2})}\right\}$$

is given by Ghosh ( 1970, pages 322-324 ).

Accept $H_0$ if $z_n(r_n) \le b^*$, reject $H_0$ if $z_n(r_n) \ge a^*$, with

$$a^* = \ln\frac{1 - \alpha_1}{\alpha_0}, \quad b^* = \ln\frac{\alpha_1}{1 - \alpha_0}.$$

otherwise, the process continues. Here again $\alpha_0$, $\alpha_1$ are the type I and type II error

probabilities. In quality control procedures only the rejection boundary is taken as the

control limit or decision boundary, the acceptance is considered along with the continue

sampling.

To illustrate this consider an example, where samples are generated using the IMSL

subroutine GGNPM (appendix C, program 5) begining with a size of $n = 3$, the sample

correlation coefficient r is calculated (subroutine RCAL) and the SPRT is constructed for

$\rho_0 = 0.1$ and $\rho_1 = 0.5$. But the samples are generated based on $\rho = 0.3, 0.5$ and 0.7. The

control limit $a^*$ is calculated for the sets of values $\alpha_0 = \alpha_1 = 0.05$ and $\alpha_0 = \alpha_1 = 0.1$,

based on theses control limits the run lengths are determined. Repeating this 100 times the

run lengths are calculated and the average run length is calculated for each case and are

tabulated in Table 3.6.

Table 3.6

| α | β | a* | ρ 0.3 | 0.5 | 0.7 |
|---|---|---|---|---|---|
| 0.05 | 0.05 | 2.944 | 29.26 | 17.20 | 12.85 |
| 0.10 | 0.10 | 2.197 | 19.56 | 13.42 | 10.39 |

The run length frequency distribution is given as a bar chart for each case with the mean (ARL) and the variance based on the 100 runlengths obtained.

(i)   $\alpha = 0.05, \ \beta = 0.05.$

Graph 1:   $\rho = 0.3,$   ARL = 29.26, Variance = 144.09.
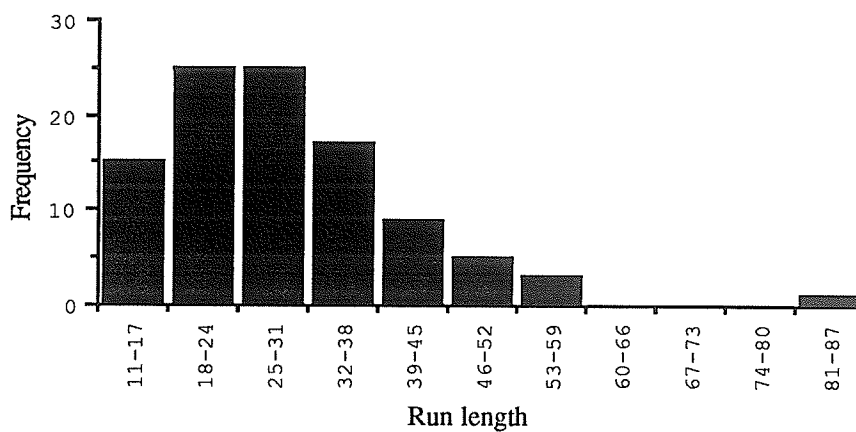


Graph 2:   $\rho = 0.5,$   ARL = 17.20, Variance = 14.83.
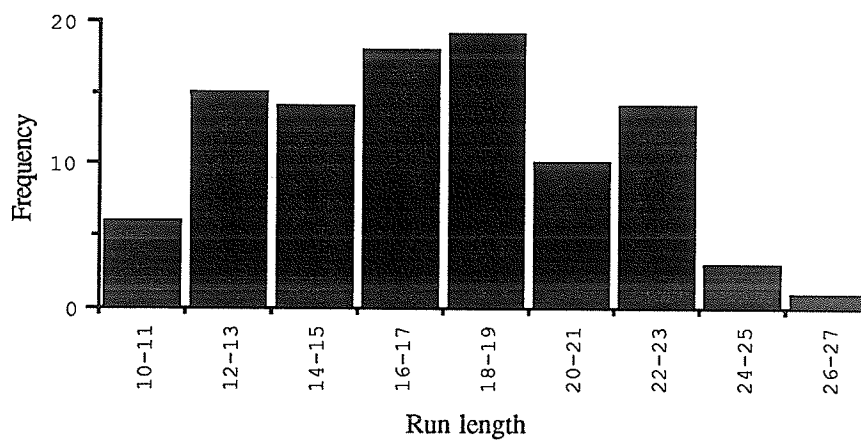
Graph 3:  ρ = 0.7,  ARL = 12.85, Variance = 2.957.



Run length

(ii) α = 0.1, β = 0.1.
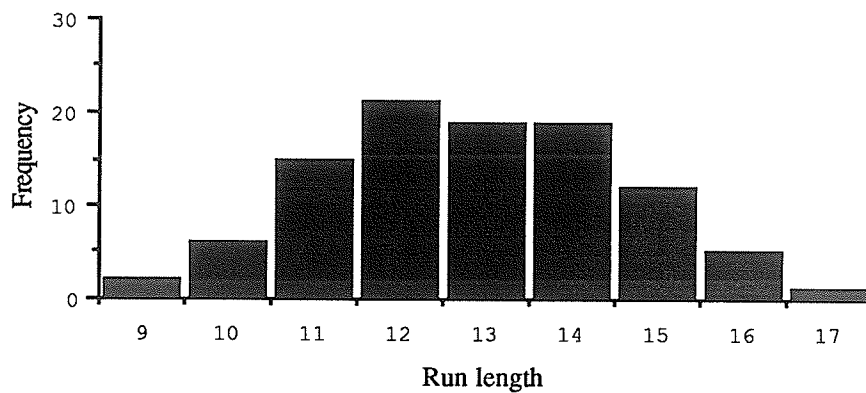
Graph 1:  ρ = 0.3,  ARL = 19.56,  Variance = 53.48.
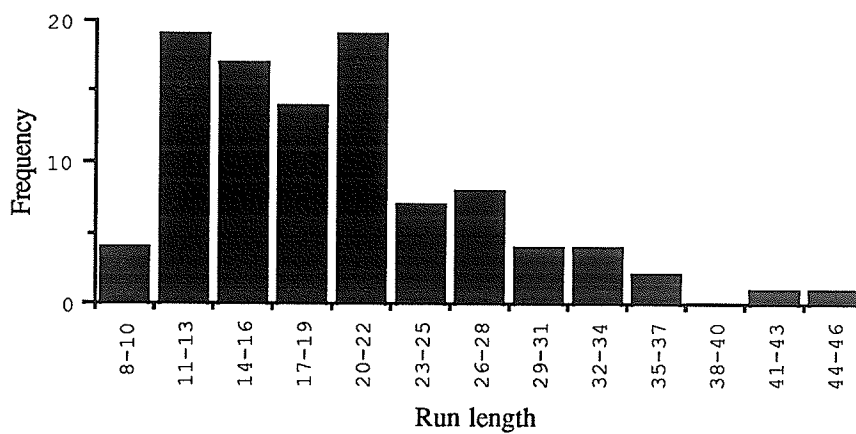


Run length

Graph 2:  ρ = 0.5,  ARL = 13.42,  Variance = 14.11.
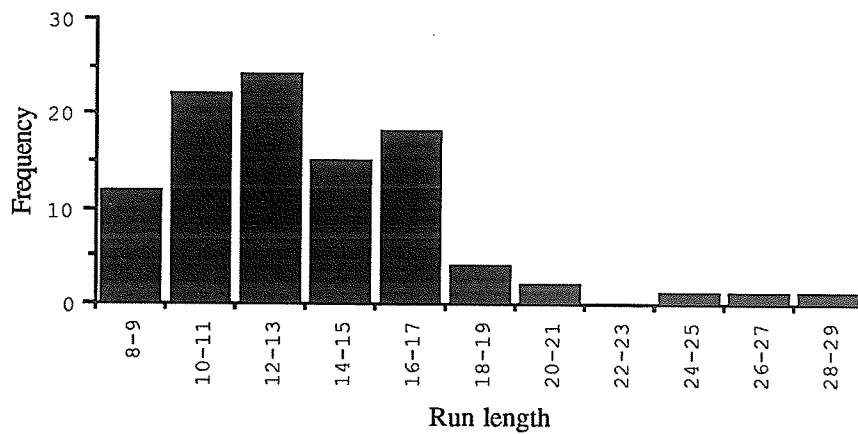


Run length

Graph 3: $\rho = 0.7$, ARL $= 10.39$, Variance $= 3.29$.
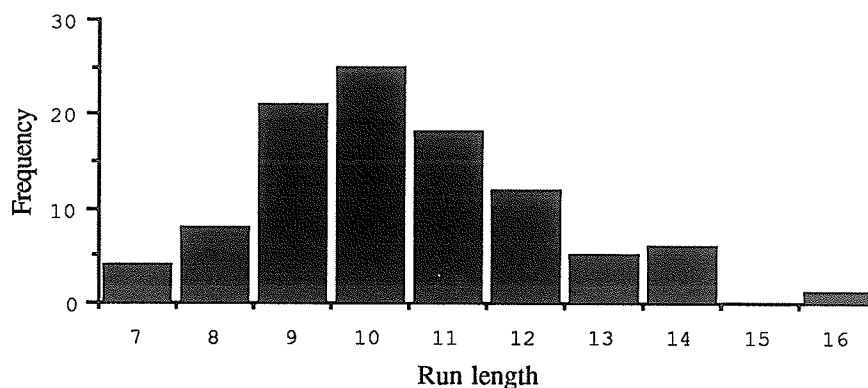


From the graphs it seems that the run length distribution may have possibly an approximate normal distribution as $\rho$ increases.

### 3.3.3 Sequential procedure using Fisher's-Z statistic.

Here again a similar kind of procedure as in section (3.3.2) is considered using the Fisher's-Z statistic,

$$u_n = \frac{1}{2} \ln \frac{1 + r_n}{1 - r_n},$$

where $r_n$ is the sample correlation coefficient based on the sample size n ($\geq 5$). Since $u_n$ has an asymptotic normal distribution with mean

$$\mu = \frac{1}{2} \ln \frac{1 + \rho}{1 - \rho}$$

and variance

$$\sigma^2 = \frac{1}{n - 3}.$$

the SPRT is based on the hypothesis

$$H_0: \mu = \eta_0 \quad \text{vs} \quad H_1: \mu = \eta_1,$$

where

$$\eta_0 = \frac{1}{2} \ln \frac{1 + \rho_0}{1 - \rho_0}$$

and

$$\eta_1 = \frac{1}{2} \ln \frac{1 + \rho_1}{1 - \rho_1}$$

as described earlier. That is the actual test is based on

$$\rho = \rho_0 \quad vs \quad \rho = \rho_1$$

Consider the density function of $u_n$

$$f(u_n) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp -\frac{1}{2}\left(\frac{u_n - \mu}{\sigma}\right)^2.$$

The likelihood ratio is,

$$\frac{L_1}{L_0} = \exp -\frac{1}{2\sigma^2}\left\{ (u_n - \eta_1)^2 - (u_n - \eta_0)^2 \right\}$$

$$= \exp -\frac{1}{2\sigma^2}\left\{(\eta_1^2 - \eta_0^2) - 2 u_n (\eta_1 - \eta_0)\right\}$$

The SPRT of

$$H_0: \mu = \eta_0 \quad vs \quad H_1: \mu = \eta_1$$

based on the log likelihood ratio

$$z_n(r_n) = \ln\frac{L_1}{L_0}$$

$$= \frac{u_n}{\sigma^2}(\eta_1 - \eta_0) - \frac{1}{2\sigma^2}(\eta_1^2 - \eta_0^2)$$

is given by Ghosh (1970).

Accept $H_0$ if $z_n(r_n) \leq b^*$, reject $H_0$ if $z_n(r_n) \geq a^*$, with

$$a^* = \ln\frac{1 - \alpha_1}{\alpha_0}, \quad b^* = \ln\frac{\alpha_1}{1 - \alpha_0}.$$

otherwise, the process continues. Here again $\alpha_0$, $\alpha_1$ are the type I and type II error probabilities. In quality control procedures only the rejection boundary is taken as the control limit or decision boundary, the acceptance is considered along with the continue sampling.

To illustrate this consider an example, where samples are generated using the IMSL subroutine GGNPM (appendix C, program 5) begining with a size of n = 5, the sample correlation coefficient r is calculated (subroutine RCAL) and the SPRT is constructed for $\rho_0 = 0.1$ and $\rho_1 = 0.5$. But the samples are generated based on $\rho = 0.2, 0.5$ and $0.8$. The control limit $a^*$ is calculated for the sets of values $\alpha_0 = \alpha_1 = 0.05$ and $\alpha_0 = \alpha_1 = 0.1$, based on these control limits the run lengths are determined. Repeating this 100 times the run lengths are calculated and the average run length is calculated for each case and are tabulated in Table 3.7.

Table 3.7

| | | | $\rho$ | | |
|---|---|---|---|---|---|
| $\alpha$ | $\beta$ | $a^*$ | 0.2 | 0.5 | 0.8 |
| 0.05 | 0.05 | 2.944 | 47.51 | 16.08 | 9.96 |
| 0.10 | 0.10 | 2.197 | 29.29 | 13.19 | 8.44 |

The run length frequency distribution is given as a bar chart for each case with the mean (ARL) and the variance based on the 100 runlengths obtained.

(i)  $\alpha = 0.05,\ \beta = 0.05.$

Graph 1:     $\rho = 0.2,$     ARL = 47.51,     Variance = 753.83.
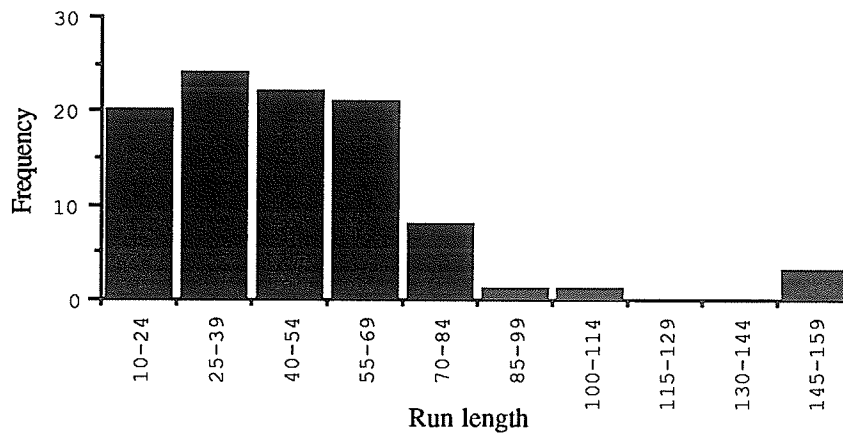


Graph 2:   $\rho = 0.5,$   ARL = 16.08,   Variance = 19.61
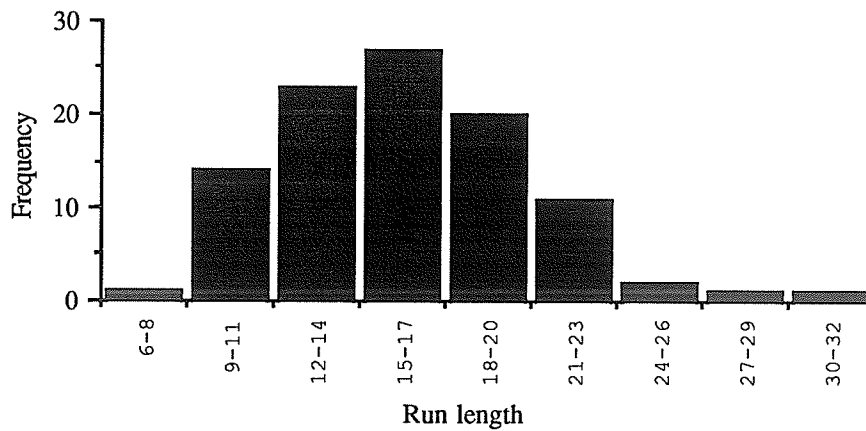


Graph 3:   $\rho = 0.8,$   ARL = 9.96,   Variance = 3.17.
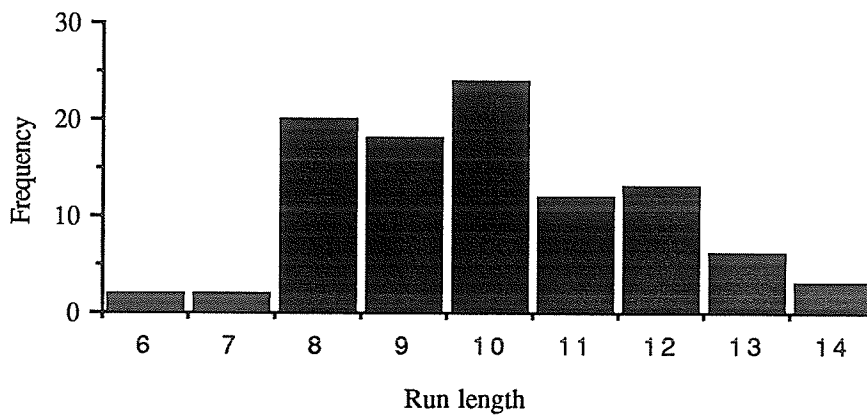
(ii) $\alpha = 0.1, \ \beta = 0.1$

Graph 1: $\rho = 0.2$, ARL = 29.29, Variance = 345.137.
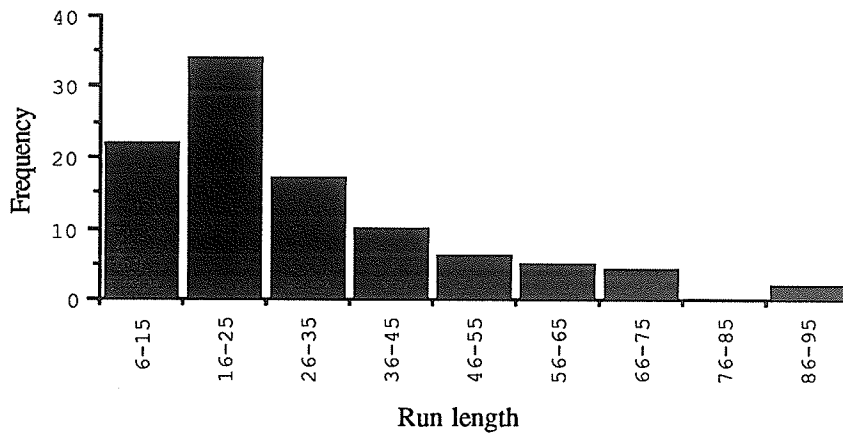


Graph 2: $\rho = 0.5$, ARL = 13.19, Variance = 14.70.
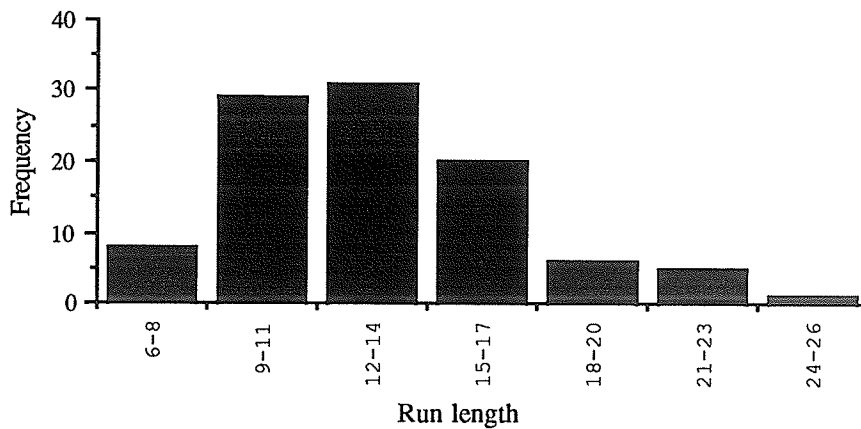


Graph 3: $\rho = 0.8$, ARL = 8.44, Variance = 2.65.
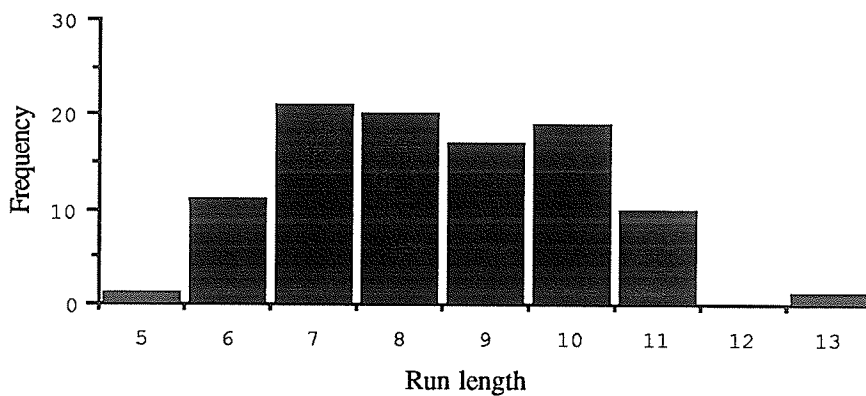
When we use Fisher's- Z statistic also we get the run length distribution close to the exact distribution run length values. This can be seen by comparing the ARL's in each case, when $\rho = 0.5$. Here also the bar charts of the run length distribution seems to tend to a normal distribution.

# Bibiliography

Alt, F. B. (1982). Bonferroni inequalities and intervals. In S. Kotz and N. L. Johnson, eds., *Encyclopedia of Statistical Sciences,* 1, 294-300. New York: John Wiley & Sons, Inc.

Alt, F. B. (1985). Multivariate Quality Control. In S. Kotz and N. L. Johnson, eds., *Encyclopedia of Statistical Sciences,* 6, 110-122. New York: John Wiley & Sons, Inc.

Alt, F. B. and Smith, N. D. (1988). Multivariate process control. In P. R. Krishnaiah and C. R. Rao, eds., *Quality Control and Reliability: Handbook of Statistics,* 7, 333-351. Amsterdam: North Holland.

Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis.* New York: John Wiley & Sons, Inc.

Armitage, P. (1950). Sequential analysis with more than two alternative hypotheses and its relation to discriminant analysis. *Journal of the Royal Statistical Society B,* 12, 137-144.

Brook, D. and Evans, D. A. (1972). An approach to the probability distribution of cusum run length. *Biometrika,* 59, 539-549.

Champ, C. W. and Woodall, W. H. (1987). Exact results for Shewart control charts with Supplementary Runs Rules. *Technometrics,* 29, 393-399.

Crosier, R. B. (1986). A new two-sided cumulative sum quality control scheme. *Technometrics,* 28, 187-194.

Crosier, R. B. (1988). Multivariate generalizations of cumulative sum quality - control schemes. *Technometrics,* 30, 291-303.

David, F. N. (1938). *Tables of the Ordinates and Probability Integral of the Distribution of the Correlation Coefficient in Small Samples,* Cambridge University, Cambridge.

Deming, W. E. (1983). *Quality, and Competitive Position.* MIT Center for Advanced Engineering study. Mass.: Cambridge.

Duncan, A. J. (1974). *Quality Control and Industrial Statistics.* Homewood, Illinois: Richard D. Irwin, Inc.

Duncan, A. J. (1982). Quality control, statistical. In S. Kotz and N. L. Johnson, eds., *Encyclopedia of Statistical Sciences,* 7, 380-393. New York: John Wiley & Sons, Inc.

Ewan, W. D. and Kemp, K. W. (1960). Sampling inspection of continuous processes with no autocorrelation between successive results. *Biometrika,* 47, 363-380.

Fellner, W. H. (1990). Algorithim AS258. Average run length for cumulative sum schemes. *Applied Statistics,* 39, 402-412.

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika,* 10, 507-521.

Ghosh, B. K. (1970). *Sequential Tests of Statistical Hypotheses.* Reading, Mass.: Addison-Wesley.

Goel, A. L. and Wu, S. M. (1971). Determination of ARL and a contour nomogram for cusum charts to control normal means. *Technometrics,* 13, 221-230.

Goldsmith, P. L. and Whitfield, H. (1961). Average run lengths in cumulative chart quality control schemes. *Technometrics,* 3, 11-20.

Healy, J. D. (1987). A note on multivariate CUSUM procedures. *Technometrics,* 29, 409-412.

Hoel, P. G. (1937). A significance test for component analysis. *Annals of Mathematical Statistics,* 8, 149-158.

Hotelling, H. (1947). Multivariate Quality Control, illustrated by the air testing of sample bombsights. In C. Eisenhart, M. Hastay, and W. A. Wallis, eds., *Techniques of Statistical Analysis*, 111-184. New York: McGraw Hill.

Hotelling, H. (1953). New light on the correlation coefficient and its transforms (with discussion). *Journal of the Royal Statistical Society B,* 15, 193-232.

IMSL (1980). *The IMSL Libraries:* FORTRAN subroutines for Mathematics and Statistics. edition 8. Houston, Texas: IMSL, Inc.

Johnson, N. L. and Leone, F. C. (1962a). Cumulative sum control charts: Mathematical principles applied to their construction and use Part I. *Industrial Quality Control,* 18, 15-21.

Johnson, N. L. and Leone, F. C. (1962b). Cumulative sum control charts: Mathematical principles applied to their construction and use Part II. *Industrial Quality Control,* 18, 29-36.

Johnson, N. L. and Leone, F. C. (1962c). Cumulative sum control charts: Mathematical principles applied to their construction and use Part III. *Industrial Quality Control,* 18, 22-28.

Kemp, K. W. (1961). The average run length of the cumulative sum chart when a V-mask is used. *Journal of the Royal Statistical Society, B,* 23, 149-153.

Kocherlakota, S., Kocherlakota, K. and Balakrishnan, N. (1983). A note on the SPRT for the correlation coefficient: Symmetric case. *Technical Report No.* 141, Department of Statistics, University of Manitoba.

Kollerstrom, J. and Wetherill, B. G. (1979). SPRT's for the normal correlation coefficient. *Journal of American Statistical Association,* 74, 815-821.

Kshirsagar, A. M. (1972). *Multivariate Analysis.* New York: Marcel Dekker, Inc.

Lucas, J. M. and Crosier, R. B. (1982). Fast initial response for CUSUM quality - control schemes: Give Your CUSUM A Head Start. *Technometrics,* 24, 199-205.

Montgomery, D. C. (1985). *Introduction to Statistical Quality Control.* New York: John Wiley & Sons, Inc.

Morrison, D. F. (1983). *Applied Linear Statistical Methods.* New Jersey: Prentice - Hall, Inc.

Page, E. S. (1954). Continuous inspection Schemes. *Biometrika,* 41, 100-115.

Pearson, E. S. (1973). Some historical reflections on the introduction of statistical methods in industry. *The Statistician,* 22, 165-179.

SAS (1984, 1986). *SAS User's Guide.* Cary, North Carolina: SAS Institute Inc.

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

Subrahmaniam, Kathleen and Subrahmaniam, Kocherlakota (1972). Some extensions to Miss. F. N. David's tables of the sample correlation coefficient: distribution and percentiles. *Sankhya B,* 45, 75-147.

Wald, A. (1947). *Sequential Analysis.* New York: John Wiley & Sons, Inc.

Woodall, W. H. (1984). On the Markov chain approach to the two-sided CUSUM procedure. *Technometrics,* 26, 41-46.

Woodall, W. H. (1985). The statistical design of quality control charts. *The Statistician,* 34, 155-160.

Woodall, W. H. and Ncube, M. M. (1985). Multivariate CUSUM quality - control procedures. *Technometrics,* 27, 285-292.

# APPENDIX A: SAS Programs of Chapter 1

Program 1 :

```
1.  // JOB
2.  // EXEC SAS
3.  OPTION LINESIZE=74;
4.  DATA BIVNOR;
5.  DO I=1 TO 20;
6.  S1=0;
7.  S2=0;
8.  DO J=1 TO 10;
9.  N1=RANNOR(6372610);
10. N2=RANNOR(3026478);
11. X1=2*SQRT(2)*N1+30;
12. X2=N1+SQRT(3)*N2+15;
13. S1=S1+X1;
14. S2=S2+X2;
15. END;
16. XB1=S1/10;
17. XB2=S2/10;
18. C1=((XB1-30)**2)/8;
19. C2=((XB2-15)**2)/4;
20. C3=((XB1-30)*(XB2-15))/(4*SQRT(2));
21. CHIS=10*(C1+C2-C3)/(1-0.5**2);
22. OUTPUT;
23. END;
24. DROP I S1 S2 J N1 N2 X1 X2 C1 C2 C3;
25. PROC PRINT;
```

Program 2:

```
1.  // JOB
2.  // EXEC SAS
3.  OPTION LINESIZE=74;
4.  DATA BIVNOR;
```

```
 5.  DO I=1 TO 20;
 6.  S1=0;
 7.  S2=0;
 8.  IF I<=5 THEN DO;
 9.              MU1=30;
10.              MU2=15;
11.              END;
12.  IF I>5 AND I<11 THEN DO;
13.              MU1=30+0.65*(I-5);
14.              MU2=15;
15.              END;
16.  IF I>10 AND I<16 THEN DO;
17.              MU1=30;
18.              MU2=15+0.3*(I-10);
19.              END;
20.  IF I>15 THEN DO;
21.              MU1=30+0.65*(I-15);
22.              MU2=15+0.3*(I-15);
23.              END;
24.  DO J=1 TO 10;
25.  N1=RANNOR(6392610);
26.  N2=RANNOR(3026471);
27.  X1=2*SQRT(2)*N1+MU1;
28.  X2=N1+SQRT(3)*N2+MU2;
29.  S1=S1+X1;
30.  S2=S2+X2;
31.  END;
32.  XB1=S1/10;
33.  XB2=S2/10;
34.  C1=((XB1-30)**2)/8;
35.  C2=((XB2-15)**2)/4;
36.  C3=((XB1-30)*(XB2-15))/(4*SQRT(2));
37.  MCHIS=10*(C1+C2-C3)/(1-0.5**2);
38.  OUTPUT;
39.  END;
40.  DROP I S1 S2 J N1 N2 X1 X2 C1 C2 C3 MU1 MU2;
```

```
41. PROC PRINT;
```

Program 3:

```
 1. // JOB
 2. // EXEC SAS
 3. OPTION LINESIZE=72;
 4. DATA BIVNOR;
 5. ARRAY X1(10) Y1-Y10;
 6. ARRAY X2(10) Y11-Y20;
 7. ARRAY XB1(10) YB1-YB10;
 8. ARRAY XB2(10) YB11-YB20;
 9. ARRAY S1(10) L1-L10;
10. ARRAY S2(10) L11-L20;
11. ARRAY S12(10) L21-L30;
12. DO I=1 TO 10;
13. T1=0; MU1=30;
14. T2=0; MU2=15;
15. IF I=3 THEN DO;
16.             MU1=33; MU2=17;
17.             END;
18. IF I=5 THEN DO;
19.             MU1=27; MU2=13;
20.             END;
21. DO J=1 TO 10;
22. N1=RANNOR(2891063);
23. N2=RANNOR(5620192);
24. X1(J)=2*SQRT(2)*N1+MU1;
25. X2(J)=N1+SQRT(3)*N2+MU2;
26. T1=T1+X1(J);
27. T2=T2+X2(J);
28. END;
29. XB1(I)=T1/10;
30. XB2(I)=T2/10;
31. T3=0; T4=0; T5=0;
32. DO K=1 TO 10;
```

```
33. T3=T3+(X1(K)-XB1(I))**2;
34. T4=T4+(X2(K)-XB2(I))**2;
35. T5=T5+(X1(K)-XB1(I))*(X2(K)-XB2(I));
36. END;
37. S1(I)=T3/9; S2(I)=T4/9; S12(I)=T5/9;
38. END;
39. T1=0; T2=0; T3=0; T4=0; T5=0;
40. DO I=1 TO 10;
41. T1=T1+XB1(I);
42. T2=T2+XB2(I);
43. T3=T3+S1(I);
44. T4=T4+S2(I);
45. T5=T5+S12(I);
46. END;
47. XBB1=T1/10; XBB2=T2/10;
48. S1BS=T3/10; S2BS=T4/10; S12B=T5/10;
49. DETSB=S1BS*S2BS-S12B**2;
51. DO I=1 TO 10;
52. M1=(XB1(I)-XBB1)**2;
53. M2=(XB2(I)-XBB2)**2;
54. M12=(XB1(I)-XBB1)*(XB2(I)-XBB2);
55. TO1=(10/DETSB)*(M1*S2BS+M2*S1BS-2*M12*S12B);
56. OUTPUT;
57. END;
58. DROP Y1-Y20 YB1-YB20 L1-L30 MU1 MU2 N1 N2;
59. DROP I J K XBB1 XBB2 S1BS S2BS S12B DETSB T1-T5 M1 M2 M12;
60. PROC PRINT;
```

Program 4:

```
1. // JOB
2. // EXEC SAS
3. OPTION LINESIZE=72;
4. DATA BIVNOR;
5. ARRAY X1(10) Y1-Y10;
6. ARRAY X2(10) Y11-Y20;
```

```
 7. ARRAY XB1(10) YB1-YB10;
 8. ARRAY XB2(10) YB11-YB20;
 9. ARRAY S1(10) L1-L10;
10. ARRAY S2(10) L11-L20;
11. ARRAY S12(10) L21-L30;
12. DO I=1 TO 10;
13. T1=0; MU1=30;
14. T2=0; MU2=15;
15. IF I=3 THEN I=I+1;
16. IF I=5 THEN I=I+1;
17. DO J=1 TO 10;
18. N1=RANNOR(2891063);
19. N2=RANNOR(5620192);
20. X1(J)=2*SQRT(2)*N1+MU1;
21. X2(J)=N1+SQRT(3)*N2+MU2;
22. T1=T1+X1(J);
23. T2=T2+X2(J);
24. END;
25. XB1(I)=T1/10;
26. XB2(I)=T2/10;
27. T3=0; T4=0; T5=0;
28. DO K=1 TO 10;
29. T3=T3+(X1(K)-XB1(I))**2;
30. T4=T4+(X2(K)-XB2(I))**2;
31. T5=T5+(X1(K)-XB1(I))*(X2(K)-XB2(I));
32. END;
33. S1(I)=T3/7; S2(I)=T4/7; S12(I)=T5/7;
34. END;
35. T1=0; T2=0; T3=0; T4=0; T5=0;
36. DO I=1 TO 10;
37. IF I=3 THEN I=I+1;
38. IF I=5 THEN I=I+1;
39. T1=T1+XB1(I);
40. T2=T2+XB2(I);
41. T3=T3+S1(I);
42. T4=T4+S2(I);
```

```
43. T5=T5+S12(I);
44. END;
45. XBB1=T1/8; XBB2=T2/8;
46. S1BS=T3/8; S2BS=T4/8; S12B=T5/8;
47. DETSB=S1BS*S2BS-S12B**2;
48. DO I=1 TO 10;
49. IF I=3 THEN I=I+1;
50. IF I=5 THEN I=I+1;
51. M1=(XB1(I)-XBB1)**2;
52. M2=(XB2(I)-XBB2)**2;
53. M12=(XB1(I)-XBB1)*(XB2(I)-XBB2);
54. TO1=(8/DETSB)*(M1*S2BS+M2*S1BS-2*M12*S12B);
55. OUTPUT;
56. END;
57. DROP Y1-Y20 YB1-YB20 L1-L30 N1 N2 MU1 MU2;
58. DROP I J K T1-T5 XBB1 XBB2 S1BS S2BS S12B DETSB M1 M2 M12;
59. PROC PRINT;
```

Program 5:

```
 1. // JOB
 2. // EXEC SAS
 3. OPTION LINESIZE=72;
 4. DATA BIVNOR;
 5. ARRAY X1(10) Y1-Y10;
 6. ARRAY X2(10) Y11-Y20;
 7. ARRAY XB1(10) YB1-YB10;
 8. ARRAY XB2(10) YB11-YB20;
 9. ARRAY S1(10) L1-L10;
10. ARRAY S2(10) L11-L20;
11. ARRAY S12(10) L21-L30;
12. DO I=1 TO 10;
13. T1=0; T2=0;
14. MU1=29.511; MU2=14.906;
15. SIG1=11.398; SIG2=5.934;
16. ROH=4.227/(SQRT(11.398*5.934));
```

```
17.  DO J=1 TO 10;
18.  N1=RANNOR(7490861);
19.  N2=RANNOR(3854509);
20.  X1(J)=SQRT(SIG1)*N1+MU1;
21.  X2(J)=ROH*SQRT(SIG2)*N1+N2*SQRT(SIG2)*SQRT(1-ROH**2)+MU2;
22.  T1=T1+X1(J);
23.  T2=T2+X2(J);
24.  END;
25.  XB1(I)=T1/10;
26.  XB2(I)=T2/10;
27.  T3=0;  T4=0;  T5=0;
28.  DO K=1 TO 10;
29.  T3=T3+(X1(K)-XB1(I))**2;
30.  T4=T4+(X2(K)-XB2(I))**2;
31.  T5=T5+(X1(K)-XB1(I))*(X2(K)-XB2(I));
32.  END;
33.  S1(I)=T3/9;  S2(I)=T4/9;  S12(I)=T5/9;
34.  END;
35.  T1=0;  T2=0;  T3=0;  T4=0;  T5=0;
36.  DO I=1 TO 10;
37.  T1=T1+XB1(I);
38.  T2=T2+XB2(I);
39.  T3=T3+S1(I);
40.  T4=T4+S2(I);
41.  T5=T5+S12(I);
42.  END;
43.  XBB1=T1/10;  XBB2=T2/10;
44.  S1BS=T3/10;  S2BS=T4/10;  S12B=T5/10;
45.  DETSB=S1BS*S2BS-S12B**2;
46.  DO I=1 TO 10;
47.  M1=(XB1(I)-XBB1)**2;
48.  M2=(XB2(I)-XBB2)**2;
49.  M12=(XB1(I)-XBB1)*(XB2(I)-XBB2);
50.  TOf=(10/DETSB)*(M1*S2BS+M2*S1BS-2*M12*S12B);
51.  OUTPUT;
52.  END;
```

```
53. DROP Y1-Y20 YB1-YB20 L1-L30 MU1 MU2 N1 N2 SIG1 SIG2 ROH;
54. DROP I J K XBB1 XBB2 S1BS S2BS S12B DETSB T1-T5 M1 M2 M12;
55. PROC PRINT;
```

Program 6:

```
1.  // JOB
2.  // EXEC SAS
3.  OPTION LINESIZE=72;
4.  DATA BIVNOR;
5.  ARRAY X1(10) G1-G10;
6.  ARRAY X2(10) H1-H10;
7.  DO I=1 TO 20;
8.  T1=0; T2=0;
9.  DO J=1 TO 10;
10. N1=RANNOR(6312510);
11. N2=RANNOR(3021471);
12. X1(J)=2*SQRT(2)*N1+30;
13. X2(J)=N1+SQRT(3)*N2+15;
14. T1=T1+X1(J);
15. T2=T2+X2(J);
16. END;
17. M1=T1/10; M2=T2/10;
18. T3=0; T4=0; T5=0;
19. DO K=1 TO 10;
20. T3=T3+(X1(K)-M1)**2;
21. T4=T4+(X2(K)-M2)**2;
22. T5=T5+(X1(K)-M1)*(X2(K)-M2);
23. END;
24. S1=T3/9; S2=T4/9;
25. SS1=SQRT(S1); SS2=SQRT(S2);
26. SD1=S1; SD2=S2;
27. S12=T5/9; S12S=S12**2;
28. DETS=S1*S2-S12S;
29. SDS=SQRT(DETS);
30. RO1=1-0.5**2;
```

```
31. TR=(S1/8+S2/4-S12/(4*SQRT(2)))/RO1;
32. W=-18-9*LOG(DETS)+9*LOG(24)+9*TR;
33. OUTPUT;
34. END;
35. DROP G1-G10 H1-H10 N1 N2 I J K T1 T2 T3 T4 T5;
36. DROP M1 M2 DETS S12 S12S RO1 TR S1 S2;
37. PROC PRINT;
```

Program 7:

```
 1. // JOB
 2. // EXEC SAS
 3. OPTION LINESIZE=74;
 4. DATA BIVNOR;
 5. ARRAY X1(10) G1-G10;
 6. ARRAY X2(10) H1-H10;
 7. DO I=1 TO 20;
 8. T1=0; T2=0;
 9. IF I<6 THEN DO;
10.               SIG1=8;
11.               SIG2=4;
12.               END;
13. IF I>5 AND I<11 THEN DO;
14.               SIG1=8*0.4*(I-5);
15.               SIG2=4;
16.               END;
17. IF I>10 AND I<16 THEN DO;
18.               SIG1=8;
19.               SIG2=4*0.2*(I-10);
20.               END;
21. IF I>15 THEN DO;
22.               SIG1=8*0.4*(I-15);
23.               SIG2=4*0.2*(I-15);
24.               END;
25. DO J=1 TO 10;
26. N1=RANNOR(3680917);
```

```
27. N2=RANNOR(1638052);

28. X1(J)=N1*SQRT(SIG1)+30;

29. X2(J)=0.5*N1*SQRT(SIG2)+0.5*SQRT(3*SIG2)*N2+15;

30. T1=T1+X1(J);

31. T2=T2+X2(J);

32. END;

33. M1=T1/10; M2=T2/10;

34. T3=0; T4=0; T5=0;

35. DO K=1 TO 10;

36. T3=T3+(X1(K)-M1)**2;

37. T4=T4+(X2(K)-M2)**2;

38. T5=T5+(X1(K)-M1)*(X2(K)-M2);

39. END;

40. S1=T3/9; S2=T4/9;

41. SS1=SQRT(S1); SS2=SQRT(S2);

42. SD1=S1; SD2=S2;

43. S12=T5/9; S12S=S12**2;

44. DETS=S1*S2-S12S;

45. SDS=SQRT(DETS);

46. RO1=1-0.5**2;

47. TR=(S1/8+S2/4-S12/(4*SQRT(2)))/RO1;

48. W=-18-9*LOG(DETS)+9*LOG(24)+9*TR;

49. OUTPUT;

50. END;

51. DROP G1-G10 H1-H10 N1 N2 I J K T1 T2 T3 T4 T5;

52. DROP M1 M2 DETS S12 S12S RO1 TR SIG1 SIG2 S1 S2;

53. PROC PRINT;
```

Program 8:

```
1. // JOB

2. // EXEC SAS

3. OPTION LINESIZE=72;

4. DATA BIVNOR;

5. ARRAY X1(10) G1-G10;

6. ARRAY X2(10) H1-H10;
```

```
 7. DO I=1 TO 20;
 8. T1=0;  T2=0;
 9. MU1=29.511;  MU2=14.906;
10. SIG1=11.398;  SIG2=5.934;
11. ROH=4.227/(SQRT(11.398*5.934));
12. DO J=1 TO 10;
13. N1=RANNOR(6312510);
14. N2=RANNOR(3021471);
15. X1(J)=SQRT(SIG1)*N1+MU1;
16. X2(J)=ROH*SQRT(SIG2)*N1+N2*SQRT(SIG2)*SQRT(1-ROH**2)+MU2;
17. T1=T1+X1(J);
18. T2=T2+X2(J);
19. END;
20. M1=T1/10;  M2=T2/10;
21. T3=0;  T4=0;  T5=0;
22. DO K=1 TO 10;
23. T3=T3+(X1(K)-M1)**2;
24. T4=T4+(X2(K)-M2)**2;
25. T5=T5+(X1(K)-M1)*(X2(K)-M2);
26. END;
27. S1=T3/9;  S2=T4/9;
28. SS1=SQRT(S1);  SS2=SQRT(S2);
29. SD1=S1;  SD2=S2;
30. S12=T5/9;  S12S=S12**2;
31. DETS=S1*S2-S12S;
32. SDS=SQRT(DETS);
33. RO1=1-0.5**2;
34. TR=(S1/8+S2/4-S12/(4*SQRT(2)))/RO1;
35. W=-18-9*LOG(DETS)+9*LOG(24)+9*TR;
36. OUTPUT;
37. END;
38. DROP G1-G10 H1-H10 N1 N2 I J K T1 T2 T3 T4 T5 S1 S2;
39. DROP M1 M2 DETS S12 S12S RO1 TR MU1 MU2 SIG1 SIG2 ROH;
40. PROC PRINT;
```

# APPENDIX B: Programs of Chapter 2

Program 1:

```
 1. // JOB 'T=2M'
 2. // EXEC WATFIV
 3. //FT05F001 DD *
 4. $JOB
 5.         DIMENSION RR(10)
 6.         DOUBLE PRECISION DSEED
 7.         DSEED=31785.D0
 8. C
 9.         N=10
10.         N1=1000
11.         FN=FLOAT(N)
12.         SIG=1.
13.         U0=0.
14. C
15.         DO 3 LOOP=1,4
16.         FL=FLOAT(LOOP)
17.         H=3.0-FL*0.5
18.         FK=0.25*FL
19. C
20.         DO 45 ID=1,5
21.         IF (ID.EQ.1) THEN DO
22.                         U=U0
23.                         GO TO 47
24.                         END IF
25.         IF (ID.EQ.2) THEN DO
26.                         U=0.5
27.                         GO TO 47
28.                         END IF
29.         DEL=FLOAT(ID)-2.
30.         U1=DEL*SIG
31.         U=U1
```

```
32.     47   WRITE (6,12) U,FK,H
33.     12   FORMAT(2X,'U =',F6.2,' K =',F6.2,' H =',F6.2)
34. C
35.          SHJ1=0.
36.          SHJ2=0.
37.          SLJ1=0.
38.          SLJ2=0.
39.          DO 60 L=1,N1
40.          SH=H/2.
41.          SL=H/2.
42. C
43. C Generating 1000 samples each of size N from a normal
44. C distribution with mean U and variance SIG.
45. C
46.          DO 10 I=1,1000
47.          T1=0.
48.          CALL GGNPM(DSEED,N,RR)
49.          DO 15 K=1,N
50.          X1=RR(1)*SQRT(SIG)+U
51.          T1=T1+X1
52.     15   CONTINUE
53.          XB=T1/FN
54.          YB=XB-U0
55.          IF (YB.GT.FK) GO TO 20
56.     10   CONTINUE
57. C
58.     20   DO 40 J=I,1000
59.          T1=0.
60.          CALL GGNPM(DSEED,N,RR)
61.          DO 35 K=1,N
62.          X1=RR(1)*SQRT(SIG)+U
63.          T1=T1+X1
64.     35   CONTINUE
65.          XB=T1/FN
66.          ASH = XB - FK + SH
67.          ASL = - XB -FK + SL
```

```
68.          SH = AMAX1(0.,ASH)
69.          SL = AMAX1(0.,ASL)
70.          IF (SH.GT.H)  GO TO 30
71.          IF (SL.GT. H) GO TO 31
72.      40  CONTINUE
73. C
74.      30  FHJ=FLOAT(J)
75.          SHJ1=SHJ1+FHJ
76.          SHJ2=SHJ2+FHJ**2
77.          GO TO 60
78.      31  FLJ=FLOAT(J)
79.          SLJ1=SLJ1+FLJ
80.          SLJ2=SLJ2+FLJ**2
81.      60  CONTINUE
82.          EHJ=SHJ1/FLOAT(N1)
83.          ELJ=SLJ1/FLOAT(N1)
84.          ARL=EHJ+ELJ
85.          WRITE(6,70) EHJ,ELJ,ARL
86.      70  FORMAT(2X,'EHJ =',F12.4,2X,'ELJ = ',F12.4,2X,
                                      'ARL = ',F12.4)
87.      45  CONTINUE
88.          PRINT,' '
89.       3  CONTINUE
90.          STOP
91.          END
92. $ENTRY
93. /*
```

Program 2:

```
1. // JOB
2. // EXEC WATFIV
3. //FT05F001 DD *
4. $JOB
5.          DIMENSION F(10),P(18),R(10,10),WR(4),CARL(15,10),G(15)
6.          REAL EI(10,10),EIR(4,4),EIRI(4,4),EMU(8,10)
```

```
  7.          INTEGER H,Y
  8. C
  9.          READ, TH, K, H, N
 10. C
 11.          DO 11 I=1,H
 12.          L=K+I-H
 13.          IF (L.EQ.0) THEN DO
 14.                            F(I)=EXP(-TH)
 15.                            GO TO 11
 16.                            END IF
 17.          CALL POPROB(TH, L, F1)
 18.          F(I) = F1
 19.     11   CONTINUE
 20. C
 21.          Y=2*H-2
 22.          DO 22 I=1,Y
 23.          L1= K+I+1-H
 24.          CALL POPROB(TH, L1, P1)
 25.          L2 = K+I-H
 26.          IF (L2.EQ.0) THEN DO
 27.                            P2=EXP(-TH)
 28.                            P(I)=P1-P2
 29.                            GO TO 22
 30.                            END IF
 31.          CALL POPROB(TH, L2, P2)
 32.          P(I) = P1 - P2
 33.     22   CONTINUE
 34. C
 35.          DO 23 I=1,H
 36.          KK1=H+1-I
 37.          R(I,1)=F(KK1)
 38.             DO 25 J=2,H
 39.             KK2=H+J-I-1
 40.             R(I,J)=P(KK2)
 41.     25      CONTINUE
 42.     23   CONTINUE
```

```
43. C

44.         PRINT, ' '

45.         PRINT, 'The Matrix R'

46.         WRITE (6,30) ((R(I,J),J=1,H),I=1,H)

47.    30   FORMAT (/, 4F10.4)

48.         PRINT, ' '

49. C

50.         DO 31 I=1,H

51.            DO 32 J=1,H

52.            IF (I.EQ.J) THEN DO

53.                               EI(I,J)=1

54.                               GO TO 32

55.                               END IF

56.            EI(I,J)=0

57.    32      CONTINUE

58.    31   CONTINUE

59. C

60.         DO 40 I=1,H

61.         S1=0.

62.               DO 42 J=1,H

63.               EIR(I,J)=EI(I,J)-R(I,J)

64.               S1=S1+EIR(I,J)

65.    42         CONTINUE

66.         CARL(1,I)=S1

67.    40   CONTINUE

68. C

69.         DO 50 K=2,N

70.            DO 51 I=1,H

71.            S2=0.

72.                  DO 52 J=1,H

73.                  S2=S2+R(I,J)*CARL(K-1,J)

74.    52            CONTINUE

75.            CARL(K,I)=S2

76.    51      CONTINUE

77.    50   CONTINUE

78. C
```

```
79.            CALL LINV1F(EIR,4,4,EIRI,1,WR,IER)
80. C
81.            DO 54 I=1,H
82.            S3=0.
83.                DO 56 J=1,H
84.                S3=S3+EIRI(I,J)
85.        56      CONTINUE
86.            EMU(1,I)=S3
87.        54  CONTINUE
88. C
89.            DO 60 K=2,4
90.                DO 62 I=1,H
91.                S4=0.
92.                    DO 58 J=1,H
93.                    S4=S4+(EIRI(I,J)-EI(I,J))*K*EMU(K-1,J)
94.        58          CONTINUE
95.                EMU(K,I)=S4
96.        62      CONTINUE
97.        60  CONTINUE
98. C
99.            DO 59 I=1,H
100.           EX1=EMU(1,I)
101.           EX2=EMU(2,I)+EX1
102.           EMU(2,I)=EX2-EX1**2
103.           EX3=EMU(3,I)+3.*EX2-2.*EX1
104.           EMU(3,I)=EX3-3.*EX2*EX1+2.*EX1**3
105.           EX4=EMU(4,I)+6.*EX3-11.*EX2+6.*EX1
106.           EMU(4,I)=EX4-4.*EX3*EX1+6.*EX2*EX1**2-3.*EX1**4
107.           EMU(5,I)=SQRT(EMU(2,I))
108.           EMU(6,I)=EMU(5,I)/EMU(1,I)
109.           EMU(7,I)=EMU(3,I)/EMU(5,I)**3
110.           EMU(8,I)=(EMU(4,I)/EMU(2,I)**2)-3.
111.       59  CONTINUE
112. C
113.           WRITE (6,57) ((EMU(K,I),K=1,8),I=1,H)
114.       57  FORMAT (/,8F9.2)
```

```
115.          PRINT,' '
116. C
117.          DO 65 I=1,H
118.          PRINT, ' '
119.          PRINT, 'The Actual ARL values'
120.          WRITE (6,63) (CARL(K,I),K=1,N)
121.     63   FORMAT(/,6F10.4)
122.          PRINT, ' '
123.          ARL=EMU(1,I)
124.          AP=1/ARL
125.          Q=1-AP
126.          PRINT, ' '
127.          PRINT, 'ARL',I,'=', ARL
128.          PRINT, ' '
129.          PRINT, 'Geometric probability = ', AP
130.          G(1)=AP
131.             DO 67 J=2,N
132.             G(J)=(Q**(J-1))*AP
133.     67      CONTINUE
134.          PRINT, ' '
135.          WRITE (6,70) (G(J), J=1,N)
136.     70   FORMAT(/, 6F10.4)
137.          PRINT, ' '
138.     65   CONTINUE
139.          STOP
140.          END
141. C
142.          SUBROUTINE POPROB(PL, X, PROB)
143.          INTEGER X
144.          S=EXP(-PL)
145.          DO 111 I=1,X
146.          CALL FACT(I, FI)
147.          PR = EXP(-PL) * (PL**I)/FI
148.          S=S+PR
149.    111   CONTINUE
150.          PROB=S
```

```
151.          RETURN
152.          END
153. C
154.          SUBROUTINE FACT(L, F)
155.          K1=1
156.          IF (L.EQ.0) THEN DO
157.                        F=1
158.                        RETURN
159.                        END IF
160.          DO 122 J=1,L
161.          K1=K1*J
162.    122   CONTINUE
163.          F=K1
164.          RETURN
165.          END
166. C
167. $ENTRY
168.    4.1, 3, 4, 12
169. /*
```

Program 3:

```
1. // JOB
2. // EXEC WATFIV
3. //FT05F001 DD *
4. $JOB
5.          DIMENSION F(10),P(10),R(10,10),WR(5),CARL(15,10),G(15)
6.          REAL EI(10,10),EIR(5,5),EIRI(5,5),EMU(8,10)
7.          INTEGER H,T,T1,T2
8. C
9.          READ, T, H, N
10. C
11.          T1=2*T-1
12.          FH=FLOAT(H)
13.          FT1=FLOAT(T1)
14.          W=2.*FH/FT1
```

```
15. C
16.         DO 5 I=1,T1
17.         FL=FLOAT(I-T)
18.         Y=(FL+1.)*W-0.5*W
19.         Z=Y-1.
20.         CALL MDNOR(Z,PR)
21.         F(I)=PR
22.      5  CONTINUE
23. C
24.         T2=T1-1
25.         DO 7 I=1,T2
26.         P(I)=F(I+1)-F(I)
27.      7  CONTINUE
28. C
29.         DO 10 I=1,T
30.         K1=T+1-I
31.         R(I,1)=F(K1)
32.            DO 9 J=2,T
33.            K2=T+J-I-1
34.            R(I,J)=P(K2)
35.      9     CONTINUE
36.     10  CONTINUE
37. C
38.         PRINT, ' '
39.         PRINT, 'The Matrix R'
40.         WRITE (6,30) ((R(I,J),J=1,T),I=1,T)
41.     30  FORMAT (/, 5F10.6)
42.         PRINT, ' '
43. C
44.         DO 31 I=1,T
45.            DO 32 J=1,T
46.            IF (I.EQ.J) THEN DO
47.                                 EI(I,J)=1
48.                                 GO TO 32
49.                                 END IF
50.            EI(I,J)=0
```

```
51.    32    CONTINUE
52.    31    CONTINUE
53. C
54.          DO 40 I=1,T
55.          S1=0.
56.             DO 42 J=1,T
57.             EIR(I,J)=EI(I,J)-R(I,J)
58.             S1=S1+EIR(I,J)
59.    42       CONTINUE
60.          CARL(1,I)=S1
61.    40    CONTINUE
62. C
63.          DO 50 K=2,N
64.            DO 51 I=1,T
65.            S2=0.
66.               DO 52 J=1,T
67.               S2=S2+R(I,J)*CARL(K-1,J)
68.    52          CONTINUE
69.            CARL(K,I)=S2
70.    51       CONTINUE
71.    50    CONTINUE
72. C
73.          CALL LINV1F(EIR,5,5,EIRI,1,WR,IER)
74. C
75.          DO 54 I=1,T
76.          S3=0.
77.            DO 56 J=1,T
78.            S3=S3+EIRI(I,J)
79.    56       CONTINUE
80.          EMU(1,I)=S3
81.    54    CONTINUE
82. C
83.          DO 60 K=2,4
84.            DO 62 I=1,T
85.            S4=0.
86.               DO 58 J=1,T
```

```
 87.              FK=FLOAT(K)
 88.              S4=S4+(EIRI(I,J)-EI(I,J))*FK*EMU(K-1,J)
 89.      58      CONTINUE
 90.            EMU(K,I)=S4
 91.      62    CONTINUE
 92.      60  CONTINUE
 93. C
 94.          DO 59 I=1,T
 95.          EX1=EMU(1,I)
 96.          EX2=EMU(2,I)+EX1
 97.          EMU(2,I)=EX2-EX1**2.
 98.          EX3=EMU(3,I)+3.*EX2-2.*EX1
 99.          EMU(3,I)=EX3-3.*EX2*EX1+2.*EX1**3.
100.          EX4=EMU(4,I)+6.*EX3-11.*EX2+6.*EX1
101.          EMU(4,I)=EX4-4.*EX3*EX1+6.*EX2*EX1**2.-3.*EX1**4.
102.          EMU(5,I)=SQRT(EMU(2,I))
103.          EMU(6,I)=EMU(5,I)/EMU(1,I)
104.          EMU(7,I)=EMU(3,I)/EMU(5,I)**3.
105.          EMU(8,I)=(EMU(4,I)/EMU(2,I)**2.)-3.
106.      59  CONTINUE
107. C
108.          WRITE (6,57) ((EMU(K,I),K=1,8),I=1,T)
109.      57  FORMAT (/,8F9.2)
110.          PRINT,' '
111. C
112.          DO 65 I=1,T
113.          PRINT, ' '
114.          PRINT, 'The Actual ARL values'
115.          WRITE (6,63) (CARL(K,I),K=1,N)
116.      63  FORMAT(/,5F10.6)
117.          PRINT, ' '
118.          ARL=EMU(1,I)
119.          AP=1./ARL
120.          Q=1.-AP
121.          PRINT, ' '
122.          PRINT, 'ARL',I,'=', ARL
```

```
123.          PRINT, ' '
124.          PRINT, 'Geometric probability = ', AP
125.          G(1)=AP
126.              DO 67 J=2,N
127.              FJ=FLOAT(J-1)
128.              G(J)=(Q**FJ)*AP
129.     67       CONTINUE
130.          PRINT, ' '
131.          WRITE (6,70) (G(J), J=1,N)
132.     70   FORMAT(/, 5F10.6)
133.          PRINT, ' '
134.     65   CONTINUE
135.          STOP
136.          END
137. C
138. $ENTRY
139.  5, 3, 15
140. /*
```

Program 4:

```
1. // JOB
2. // EXEC WATFIV
3. //FT05F001 DD *
4. $JOB
5.          DIMENSION F(10),P(10),R(10,10),WR(4),CARL(15,10),G(15)
6.          REAL EI(10,10),EIR(4,4),EIRI(4,4),EMU(4,10)
7.          INTEGER H,T,T1,T2,T3
8. C
9.          READ, T, IS, N
10. C
11.          T1=2*T-1
12.          T2=2*T-3
13.          FT=FLOAT(T)
14.          FIS=FLOAT(IS)
15.          FT1=FLOAT(T1)
```

```
16.          FT2=FLOAT(T2)
17.          FH=FT1*FIS/(4*(FT-1))
18.          W=2.*FH/FT1
19.          FK=FT2*FH/FT1
20.          PRINT,'K = ',FK,'    H = ',FH
21. C
22.          DO 5 I=1,T1
23.          FL=FLOAT(I-T)
24.          Y=FK+FL*W+0.5*W
25.          F(I)=1.-EXP(-Y)
26.       5  CONTINUE
27. C
28.          T3=T1-1
29.          DO 7 I=1,T3
30.          P(I)=F(I+1)-F(I)
31.       7  CONTINUE
32. C
33.          DO 10 I=1,T
34.          K1=T+1-I
35.          R(I,1)=F(K1)
36.             DO 9 J=2,T
37.             K2=T+J-I-1
38.             R(I,J)=P(K2)
39.       9     CONTINUE
40.      10  CONTINUE
41. C
42.          PRINT, ' '
43.          PRINT, 'The Matrix R'
44.          WRITE (6,30) ((R(I,J),I=1,T),J=1,T)
45.      30 FORMAT (/, 4F10.5)
46.          PRINT, ' '
47. C
48.          DO 31 I=1,T
49.             DO 32 J=1,T
50.             IF (I.EQ.J) THEN DO
51.                                 EI(I,J)=1.
```

```
52.                                    GO TO 32
53.                                    END IF
54.           EI(I,J)=0.
55.     32    CONTINUE
56.     31  CONTINUE
57. C
58.         DO 40 I=1,T
59.         S1=0.
60.             DO 42 J=1,T
61.             EIR(I,J)=EI(I,J)-R(I,J)
62.             S1=S1+EIR(I,J)
63.     42      CONTINUE
64.         CARL(1,I)=S1
65.     40  CONTINUE
66. C
67.         DO 50 K=2,N
68.           DO 51 I=1,T
69.           S2=0.
70.               DO 52 J=1,T
71.               S2=S2+R(I,J)*CARL(K-1,J)
72.     52        CONTINUE
73.           CARL(K,I)=S2
74.     51      CONTINUE
75.     50  CONTINUE
76. C
77.         CALL LINV1F(EIR,4,4,EIRI,1,WR,IER)
78. C
79.         DO 54 I=1,T
80.         S3=0.
81.           DO 56 J=1,T
82.           S3=S3+EIRI(I,J)
83.     56      CONTINUE
84.         EMU(1,I)=S3
85.     54  CONTINUE
86. C
87.         DO 60 K=2,4
```

```
88.              DO 62 I=1,T
89.                S4=0.
90.                  DO 58 J=1,T
91.                  FK=FLOAT(K)
92.                  S4=S4+(EIRI(I,J)-EI(I,J))*FK*EMU(K-1,J)
93.     58           CONTINUE
94.                EMU(K,I)=S4
95.     62        CONTINUE
96.     60    CONTINUE
97. C
98.           DO 59 I=1,T
99.           EX1=EMU(1,I)
100.          EX2=EMU(2,I)+EX1
101.          EMU(2,I)=EX2-EX1**2.
102.          EX3=EMU(3,I)+3.*EX2-2.*EX1
103.          EMU(3,I)=EX3-3.*EX2*EX1+2.*EX1**3.
104.          EX4=EMU(4,I)+6.*EX3-11.*EX2+6.*EX1
105.          EMU(4,I)=EX4-4.*EX3*EX1+6.*EX2*EX1**2.-3.*EX1**4.
106.    59    CONTINUE
107. C
108.          WRITE (6,57) ((EMU(K,I),K=1,4),I=1,T)
109.    57    FORMAT (/,4F15.2)
110.          PRINT,' '
111. C
112.          DO 65 I=1,T
113.          PRINT, ' '
114.          PRINT, 'The true distribution of run length'
115.          WRITE (6,63) (CARL(K,I),K=1,N)
116.    63    FORMAT(/,6F10.4)
117.          PRINT, ' '
118.          ARL=EMU(1,I)
119.          AP=1./ARL
120.          Q=1.-AP
121.          PRINT, ' '
122.          PRINT, 'ARL',I,'=', ARL
123.          PRINT, ' '
```

```
124.          PRINT, 'Geometric probability = ', AP
125.          G(1)=AP
126.            DO 67 J=2,N
127.            FJ=FLOAT(J-1)
128.            G(J)=(Q**FJ)*AP
129.    67      CONTINUE
130.          PRINT, ' '
131.          WRITE (6,70) (G(J), J=1,N)
132.    70  FORMAT(/, 6F10.4)
133.          PRINT, ' '
134.    65  CONTINUE
135.          STOP
136.          END
137. C
138. $ENTRY
139.   4, 5, 12
140. /*
```

# APPENDIX C: Programs of Chapter 3

Program 1:

```
1.  // JOB
2.  // EXEC WATFIV
3.  //FT05F001 DD *
4.  $JOB
5.          DIMENSION X1(50),X2(50),RR(2)
6.          DIMENSION UC(9),CL(9),CO1(9)
7.          DOUBLE PRECISION DSEED
8.          DSEED=31785.D0
9.          READ,N,TU,TL
10. C
11.          PRINT, '    ROH  ','    T   ','    R',' '
12. C
13.          DO 5 J=1,9
14.          ROH=0.1*FLOAT(J)
15. C
16.          CO1(J)=0.
17. C
18. C Generating 1000 samples each of size N from a bivariate
19. C normal distribution with parameters U1,U2,SIG1,SIG2,ROH.
20. C
21.          DO 10 I=1,1000
22.          U1=30.
23.          U2=15.
24.          SIG1=8.
25.          SIG2=4.
26.          T1=0.
27.          T2=0.
28. C
29.          DO 15 K=1,N
30.          CALL GGNPM(DSEED, 2,RR)
```

```
31.          X1(K)=RR(1)*SQRT(SIG1)+U1
32.          X2(K)=ROH*RR(1)*SQRT(SIG2)
                              +SQRT(1.- ROH**2)*SQRT(SIG2)*RR(2)+U2
33.          T1=T1+X1(K)
34.          T2=T2+X2(K)
35.    15    CONTINUE
36.          FN=FLOAT(N)
37.          XB1=T1/FN
38.          XB2=T2/FN
39.          T3=0.
40.          T4=0.
41.          T5=0.
42. C
43.          DO 20 K=1,N
44.          T3=T3+(X1(K)-XB1)**2
45.          T4=T4+(X2(K)-XB2)**2
46.          T5=T5+(X1(K)-XB1)*(X2(K)-XB2)
47.    20    CONTINUE
48. C
49.          S1=T3/(FN-1)
50.          S2=T4/(FN-1)
51.          S1S=SQRT(S1)
52.          S2S=SQRT(S2)
53.          S12=T5/(FN-1)
54.          R=S12/(S1S*S2S)
55.          T=R*SQRT(FN-2.)/SQRT(1.-R**2)
56. C
57. C   Checking the statistic T, with the control limits.
58. C
59.          IF((T.GT.TU) .OR.(T.LT.TL)) CO1(J)=CO1(J)+1.0
60.    10    CONTINUE
61. C
62.          C1=CO1(J)/10.
63.          WRITE(6,25) ROH,C1
64.    25    FORMAT(2F8.1)
65.     5    CONTINUE
```

```
66.          STOP
67.          END
68. $ENTRY
69.  10, 2.3060, -2.3060,
70. /*
```

Program 2:

```
 1. // JOB
 2. // EXEC WATFIV
 3. //FT05F001 DD *
 4. $JOB
 5.          DIMENSION X1(50),X2(50),RR(2)
 6.          DIMENSION UC(9),CL(9),CO1(9)
 7.          DOUBLE PRECISION DSEED
 8.          DSEED=31785.D0
 9.          ROH=0.5
10. C
11.          PRINT, '     N  ','    T   ','  '
12. C
13.          DO 5 J=1,5
14.          READ,N,TU
15.          TL=-TU
16. C
17.          CO1(J)=0.
18. C
19. C    Generating 1000 samples of size N from a bivariate normal
20. C    distribution with parameters U1, U2, SIG1, SIG2 and ROH.
21. C
22.          DO 10 I=1,1000
23.          U1=30.
24.          U2=15.
25.          SIG1=8.
26.          SIG2=4.
27.          T1=0.
28.          T2=0.
```

```
29. C

30.          DO 15 K=1,N

31.          CALL GGNPM(DSEED, 2,RR)

32.          X1(K)=RR(1)*SQRT(SIG1)+U1

33.          X2(K)=ROH*RR(1)*SQRT(SIG2)
                     +SQRT(1.- ROH**2)*SQRT(SIG2)*RR(2)+U2

34.          T1=T1+X1(K)

35.          T2=T2+X2(K)

36.    15    CONTINUE

37.          FN=FLOAT(N)

38.          XB1=T1/FN

39.          XB2=T2/FN

40.          T3=0.

41.          T4=0.

42.          T5=0.

43. C

44.          DO 20 K=1,N

45.          T3=T3+(X1(K)-XB1)**2

46.          T4=T4+(X2(K)-XB2)**2

47.          T5=T5+(X1(K)-XB1)*(X2(K)-XB2)

48.    20    CONTINUE

49. C

50.          S1=T3/(FN-1)

51.          S2=T4/(FN-1)

52.          S1S=SQRT(S1)

53.          S2S=SQRT(S2)

54.          S12=T5/(FN-1)

55.          R=S12/(S1S*S2S)

56.          T=R*SQRT(FN-2.)/SQRT(1.-R**2)

57. C

58. C   Checking the control limits.

59. C

60.          IF((T.GT.TU) .OR.(T.LT.TL)) CO1(J)=CO1(J)+1.0

61.    10    CONTINUE

62. C

63.          C1=CO1(J)/10.
```

```
64.         WRITE(6,25) FN,C1
65.    25   FORMAT(2F8.1)
66.     5   CONTINUE
67.         STOP
68.         END
69. $ENTRY
70.     5, 3.1824,
71.    15, 2.1604,
72.    20, 2.1009,
73.    30, 2.0484,
74.    40, 1.96,
75. /*
```

Program 3:

```
1.  // JOB 'T=6M'
2.  // EXEC WATFIV
3.  //FT05F001 DD *
4.  $JOB
5.         DIMENSION X1(50),X2(50),RR(2)
6.         DIMENSION UC1(9),CL1(9),UC2(9),CL2(9),CO1(9),CO2(9)
7.         DOUBLE PRECISION DSEED
8.         DSEED=31785.D0
9.         ROH=0.5
10. C
11.        DO 3 L=1,5
12.        READ(5,100) N
13.   100  FORMAT(I3)
14. C
15.        PRINT,'  ROH  ','    C1','  '
16. C
17.        DO 5 J=1,9
18.        RO=0.1*FLOAT(J)
19. C
20.        READ(5,101) UC1(J),CL1(J)
21.   101  FORMAT(2F7.4)
```

```
22.          CO1(J)=0.
23. C
24. C  Generating 1000 samples of size N from a bivariate normal
25. C    distribution with parameters U1, U2, SIG1, SIG2 and ROH.
26. C
27.          DO 10 I=1,1000
28.          U1=30.
29.          U2=15.
30.          SIG1=8.
31.          SIG2=4.
32.          T1=0.
33.          T2=0.
34. C
35.          DO 15 K=1,N
36.          CALL GGNPM(DSEED, 2,RR)
37.          X1(K)=RR(1)*SQRT(SIG1)+U1
38.          X2(K)=ROH*RR(1)*SQRT(SIG2)
                     +SQRT(1.-ROH**2)*SQRT(SIG2)*RR(2)+U2
39.          T1=T1+X1(K)
40.          T2=T2+X2(K)
41.     15   CONTINUE
42.          FN=FLOAT(N)
43.          XB1=T1/FN
44.          XB2=T2/FN
45.          T3=0.
46.          T4=0.
47.          T5=0.
48. C
49.          DO 20 K=1,N
50.          T3=T3+(X1(K)-XB1)**2
51.          T4=T4+(X2(K)-XB2)**2
52.          T5=T5+(X1(K)-XB1)*(X2(K)-XB2)
53.     20   CONTINUE
54. C
55.          S1=T3/(FN-1.)
56.          S2=T4/(FN-1.)
```

```
57.           S1S=SQRT(S1)
58.           S2S=SQRT(S2)
59.           S12=T5/(FN-1.)
60.           R=S12/(S1S*S2S)
61.           T=R*SQRT(FN-2.)/SQRT(1.-R**2)
62. C
63. C    Checking the statistic R with the control limits.
64. C
65.           IF(( R.GT.UC1(J)).OR.(R.LT.CL1(J)))   CO1(J)=CO1(J)+1.0
66.
67.      10  CONTINUE
68. C
69.           C1=CO1(J)/10.
70.           WRITE(6,25) RO,C1
71.      25  FORMAT(2F8.1)
72.       5  CONTINUE
73.       3  CONTINUE
74.           STOP
75.           END
76. $ENTRY
77.  5
78.  0.9026-0.8482
79.  0.9223-0.8099
80.  0.9386-0.7606
81.  0.9521-0.6955
82.  0.9635-0.6072
83.  0.9732-0.4839
84.  0.9814-0.3056
85.  0.9885-0.0384
86.  0.9947 0.3733
87.  10
88.  0.6917-0.5633
89.  0.7440-0.4839
90.  0.7901-0.3916
91.  0.8310-0.2835
92.  0.8673-0.1556
```

```
 93.  0.8997-0.0034
 94.  0.9288 0.1793
 95.  0.9549 0.4005
 96.  0.9786 0.6700
 97.  15
 98.  0.5865-0.4331
 99.  0.6517-0.3427
100.  0.7107-0.2410
101.  0.7641-0.1263
102.  0.8127 0.0037
103.  0.8570 0.1518
104.  0.8974 0.3211
105.  0.9345 0.5157
106.  0.9686 0.7401
107.  25
108.  0.4788-0.3063
109.  0.5553-0.2086
110.  0.6259-0.1021
111.  0.6915 0.0142
112.  0.7523 0.1416
113.  0.8089 0.2814
114.  0.8617 0.4352
115.  0.9109 0.6048
116.  0.9569 0.7923
117.  50
118.  0.3693-0.1828
119.  0.4551-0.0812
120.  0.5362 0.0265
121.  0.6131 0.1409
122.  0.6861 0.2624
123.  0.7553 0.3916
124.  0.8211 0.5293
125.  0.8837 0.6760
126.  0.9433 0.8327
127.  /*
```

Program 4:

```
 1. // JOB 'T=6M'
 2. // EXEC WATFIV
 3. //FT05F001 DD *
 4. $JOB
 5.          DOUBLE PRECISION DSEED
 6.          COMMON DSEED
 7.          DSEED=31785.D0
 8.          N=10
 9.          U0=0.5*ALOG((1.+0.1)/(1.-0.1))
10.          N1=1000
11.          FN=FLOAT(N)
12.          SIG=1./SQRT(FN-3.)
13.          WRITE (6,12) U0,SIG
14.     12   FORMAT(2X,'U0 =',F6.2,'   SIG =',F6.2)
15. C
16.          DO 3 LOOP=1,3
17.          FL1=FLOAT(LOOP)
18.          FK=0.5*FL1
19.          H=3-0.75*FL1
20.          WRITE(6,9) FK,H
21.      9   FORMAT(2X,' K =',F6.2,' H =',F6.2)
22. C
23.          DO 5 LR=1,3
24.          FLR=FLOAT(LR)
25.          ROH=0.1+(FLR-1.)*0.4
26.          U=0.5*ALOG((1.+ROH)/(1.-ROH))
27.          WRITE(6,11) ROH,U
28.     11   FORMAT(2X,'ROH = ',F5.1,' U =',F10.3)
29. C
30.          SHJ1=0.
31.          SLJ1=0.
32.          DO 60 L=1,N1
33.          SH=H/2.
34.          SL=H/2.
```

```
35. C
36.            DO 10 I=1,1000
37.            CALL ZCAL(N,ROH,Z)
38.            Y=(Z-U0)/SIG
39.            IF (Y.GT.FK) GO TO 23
40.     10     CONTINUE
41.     23     DO 42 J=I,1000
42.            ASH = Y - FK + SH
43.            ASL = - Y -FK + SL
44.            CALL ZCAL(N,ROH,Z)
45.            Y=(Z-U0)/SIG
46.            SH = AMAX1(0.,ASH)
47.            SL = AMAX1(0.,ASL)
48.            IF (SH.GT.H)  GO TO 30
49.            IF (SL.GT. H) GO TO 31
50.     42     CONTINUE
51. C
52.     30     FHJ=FLOAT(J)
53.            SHJ1=SHJ1+FHJ
54.            GO TO 60
55.     31     FLJ=FLOAT(J)
56.            SLJ1=SLJ1+FLJ
57.     60     CONTINUE
58.            EHJ=SHJ1/FLOAT(N1)
59.            ELJ=SLJ1/FLOAT(N1)
60.            ARL=EHJ+ELJ
61.            WRITE(6,70) EHJ,ELJ,ARL
62.     70     FORMAT(2X,'EHJ =',F15.6,2X,'ELJ = ',F15.6,2X,'ARL =
    ',F15.6)
63.            PRINT,' '
64. C
65.      5     CONTINUE
66.            PRINT,' '
67.      3     CONTINUE
68.            STOP
69.            END
```

```
70.  C

71.  C    Subroutine to generate samples of size N from a
                                    bivariate normal
72.  C    distribution and to calculate the Fisher's Z
                                    statistic.
73.  C

74.        SUBROUTINE ZCAL(N,ROH,Z1)
75.        DIMENSION X1(50),X2(50),RR(2)
76.        DOUBLE PRECISION DSEED
77.        COMMON DSEED
78.        U1=30.
79.        U2=15.
80.        SIG1=8.
81.        SIG2=4.
82.        T1=0.
83.        T2=0.
84.  C

85.        DO 15 K=1,N
86.        CALL GGNPM(DSEED, 2,RR)
87.        X1(K)=RR(1)*SQRT(SIG1)+U1
88.        X2(K)=ROH*RR(1)*SQRT(SIG2)
                  +SQRT(1.-ROH**2)*SQRT(SIG2)*RR(2)+U2
89.        T1=T1+X1(K)
90.        T2=T2+X2(K)
91.    15  CONTINUE
92.        FN=FLOAT(N)
93.        XB1=T1/FN
94.        XB2=T2/FN
95.        T3=0.
96.        T4=0.
97.        T5=0.
98.  C

99.        DO 20 K=1,N
100.       T3=T3+(X1(K)-XB1)**2
101.       T4=T4+(X2(K)-XB2)**2
102.       T5=T5+(X1(K)-XB1)*(X2(K)-XB2)
```

```
103.    20  CONTINUE
104. C
105.        S1=T3/(FN-1.)
106.        S2=T4/(FN-1.)
107.        S1S=SQRT(S1)
108.        S2S=SQRT(S2)
109.        S12=T5/(FN-1.)
110.        R=S12/(S1S*S2S)
111.        Z1=0.5*ALOG((1.+R)/(1.-R))
112.        RETURN
113.        END
114. C
115. $ENTRY
116. /*
```

Program 5:

```
 1. // JOB 'T=8M'
 2. // EXEC WATFIV
 3. //FT05F001 DD *
 4. $JOB
 5.        DIMENSION NN(100)
 6.        DOUBLE PRECISION DSEED
 7.        COMMON DSEED
 8.        DSEED=23512.D0
 9. C
10.        RH0=0.1
11.        RH1=0.5
12.        N2=300
13.        N3=100
14.        WRITE(6,2) RH0,RH1
15.     2  FORMAT(2X,'  RH0=',F5.1,'  RH1=',F5.1)
16. C
17.        DO 7 LA=1,2
18.        FLA=FLOAT(LA)
19.        ALFA=0.05*FLA
```

```
20.          BETA=0.05*FLA
21.          AS=ALOG((1.-BETA)/ALFA)
22.          WRITE(6,6) ALFA,BETA,AS
23.     6    FORMAT(2X,'ALFA = ',F5.2,' BETA = ',F5.2,' AS
=',F6.3)
24. C
25.          DO 9 LR=1,3
26.          FLR=FLOAT(LR)
27.          RH=0.2*(FLR-1.)+0.3
28.          WRITE (6,8) RH
29.     8    FORMAT(2X,'ROH =',F5.1)
30. C
31.          CO1=0.
32.          CO2=0.
33.          DO 3 L=1,N3
34.          DO 5 N=3,N2
35.          FN=FLOAT(N)
36.          CALL RCAL(N,RH,R)
37.          Y=(1.+RH1*R)/2.
38.          CALL F(N,Y,FR1)
39.          Y=(1.+RH0*R)/2.
40.          CALL F(N,Y,FR2)
41.          Z=((FN-1.)/2.)*ALOG((1.-RH1**2)/(1.-RH0**2))
42.     1        -(FN-1.5)*ALOG((1.-RH1*R)/(1.-RH0*R))
                                   +ALOG(FR1/FR2)
43.          IF (Z.GE.AS)  THEN DO
44.                            GO TO 25
45.                            END IF
46.     5    CONTINUE
47.    25    NN(L)=N
48.          CO1=CO1+FLOAT(N)
49.          CO2=CO2+FLOAT(N)**2
50.     3    CONTINUE
51. C
52.          WRITE(6,23)(NN(I),I=1,100)
53.    23    FORMAT(5I5)
```

```
54. C
55.        FN3=FLOAT(N3)
56.        ARL=CO1/FN3
57.        VN=(CO2-CO1**2/FN3)/(FN3-1)
58.        WRITE(6,27) ARL,VN
59.    27  FORMAT(2X,' ARL =',F9.3,'  Variance =',F9.3)
60.     9  CONTINUE
61.        PRINT,' '
62.     7  CONTINUE
63.        STOP
64.        END
65. C
66. C Subroutine to calculate the Gaussian hypergeometric
                                              function.
67. C
68.        SUBROUTINE F(N,Y,TS)
69.        FN=FLOAT(N)
70.        GAM= FN - 0.5
71.        N1=100
72.        TSU=1.
73.        DO 10 J=1,5
74.        FJ=FLOAT(J)
75.        TJ1=2.*ALGAMA(0.5+FJ)+ALGAMA(GAM)+FJ*ALOG(Y)
76.        TJ2=2.*ALGAMA(0.5)+ALGAMA(GAM+FJ)+ALGAMA(FJ+1.)
77.        TJ=EXP(TJ1-TJ2)
78.        TSU=TSU+TJ
79.    10  CONTINUE
80. C
81.        DO 12 J=6,N1
82.        FJ=FLOAT(J)
83.        TJ1=2.*ALGAMA(0.5+FJ)+ALGAMA(GAM)+FJ*ALOG(Y)
84.        TJ2=2.*ALGAMA(0.5)+ALGAMA(GAM+FJ)+ALGAMA(FJ+1.)
85.        TJ=EXP(TJ1-TJ2)
86.        IF (TJ.LT.1.E-07) GO TO 16
87.        TSU=TSU+TJ
88.    12  CONTINUE
```

```
89.     16   TS=TSU
90.          RETURN
91.          END
92. C
93. C Subroutine to generate samples of size N from a
                                      bivariate normal
94. C    distribution and to calculate the correlation
                                         coefficient R.
95. C
96.          SUBROUTINE RCAL(N,ROH,R)
97.          DIMENSION X1(300),X2(300),RR(2)
98.          DOUBLE PRECISION DSEED
99.          COMMON DSEED
100.         U1=0.
101.         U2=0.
102.         SIG1=1.
103.         SIG2=1.
104.         T1=0.
105.         T2=0.
106. C
107.         DO 15 K=1,N
108.         CALL GGNPM(DSEED, 2,RR)
109.         X1(K)=RR(1)*SQRT(SIG1)+U1
110.         X2(K)=ROH*RR(1)*SQRT(SIG2)+SQRT(1.-ROH**2)
                                   *SQRT(SIG2)*RR(2)+U2
111.         T1=T1+X1(K)
112.         T2=T2+X2(K)
113.    15   CONTINUE
114.         FN=FLOAT(N)
115.         XB1=T1/FN
116.         XB2=T2/FN
117.         T3=0.
118.         T4=0.
119.         T5=0.
120. C
121.         DO 20 K=1,N
```

```
122.          T3=T3+(X1(K)-XB1)**2

123.          T4=T4+(X2(K)-XB2)**2

124.          T5=T5+(X1(K)-XB1)*(X2(K)-XB2)

125.    20    CONTINUE

126. C

127.          S1=T3/(FN-1.)

128.          S2=T4/(FN-1.)

129.          S1S=SQRT(S1)

130.          S2S=SQRT(S2)

131.          S12=T5/(FN-1.)

132.          R=S12/(S1S*S2S)

133.          RETURN

134.          END

135. C

136. $ENTRY

137. /*
```

Program 6:

```
1. // JOB 'T=8M'

2. // EXEC WATFIV

3. //FT05F001 DD *

4. $JOB

5.          DIMENSION NN(100)

6.          DOUBLE PRECISION DSEED

7.          COMMON DSEED

8.          DSEED=23512.D0

9. C

10.         RH0=0.1

11.         RH1=0.5

12.         N2=300

13.         N3=100

14.         WRITE(6,2) RH0,RH1

15.    2    FORMAT(2X,'RH0=',F5.1,' RH1=',F5.1)

16. C

17.         DO 3 LA=1,2
```

```
18.           FLA=FLOAT(LA)
19.           ALFA=0.05*FLA
20.           BETA=0.05*FLA
21.           AS=ALOG((1.-BETA)/ALFA)
22.           WRITE(6,6) ALFA,BETA,AS
23.      6    FORMAT(2X,' ALFA = ',F5.2,' BETA = ',F5.2,
                                  ' AS = ',F6.3)
24. C
25.           DO 5 LR=1,3
26.           FLR=FLOAT(LR)
27.           RH=0.2+(FLR-1.)*0.3
28.           WRITE(6,8) RH
29.      8    FORMAT(2X,'RH =',F5.1)
30. C
31.           FMU1=0.5*ALOG((1.+RH1)/(1.-RH1))
32.           FMU0=0.5*ALOG((1.+RH0)/(1.-RH0))
33.           CO1=0.
34.           CO2=0.
35.           DO 7 L=1,N3
36.           DO 9 N=5,N2
37.           FN=FLOAT(N)
38.           CALL RCAL(N,RH,R)
39.           U=0.5*ALOG((1.+R)/(1.-R))
40.           FN=FLOAT(N)
41.           Z=U*(FMU1-FMU0)*(FN-3.)-0.5*(FMU1**2-FMU0**2)
                                  *(FN-3.)
42.           IF (Z.GE.AS)   THEN DO
43.                              GO TO 25
44.                              END IF
45.      9    CONTINUE
46.      25   NN(L)=N
47.           CO1=CO1+FLOAT(N)
48.           CO2=CO2+FLOAT(N)**2
49.      7    CONTINUE
50.           WRITE(6,23) (NN(I),I=1,100)
51.      23   FORMAT(5I5)
```

```
52. C

53.          FN3=FLOAT(N3)

54.          ARL=CO1/FN3

55.          VN=(CO2-CO1**2/FN3)/(FN3-1)

56.          WRITE(6,27) ARL,VN

57.     27   FORMAT (2X,' ARL =',F9.3,' Variance =',F9.3)

58.      5   CONTINUE

59.          PRINT,' '

60.      3   CONTINUE

61.          STOP

62.          END

63. C

64. C   Subroutine to generate samples of size N from a
                                          bivariate normal
65. C   distribution and to calculate the correlation
                                          coefficient R.
66. C

67.          SUBROUTINE RCAL(N,ROH,R)

68.          DIMENSION X1(300),X2(300),RR(2)

69.          DOUBLE PRECISION DSEED

70.          COMMON DSEED

71.          U1=0.

72.          U2=0.

73.          SIG1=1.

74.          SIG2=1.

75.          T1=0.

76.          T2=0.

77. C

78.          DO 15 K=1,N

79.          CALL GGNPM(DSEED, 2,RR)

80.          X1(K)=RR(1)*SQRT(SIG1)+U1

81.          X2(K)=ROH*RR(1)*SQRT(SIG2)
                 +SQRT(1.-ROH**2)*SQRT(SIG2)*RR(2)+U2

82.          T1=T1+X1(K)

83.          T2=T2+X2(K)

84.     15   CONTINUE
```

```
 85.          FN=FLOAT(N)
 86.          XB1=T1/FN
 87.          XB2=T2/FN
 88.          T3=0.
 89.          T4=0.
 90.          T5=0.
 91. C
 92.          DO 20 K=1,N
 93.          T3=T3+(X1(K)-XB1)**2
 94.          T4=T4+(X2(K)-XB2)**2
 95.          T5=T5+(X1(K)-XB1)*(X2(K)-XB2)
 96.     20   CONTINUE
 97. C
 98.          S1=T3/(FN-1.)
 99.          S2=T4/(FN-1.)
100.          S1S=SQRT(S1)
101.          S2S=SQRT(S2)
102.          S12=T5/(FN-1.)
103.          R=S12/(S1S*S2S)
104.          RETURN
105.          END
106. C
107. $ENTRY
108. /*
```