A REVIEW OF CUSUM SEQUENTIAL CONTROL PROCEDURES
WITH SPECIAL EMPHASIS ON
THE RUN-LENGTH DISTRIBUTION OF A CUSUM STATISTIC

by

Al Byrd

B.Sc., Simon Fraser University, 1978

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
in the Department
of
Mathematics

© Al Byrd 1980
SIMON FRASER UNIVERSITY
MARCH 1980

All rights reserved. This work may not be reproduced in whole or in part, by photocopy or other means, without permission of the author.
APPROVAL

Name: Al Byrd

Degree: Master of Science

Title of Thesis: A Review of CUSUM Sequential Control Procedures with Special Emphasis on the Run-length Distribution of a CUSUM Statistic

Examining Committee:

Chairperson: S.K. Thomason, Professor

---------------------------------------------
Dr. M.A. Stephens, Professor
Senior Supervisor

---------------------------------------------
Dr. D. Eaves, Assistant Professor

---------------------------------------------
Dr. P. Kennedy, Professor (Economics Dept.)

---------------------------------------------
Dr. A. Beckenbach, Assistant Professor
Biology Department, Simon Fraser University
External Examiner

April 15, 1980
Date Approved:  

(ii)
PARTIAL COPYRIGHT LICENSE

I hereby grant to Simon Fraser University the right to lend my thesis or dissertation (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this thesis for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Title of Thesis/Dissertation:

A REVIEW OF CUSUM SEQUENTIAL CONTROL PROCEDURES WITH SPECIAL EMPHASIS ON THE RUN-LENGTH DISTRIBUTION OF A CUSUM STATISTIC

Author: ____________________________

(signature)

AL BYRD

(name)

15 APR 80

(date)
ABSTRACT

In this thesis we review a number of the most important methods of controlling a parameter of a sequence of random variables, with special emphasis on CUSUM control procedures. We focus our attention on the problem of deriving or approximating the average run-length and run-length distribution of a CUSUM statistic. The Bagshaw-Johnson (1974b, 1975a) approximations are compared, from both a theoretical and practical point of view, with the approximations or exact expressions obtained by other authors; in particular, those of Page (1954), Ewan and Kemp (1960) and Brook and Evans (1972). We also discuss the computational and numerical problems associated with these approximations and exact expressions. A Monte Carlo simulation study is carried out to empirically determine the goodness-of-fit of the Bagshaw-Johnson approximations. In addition, a few of the results of Page, Ewan and Kemp and Brook and Evans are linked together, several new results are introduced, and a new approximation to the run-length distribution of a CUSUM statistic is derived for the case in which the observations are autoregressive and the process is out of control.
ACKNOWLEDGEMENTS

I would like to thank the members of my supervisory committee, especially Dr. M. A. Stephens, for their helpful suggestions. Special thanks also go to Dr. R. Lockhart, Dr. C. Kim, Dr. C. Villegas and Dr. L. Weldon for their assistance to me at various times over the last few months. Last but not least, I would like to thank Jorge Holguín, Dora Hutchinson and John Spinelli for putting up with me all these months while I drafted and re-drafted and ... and re-drafted my thesis.
# Table of Contents

<table>
<thead>
<tr>
<th>Approval</th>
<th>ii</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>iii</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>iv</td>
</tr>
<tr>
<td>List of Figures</td>
<td>vi</td>
</tr>
</tbody>
</table>

## Chapter 1: Introduction to Sequential Control Procedures

1.1: The Sequential Change-point Detection Problem  
1.2: Applications of Sequential Control Procedures  
1.3: Sequential Control Procedures: an Overview  

## Chapter 2: Results of Page, Ewan and Kemp and Brook and Evans

2.1: Results of Page  
2.2: Results of Ewan and Kemp  
2.3: Results of Brook and Evans  

## Chapter 3: Results of Bagshaw and Johnson

3.1: Review of some Topics from Time Series Analysis  
3.2: Controlling the Mean of an ARMA Process  
3.3: Controlling an Autoregressive or Moving Average Parameter of an ARMA process  

## Chapter 4: Analysis of Results

4.1: Analysis of Bagshaw-Johnson Results  
4.2: Analysis of Results of Page, Ewan and Kemp and Brook and Evans  
4.3: Some Practical Problems Associated with the Computation of the Run-length Distribution of a CUSUM scheme  

## References:  
152
<table>
<thead>
<tr>
<th>Figure 1</th>
<th>141</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 2</td>
<td>142</td>
</tr>
<tr>
<td>Figure 3</td>
<td>143</td>
</tr>
<tr>
<td>Figure 4</td>
<td>144</td>
</tr>
<tr>
<td>Figure 5</td>
<td>145</td>
</tr>
<tr>
<td>Figure 6</td>
<td>146</td>
</tr>
<tr>
<td>Figure 7</td>
<td>147</td>
</tr>
<tr>
<td>Figure 8</td>
<td>148</td>
</tr>
<tr>
<td>Figure 9</td>
<td>149</td>
</tr>
<tr>
<td>Figure 10</td>
<td>150</td>
</tr>
</tbody>
</table>
INTRODUCTION TO SEQUENTIAL CONTROL PROCEDURES

INTRODUCTION

Section 1 of this chapter describes the general class of problems which we are concerned with in this thesis, and the particular member of this class which we focus our attention on (the SEQUENTIAL CHANGE-POINT DETECTION PROBLEM). The second section contains a number of examples of sequential change-point detection problems, and illustrates the application of sequential control procedures to these problems. The last section reviews some of the most important types of sequential control procedures.
SECTION 1.1: THE SEQUENTIAL CHANGE-POINT DETECTION PROBLEM

It is well-known in the field of Statistics that the problem of finding a parametric model for a non-stationary stochastic process is quite difficult, due to the complexity of the relationships which exist between dependent random variables. At present there is no unified, comprehensive class of methods for dealing with this problem, although there are methods available for handling certain special cases.

One of the more interesting and innovative methods to come to the attention of this author is based on a so-called 'dynamic linear model' for the stochastic process. This model, introduced by P.J. Harrison and C.F. Stevens (1971, 1976), generalizes the ordinary linear model by allowing the parameters of the model to vary stochastically. To be specific, the parameter vector follows a multi-dimensional random walk. Their approach is based on the principles of Bayesian inference, and their chief results are modifications of the Kalman filtering algorithm (Jaswinski, 1970).

In a recent paper (Ledolter, 1979), J. Ledolter employs a similar technique, also based on the Kalman filtering algorithm, for recursively estimating the parameters of an ARIMA time series model with stochastically-varying coefficients.
A frequency-domain approach to this problem has been adopted by M.Y. Hussain and T. Subba Rao (1976), who have developed a method of obtaining point estimates of the parameters of an ARMA stochastic process with time-dependent coefficients. Their method is based on a rather sophisticated technique called 'evolutionary spectral analysis'. A general description of this approach and a number of key references are to be found in the discussion section of a paper by R.L. Brown, J. Durbin and J.M. Evans (1975).

For an up-to-date review of recent progress in the field of time series analysis (including methods of modelling non-stationary time series), the reader is referred to the expository paper by S. Makridakis (1978).

In contrast to the non-stationary case, the theory and methodology associated with stationary time series models is well-developed. Numerous books have been written on the theoretical and/or practical aspects of modelling stationary time series (e.g. Box and Jenkins, 1976; Jenkins and Watts, 1968). In light of this fact, it would seem reasonable to attempt to model a non-stationary time series by applying the theory and methodology which have been developed for stationary time series. One way of achieving this aim, of course, would be to transform the time series to stationarity.
This is a well-established procedure, and both of the above-mentioned books describe methods of doing this.

Another way of achieving this aim would be to assume that the index set of the time series can be partitioned into a number of disjoint subsets in such a way that the time series can be treated as if it were stationary over each subset of the partition, but not necessarily between subsets of the partition or close to the boundaries of the subsets. We shall call a time series with this characteristic a PIECEWISE STATIONARY TIME SERIES. This definition is analogous to the definition of a piecewise linear or a piecewise continuous function in real analysis.

Consider now the following set of conditions: let the subsets of the partition referred to above be defined by

\[ T(1) = \{ t: -\infty < t < t'(1) \} \]

\[ T(2) = \{ t: t'(1) \leq t < t'(2) \} \]

\[ \ldots \]

\[ T(i) = \{ t: t'(i-1) \leq t < t'(i) \} \]

\[ \ldots \]
$T(n+1) = \{ t: t'(n) \leq t < \infty \}$.

The points $t'(1), t'(2), \ldots, t'(n)$ will be called CHANGE-POINTS. Let $X[i], i \in I$, be a piecewise stationary stochastic process defined on the set of integers, $I$, and let $x(1), x(2), \ldots$, be a sequence of sample values of $X$. We are concerned with the problem of using this sequence of observations to detect a change from one subset of the partition to another. We shall call this the GENERAL SEQUENTIAL CHANGE-POINT DETECTION PROBLEM.

The special case of this problem which we shall focus our attention on is the one such that:

1. **WITHIN** each subset of the partition, the random variables, $X[i], i \in I$, are either
   a.) independent and identically-distributed, or
   b.) they follow an autoregressive-moving average (ARMA) model.

2. **BETWEEN** each subset of the partition, the distributions of the random variables differ by a change in
   a.) the mean of a distribution function, which may or may not be normal, or
   b.) one or more of the autoregressive or moving average parameters of an ARMA model.
3. There is only one change-point, say $t'$, and this change-point is unknown. The procedures discussed may be re-applied to detect successive change-points if they are relatively far apart.

4. There is a fully-specified mathematical model for the stochastic process, and the purpose of monitoring the observations is to detect significant deviations of the process from this model.

From now on, whenever we refer to a sequential change-point detection problem, we shall be referring to the above-defined problem. Note that we are NOT concerned with the problem of estimating the change-points, but rather with detecting a change from one subset of the partition to another. Page (1955) and Hinkley (1971) are two authors who have worked on the change-point estimation problem. Note also that we are concerned with sequential procedures and NOT fixed-sample-size procedures. Many authors have adopted a fixed-sample-size approach to this problem; for example, Brown, Durbin and Evans (1975) and Schweder (1976). Illustrative examples of change-point detection problems, and the application of sequential control procedures to these problems, will be given in the next section.
SECTION 1.2: APPLICATIONS OF SEQUENTIAL CONTROL PROCEDURES

1. Industrial Quality Control

Perhaps the most obvious application of sequential control procedures is in the field of industrial quality control. Suppose that a manufacturing process produces a sequence of batches of items (e.g., handbags, chemical fertilizer, bubblegum), and suppose further that it is possible to assign to every item in every batch a number which can be used to assess the quality of the item according to a specified set of criteria. A 'quality number' such as this may simply be some measured or counted characteristic of the item (e.g., length, weight, number of defects per square meter, etc.), or it may be an assigned quantity representing the level of quality of the item on a nominal scale (e.g., good = 1, bad = 0, etc.). The chief task of industrial quality control is to select a sequence of samples of items (one sample from each batch), compute a corresponding sequence of statistics, $x(1), x(2), \ldots$, and based on an analysis of this information, determine if the level of quality of the batches produced by the process has changed significantly from some TARGET VALUE.

If the procedure indicates that a significant
change has occurred, then action is taken to determine if the change is due to random variation or a real change in the nature of the production process. In the former case the signal is said to be a FALSE ALARM and the control procedure is re-started, while in the latter case the process is said to be OUT OF CONTROL and measures are initiated to bring the process back into control, after which the control procedure is re-started. In this sense the process is sequentially 'controlled' by the statistical procedure (hence the name SEQUENTIAL CONTROL PROCEDURE).

One way of attacking the problem faced in industrial quality control is to assume that the sequence of observations, \( x(1), x(2), \ldots \) are sample values of a sequence of random variables, \( X[1], X[2], \ldots \), and that a change in the level of quality of the industrial process is equivalent to a change in the joint distribution of these random variables. For example, we could assume that, when the level of quality of the batches produced by the process is satisfactory, the \( X[i], i = 1, 2, \ldots \) form a sequence of independent, identically-distributed normal random variables with a known mean and standard deviation. Depending on the circumstances, it might be reasonable to assume that a change in the level of quality of the batches is due to a change in the mean, standard deviation, serial correlation coefficient, or some
other parameter (possibly vector-valued), associated with their joint distribution. A value of this parameter which results in an acceptable level of quality will be called an ACCEPTABLE PARAMETER VALUE (APV), and a value which results in a rejectable level of quality will be called a REJECTABLE PARAMETER VALUE (RPV). The difference between the target value and the actual level of quality of a batch will be referred to as the ERROR or DEVIATION of that batch.

At this point we should distinguish between the type of statistical control procedure which initiates action after each observation, and the type which initiates action only after evidence of a possibly significant departure from the target value. The former type of control procedure is often implemented on an automatic device which is directly connected to the production machinery, such as a computer-controlled servo-mechanism, and for this reason will be referred to as a NO-DELAY CONTROL PROCEDURE. In contrast, the latter type of control procedure is typically implemented by a human, such as a quality control inspector, or by a piece of auxiliary equipment, such as an emergency warning device, and consequently will be referred to as a DELAYED CONTROL PROCEDURE. No-delay control procedures are most profitably applied to situations in which the possible sources of errors, and the corresponding appropriate corrective actions,
are well-known, and the cost of taking the corrective actions is relatively small. On the other hand, delayed control procedures are most profitably applied when one or more of these conditions do not hold, or as back-up systems for no-delay control procedures. The adaptive feedback control procedures described in the book by Box and Jenkins, (1976, part 4), are examples of no-delay control procedures, whereas the control chart procedures described in the review paper by S.W. Roberts (1966) are examples of delayed control procedures. In this thesis we are mainly concerned with delayed sequential control procedures.

2. Remote Control and Tracking of Moving Objects

Consider now the problem of controlling the position of an object in two dimensions. The observations, \( x(1), x(2), \ldots \), correspond to 'tracking signals' which measure the deviation of the object above or below a target path, say \( f(t) \), where \( f(t) \) is the preferred position of the object at time \( t \). The purpose of the control procedure is to detect significant departures from the target path, so that action can be taken to correct the trajectory of the object by remote control.

The problem posed in this application differs in at least two respects from the problem posed in the
quality control application. The first difference is that in this application the index space of the underlying stochastic process is continuous, since the position of the object at any point in time is well-defined, whereas in the quality control application the index set of the underlying stochastic process is discrete, since batches of items are discrete entities. Secondly, in this application the target value varies as a function of the index variable, whereas in the quality control application the target value is independent of the index. However, these are not major differences from our point of view, as the random variables, $X[1], X[2], \ldots$, corresponding to the tracking signals, $x(1), x(2), \ldots$, do have a discrete index space, and are centered about a constant value of zero when the trajectory of the object is under control.

We realize that a no-delay control procedure is probably required to directly control the trajectory of the object, but a delayed control procedure may be useful as a back-up system.

3. Economic Impact Studies

Another interesting application of sequential control procedures is to economic impact studies. The major aims of an economic impact study are to determine
if a certain event with economic ramifications (e.g.,
the introduction of a new tax measure; the unfolding of a
political scandal; the announcement of a
multi-million dollar trade agreement; etc) has had, or
is having, a significant impact on the economy, and if
so, the extent and time of onset of the impact. The
extent of the impact is usually measured by changes in
one or more economic indicators, such as the consumer
price index, or the gross national product. We will
only consider the case of a single economic indicator.
It should be emphasized that even though the time of
the economic event might be known, the time of the
onset of the impact might still be unknown, except, of
course, that if the impact does occur, it must
certainly occur after the economic event.

The problem of detecting a significant impact may
be dealt with by fitting a time series model to data
prior to the occurrence of the economic event, and then
monitoring the sequence of one-step-ahead forecast
errors (deviations from the values of the economic
indicator forecast by the model) in search of an
otherwise unexplainable change in their joint
distribution.

4. Drug Response Studies

It would not be very difficult to re-word the
above problem so that it refers to the problem of sequentially monitoring values of a physiological index in order to determine if a certain drug has had, or is having, a significant impact on the health of a patient. Thus, this type of problem may also be treated as a sequential change-point detection problem.

5. Emergency Warning Systems

Suppose that the occurrence of a certain event is considered extremely undesirable, and that it is known that the event is usually preceded by changes in the structure of one or more stochastic processes. Whether the changes cause the onset of the event or vice versa is immaterial; the important thing is that the changes usually precede the event. In this sort of situation it might be useful to construct an emergency warning system by monitoring a sequence of observations generated by these stochastic processes and issuing a warning signal whenever there is sufficient evidence of one or more changes in their structure. Some of the areas in which such a system might be useful are meteorology, pollution monitoring and industrial accident prevention.

We could supply the reader with many more examples, but we hope that the above examples are
sufficient to give the reader an idea of the scope of the applications of sequential control procedures. It is possible that some of the control procedures to be discussed in this thesis could be applied to some of the above-mentioned applications, but it is unlikely that any of the control procedures could be applied successfully to all of the applications. In general, the control procedure which is appropriate for a particular application will depend on the unique characteristics of that application. In other words, there is no such thing as an all-purpose sequential control procedure.
SECTION 1.3: SEQUENTIAL CONTROL PROCEDURES - AN OVERVIEW

The vast majority of the work on the development of sequential control procedures has been done in the field of industrial quality control, and so we have decided to discuss them in this context. The special case of the sequential change-point detection problem studied most frequently in this field is that in which the observations, which measure the level of quality of the batches of items produced by the process, are treated as values of independent random variables. Typically, the common distribution function of the random variables is fully-specified when the process is in control, and only partially-specified when the process is out of control. An out of control condition is assumed to be caused by a change in the value of a parameter of the distribution function, usually the mean or standard deviation. The distribution function itself is usually assumed to be of some standard form, such as that given by the normal curve. We shall begin our discussion in this section by a review of the some of the methods which have been developed for this special case.

One of the oldest and most widely-applied sequential control procedures was introduced by W.A. Shewhart (1931) in the United States and B.P. Dudding
and W.J Jennet (1942) in Great Britain. The essential features of the original form of this procedure are outlined in the following steps: given the sequence of batches of items produced by the manufacturing process, and a target value, \( m' \), for the average acceptable quality level,

1. Take a sample of \( N \) items from the \( i \)-th batch, for \( i = 1, 2, \ldots \), and compute the quality level,

\[
x(i, j), \ j = 1, 2, \ldots, N
\]

of each item in the batch.

2. Compute the mean quality level of the items in each batch,

\[
x(i) = \frac{x(i, 1) + \ldots + x(i, N)}{N}
\]

and plot the points \((i, x(i))\) on a graph, for \( i = 1, 2, \ldots \).

3. If the mean of the \( i \)-th sample falls more than \( g \) standard deviations (where \( g > 0 \) is a specified number, usually 3) above or below the target value, signal that the process is out of control and take appropriate action; otherwise continue sampling.
The graph of the points \((i, x(i))\), \(i = 1, 2, \ldots\) together with the super-imposed TARGET LINE (or center line),

\[ x = m' , \]

and ACTION LINES (or \(g\)-sigma limits),

\[ x = m' + g \cdot s' \]

\[ x = m' - g \cdot s' , \]

is called an XBAR, or SHEWHART, CONTROL CHART.

One disadvantage of this scheme which is immediately obvious is that only the current mean, \(x(i)\), is used at the \(i\)th stage of the procedure to decide if the process is in control or out of control. No use if made of the previous means, \(x(i-j)\), \(j = 1, 2, \ldots\), and in this sense the procedure does not have a 'memory'. In fact, the standard xbar control chart is equivalent to a sequence of independently-applied standard normal tests (often called Z-tests). Although this procedure is sensitive to relatively large changes in the average quality level, (e.g. more than \(g\) standard deviations from the target value), it is insensitive to smaller changes.
Dudding and Jennet (1942), proposed that WARNING LINES,

\[ x = m' + g(w) \cdot s', \quad 0 < g(w) < g \]

\[ x = m' - g(w) \cdot s', \]

be placed on opposite sides of the target line and inside of the action lines, and that a specified number of points between the warning and action lines, or a single point outside of the action lines, should cause an out of control signal to be issued. Several other authors, (e.g. Weiler, 1953, Moore, 1958, as reported in Page, 1961), considered similar schemes involving runs of points within certain horizontal regions of a control chart. Schemes such as this are known as RUN TESTS.

An interesting variant of a run test, called a RUN SUM TEST, was discussed by Roberts (1966). The control chart is partitioned into a number of disjoint zones, symmetrically situated on either side of the center line, and a non-zero integer score is assigned to each of these zones. Zones above the center line are assigned a positive score; those below the center line are assigned a negative score; and the magnitudes of the scores increase as a function of the distance of the zones from the center line. The scores are
accumulated on positive and on negative runs, and when the magnitude of the cumulative sum exceeds a specified value, an out of control signal is produced. As we shall see, this type of procedure may be treated as a restricted form of a more general cumulative sum procedure.

Two other techniques which are designed to make greater use of the available information are based on the construction of moving average and geometric moving average charts (Roberts, 1959, 1966). A moving average of order $p$ is simply the arithmetic mean of the $p$ most recent sample means, $x(i), x(i-1), \ldots, x(i-p+1)$, i.e.,

$$m(i,p) = \frac{x(i) + \ldots + x(i-p+1)}{p},$$

The points $(i, m(i,p))$ are plotted on a graph, as with the Shewhart chart, and an out of control condition is signalled if the most recent point falls outside of a zone determined by two horizontal lines super-imposed on the graph. Although the moving average of order $p$ obviously uses more of the available information than the simple Shewhart chart, it still does not utilize all of the information available. Moreover, the number of terms to be used in a moving average is a thorny problem in practice, as too few terms make the procedure insensitive to small changes.
in the mean, whereas too many terms make the procedure insensitive to large changes. Indeed, for \( p = 1 \), a moving average chart reduces to an (unmodified) Shewhart chart, and as \( p \) approaches infinity, recent observations receive increasingly smaller weights.

A GEOMETRIC MOVING AVERAGE (also called an EXPONENTIALLY-WEIGHTED MOVING AVERAGE) with weight \( a \), \( 0 < a < 1 \) is given by the recursive formula, for \( i = 1, 2, \ldots \)

\[
g(i) = (1-a) \cdot g(i-1) + a \cdot x(i), \quad g(0) = m'.
\]

A geometric moving average control chart is similar in construction to the other control charts mentioned previously. Note that although a geometric moving average uses all of the available information at each stage of the procedure, it places successively smaller weights on the information contained in past samples. This may or may not be a good feature, depending on the importance of the information contained in the previous samples. Also, the performance of this method is quite sensitive to the choice of the weighting factor, \( a \), which determines the extent to which information from the past is discounted, and it is not always easy to choose a reliable value of this parameter.

It is possible to use a SEQUENTIAL PROBABILITY
RATIO TEST, or SPRT, as a sequential control procedure, although it was not designed for this purpose. The $i$th SPRT statistic, $w(i)$, is the sum of the logarithms of a sequence of probability ratio statistics. If the observations are independent and identically-distributed with density $f(x;c)$, and if $c = c(0)$ under the null hypothesis and $c = c(1)$ under the alternate hypothesis, then

$$w(i) = w(i-1) + \ln[ f(x;c(1))/f(x;c(0)]$$

Sampling continues as long as $b < w(i) < a$, with acceptance of the null hypothesis if $w(i) \leq b$ and rejection if $w(i) \geq a$. We shall call any test of this sort a WALK TEST, whether or not $w(i) - w(i-1)$ is a probability ratio statistic. See the book by Ghosh (1970) for a comprehensive account of the theory of sequential tests of hypotheses.

In this situation, an out of control signal is produced if the null hypothesis is either accepted or rejected. That is, the union of the usual acceptance and rejection regions constitutes the out of control region of the scheme. This procedure uses all of the available information, and is uniformly most efficient under certain conditions. Bagshaw (1974) developed a one-sided version of this scheme and obtained a Wiener process approximation to some of its properties.
He showed, however, that sequential control procedures based on SPRT's may not be as efficient as CUSUM control procedures (discussed below) in detecting a small change in the mean of a process when the change-point, $t^*$, is unknown, which is the case we are interested in.

The cumulative sum (CUSUM) procedures introduced by E.S. Page in 1954 make use of all the available information and are reasonably sensitive to both small and large changes in the mean. Page discussed a number of different rules for controlling a process with a cumulative sum procedure, and we shall discuss most of these in the next chapter. The $i$th cumulative sum is usually computed using a recursive formula of the form:

$$s(i) = s(i-1) + (x(i) - k), \quad s(0) = 0.$$  

The quantity $k$ is called a REFERENCE VALUE and is usually chosen to give the graph of the points, $(i, s(i))$, $i = 1, 2, \ldots$ a downward drift when the quality of the products being produced by the process is satisfactory, and an upward drift when the quality is unsatisfactory. One of the rules suggested by Page was that if the graph of the points rises a certain specified amount, say $h$ ($h > 0$), above its previous minimum, the process should be considered to be out of control. In other words, an out of control signal
should be produced whenever

\[ s(i) - \min[s(j)] \geq h, \quad j \leq i. \]

The quantity \( h \) is referred to as the DECISION INTERVAL or the CRITICAL VALUE of the scheme. This type of CUSUM scheme is appropriate for detecting an increase in the mean of the process; other, more general, types can easily be constructed.

In contrast to the standard Shewhart chart, a cumulative sum chart describes the cumulative effect of all the sample means produced by the process in the past. In this sense a cumulative sum procedure has an infinite memory. Thus, if the true mean of the process has increased, this will be reflected in the cumulative sum chart by a pronounced upward drift in the plotted points. Also, in a cumulative sum chart, a rough estimate of the time at which the increase occurred is the time of the previous minimum of the cumulative sum. In fact, as mentioned by Page, under certain conditions this time is the maximum-likelihood estimator of the change-point.

As might be expected, a fair number of papers have been written comparing various control chart procedures. Almost all of the references listed at the end of this thesis pertaining to CUSUM procedures contain comparative studies. We particularly recommend
the papers by W.D. Ewan (1963) and S.W. Roberts (1966).

The usual way of comparing control chart procedures is in terms of their average run-length curves. In this context, the AVERAGE RUN-LENGTH of a procedure is the average number of samples taken before an out of control signal is produced, and an AVERAGE RUN-LENGTH CURVE is a graph of the average run-length of a procedure as a function of the true mean of the samples. One reason for the importance of the average run-length curve in practice is that the cost of controlling a manufacturing process is often an increasing function of both the number of false alarms when the process is in control and the number of batches produced after the process has gone out of control. If the average run-length is long when the process is in control, there will be very few false alarms, and if the average run-length is short when the process is out of control, there will be very few batches produced after the process has gone out of control. Thus, the cost of controlling a process is often a monotonic function of a pair of quantities which are directly related to the average run-length curve.

Another reason for the popularity of the average run-length curve as a basis for comparing statistical control procedures is that it apparently has intuitive
appeal to people working in the field. Probably one of
the most compelling reasons that statisticians have
elected to deal almost exclusively with average
run-length curves is that in most cases it appears to
be very difficult to derive the run-length
distributions in an explicit form. In certain cases
the run-length probabilities can be computed
recursively (Ewan and Kemp, 1960), but the required
computer procedure is computationally expensive and
numerically troublesome. This is an unfortunate
circumstance, as it has long been appreciated (e.g.
Barnard, 1959), that since run-length distributions
tend to be highly right-skewed, the average run-length
of a procedure is not necessarily a very informative
measure of the performance of the procedure.

Predictably, CUSUM control procedures have proven
to be more sensitive to small shifts in the mean of a
process than standard Shewhart control charts, with the
reverse being true for large shifts in the mean (Ewan,
1963; Roberts, 1966). CUSUM control procedures also
appear to be as good as or better than most other
control procedures with respect to the detection of
small shifts in the mean of a process. However, the
remarks of both Ewan and Roberts suggest that the
differences between the average run-length curves of
most of the established sequential control procedures
are probably of more interest to mathematicians than to
people working in the field of quality control. Indeed, in some situations it may be undesirable to use a control procedure which is too sensitive to small changes in the mean of the process, especially if a fairly broad range of means is considered acceptable or the cost of taking corrective action is quite high. If these conditions prevail, the standard Shewhart chart is probably just as useful as any of the more recently-developed procedures.

As might be inferred from the above comments, we do not claim that CUSUM procedures are in any way optimal, but rather that they are competitive, in terms of their average run-length curves, with most of the other common control chart procedures, and that any differences between a well-chosen CUSUM procedure and a hypothetical optimal procedure are likely to be unimportant from a practical point of view.
CHAPTER 2

RESULTS OF PAGE, EWAN AND KEMP
AND BROOK AND EVANS

INTRODUCTION

In this chapter we review some well-known results concerning CUSUM sequential control procedures. The results of Page (1954), Ewan and Kemp (1960) and Brook and Evans (1972) are reproduced in sections 1, 2 and 3, respectively. Of course, many other authors could have been reviewed (e.g., Barnard, 1959; Bissel, 1968; etc.), but we have decided to concentrate on the afore-mentioned authors because their results are the most relevant to the chief topic of our thesis - the run-length distribution of a CUSUM sequential control procedure.
SECTION 2.1: RESULTS OF PAGE

In 1954 E.S. Page introduced a class of sequential control procedures which since then have come to be known as cumulative sum, or CUSUM, schemes. He defined the schemes by listing a number of rules, reproduced below, which described how the schemes were to be applied:

RULE 1:

Take samples of fixed size $N$ at regular intervals; assign a SCORE, $x(i)$, to the $i$th sample and plot the cumulative sum

$$s(n) = x(1) + \ldots + x(n)$$

on a chart. Take action if

$$s(n) - \min[s(i)] \geq h,$$

where $h > 0$ and $i \leq n$.

The quantity $h$ is currently referred to as the DECISION INTERVAL or the CRITICAL VALUE of the scheme. This procedure was designed to detect changes in only one direction (e.g., an increase in the mean of the
process), and is therefore sometimes referred to as a ONE-SIDED sequential control scheme. Page suggested that the sequence of scores, \( x(1), x(2), \ldots \) be so chosen that the sample path of the cumulative sums slopes downward when the level of quality of the items produced by the process is acceptable, and upward when the level of quality is rejectable. In the case of controlling the mean of a process, it is now common practice to produce this type of sample path by subtracting a constant, \( k \), called a REFERENCE VALUE, from each sample mean. It is also common practice to scale the cumulative sums by some multiple of the standard deviation, which is assumed to be known.

**NOTE:** In order to simplify our notation, we generally assume that \( k = 0 \), and most of the formulas in this and the remaining chapters are based on this assumption. However, this does not present any theoretical or practical difficulties, as it is only necessary to make the transformation \( X \rightarrow X - k \) in order to derive the appropriate formulas when \( |k| > 0 \).

Page derived the properties of this scheme as special cases of the properties of the schemes described in rules 2 and 3.

**RULE 2:**

Define
The above stopping rule is then equivalent to the rule: take action if

\[ s'(n) \geq h. \]

The schemes described by rules 1 and 2 are equivalent because the amount by which \( s'(n) \) exceeds 0 is always equal to the amount by which \( s(n) \) exceeds \( \min[s(i)] \) for \( i \leq n \), and thus \( s'(n) \geq h \) iff \( s(n) - \min[s(i)] \geq h \).

Page stated that rule 2 (and hence rule 1) is equivalent to a sequence of Wald tests with common boundaries at \([0, h]\) and common initial values of zero. Recall from chapter 1 that a Wald test is a procedure in which sampling continues as long as \( b < w(i) < a \), where \( w(i) = w(i-1) + x(i), w(0) = w' \), with the null hypothesis being accepted if \( w(i) \leq b \) and rejected if \( w(i) \geq a \). In the case of a CUSUM scheme, \( b = 0 \) and \( a = h \). If the first test ends on the lower boundary, the next test starts at the lower boundary, and if the first test ends on the upper boundary, action is taken. Similarly, if the second test ends on
the lower boundary, the next test starts at the lower boundary, and if the second test ends on the upper boundary, action is taken, etc.

Using this equivalence between a CUSUM scheme and a sequence of Wald tests (which we shall call the WALD EQUIVALENCE), Page was able to obtain an expression for the average run-length of a CUSUM scheme in terms of the average run-length of a Wald test and the probability that a Wald test will end on the lower boundary.

Let $N'(s)$ be the average run-length of a CUSUM procedure starting from $s$, and let $Nw'(s)$, $Nw'(s|0)$ and $Nw'(s|h)$ be the average run-length of a Wald test starting from $s$, unconditional, conditional upon the test ending on the lower boundary, and conditional upon the test ending on the upper boundary, respectively. Also let $Pw(s|0)$ be the probability that the test ends on the lower boundary. For notational convenience put $N' = N'(0)$, $Nw' = Nw'(0)$ and $Pw = Pw(0|0)$.

Since the Wald tests are applied independently, the probability that $r$ acceptance tests occur before the first rejection test is a geometric random variable with probability mass function

$$P[R = r] = Pw(1 - Pw).$$
The expected value of $R$ is thus

$$R' = \frac{P_w}{1 - P_w}.$$ 

It follows that the average run-length of a CUSUM scheme is given by

$$N' = R' \cdot N_w'(0|0) + N_w'(0|h).$$

In words, $N'$ is equal to the average number of acceptance tests before the first rejection test, times the average run-length of an acceptance test, plus the average run-length of a rejection test. Using the fact that

$$N_w' = P_w \cdot N_w'(0|0) + (1 - P_w) \cdot N_w'(0|h),$$

the above equation can be re-expressed as

$$N' = \frac{N_w'}{(1 - P_w)}.$$

Thus $N'$ can be computed if $N_w'$ and $P_w$ are obtainable. It is known (see, for example, van Dobben de Bruyn, 1968), that $N_w'$ and $P_w$ can be obtained as solutions of a pair of integral equations. Let $F(x)$ be the distribution function of the observations, which may be discrete or continuous. Then $N_w'(s)$ satisfies
the integral equation
\[ Nw'(s) = 1 + \int_0^h Nw'(x) \, dF(x - s) \]
and \( Pw(s) \) satisfies the integral equation
\[ Pw(s) = P(-s) + \int_0^h Pw(x) \, dF(x - s) \]
where \( 0 \leq s < h \) and the integral is a Stieltjes integral. These equations can be derived by conditioning on the first observation.

Page also obtained an approximation to the run-length distribution of a CUSUM test when \( Pw \) is nearly 1 (i.e. when there is a high probability that the component Wald tests will end on the lower boundary). If \( N[s] \) is the random variable representing the run-length of a CUSUM scheme starting from \( s \) (with \( N = N[0] \)), and if \( G(n;s) = P[N[s] \leq n] \) (with \( G(n) = P[N \leq n] \)), then Page showed that

\[ G(n) \rightarrow 1 - Pw^n \]
as \( Pw \rightarrow 1 \), where

\[ n' = 1 + n/Nw'(0|0) . \]

Another result which Page stated is that if two
independent CUSUM schemes of this type are operated simultaneously, then, using the above approximation to the run-length distribution of a CUSUM scheme, the run-length of the combined scheme is distributed as the minimum of two independent random variables. Using this result, it can be shown that the following relationship must hold between the average run-length of the combined scheme, say \( N' \), and the average run-lengths of the component schemes, say \( N'(1) \) and \( N'(2) \):

\[
\frac{1}{N'} = \frac{1}{N'(1)} + \frac{1}{N'(2)}.
\]

Page introduced the following rule in order to derive an alternative integral equation for the average run-length of a CUSUM scheme.

**RULE 3:**

Compute \( s(n) \) as in rule 1, except set \( s(0) = s \), where \( 0 \leq s < h \). Take action if either

(a) \( s(n) \geq h \) and \( s(i) > 0 \), for \( i \in [1,n-1] \),

or (b) \( s(n) - \min[s(i)] \geq h \).

An integral equation for this scheme, which
generalizes rule 1, is given by

$$N'(s) = 1 + N'(0).F(-s) + \int_{0}^{h} N'(x).dF(x-s).$$

This integral equation can also be derived by conditioning on the first observation.

The following rule is a TWO-SIDED CUSUM scheme; that is, it can be used to detect both an increase and a decrease in the parameter being controlled.

RULE 4:

Define $s(n)$ as in rule 1 and take action if either

$$s(n) - \min[s(i)] >= h'$$

or

$$\max[s(i)] - s(n) >= h'''.$$

Thus, the two-sided scheme is composed of two one-sided schemes, and an out of control signal is produced if either the cumulative sum rises above its previous minimum by an amount $h'$, or falls below its previous maximum by an amount $h'''$. The properties of this CUSUM scheme were not investigated by Page, but he stated that it is possible to derive integral equations for the average run-length
directly and in terms of component Wald tests, as in the case of a one-sided scheme. In this thesis we shall concentrate mainly on one-sided CUSUM schemes.

An interesting observation made by Page is that if the scores, \( x(1), x(2), \ldots \) are log-probability ratio statistics,

\[
x(i) = \log\left(\frac{f(x;c(1))}{f(x;c(0))}\right),
\]

and if the parameter \( c \) increases from \( c = c(0) \) to \( c = c(1) \) at the point \( t' \), then the time corresponding to the minimum of the cumulative sum, \( s(n) \), is the maximum-likelihood estimator of the change-point \( t' \).

In addition to characterizing a CUSUM scheme as a sequence of Wald tests, Page also recognized that a CUSUM scheme can be treated as a random walk between an absorbing barrier at \( h \) and a holding (reflecting) barrier at zero. This equivalence has proven to be a useful analytical tool, and we shall refer to it as the RANDOM WALK EQUIVALENCE.

Page also made a number of other significant contributions to both the theory and practice of cumulative sum tests (most of his important papers are listed in the bibliography), but we do not intend to focus on his other results in this thesis.
SECTION 2.2: RESULTS OF EWAN AND KEMP

In a series of papers (Ewan and Kemp, 1960; Ewan, 1963; Kemp, 1961-1971), W.D. Ewan and K.W. Kemp extended both the theory and methodology of CUSUM sequential control procedures. The 1963 paper by Ewan is recommended as an introduction to the methodology, and the 1971 paper by Kemp is recommended as a rigorous treatment of the theory. Most of the results which we review in this section are contained in their joint 1960 paper.

Starting with the following recursive equation for the run-length probabilities

\[ P(n; s) = P(n-1; 0) \cdot F(-s) + \int_0^h P(n-1; x) \cdot dF(x-s), \]

\[ P(1; s) = 1 - F(h-s), \]

which was obtained by conditioning on the first observation, Ewan and Kemp showed that the moment generating function, \( M'(t; s) \), of the random variable \( N[s] \) satisfies the recursive equation

\[ M'(t; s) = \exp[t] \cdot [1 - F(h-s) + M'(t; 0) \cdot F(-s) \]

\[ + \int_0^h M'(t; x) \cdot dF(x-s) ]. \]
It is easy to establish this result by noting that the definition of \( \mathcal{M}'(t; s) \) is

\[
\mathcal{M}'(t; s) = \sum_{n=1}^{\infty} p(n; s) \cdot \exp[n \cdot t].
\]

If both sides of the recursive equation for \( p(n; s) \) are multiplied by \( \exp[n \cdot t] \) and then summed over \( n \), the recursive equation for \( \mathcal{M}'(t; s) \) is obtained.

By repeatedly differentiating the recursive equation for \( \mathcal{M}'(t; s) \) with respect to \( t \) and setting \( t = 0 \), it is possible to derive integral equations which the moments of \( N[s] \) must satisfy. For example, differentiating once and setting \( t = 0 \) we get

\[
N'(s) = 1 + N'(0) \cdot F(-s) + \int_{0}^{h} N'(x) \cdot dF(x-s),
\]

which is the integral equation for \( N'(s) \) obtained by Page. Similarly, if \( N''(s) \) is the second moment of \( N[s] \), we must have

\[
N''(s) = 2 \cdot N'(s) - 1 + N'(0) \cdot F(-s) + \int_{0}^{h} N''(x) \cdot dF(x-s).
\]

An interesting observation made by Ewan and Kemp is that if we are using a CUSUM control procedure to detect an increase in the mean, \( m' \), of the
observations, and if the mean does increase, a rough approximation to the average run-length of the CUSUM scheme is given by

\[ N' = 1 + h/m', \text{ (approximately).} \]

They did not attempt to justify this assertion.

By solving the integral equation for \( N' \) with a number of different parameter values, Euan and Kemp constructed a nomogram which can be used to determine the average run-length of a CUSUM scheme, given the parameter values. They assumed that the observations are normally distributed. A table of the average run-lengths of CUSUM schemes for selected parameter values (assuming normality and a standard deviation of 1) is available in the book by van Dobben de Bruyn (1969).

Based on an analysis of their nomogram, Euan and Kemp observed that if a reference value, \( k \), is subtracted from each observation, and if we want the average run-length at \( m' = m'(a) \) to be fixed at some value, say \( C \), then the average run-length at \( m' = m'(r) \) is minimized by choosing

\[ k = \left( m'(a) + m'(r) \right)/2. \]

In the above, \( m'(a) \) is an acceptable parameter
value, \( m'(r) \) is a rejectable parameter value, and for convenience we assume \( m'(r) > m'(a) \). This observation has been theoretically justified by both Reynolds (1975) and Bagshaw and Johnson (1975c).

It has been pointed out by Bissell (1968) that this value of \( k \) is not necessarily optimal (even approximately) if the variance of the observations changes when the mean changes.

It is also possible to express \( M'(t;0) \) as a function of the moment generating function, \( Mw'(t;0|0) \) of a Wald test which, starting from zero, ends on the lower boundary, and the moment generating function, \( Mw'(t;0|h) \), of a Wald test which, starting from zero, ends on the upper boundary. The functional relationship is

\[
M'(t;0) = \frac{1 - Pw}{1 - Pw}\cdot\frac{Mw'(t;0|h)}{1 - Pw\cdot Mw'(t;0|0)}.
\]

The most straightforward way of obtaining the above result is to make use of another functional relationship which Ewan and Kemp established. Let \( Pw(m;0|0) \) be the probability that a simple Wald test starting from zero will end on the lower boundary in \( m \) steps, and let \( Pw(m;0|h) \) be the corresponding probability that it ends on the upper boundary in \( m \) steps. Then by enumerating all the possible outcomes of the Wald tests of which the CUSUM control
procedure is composed, it follows that

\[ P(n;0) = Pw(n;0|0) + \sum_{m=1}^{n-1} Pw(m;0|0) \cdot Pw(n-m;0|0). \]

That is, the event that the CUSUM control procedure, starting from zero, ends on the upper boundary in \( n \) steps is equivalent to the event that a Wald test starting from zero ends on the upper boundary in \( n \) steps, or it ends on the lower boundary in \( m \) steps and is followed by a Wald test which ends on the upper boundary in \( n-m \) steps, for some \( m \in [1, n-1] \).

If we multiply the above equation by \( \exp[n \cdot t] \) and sum over \( n \), we obtain the functional relationship between \( M'(t;0) \), \( MW'(t;0|0) \) and \( MW'(t;0|0) \). We note in passing that it must be possible to derive the relationship between \( N' \) and \( NW' \) obtained by Page and mentioned in section 1 by differentiating \( M'(t;0) \) with respect to \( t \) and setting \( t = 0 \).

We see from the above result that one way to find \( P(n;0) \) is to first find \( Pw(m;0|0) \) and \( Pw(m;0|0) \), and then to employ the relationship between these quantities. Ewan and Kemp showed that \( Pw(m;0|0) \) and \( Pw(m;0|0) \) satisfy a pair of integral equations, which we shall reproduce below. Let

\[ H(n,x;s) = P[S[n] \leq x|S[i] \in (0,h), i < n] \]
where \( S[0] = s \) and \( n \geq 1 \).

By conditioning on the first observation after \( S[n-1] \) (i.e. \( X[n] \)), we find that the following recursive equation holds

\[
H(n, x; s) = \int_0^h F(x-u) \cdot dH(n-1, u; s).
\]

Given this method of computing \( H(n, x; s) \) recursively, the quantities \( P_w(m; 0|0) \) and \( P_w(m; 0|h) \) can be found by means of the equations

\[
P_w(m; 0|0) = \int_0^{dH(m, u; 0)} \text{and}
\]

\[
P_w(m; 0|h) = \int_h^{dH(m, u; 0)}
\]

respectively. The first equation computes all the possible ways in which a Wald test, starting from 0, can reach the lower boundary in \( m \) steps, while the second equation computes all the possible ways in which the upper boundary can be reached in \( m \) steps.

Besides demonstrating several ways in which the exact run-length distribution of \( N \) can be obtained, Ewan and Kemp also derived an asymptotic approximation to the distribution. By repeatedly differentiating the moment-generating function of \( N' \), \( M'(t; 0) \), they were able to show that if \( N' \) is large, the variance of \( N \) is approximately the square of \( N' \), so that it is
reasonable to anticipate that, except for small values of \( n \),

\[
G(n) = 1 - \exp\left(-\frac{(n-1)}{N'}\right) \quad \text{(approximately)}.
\]

An important implication of this result, which Ewan and Kemp pointed out, is that when \( N' \) is large, the distribution of \( N \) is completely determined by \( N' \). This fact re-enforces, and justifies, in an asymptotic sense, the statement made in the previous chapter that the average run-length is a very important quantity in the study of sequential control procedures.

In his 1961 paper, Kemp established that a useful relationship exists between \( N' \), \( N'(1) \) and \( N'(2) \), where \( N' \) is the average run-length of Page's two-sided CUSUM procedure, and \( N'(1) \) and \( N'(2) \) are the average run-lengths of the two one-sided procedures of which it is composed. Recall that Page's two-sided procedure signals at the first \( n \) such that either

\[
s(n) - \min[s(i)] \geq h'
\]

or \( \max[s(i)] - s(n) \geq h'' \).

Kemp showed that if \( h' = h'' = h \), then

\[
1/N' = 1/N'(1) + 1/N'(2).
\]
This result was later proved by Reynolds (1975).

The following alternative methods of computing $P(n;0)$, $Pw$ and $N'$ were provided by Kemp in his 1971 paper: \( H_\infty^{(m)} = H(n,\infty;0) \)

\[
P(n;0) = H_\infty^{(m)} - H(n,h;0) + \sum_{i=0}^{n-1} H(i,0;0) \cdot P(n-i;0),
\]

\[
Pw = \sum_{i=1}^{\infty} H(i,0;0), \text{ and}
\]

\[
N' = \sum_{i=1}^{\infty} i \cdot [H_\infty^{(m)} - H(i,h;0) + H(i,0;0)]/[1 - Pw].
\]

The first two equations are established by enumeration and the last equation follows from the fact that

\[
N' = Nw'/(1 - Pw),
\]

and from the definition of $Nw'$.

Ewan and Kemp also made a number of other contributions to the theory and practice of CUSUM control procedures, but we shall not discuss these results in this thesis as they are not pertinent to our main topic.
SECTION 2.3: RESULTS OF BROOK AND EVANS

D. Brook and D.A. Evans (1972) exploited the random walk equivalence to develop methods of obtaining the moments and run-length distribution of a CUSUM scheme. Their method is theoretically applicable to any sequence of cumulative sums of i.i.d. discrete random variables, and as an approximation to i.i.d. continuous random variables as well. They treated the different possible values of the cumulative sums as states of a Markov chain, with h being an absorbing state and zero being a reflecting state. They derived the transition probability matrix of the chain, and based on this they determined the moments and run-length distribution of the CUSUM scheme. They also obtained an asymptotic approximation to the run-length distribution.

Let

\[ r(0) = (-\infty, 0] \]

\[ r(i) = ((i-1)w, i\cdot w], \quad (i = 1, 2, \ldots, m) \]

\[ r(m+1) = [h, \infty) \]

where \( 0 < w < h/m, \quad r(m+1) = r(h) \) by definition, and \( r(0), \ldots, r(m+1) \) form a partition of the real line.
The transition probabilities may be defined in the following manner:

\[ P(i,j) = P[S[n] E r(j) | S[n-1] E r(i)] \]

where \( i, j = 0, 1, \ldots, m+1 \) and \( P(h, j) = P(m+1, j) \), \( P(i, h) = P(i, m+1) \) by definition. Thus, the transition probability matrix, \( A \), is given by

\[
\begin{align*}
P(0,0), & \quad P(0,1), \quad \ldots \quad , \quad P(0,m), \quad P(0,h) \\
& \\
& \\
& \\
& \\
& \\
& \\
& \\
& \\
& \\
& \\
& \\
P(m,0), & \quad P(m,1), \quad \ldots \quad , \quad P(m,m), \quad P(m,h) \\
& \\
\end{align*}
\]

It is immediately obvious that

\[ P(h,i) = 0 \quad (i = 0, 1, \ldots, m), \]

\[ P(h,h) = 1, \]
Using the midpoint of the interval \((i-1).w, i.w]\) as a token point for the interval, we find that the remaining transition probabilities are approximated by

\[
P(i, j) = P[X + i.w - w/2 E r(j)],
\]

\(i = 1, 2, \ldots, m\) and \(j = 0, 1, \ldots, m+1\).

We shall now discuss some of the results of Brook and Evans. Let \(N[i]\) be the run-length of \(S[n]\) starting from state \(i\), let \(N'(i)\) be the average run-length, and let \(N(k;i)!\) be the \(k\)th factorial moment of \(N[i]\); that is,

\[
N(k;i)! = E[N[i].(N[i]-1) \cdot \ldots \cdot (N[i]-k+1)].
\]

In particular, \(N(1;i)! = N'(i)\).

If \(B\) is the matrix obtained by deleting the last row and last column of \(A\), and if \(N'\) is an \((m+1 \times 1)\) column vector with components \(N'(0), N'(1), \ldots, N'(m)\), then Brook and Evans showed that a matrix equation for \(N'\) is given by

\[
(I - B) \cdot N' = \lambda,
\]

where \(I\) is an \((m+1 \times m+1)\) identity matrix and \(\lambda\) is
an \((m+1 \times 1)\) column vector of ones.

Note the similarity between this formula and the formula for the mean of a geometric distribution. If we observe a sequence of independent trials with a fixed probability, \(p\), of success on each trial, then the probability that the first failure occurs on trial \(t\) is given by

\[
P[T = t] = p \cdot (1-p)^{t-1}
\]

The mean of \(T\), say \(T'\), is given by

\[(1-p) \cdot T' = 1\]

The similarity referred to above should now be obvious. We shall see that there are more similarities between the properties of the run-length distribution of a CUSUM scheme and the geometric distribution.

NOTE: The variables \(R\) and \(T\) are both geometric random variables, but they have different means because they have been defined differently. \(P[R=r]\) is the probability of obtaining the first failure on the \((r+1)\)st trial, and \(P[T=t]\) is the probability of obtaining the first failure on trial \(t\). Thus,

\[
P[R=r] = p \cdot P[T=r].
\]
If we define $\mathbf{N}(k)!$ to be the $(m+1 \times 1)$ column vector with components $N(k;0)!$, $N(k;1)!$, ..., $N(k;m)!$, then Brook and Evans showed that

$$(I - B).\mathbf{N}(k)! = k! \cdot B.\mathbf{N}(k-1)!.$$  

Once again, there is a striking similarity between this recursive relationship and the one that holds between the factorial moments of a geometric random variable. If $T(k)!$ is the $k$th factorial moment of a geometric random variable,

$$(1-p).T(k)! = k! \cdot p.T(k-1)!.$$  

which is obviously of the same form as the preceding matrix equation.

By successive substitution we obtain the following explicit form for the $k$th factorial moment of $\mathbf{N}$, where

$$\mathbf{N} = [N[0], ..., N[m]]:$$

$$(I - B).\mathbf{N}(k)! = k! \cdot B.\mathbf{N}^{k-1}.$$  

This equation is analogous to the following equation which holds for the $k$th factorial moment of a geometric random variable:

$$(1-p).T(k)! = k! \cdot p^{k-1}.$$
Let \( \underline{p}(n) \) be the \((m+1 \times 1)\) column vector with components \( p(n;0), p(n;1), \ldots, p(n;m) \). Then Brook and Evans showed that a recursive equation for \( \underline{p}(n) \) is given by

\[
\underline{p}(1) = (I - B) \cdot \underline{p}
\]

\[
\underline{p}(n) = B \cdot \underline{p}(n-1),
\]

for \( n = 2, 3, \ldots \). By successive substitutions we discover that

\[
\underline{p}(n) = B \cdot (I - B) \cdot \underline{p} \cdot n-1
\]

This is obviously of the same form as the probability function for the geometric random variable, \( T \) (see above).

If we define \( \underline{g}(n) \) to be the \((m+1 \times 1)\) column vector with components \( g(n;0), g(n;1), \ldots, g(n;m) \), then Brook and Evans state that \( \underline{g}(n) \) satisfies the following matrix equation:

\[
\underline{g}(n) = (I - B) \cdot \underline{g}, \quad n = 1, 2, \ldots
\]

The geometric random variable \( T \) satisfies a univariate analogue of the above equation.
Brook and Evans used the following argument to establish their asymptotic results:

By the Perron-Frobenius Theorem (Cox and Miller, 1965, p120), if B is irreducible and primitive, then B has a unique maximum eigenvalue \( q > 0 \), and corresponding to q there exists two positive eigenvectors, \( \mathbf{v}_1 \) and \( \mathbf{v}_2 \), such that

\[
B \cdot \mathbf{v}_1 = q \cdot \mathbf{v}_1.
\]

\[
\mathbf{v}_2^* \cdot B = q \cdot \mathbf{v}_2^*.
\]

If \( \mathbf{v}_1 \) and \( \mathbf{v}_2 \) are normalized such that \( \mathbf{v}_2^* \cdot \mathbf{v}_1 = 1 \), then (Cox and Miller, 1965, page 123)

\[
\lim_{n \to \infty} \frac{B}{q^n} \to \mathbf{v}_1 \cdot \mathbf{v}_2^* \quad (n \to \infty).
\]

Let \( v_1(i), v_2(i), i = 0, 1, \ldots, m \) be the components of \( \mathbf{v}_1 \) and \( \mathbf{v}_2 \), respectively, and let

\[
C = \sum_{i=0}^{m} v_2(i) \cdot \mathbf{v}_1.
\]

Then by making use of the above facts, it can be shown that the above matrix equations for \( \mathbf{N}' \), \( \mathbf{N}(k) \), \( \mathbf{P}(n) \) and \( \mathbf{G}(n) \) converge to

\[
\mathbf{N}' \to C(1-q).
\]
\[ N(k) \rightarrow \sum_{k=0}^{\infty} \frac{k!}{(1-q)^k} q^k \]

\[ P(n) \rightarrow \sum_{n=0}^{\infty} (1-q)^n q^n \]

\[ G(n) \rightarrow 1 - \sum_{n=0}^{\infty} q^n \]

as \( n \rightarrow \infty \).
CHAPTER 3

RESULTS OF BAGSHAW AND JOHNSON

INTRODUCTION

In section 1 of this chapter we review some topics from time series analysis which serve as background material for the subjects covered in the following two sections. Specifically, we look at the so-called Box-Jenkins method of modelling ARIMA time series, a number of properties of Wiener processes, and a functional central limit theorem. In the next section we review the work of Bagshaw and Johnson on the problem of detecting an increase in the mean of an ARMA process. The last section deals with the complimentary problem of detecting a change in an autoregressive or moving average parameter. The work of some other authors on related problems is also discussed.
SECTION 3.1: REVIEW OF SOME TOPICS FROM

TIME SERIES ANALYSIS

PART 1: The Box-Jenkins Method

Much of the recent work of Bagshaw and Johnson (e.g. 1975a, 1977), is based on the assumption that the stochastic process being controlled follows an ARMA model. Since we intend to review their work later on in this chapter, it is appropriate that we first discuss the class of ARMA time series models. The most popular method of fitting ARMA models to time series data is undoubtedly the so-called Box-Jenkins method, so we have decided to present our discussion of ARMA models in the context of this method.

Box and Jenkins (1976) have developed a method of forecasting time series which theoretically produces uncorrelated one-step-ahead forecast errors. In this section we shall only describe those aspects of the Box-Jenkins method which are relevant to the ensuing discussion.

Given a time series, \( X[t], t = 1, 2, \ldots \), the Box-Jenkins method consists essentially of transforming this series into a stationary time series, \( X[t], t = 1, 2, \ldots \), fitting an ARMA model to the transformed series, checking to determine if the data fits the model adequately, and either using the model to produce
forecasts or choosing another model and repeating the above steps, depending on whether the fit is adequate or not.

Box and Jenkins suggest differencing the time series a sufficient number of times to insure stationarity. That is, set the \( X[t] \) series equal to the \( d \)th difference of the \( X_1[t] \) series,

\[
X[t] = D[d]:X_1[t].
\]

\( D[d] \): is the DIFFERENCE OPERATOR and can be defined recursively in the following manner:

\[
D[1]:X_1[t] = X_1[t] - X_1[t-1],
\]

\[
D[2]:X_1[t] = D[1]:X_1[t] - D[1]:X_1[t-1],
\]

\[
\ldots
\]

\[
D[d]:X_1[t] = D[d-1]:X_1[t] - D[d-1]:X_1[t-1].
\]

Repeated differences can also sometimes be used to eliminate a seasonal effect, although in this case the periodicity of the time series must be taken into account. The number of differences, \( d \), is usually chosen by repeatedly differencing the \( X_1[t] \) time series until the auto-correlations generated by the \( X[t] \) time
series appear to be independent of the time origin (that is, until they appear to depend only on time differences), which is a property of a stationary time series (Box and Jenkins, 1976, chapter 1).

Of course, other types of transformations, such as the Box-Cox transformations (1964), may be useful in this regard. A comprehensive review of statistical transformation techniques is available in the paper by M.H. Hoyle (1973).

An AUTOREGRESSIVE-MOVING AVERAGE (ARMA) time series has the following form:

\[ X[t] - a(1)X[t-1] - \ldots - a(p)X[t-p], \]

\[ = E[t] - b(1)E[t-1] - \ldots - b(q)E[t-q], \]

where \( A_p = [a(1), \ldots, a(p)] \) and \( B_q = [b(1), \ldots, b(q)] \) are two vectors of constants, and where \( E[t] \) is a white-noise stochastic process (usually a sequence of i.i.d. normal random variables with a mean of zero and a variance of \( \sigma^2 \)). When we wish to emphasize the dependence of an ARMA model on the parameters \( p, q, A_p \) and \( B_q \) we shall use the notation \( \text{ARMA}[p,q,A_p,B_q] \).

The above model can be conveniently represented by the operator equation:

\[ L[A_p]:X[t] = R[B_q]:E[t], \]
where

\[ L[Ap] = 1 - a(1) \cdot T[1] - \cdots - a(p) \cdot T[p], \]

\[ R[Bq] = 1 - b(1) \cdot T[1] - \cdots - b(q) \cdot T[q], \]

and

\[ T[m] : X[t] = X[t-m] \text{ is the BACKSHIFT OPERATOR.} \]

If the \( X[t] \) time series follows an \( ARMA[p,q,Ap,Bq] \) model, and if it was obtained from the \( X1[t] \) time series by differencing the latter time series \( d \) times, then the \( X1[t] \) time series is said to follow an AUTOREGRESSIVE-INTEGRATED-MOVING AVERAGE, \( ARIMA[p,d,q,Ap,Bq] \) model.

Box and Jenkins (1976) show that if the roots of the operator equation,

\[ L[Ap] = 0, \]

(treating \( T[m] \): as the \( m \)th power of a continuous variable, \( T \)), all lie outside of the unit circle, the ARMA process is STATIONARY and may be represented by the GENERAL LINEAR PROCESS (GLP),
In the above equation,

\[ \text{GLP}[C] : = 1 + c(1) \cdot T[1] : + c(2) \cdot T[2] : + \ldots, \]

and the sequence of constants, \( C = [c(1), c(2), \ldots] \)
can be found by expanding \( R[Bq] : / L[Ap] : \) in ascending values of \( T[m] : , m = 1, 2, \ldots \) and equating corresponding terms in


Similarly, if the roots of the operator equation,

\[ R[Bq] = 0, \]

all lie outside of the unit circle, the ARMA process is said to be INVERTIBLE and may be represented by the GENERAL AUTOREGRESSIVE PROCESS (GAP),

\[ \text{GAP}[D]: X[t] = E[t], \]

where

\[ \text{GAP}[D] : = 1 + d(1) \cdot T[1] : + d(2) \cdot T[2] : + \ldots, \]
and the sequence of constants, \( D = [d(1), d(2), \ldots] \) can be found by equating corresponding terms in

\[
\]

Box and Jenkins generally assume that the ARMA model for the time series, \( X[t] \), is both stationary and invertible.

The order of the ARMA model can be chosen by examining the auto-correlation and partial auto-correlation functions of the time series, as suggested by Box and Jenkins (chapter 6), or by some other method (see the papers by Ozaki, 1977; McClave, 1975; Gray, Kelley and McIntire, 1978). Box and Jenkins (chapter 7) discuss a number of estimation methods, including a Bayesian and an approximate maximum likelihood approach. Ansley (1979) has worked out the details of an exact maximum likelihood approach, and Ansley et al. (1977), have developed an algorithm which can be used to simultaneously find a suitable Box-Cox transformation of the data and to estimate the parameters of the model.

Let \( X'[t; t_0] \) and \( E'[t; t_0] \) be the conditional expectations of \( X[t] \) and \( E[t] \), respectively, given information up to time \( t_0 \). Then Box and Jenkins show in chapter 5 that the minimum mean-square error forecast, \( X_f[t; t_0] \), of \( X[t] \) at time \( t_0 \) is given by
\[ \text{where for } t-t_0 > 0, \]

\[ x'[t; t_0] = a(1) \cdot x'[t-1; t_0] + \ldots + a(p) \cdot x'[t-p; t_0] \]

\[ -b(1) \cdot e'[t-1; t_0] - \ldots - b(q) \cdot e'[t-q; t_0] \]

and

\[ x'[t; t_0] = x[t], t-t_0 \leq 0, \]

\[ e'[t; t_0] = x[t] - x_f[t; t-1], t-t_0 \leq 0, \]

\[ e'[t; t_0] = 0, t-t_0 > 0. \]

The above recursion relationships are usually initiated by substituting zero for unknown values of \( x'[t; t_0] \) and \( e'[t; t_0] \). The effect of these approximations on the forecasts is negligible for sufficiently long time series.

The authors also show that the corresponding forecast error, \( E_f[t; t_0] \), is given by

\[ E_f[t; t_0] = E[t] + c(1) \cdot E[t-1] + \ldots + c(t-t') \cdot E[t'], \]
where \( t' = t_0 + 1 \). In particular, the one-step-ahead forecast error is simply

\[ E[f[t; t-1]] = E[t], \]

and since \( E[i], E[j] \), are independent for \( i \) not equal to \( j \), the one-step-ahead forecast errors are uncorrelated.

The above results concerning the forecast errors are based on the assumption that the model used to forecast the time series is the same as the model used to generate the time series. In practice, this will seldom if ever be the case. A more general relationship between the one-step-ahead forecast errors and the white-noise sequence, when the forecasting and generating models are not the same, was given by Bagshaw and Johnson (1977) and is reproduced below.

Suppose that a stochastic process, \( X[t] \), is generated by the stationary and invertible ARMA[\( p, q, A_p, B_q \)] model

\[ L[A_p]:X[t] = R[B_q]:E[t]. \]

Suppose further that the following stationary and invertible ARMA[\( p(1), q(1), A_p(1), B_q(1) \)] model, which we shall call MODEL 1, is fitted to the data:
\[ L[Ap(1)]:X[t] = R[Bq(1)]:E1[t], \]

where

\[ L[Ap(1)]: = 1-a1(1).T[1]:-...-a1(p(1)).T[p(1)]:, \]

\[ R[Bq(1)]: = 1-b1(1).T[1]:-...-b1(q(1)).T[q(1)]:, \]

and \( E1[t] \) is the sequence of one-step-ahead forecast errors generated by this model. Then by a straightforward substitution,

\[ L[Ap(1)]: (R[Bq]:/L[Ap]):E[t] = R[Bq(1)]:E1[t], \]

which implies

\[ L[Ap]:R[Bq(1)]:E1[t] = L[Ap(1)]:R[Bq]:E[t]. \]

Defining

\[ L[Ap'] = L[Ap]:R[Bq(1)]: \]

\[ = 1 - a'(1).T[1]: - ... - a'(p').T[p']:, \]

and

\[ R[Bq'] = L[Ap(1)]:R[Bq]: \]
where \( p' \leq p + q(1) \) and \( q' \leq q + p(1) \), we see that the \( E_t \) and \( E_t \) sequences are related by the ARMA\([p', q', Ap', Bq']\) model.

\[
\]

If the fit is exact; that is, if model 1 is identical to the true model, then \( L[Ap] : = L[Ap(1)] : \) and \( R[Bq] : = R[Bq(1)] : \), so that

\[
E_t : = E_t.
\]

Thus, the one-step-ahead forecast errors for the fitted model will be uncorrelated if and only if the fit is exact; otherwise they will follow an ARMA model.

PART 2: Properties of Wiener Processes and a Functional Central Limit Theorem

Most of the results presented in the remainder of this chapter are based on the fact that under a fairly broad set of conditions a normalized sequence of partial sums of random variables converges to a Wiener process. Consequently, in this part of section 1 we
briefly review some useful background material concerning Wiener processes and discuss a functional central limit theorem.

Let \( W[t], t \in [0, \infty) \), be a WIENER PROCESS with MEAN PARAMETER \( w' \), VARIANCE PARAMETER \( w'' \) and INITIAL VALUE \( w \) (if \( w' = 0 \), \( w'' = 1 \), this is called a STANDARD WIENER PROCESS). Then \( W[t] \) has the following properties (see, for example, Cox and Miller, 1965):

\[
\begin{align*}
1. \quad & E[W[t]] = w + w'.t \\
2. \quad & \text{var}[W[t]] = w''.t \\
3. \quad & \text{cov}[W[s], W[t]] = w''.\min[s, t] \\
4. \quad \text{The increments } W[t(1)] - W[t(2)], W[t(3)] - W[t(4)] \text{ are independent if the intervals } (t(1), t(2)), (t(3), t(4)) \text{ do not overlap.}
\end{align*}
\]
For fixed $t$, $W[t]$ is normally distributed, and the joint distribution of $W[t(1)]$, ..., $W[t(n)]$, where $t(1)$, ..., $t(n)$ are fixed numbers in $(0, \infty)$, is multivariate normal.

(6.)

The transition probability density of $W[t]$, $T(W[t];W)$, satisfies the partial differential equation,

$$\left(\frac{W''}{2}\right) \frac{\partial^2 T}{\partial W^2} - W' \frac{\partial T}{\partial W} = \frac{\partial T}{\partial t}.$$

(7.)

If $W[t]$ has an absorbing barrier at $h$ and a reflecting barrier at zero, $T(W[t];W)$ is subject to the constraints

$$T(W[0];W) = 1 \ (W[0] = w),$$

$$= 0 \ (W[0] < w \text{ or } W[0] > w),$$

$$\frac{\partial T}{\partial W} = (2 \cdot W' / W'') \cdot T(W[t];W) = 0 \ (W[t] = 0).$$
Under the above-stated conditions, the probability density of the first passage time (FPT) of $W[t]$ to the absorbing barrier $h$ is given by

$$u(t; w) = -\frac{dv}{dt},$$

where

$$v = \int_0^h T(W[t]; w) \, dw.$$ 

(9.)

The Laplace transform, $L(x; w)$, of the first passage time distribution satisfies the differential equation

$$\left(\frac{d^2 L}{dw^2} + w \cdot \frac{dL}{dw}\right) = x \cdot L,$$

where $x$ is the dummy variable. The boundary conditions are

$$\frac{dL}{dw} = 0 \ (w = 0).$$
Before we introduce the central limit theorem, it is first necessary to define a few terms. A stochastic process is said to be **strictly stationary** if for any fixed numbers, $t(1), t(2), ..., t(n)$, in the domain of $X[t]$, the joint distribution of $X[t(1)], X[t(2)], ..., X[t(n)]$ depends only on the differences, $t(i) - t(j)$, $i, j \in \{1, n\}$. In other words, if all the finite joint probability distributions of $X[t]$ are independent of the time origin, then $X[t]$ is strictly stationary.

Let $X[t]$ be a strictly stationary stochastic process, and suppose that there exists a sequence of constants, $p'(0), p'(1), ..., p'(n)$, converging to zero, such that

$$\left| P[X[t] \in r(i) | X[t-n] \in r(j)] - P[X[t] \in r(i)] \right|$$

is bounded by $p'(n)$, where $r(i)$ and $r(j)$ are arbitrary non-empty subsets in the range of $X[t]$. Then $X[t]$ is said to be a **phi-mixing** stochastic process. The implication of this definition is that, for sufficiently large $n$, the 'present' event, $X[t] \in r(i)$, is effectively independent of the 'past' event, $X[t-n] \in r(j)$.

The following theorem may be found in Billingsley (1968, pages 182-190).
THEOREM 3.1.1: Function Central Limit Theorem

Let $X[i], i = ..., -1, 0, 1, ...$, be a phi-mixing stochastic process, and let $Y[i], Y[i,j]$ be stochastic processes defined by

$$Y[i] = Y[... , X[i-1], X[i], X[i+1], ...],$$

$$Y[i,j] = Y[ X[i-j], ..., X[i], ..., X[i+j]].$$

That is, $Y[i]$ is a real-valued function of $X[i]$, for all $i$, while $Y[i,j]$ is a real-valued function of $X[k]$, for $k \in [i-j, i+j]$.

Further, let


$$W'' = E[Y[0]] + 2 \sum_{i=1}^{\infty} E[Y[0], Y[i]],$$

$$S'[t] = S[n'] / \sqrt{n' W''}.  \tag{3.1.7}$$

where $n'$ is the greatest integer in $nt$ and

$0 \leq t < 1$.

Now suppose that the following conditions hold:
\[
\sum_{i=0}^{\infty} (p'(i))^{k_2} < \infty,
\]

(2.)

\[
E[\psi(n)] = 0,
\]

(3.)

\[
0 < W^{t_1} < \infty,
\]

(4.)

\[
\sum_{j=0}^{\infty} \left( E\left[ (Y[0] - Y[0,j])^2 \right] \right)^{k_2} < \infty.
\]

Then as \( n \to \infty \), \( S'[t] \) converges in distribution to a standard Wiener process, \( W[t] \).
SECTION 3.2: CONTROLLING THE MEAN OF AN ARMA PROCESS

A common assumption of most sequential control procedures is that the random variables generated by the stochastic process which is being controlled are independent and identically-distributed. This is not always a reasonable assumption. The stochastic processes in which we are usually interested are mathematical models for observed phenomena (e.g., industrial processes or diffusion processes or economic processes, etc.) and as such should behave in a manner which coincides with our observations, at least in an approximate sense. Since many real-world phenomena, by their very nature, or by some artifact introduced by the observer, generate auto-correlated observations, it would therefore seem unreasonable to model these phenomena as white-noise processes.

In a series of publications between 1974 and 1977, M. Bagshaw and R.A. Johnson obtained a number of results which may prove to be useful in the analysis of CUSUM schemes with dependent random variables. More specifically, they assumed that the random variables being accumulated follow a specified ARMA time series model, and based on this assumption and the work of Billingsley (1968), Darling and Siegert (1953) and Sweet and Harden (1970), they showed that the sequence of cumulative sums converges to a Wiener process, and
they obtained asymptotic approximations to the run-length distribution and the average run-length.

In this section we consider the problem of detecting an increase in the mean, m', of a sequence of random variables which follows the ARMA[p,q,Ap,Bq] model,

\[ L[Ap]: (X[t] - m'(1)) = R[Bq]: E[t], \]

The parameter m'(1) is the target value of m'. The operators L[Ap] and R[Bq] are defined as in section 1; that is,

\[ L[Ap]: = 1 - a(1) \cdot T[1]: - \ldots - a(p) \cdot T[p]:, \]

\[ R[Bq]: = 1 - b(1) \cdot T[1]: - \ldots - b(q) \cdot T[q]:, \]

where \( T[m]: X[t] = X[t-m] \) is the backshift operator and \( Ap = [a(1), \ldots, a(p)] \), \( Bq = [b(1), \ldots, b(q)] \) are two vectors of constants. \( E[t] \) is a sequence of i.i.d random variables with a mean of zero and a variance of \( E'' \).

Suppose that we apply Page's one-sided CUSUM procedure to control the mean of this process. Recall that this procedure signals that the process is out of control at the first \( n \) such that
where \( n \geq 1, \; i \leq n, \; h > 0 \) and \( S[n] \) is given by

\[
S[n] = X[1] - m'(1) + \ldots + X[n] - m'(1),
\]

\( S[0] = 0 \) by definition.

Note that we are subtracting \( m'(1) \) from each observation, and we are assuming that the reference value, \( k \), is equal to zero, which implies that the expected value of \( S[n] \) will be zero. We consider the more general (and useful) case when \( k > 0 \) subsequently.

One possible way of obtaining an asymptotic approximation to the run-length distribution of this CUSUM procedure is to first obtain an asymptotic approximation to \( S[n] \), say \( S'[c] \), and then to find the exact run-length distribution of \( S'[c] \). This is in fact the way that Bagshaw and Johnson obtained their asymptotic approximation. In Appendix 2 of their 1974 paper, they showed that if we define \( n' \) to be the largest integer in \( nt \), where \( 0 \leq t \leq 1 \), and

\[
B'' = [1 - b(1) - \ldots - b(q)]^2,
\]

\[
A'' = [1 - a(1) - \ldots - a(p)]^2,
\]
\[ W'' = E'' \cdot B''/A'' \]

\[ S'[t] = S[n]/\sqrt{n} \]

then as \( n \to \infty \), \( S'[t] \) converges to a Wiener process, \( W[t] \), with a mean parameter of zero and a variance parameter of \( W'' \).

In order to prove this result they showed that \( S'[n] \) satisfies the conditions of Billingsley's functional central limit theorem (1968, page 182-190). Their result holds true for a general white-noise process (not necessarily normal). Note that although \( S[n] \) is defined only on the non-negative integers, the limiting Wiener process, \( W[t] \), is defined for all real \( 0 \leq t \leq 1 \).

Having obtained an asymptotic approximation to \( S[n] \), they then considered the problem of obtaining the exact run-length distribution and average run-length of this approximation. As a first step in this direction, they noted that when the mean of the Wiener process is zero,

\[ W[t] - \infimum[W[s]] \quad (s \leq t) \]

and \( \|W[t]\| \)
have the same distribution. This result can be found in Karlin (1966, page 281). Thus, the run-length (first passage time) distribution of \( W[t] - \inf \{ W[s] \} \) to \( h \), which we shall call \( u(t;0) \), is equal to the distribution of the first time that \( |W[t]| \) crosses \( h \).

Darling and Siegert (1953) obtained the Laplace transform of \( u(t;0) \) and inverted this transform to obtain \( u(t;0) \). The transform, \( L(x;0,h) \), (\( x \) is the dummy variable) with initial value zero is given by

\[
L(x;0,h) = \frac{1}{\cosh[h(2x/W'')]},
\]

and the first passage time distribution is given by

\[
u(t;0) = \sum_{i=0}^{\infty} u_0(i) \cdot u_1(i) \cdot \exp[(-t/2) \cdot u_1(i)],
\]

where

\[
u_0(i) = (W'')^{i/2}h, \quad (i \text{ even}),
\]

\[-(W'')^{i/2}h, \quad (i \text{ odd}),
\]

\[
u_1(i) = \Pi_{i+1/2} (W'')^{i/2}h.
\]

By differentiating the negative of this transform
and setting \( x = 0 \), Bagshaw and Johnson found that the expected first passage time of \( |W[t]| \) to \( h \), say \( U' \), is equal to

\[ U' = \frac{h}{W''}. \]

Of course, the higher order moments of \( u(t;0) \) could also be found in a similar manner.

Consider now the problem in which \( S[n] \) is re-defined as

\[ S[n] = X[1] - k + \ldots + X[n] - k, \]

where \( k > 0 \). If we set \( S'[r] = S[n']/(n) \), then the Wiener process approximation still holds, except that the process now has a variance parameter of \( W'' \) and a mean parameter of

\[ W' = m'(1) - k \]

when the process is in control, and

\[ W' = m' - k \]

when the process is out of control.

In this case it is no longer true that \( W[t] - \infimum[W[s]] \) and \( |W[t]| \) have the same distribution, so
that a new approach is necessary. Since, as pointed out by Page (1954), a CUSUM procedure is equivalent to a random walk between a reflecting barrier at zero and an absorbing barrier at \( h \), it is not unreasonable to expect that the corresponding procedure involving a Wiener process is equivalent to a continuous random walk (i.e., diffusion process) between the same barriers. Bagshaw and Johnson proved that this is in fact true in Appendix 1 of their 1975a paper. So they reduced the problem of finding the first passage time distribution of \( W[t] - \inf\{W[s]\} \) to \( h \) when \( k > 0 \) to the problem of finding the first passage time distribution of a Wiener process between a reflecting barrier at zero and an absorbing barrier at \( h \).

This latter problem has been solved by Sweet and Harden (1970), although, as pointed out by Bagshaw and Johnson, there is an incorrect constant in their expression. The correct equation, taken from Bagshaw and Johnson (1975a), is given by

\[
u(t; 0) = Q_0(t) \left[ \sum_{i=0}^{\infty} Q_1(t, i)/Q_2(i) + Q_3(t) \right],
\]

where

\[
Q_0(t) = \left(\frac{\omega''}{h}\right) \exp\left[\left(\frac{\omega'}{\omega''}\right) \cdot h \right] - \left(\frac{\omega'}{\omega''}\right) \cdot t; (2, \omega''),
\]

\[
Q_1(t, i) = q(i) \cdot \sin(q(i) \cdot h) \cdot \exp\left[-(q(i))^2 \cdot \omega'' \cdot t/2\right],
\]
\[ Q_2(i) = 1 + \left( W'/hW'' \right) \cdot \left( \sin(q(i)h)/q(i) \right)^2, \]

\[ Q_3(t) = 0 \quad (-W'h/W'' < 1), \]

\[ = 3/(2h) \quad (-W'h/W'' = 1), \]

\[ = Q_{31}(t)/Q_{32}(t) \quad (-W'h/W'' > 1), \]

\[ Q_{31}(t) = -q'sinh(q'h) \cdot \exp[q'W''t/2] \]

\[ Q_{32}(t) = 1 + \left( W'/hW'' \right) \cdot \left( \sinh(q'h)/q' \right)^2, \]

and \( q(1) < q(2) < \ldots \) are the solutions of \( \tan(q'h) = -qW'/W' \) and \( q' \) is the solution of \( \tanh(q'h) = -q'W'/W' \).

As mentioned in section 1, the moment generating function of the Wiener process can be obtained as the solution of a differential equation with appropriate boundary conditions. Bagshaw and Johnson (1975a, Appendix 2) solved this differential equation, following the treatment in Cox (1968). The solution (with initial value zero) is

\[ L(x; 0, h) = L_1(x)/L_2(x), \]

where
\[ L_1(x) = L_3(x) - L_4(x), \]

\[ L_2(x) = L_3(x) \cdot \exp[hL_4(x)] - L_4(x) \cdot \exp[hL_3(x)], \]

\[ L_3(x) = \left[ -u' - (u'^2 + 2xu') \right] / u''. \]

\[ L_4(x) = \left[ -u' + (u'^2 + 2xu') \right] / u''. \]

The expected first passage time of \( W[t] \) can be obtained from this transform, and Bagshaw and Johnson found that it is

\[ U' = \frac{h}{W'} \cdot (1/c) \cdot [\exp[-c] - 1 + c], \]

where

\[ c = 2W' \cdot \frac{h}{W''}. \]

This latter result was obtained independently by Reynolds (1975), although he assumed that the observations are independent and hence did not obtain the same value of \( W' \).

Baghsaw and Johnson (1975c) used the above expression for the expected first passage time of \( W[t] \) to show that if \( h \) and \( k \) are chosen such that \( U' \) is equal to a prescribed constant, say \( A' \), when \( m' = \)
m'(1), then the value of U' corresponding to any other
value of m' > m'(1), say m'(2), is minimized by taking

\[ k = \frac{(m'(1) + m'(2))}{2}, \]

and choosing h such that U' = A' at this value of k. Reynolds (1975) showed that a similar result holds
approximately for a two-sided CUSUM scheme. Both of
these results were obtained by constrained minimization
of the expression for U'. Recall from chapter 2 that
Ewan and Kemp (1960) obtained the same result by an
analysis of their nomogram for the average run-length
of a CUSUM scheme with independent normal random
variables.

Bagshaw and Johnson (1975c) also studied the
effect of estimating the standard deviation on the
average run-length of a CUSUM scheme. Their analysis
showed that the average run-length is sensitive to
minor variations in the value of the standard deviation
and that this sensitivity is greater when k > 0 than
when k = 0. However, by averaging U' over the
distribution of the estimates (assuming i.i.d. normal
observations), they discovered that even though the
values of U' obtained by using estimates of the
standard deviation differed significantly from the true
values of U', when k > 0 these differences appeared to
be in the 'right direction'. For small values of m' -
m'(1), the estimated values of U' tended to
over-estimate the true values, with the reverse case
being true for large values of m' - m'(1). In quality
control applications, it is generally desirable to have
large average run-lengths when the process is in
control, and small average run-lengths when the process
is out of control, so that from this point of view the
differences are in the 'right direction'.

Before concluding this section we should mention
that Nadler and Robbins (1971) obtained asymptotic
Wiener process approximations to the run-length
distribution and the average run-length of Page's
two-sided CUSUM procedure. They assumed that the
observations are independent, and that the scheme is
symmetric; that is, h' = h'' = h. Under these
conditions, they proved that Page's two-sided procedure
is equivalent to the procedure which signals at the
first n such that

$$\max[S[i]] - \min[S[i]] \geq h \ (i \leq n).$$

Following the same line of reasoning as Bagshaw
and Johnson, they approximated S[n] with a Wiener
process, and then obtained the transition probability
density of the range of a Wiener process, from whence
they obtained the first passage time distribution.
They also derived the moment generating function of the
first passage time distribution, and from this they deduced the expected first passage time.
SECTION 3.3: CONTROLLING AN AUTOREGRESSIVE OR MOVING AVERAGE PARAMETER OF AN ARMA PROCESS

In this short section we shall consider the problem of sequentially controlling an autoregressive or moving average parameter of an ARMA process. Very little literature exists on methods of handling this problem. Phatarfod (1971) developed an SPRT for testing the autoregressive parameter of a first order autoregressive, AR(1), process, and he derived asymptotic expressions for its moment generating function and operating characteristic curve. As pointed out in chapter 1, an SPRT can be used as a two-sided sequential control procedure, so Phatarfod's SPRT can be applied to control the autoregressive parameter of an AR(1) process. In an unpublished paper, Box and Jenkins (1966) (also see chapter 4 of Bagshaw, 1974a) described an approximate SPRT for testing the moving average parameter of a first order integrated moving average, MA(1,1), process. The method described below applies to both of these cases, and in addition to many cases in which they do not apply. Bagshaw and Johnson (1977) introduced a CUSUM procedure which can (theoretically, at least) be used to control an autoregressive or moving average parameter of a stationary and invertible ARMA process. Actually, their method applies to an arbitrary ARIMA
process as well, although it is necessary to make certain restrictive assumptions in this case.

Suppose that a time series, $X[t]$, follows the stationary and invertible ARIMA$(p,d,q,Ap,Bq)$ model

$$L[Ap]: Y[t] = R[Bq]: E[t],$$

where $L[Ap]$: and $R[Bq]$: are the operators described in the previous section, $E[t]$ is a sequence of i.i.d. random variables with a mean of zero and a variance of $E''$, and $Y[t]$ is the $d$th difference of $X[t]$; that is

$$Y[t] = D[d]: X[t].$$

Suppose further that two alternative stationary and invertible ARIMA models for the process,

$\text{ARIMA}[ p(1), d, q(1), Ap(1), Bq(1) ],$

$\text{ARIMA}[ p(2), d, q(2), Ap(2), Bq(2) ],$ are given by

MODEL 1: $L[Ap(1)]: Y[t] = R[Bq(1)]: E1[t],$

MODEL 2: $L[Ap(2)]: Y[t] = R[Bq(2)]: E2[t],$

where for $i = 1, 2$,

$$L[Ap(i)]: = 1 - ai(1).T[1] - \cdots - ai(p(i)) . T[p(i)].$$
\[ R[Bq(i)] = 1 - bi(1) \cdot T[1] - \ldots - bi(q(i)) \cdot T[q(i)] \]

and

\[ A_p(i) = [ai(1), \ldots, ai(p(i))] \]

\[ Bq(i) = [bi(1), \ldots, bi(q(i))] \]

and \( E1[t] \) and \( E2[t] \) are the one-step-ahead forecast errors generated by the models. Note that it is assumed that the degree of differencing, \( d \), is the same in all three models. Now define

\[ L[A_p'] = L[A_p] \cdot R[Bq(1)] \]

\[ L[A_p''] = L[A_p] \cdot R[Bq(2)] \]

\[ R[Bq'] = L[A_p(1)] \cdot R[Bq] \]

\[ R[Bq''] = L[A_p(2)] \cdot R[Bq] \]

\[ A_p' = [a'(1), \ldots, a'(p')] \]

\[ A_p'' = [a''(1), \ldots, a''(p'')] \]

\[ Bq' = [b'(1), \ldots, b'(q')] \]
\[ Bq''' = [b''(1), \ldots, b''(q'')] \]

where \( p' = p + q(1), p'' = p + q(2), q' = p(1) + q, q'' = p(2) + q \), and the elements of \( Ap', Ap'', Bq', Bq''' \) are determined by equating coefficients in the above definitions.

Substituting for \( Y[t] \) in models 1 and 2, it is not difficult to show that the following relationships must exist between \( E[t], E1[t] \) and \( E2[t] \):

\[ L[Ap']E1[t] = R[Bq']E[t], \]

\[ L[Ap''']E2[t] = R[Bq''']E[t]. \]

That is, \( E1[t] \) and \( E2[t] \) are both ARMA processes related to the same white-noise process, but with different operators. Moreover, since the \( Y[t] \) time series is stationary and invertible, the \( E1[t] \) and \( E2[t] \) time series will also be stationary and invertible. This in turn implies (Box and Jenkins, 1976, chapter 1) that \( E1[t] \) and \( E2[t] \) have the general linear process representations:

\[ E1[t] = GLP[C']E[t], \]

\[ E2[t] = GLP[C''']E[t], \]
where

$$\text{GLP}[C'] = 1 - c'(1) \cdot T[1] - c'(2) \cdot T[2] - \ldots,$$

$$\text{GLP}[C''] = 1 - c''(1) \cdot T[1] - c''(2) \cdot T[2] - \ldots,$$

and

$$C' = [c'(1), c'(2), \ldots],$$

$$C'' = [c''(1), c''(2), \ldots].$$

The elements of $C'$ and $C''$ are obtained by equating coefficients in $\text{GLP}[C'] = R[Bq']:/L[Ap']$ and $\text{GLP}[C''] = R[Bq'']:/L[Ap'']$, respectively.

Box and Jenkins (1976, page 134) show that the elements of $C'$ can be determined recursively in the following way:

$$c'(j) = a'(1) \cdot c'(j-1) + \ldots + a'(p) \cdot c'(j-p) - b'(j),$$

where $c'(j) = 0$, $j < 0$, $c'(0) = 1$ and $b'(j) = 0$, $1 > j > q$. The elements of $C''$ can be determined in the same way.

Note that if $L[Ap(1)] = L[Ap]$ and $R[Bq(1)] = R[Bq]$, then $E'[t] = E[t]$, for all $t$, with a similar
result holding for \( E_2[t] \). Otherwise, both sequences will be autocorrelated and crosscorrelated.

Let \( H_1 \) be the hypothesis that \( A_p = A_p(1), B_q = B_q(1) \), and let \( H_2 \) be the hypothesis that \( A_p = A_p(2), B_q = B_q(2) \). Define the cumulative sum, \( S_z[n] \), as

\[
\]

where

\[
Z[t] = E_1[t] - E_2[t].
\]

Bagshaw and Johnson suggested that if it is desired to sequentially detect a change in the parameter vector \((A_p, B_q)\) away from \((A_p(1), B_q(1))\) and towards \((A_p(2), B_q(2))\), then Page's rule 1 CUSUM control procedure should be applied to \( S_z[n] \).

In order to approximate the run-length distribution of this CUSUM scheme, Bagshaw and Johnson obtained a Wiener process approximation to \( S_z[n] \), and then applied the results described in the previous section. Specifically, in the Appendix of their 1977 paper they showed that if \( n' \) is the largest integer in \( n t, 0 \leq t < \infty \), and

\[
W_{z'} = E[Z[t]],
\]
\[ Wz'' = \text{Var}[Z(t)] + 2 \sum_{k=1}^{\infty} \text{Cov}[Z(t), Z(t+k)], \]

\[ S_z'[t] = S_z[n'] / (n^2), \]

then \( S_z'[t] \) converges in distribution to a Wiener process, \( Wz[t] \), with a mean parameter of \( Wz' \) and a variance parameter of \( Wz'' \).

In order to obtain their expressions for \( Wz' \) and \( Wz'' \), Bagshaw and Johnson made use of the following properties of the moments of \( E1[t] \) and \( E2[t] \):

(1)

\[ E[E1[t]] = E[E2[t]] = 0, \]

(2)

\[ E[E1[t]] = E'' \cdot \sum_{i=0}^{\infty} [c'(i)]^2, \]

\[ E[E2[t]] = E'' \cdot \sum_{i=0}^{\infty} [c''(i)]^2, \]

(3)

\[ E[E1[t] \cdot E2[t]] = E'' \cdot \sum_{i=0}^{\infty} [c'(i) \cdot c''(i)], \]
If the $E[t]$ sequence is normal,

\[ E[E_1[t], E_2[t+k]] = E'' \cdot \sum_{i=0}^{8} [c'(i) \cdot c'(i+k)], \]

\[ E[E_2[t], E_2[t+k]] = E'' \cdot \sum_{i=0}^{8} [c''(i) \cdot c''(i+k)], \]

\[ E[E_1[t], E_2[t+k]] = E'' \cdot \sum_{i=0}^{8} [c'(i) \cdot c''(i+k)], \]

\[ E[E_2[t], E_1[t+k]] = E'' \cdot \sum_{i=0}^{8} [c''(i) \cdot c'(i+k)]. \]

(5) If the $E[t]$ sequence is normal,

\[ E[E_1[t]] = 3 \cdot (E[E_1[t]])^2, \]

\[ E[E_2[t]] = 3 \cdot (E[E_2[t]])^2. \]

(6) If the $E[t]$ sequence is normal,

\[ E[E_1[t], E_2[t]] = E[E_1[t]].E[E_2[t]] + 2 \cdot (E[E_1[t], E_2[t]])^2. \]

(7) If the $E[t]$ sequence is normal,

\[ E[E_1[t], E_2[t+k]] = E[E_1[t]].E[E_2[t+k]] + 2 \cdot (E[E_1[t], E_2[t+k]])^2, \]

\[ E[E_2[t], E_1[t+k]] = E[E_2[t]].E[E_1[t+k]] + 2 \cdot (E[E_2[t], E_1[t+k]])^2. \]
The first property is a result of the fact that the \( E[t] \) sequence has a mean of zero, and the second, third and fourth properties hold because the \( E[t] \) are i.i.d. with a variance of \( E'' \) (which implies that \( E[E[t], E[t-i]] = 0 \) for \( i \) not equal to 0). The fifth, sixth and seventh properties are well-known properties of bivariate normal random variables (see, for example, Kendall and Stuart, Volume 1, page 83), and if the \( E[t] \) sequence is normal, then for fixed \( t \), \( E1[t] \) and \( E2[t] \) are bivariate normal random variables.

For notational convenience in the following presentation, define, for \( i, j = 1, 2 \),

\[
M_i(n) = E[E_i[t]],
\]

\[
M_{ij}(n, m) = E[E_i[t], E_j[t]],
\]

\[
M_{ij}(k; n, m) = E[E_i[t], E_j[t+k]],
\]

Thus, for example,

\[
M1(2) = E[E1[t]],
\]

\[
M22(2, 2) = E[E2[t], E2[t]] = E[E2[t]],
\]

\[
M12(k; 2, 2) = E[E1[t], E2[t+k]].
\]
With these conventions, the formulas for $Wz'$ and $Wz''$ derived by Bagshaw and Johnson (assuming normality) are given by

$$Wz' = M1(2) - M2(2).$$

$$Wz'' = 2 \left( \lambda \frac{(M1(2))}{2} + (M2(2)) - 2 \frac{(M2(1, 1))}{2} \right. + 2 \sum_{k=1}^{\infty} \frac{(M1(k; 1, 1))}{2} + (M2(k; 1, 1)) \right. - \left( (M1(k; 1, 1)) - (M2(k; 1, 1)) \right).$$

In the special case that model 1 is correct (i.e. $C' = 0$), these equations reduce to

$$Wz' = E' - M2(2),$$

$$Wz'' = 2 \left( \lambda - 2 \sum_{k=1}^{\infty} (M2(k; 1, 1)) \right).$$

If model 2 is correct, the corresponding equations are

$$Wz' = M1(2) - E',$$

$$Wz'' = 2 \left( \lambda - 2 \sum_{k=1}^{\infty} (M2(k; 1, 1)) \right).$$

Since $M1(2) \geq E'$ and $M2(2) \geq E'$, it is clear from the above equations that $Wz'$ will be negative when model 1 is correct and positive when model 2 is
correct. It is also clear from the above equations that $Wz''$ depends on $Wz'$. In the degenerate case when both model 1 and model 2 are correct, then obviously $Wz' = Wz'' = 0$.

Before concluding this section, we should point out that the Wiener process approximation obtained by Bagshaw and Johnson holds for any white-noise sequence $E[t]$ with a finite fourth moment and $E' = 0$, $E'' > 0$. Also, it may be possible to improve the performance of the scheme by subtracting a reference value from $Sz[n]$. However, since the variance parameter of the process changes when the mean parameter changes, the optimal reference value is not necessarily $k = (Wz'(1) + Wz'(2))/2$. 
SECTION 1 of this chapter contains an analysis of some of the results of Bagshaw and Johnson. We show how the approach which they adopted in their 1977 paper can be applied to the problem of detecting an increase in the mean of an ARMA process, and we establish a number of theoretical properties of the first passage time distribution of a Wiener process. In the next section we look at the work of Page, Brook and Evans, and Ewan and Kemp. The similarities between their apparently different approaches are pointed out, and we demonstrate how their asymptotic approximations to the run-length distribution of a CUSUM scheme are inter-related. A new approximation is derived for the run-length distribution of a CUSUM scheme for the case in which the observations are autoregressive and the process is out of control. In the last section we discuss some computational and numerical problems and the results of a Monte Carlo goodness-of-fit study of the Wiener process approximation.
SECTION 4.1: ANALYSIS OF BAGSHAW-JOHNSON RESULTS

In this section we analyze some of the results established by Bagshaw and Johnson in their 1975a and 1977 papers. The analysis is relatively simple and straightforward, and may be useful to someone who wishes to apply the methods or study them theoretically.

Part 1: An Equivalent CUSUM Scheme for Detecting a Change in the Mean of an ARMA Process

It is possible to modify the method described in Bagshaw and Johnson (1977) to develop an CUSUM scheme for detecting a change in the mean of an ARMA process which is equivalent to the scheme which they advanced in their 1975a paper. The method proposed by Bagshaw and Johnson in their 1975a paper is to apply Page's rule 1 CUSUM procedure and to approximate its run-length distribution and average run-length with the corresponding first passage time distribution and expected first passage time of a Wiener process approximation to the cumulative sum

\[ S[n] = (X[1] - k) + \ldots + (X[n] - k). \]

On the other hand, in their 1977 paper, in which
they studied the problem of detecting a change in an autoregressive or moving average parameter of an ARMA process, they proposed that Page's rule 1 CUSUM procedure should be applied to

\[ S_z[n] = Z[1] + \ldots + Z[n], \]

where

\[ Z[t] = E_1[t] - E_2[t] \]

and \( E_1[t], E_2[t] \) are the one-step-ahead forecast errors from models 1 and 2, respectively. It is natural to wonder, therefore, whether the approach adopted in the 1977 paper can be applied to the problem attacked in the 1975a paper. We shall show that this can be done, and that in fact the resulting schemes are equivalent.

If we are concerned with detecting an increase in the mean of an ARMA process, and if \( m'(1) \) and \( m'(2) \) are the assumed values of \( m' \) under models 1 and 2, respectively \((m'(2) > m'(1))\), then \( m'(1) \) and \( m'(2) \) are two simple one-step-ahead forecasts of \( X[t] \) for all \( t \). Since \( E[X[t]] = m'(1) \) if \( m' = m'(1) \) and \( E[X[t]] = m'(2) \) if \( m' = m'(2) \), the forecasts are unbiased if the corresponding models are correct. The associated one-step-ahead forecast errors at time \( t \) are \( X[t] - m'(1) \)
and $X(t) - m'(2)$. This suggests that a simple method of detecting an increase in the mean of an ARMA process is to apply Page's rule 1 CUSUM procedure to the cumulative sum

$$SSD[n] = SD[1] + \ldots + SD[n],$$

where

$$SD[t] = (X[t] - m'(1))^2 - (X[t] - m'(2))^2.$$

If we define

$$SS1[n] = (X[1] - m'(1))^2 + \ldots + (X[n] - m'(1))^2$$

$$SS2[n] = (X[1] - m'(2))^2 + \ldots + (X[n] - m'(2))^2$$

then $SSD[n]$ can be re-expressed as

$$SSD[n] = SS1[n] - SS2[n].$$

Thus, $SSD[n]$ is just the difference between the sum-of-squares under models 1 and 2, respectively (hence $SS1$, $SS2$ and $SSD$), and consequently $SSD[n]$ is proportional to the difference between the corresponding sample variances. This is an intuitively appealing result, as it is normally considered
desirable to choose the model which produces the smallest sample variance. Moreover, if model 1 is the correct model, \( E[SD[t]] < 0 \), and if model 2 is correct, \( E[SD[t]] > 0 \), so that the sample path of \( SSD[n] \) will behave in the manner suggested by Page.

It is fairly easy to show that \( SSD[n] \) converges to a Wiener process, and to determine the mean and variance parameters of this approximation. If we expand \( SD[t] \) and collect powers of \( X[t] \), the squared terms drop out and we get

\[
SD[t] = d^2 + 2.d.(X[t] - m'(2)) = -d^2 + 2.d.(X[t] - m'(1)) = 2.d.(X[t] - k')
\]

where

\[
d = m'(2) - m'(1)
\]

and

\[
k' = (m'(1) + m'(2))/2.
\]

Since \( SD[t] \) is a linear function of \( X[t] \), and Bagshaw and Johnson showed that \( S[n] \) converges to a Wiener process, therefore \( SSD[n] \) must also converge to a Wiener process. The mean parameter of the Wiener
process approximation is

\[ W'(SD) = E[SD(t)] = 2.d.W' \]

and the variance parameter is

\[ W''(SD) = \text{Var}[SD(t)] + 2 \sum_{k=1}^{\infty} \text{Cov}[SD(t), SD(t+k)] = (2.d^2)W'' \]

where \( W' \) and \( W'' \) are the mean and variance parameters of the Wiener process approximation to \( S[n] \). Thus, except for a scale factor, the SSD[n] procedure is identical to the \( S[n] \) procedure with \( k = k' \).

If model 1 is correct,

\[ E[SD(t)] = -d^2 \]

and if model 2 is correct,

\[ E[SD(t)] = d^2. \]

It follows from the above two facts, and the result of Bagshaw and Johnson (1975c), that the asymptotically optimum reference value for this CUSUM scheme is
\[ k = \left[ d^2 - \frac{2}{d^2} \right] / 2 = 0. \]

In other words, for any choice of \( m'(1) \) and \( m'(2) \), this scheme is asymptotically optimal, in the sense of minimizing the expected first passage time under the hypothesis \( W'(\text{SD}) = d^2 \), subject to a fixed expected first passage time under the hypothesis \( W'(\text{SD}) = -d^2 \).

We see, therefore, that the Bagshaw-Johnson CUSUM statistic, \( S[n] \), is optimal iff it is equivalent to \( \text{SSD}[n] \).

**Part 2: The Asymptotic Average Run-length of a CUSUM Scheme and Bounds for the Expected First Passage Time of a Wiener Process**

As stated in chapter 2, Ewan and Kemp (1960) found that if \( m' > 0 \), a rough approximation to the average run-length of a CUSUM scheme is given by

\[ N' = 1 + h/m', \quad \text{(approximately)} \]

They did not attempt to justify this assertion, but it is not difficult to do so in a heuristic manner.

If each observation had the value \( m' \), then the CUSUM scheme would signal at the first integer \( n \) such that \( n.m' \geq h \), and the average run-length of the procedure in this case would be \( N' = n \).
n ≥ h/m', since (h/m').m' = h, and n < 1 + h/m', since (n-1).m' < h implies n < 1 + h/m'. Thus, we must have that

h/m' ≤ N' < 1 + h/m'.

It therefore makes sense that as the standard deviation of the observations decreases relative to m', N' should converge to a point somewhere in the interval [h/m', 1+ h/m']. Also, since N ≥ 1 and run-length distributions are often highly right-skewed, it is not surprising that Ewan and Kemp found that N' is closely approximated by 1+h/m', especially for small values of h. We have found that when h is fairly large and the run-length distribution is not highly right-skewed, h/m' is often a better approximation to N' than 1+h/m'.

Note that if N were a continuous random variable rather than a discrete random variable, the restriction (n-1).m' < h would no longer be valid, and we would therefore expect N' to converge to a point in the 'interval' [h/m',h/m'] (i.e. N' -> h/m').

We now show that the expected first passage time of a Wiener process converges to h/W' as W'' -> 0, which is in accord with the above heuristic argument.

Recall that the expected first passage time,
$U'$, satisfies the equation
\[ U' = \frac{h}{W'} \cdot \left( \frac{1}{c} \right) \cdot \left[ \exp(-c) - 1 + c \right], \]
where
\[ c = 2 \cdot \frac{W'}{h/W'}. \]
If $c > 0$, then as $c \to \infty$,
\[ U' \to \frac{h}{W'}. \]
The result follows directly.

An alternative way of proving this result is to
take the limit as \( \mathbb{W}' \to 0 \) in the expression given in chapter 3 for the Laplace transform of the first passage time distribution. If this is done (it is necessary to apply L'Hopital's rule here), the transform converges to \( \exp[-h \cdot x/W'] \). The inverse of this transform is \( h/W' \), which establishes the result.

We now proceed to demonstrate how simple upper and lower bounds for the expected first passage time of the Wiener process approximation to the sequence of cumulative sums can be constructed. In the above expression for \( U' \), note that \(-1 + c\) are the first two terms in the Taylor's series expansion of \( \exp[-c] \). Therefore, using Lagrange's form for the remainder, we find for \( c < 0 \),

\[
\exp[-c] - 1 + c = \left(\frac{c^2}{2}\right) \cdot \exp[-c'] \quad (c < c' < 0)
\]

But this implies that

\[
U' = (h/W') \cdot (1/c) \cdot \left[ \exp[-c] - 1 + c \right]
\]

\[
\leq \frac{1}{(1/c)} \cdot \left[ h \cdot \left(\frac{c^2}{2}\right) \cdot \exp[-c] \right]
\]

\[
\leq \left(\frac{c^2}{2}\right) \cdot \exp[-c].
\]

Also, substituting \( c' = 0 \) in the above Taylor series we obtain
\[ \exp[-c] - 1 + c \geq \left(\frac{c}{2}\right)^2, \]

which implies

\[ u' \geq \left[\frac{h}{w'}\right] \cdot \left\{1/c\right\} \cdot \left[\frac{c}{2}\right] \geq \left(\frac{h}{w''}\right)^2. \]

Combining these results we obtain

\[ \left(\frac{h}{w''}\right)^2 \leq u' \leq \left(\frac{h}{w''}\right)^2 \cdot \exp[-c] \quad (c < 0). \]

The term \( \left(\frac{h}{w''}\right)^2 \) is the expected first passage time when \( w' = 0 \), so the above inequalities show that as \( w' \to 0, u' \to \left(\frac{h}{w''}\right)^2 \) at least as fast as \( \exp[-c] \to 1 \).

An analogous argument can be used to show that when \( c > 0 \),

\[ \left(\frac{h}{w''}\right)^2 \cdot \exp[-c] \leq u' \leq \left(\frac{h}{w''}\right)^2 \quad (c > 0). \]

These bounds can in turn be used to bound the ratio of the expected first passage time, \( u'(2) \), when the process is out of control to the expected first passage time, \( u'(1) \), when it is in control. Let \( w'(1), w'(2) \) be the mean parameters of the process when it is in control and out of control, respectively, and let \( w''(1) \) and \( w''(2) \) be the corresponding variance.
parameters. Define $c(1)$ and $c(2)$ in the obvious manner. Then, since $W'(1) < 0$ and $W'(2) > 0$,

$$U'(1) \geq \left(\frac{2}{W''(1)}\right).$$

$$U'(2) \leq \left(\frac{2}{W''(2)}\right).$$

from which it follows that

$$U'(2)/U'(1) \leq W''(1)/W''(2).$$

By similar reasoning

$$U'(1) \leq \left(\frac{2}{W''(1)}\right).\exp[-c(1)].$$

$$U'(2) \geq \left(\frac{2}{W''(2)}\right).\exp[-c(2)].$$

so that

$$U'(2)/U'(1) \geq \left(\frac{W''(1)}{W''(2)}\right).\exp[-(c(2)-c(1))].$$

Taken together these results lead to the following inequalities for $U'(2)/U'(1)$:

$$W(1,2).\exp[-c(1,2)] \leq U'(2)/U'(1) \leq W(1,2),$$

where
\[ W(1,2) = W''(1)/W''(2) \]

\[ c(1,2) = c(2) - c(1) = 2.W'(2).h/W''(2) - 2.W'(1).h/W''(1) = 2.[W'(2).W''(1) - W'(1).W''(2)].h/W''(1).W''(2). \]

Putting

\[ W'(1,2) = W'(2).W''(1) - W'(1).W''(2) \]

we can re-write \( c(1,2) \) as

\[ c(1,2) = 2.W'(1,2).h/W''(1,2). \]
SECTION 4.2: ANALYSIS OF RESULTS OF PAGE, EWAN AND KEMP, AND BROOK AND EVANS

The purpose of this section is to point out the similarities in the work of Page (1954), Ewan and Kemp (1960) and Brook and Evans (1972). We also show that it is possible to extend some of their results. Specifically, we derive recursive approximations to the run-length distribution, average run-length and moment generating function of a CUSUM scheme for an autoregressive process which is out of control. Further, we discuss some of the mathematical problems involved in obtaining an approximation for the case in which the process is in control.

Part 1: Connection between the Brook-Evans Matrix Equations and the Integral Equations of Page, Ewan and Kemp

Although the approaches adopted by Page and Ewan and Kemp, on the one hand, and Brook and Evans, on the other, are apparently dissimilar, we now demonstrate that they are in fact very similar. To begin with, consider the Brook-Evans matrix equation for the average run-length of a CUSUM scheme:

\[(I - B) \cdot \hat{N}^* = 1\]
If we write down the $i$th row of this equation and re-arrange terms, we obtain

$$N'(i) = 1 + \sum_{j=0}^{m} N'(j) \cdot P(i,j)$$

Letting the number of partitions, $m \to \infty$, and the partition width, $w \to 0$, while $m\cdot w \to h$, the continuous form of the Page-Ewan-Kemp equation is obtained.

The Brook-Evans matrix equations for the run-length distribution and factorial moments of a CUSUM scheme can also be re-written as integral equations. The latter integral equation was not obtained by either Page or Ewan and Kemp. The integral equation for the $k$th factorial moment of a CUSUM scheme with initial value $s$ is given by

$$N(k; s)! = \left[ N(k; 0)! + k \cdot N(k-1; 0)! \right] \cdot P(-s)$$

Note that, due to the well-known relationships between the factorial moments and moments about zero of a distribution (e.g. Johnson and Kotz, 1969, page 19), the above equation can be used to produce an integral
equation for the moments about the origin of the run-length distribution. For example, if \( N'(i;s) \) is the \( i \)th moment about zero, for \( i = 1, 2, \ldots \), we have:

\[
\begin{align*}
N'(1;s) &= N(1;s)!, \\
N'(2;s) &= N(2;s)! + N(1;s)!, \\
N'(3;s) &= N(3;s)! + 3N(2;s)! + N(1;s)!, \\
N'(4;s) &= N(4;s)! + 6N(3;s)! + 7N(2;s)! + N(1;s)!
\end{align*}
\]

Thus, it is possible to obtain the moments about zero (and from these the moments about the mean, kurtosis, etc.) of the run-length distribution without repeatedly differentiating the expression given by Ewan and Kemp for the moment generating function.

Although the integral equation approach is more general than the matrix equation approach, we believe that both approaches are useful and should be used in conjunction with one another. The connection between the run-length distribution of a CUSUM scheme and the geometric distribution is obvious in the matrix equation formulation, but it is not so apparent in the integral equation formulation. Moreover, the techniques of theoretical and numerical linear algebra
can be applied to the problem of analyzing the properties of a CUSUM scheme using the matrix equation approach, whereas the more sophisticated techniques of the theoretical and numerical analysis of integral equations must be used in the latter approach. Since statisticians (at least this would-be statistician) are generally more familiar with the geometric distribution and linear algebra than they are with run-length distributions and integral equations, these two facts alone would seem to justify the matrix equation approach.

Part 2: Connection between the Asymptotic Results of Page, Brook and Evans, and Ewan and Kemp

It is also possible to establish a connection between the asymptotic results of Page, Brook and Evans, and Ewan and Kemp. Again, the dissimilarity between these results is more apparent than real.

In 1961 Page showed that his asymptotic expression for the run-length distribution function of a CUSUM scheme converges to the corresponding expression given by Ewan and Kemp (except for a term which is asymptotically negligible). He obtained his asymptotic approximation by repeatedly differentiating the characteristic function of the total number of observations in the Wald acceptance tests and letting
the probability of an acceptance test, $P_w$, approach 1.

We now derive a slightly modified version of Page's approximation in a completely different manner, and prove that it converges exactly to the Ewan-Kemp approximation.

Let $N_{wi}(0|0)$ be the run-length of the $i$th acceptance test in a CUSUM scheme. Then by the Wald equivalence we know that

$$N = N_{w1}(0|0) + N_{w2}(0|0) + \ldots + N_{wr}(0|0) + N_{w}(0|h).$$

As $P_w \to 1$, $N_{wi}(0|0) \to 1$ and $N_{w}(0|h) \to 1$ in probability, so that

$$N \to R + 1.$$ 

From this it follows that

$$G(n) = P[N \leq n] \to P[R \leq n-1] \quad \text{as} \quad P_w \to 1.$$
Since $R$ is a geometric random variable, this probability is easy to compute and we get

$$P[R <= n-1] = Pw \cdot (1 - Pw)^{n-1},$$

$$= 1 - Pw^{n-1} \quad (Pw \to 1).$$

Now we make use of the facts that

$$N' = Nw'/(1 - Pw) \to 1/(1 - Pw) \quad (Pw \to 1)$$

and

$$\frac{1}{x} (1 + x) \to e \quad (x \to 0)$$

(put $x = Pw - 1$), we find that

$$G(n) \to 1 - \exp[-(n-1)/N'],$$

which is exactly the expression obtained by Ewan and Kemp. Page found that

$$G(n) = 1 - \exp[-n/N'] \text{ (approximately)}. $$

This difference is unimportant, however, since as $Pw \to 1, N' \to \infty$ so that $1/N' \to 0$, and the
results are therefore asymptotically equivalent.

So we have established a connection between Page's asymptotic approximation to the run-length distribution of a CUSUM scheme and that of Evans and Kemp. It remains to tie in the asymptotic approximation of Brook and Evans.

Recall that Brook and Evans stated that for large $n$

$$G(n) = \lambda - \xi q^n$$

$$\xi' = \xi/(1 - q),$$

where $G(n) = [G(n; 0), \ldots, G(n; m)]$ is the vector of cumulative run-length probabilities corresponding to the initial states 0, 1, ..., $m$. $\xi'$ is the associated vector of average run-lengths, $q$ is the largest eigenvalue of $B$, and

$$C = \left[ \sum_{i=0}^{m} V_2(i) \right] V_1.$$

$V_1$ and $V_2$ are the positive right and left-hand eigenvectors corresponding to $q$ (normalized so that $\sum_{i=0}^{m} V_1(i) V_2(i) = 1$).

The matrix $B$ is obtained from the transition probability matrix, $A$, of the CUSUM scheme by deleting the last row and last column of $A$. Since $A$ is a
transition probability matrix, the rows of $A$ add to one, which implies that $A$ satisfies the matrix equation

$$A \mathbf{1} = \mathbf{1}.$$  

Thus, the vector $\mathbf{1}$ is an eigenvector of $A$ belonging to the eigenvalue 1. Moreover, 1 is the maximum eigenvalue of $A$, since the max-norm of $A$ is the maximum row-sum of $A$, which is equal to 1, and no eigenvalue of a matrix can be greater in modulus than any norm of a matrix (Conte and de Boor, 1972, page 172).

As the probabilities in the last column of $A$ approach zero, the row-sums of $B$ must approach 1, since $A$ is a stochastic matrix. This implies that $B$ must approach a stochastic matrix, which in turn implies that the maximum eigenvalue of $B$, which is $\lambda$, must approach 1, and $C$ must approach 1. Using these facts, and the fact that

$$\sum_{k=0}^{K} \frac{(1 + x)^k}{k!} \to e (x \to 0),$$  

(put $x = \lambda - 1$) we find that, for $i = 1, 2, ..., m$,

$$G(n; i) \to 1 - \exp[-n/N^*[i]],$$

which is asymptotically equivalent to the results
of Page and Ewan and Kemp.

Part 3: Recursive Approximations to the Run-length Distribution, Average Run-length and Moment Generating Function of an Autoregressive Process which is Out of Control

The Brook-Evans matrix equations and the Page-Ewan-Kemp integral equations are based on the assumption that the stochastic process being controlled is a white-noise process. The Bagshaw-Johnson approximations, on the other hand, are less restrictive in the sense that the process can follow an arbitrary stationary and invertible ARMA model, provided that the sequence of cumulative sums converges to a Wiener process. The question which we have asked ourselves is: Is it possible to derive integral equation approximations for an ARMA process? The answer is a qualified yes. If the ARMA process is purely autoregressive, and if the process is out of control, then it is possible to derive such approximations.

We now proceed to justify this assertion.

We begin our analysis by deriving the run-length distribution of a CUSUM scheme with positive first-order autoregressive observations. Let

\[ X[i] = a \cdot X[i-1] + E[i] \quad (i = -1, 0, 1, \ldots) \]
where, \( 0 < a < 1 \), and the \( E[i] \) are independent and identically-distributed positive random variables with distribution function \( F(e) \). Also, let

\[
S[n] = S[n-1] + X[n] \quad (n = -1, 0, 1, \ldots)
\]

\[
= 0 \quad (n < -1),
\]

and

\[
P[N = n] = P[S[n] \geq h \text{ and } S[m] < h, m < n].
\]

Finally, let

\[
R(0) = (-\infty, 0],
\]

\[
R(r) = ((r-1)w, r, w) \quad (r = 1, 2, \ldots, m),
\]

\[
R(m+1) = [h, +\infty),
\]

where \( 0 < w < h/m \) and \( R(r), r = 0, 1, \ldots, m+1 \) form a partition of the real line.

Now for any events \( A \) and \( B \), and any set of mutually exclusive and exhaustive events \( C(r), r = 0, 1, \ldots, m+1 \).
Let $s(0)$ and $s(-1)$ be two numbers in the interval $(0, h)$. Then since $S[1] \in R(r), r = 0, 1, \ldots, m+1$ form a set of mutually exclusive and exhaustive events,

$$P[N = n | S[0] = s(0), S[-1] = s(-1)] = \sum_{r=0}^{\infty} P[N = n | S[1] \in R(r), S[0] = s(0), S[-1] = s(-1)] \cdot P[S(1) \in R(r) | S[0] = s(0), S[-1] = s(-1)]$$

for all $n \geq 1$. Our task now is to express the terms on the right-hand-side of the above equation in a form which allows the run-length probabilities to be computed recursively.

It follows from the assumption that $E[i] > 0$ that

$$P[S[1] \in R(0) | S[0] = s(0), S[-1] = s(-1)] = 0.$$

Also, since $S[1] \in R(m+1)$ implies $S[1] \geq h$,

$$P[N = n | S[1] \in R(m+1), S[0] = s(0), S[-1] = s(-1)] =
\begin{align*}
&= 1 \quad (n = 1) \\
&= 0 \quad (n > 1).
\end{align*}$$

Before we can proceed any further we need to establish the following fact:
for all \( m < n \). This fact is easily established by induction. For \( m = n - 1 \) we know that

\[
\]

and that

\[
S[n-1] = S[n-2] + X[n-1].
\]

Therefore \( a \cdot (S[n-1] - S[n-2]) = a \cdot X[n-1] \), from which it follows that

\[
S[n] = S[n-1] + a \cdot (S[n-1] - S[n-2]) + E[n],
\]

which is of the required form. Suppose now that it is true for \( m = m' \); that is,

\[
S[n] = S[m'] + \left[ \sum_{i=1}^{n-m'} a^i \right] (S[m'] - S[m'-1])
\]

\[
+ \sum_{j=m'+1}^{n} a^j E[j].
\]

If we substitute \( S[m'-1] + a \cdot (S[m'-1] - S[m'-2]) + E[m'] \) for \( S[m'] \) in the above expression and re-arrange terms, collecting powers of \( a \), the result is obtained.

The important thing to notice about the above
expansion for \( S[n] \) is that, if \( S[m] \) and \( S[m-1] \) are given, then the distribution of \( S[n] \) is determined by the independent random variables \( E[m+1] + \ldots + E[n] \), and does NOT depend on \( E[i] \) for \( i < m+1 \). In other words, the sequence of cumulative sums is Markovian.

We can now calculate some probabilities. Firstly,

\[
P[N = 1 | S[0] = s(0), S[-1] = s(-1)]
\]

\[= P[S[1] >= h | S[0] = s(0), S[-1] = s(-1)]
\]

\[= P[S[0] + a.(S[0] - S[-1]) + E[1] >= h | S[0] = s(0), S[-1] = s(-1)]
\]

\[= P[S(0) + a.(S(0) - s(-1)) + E[1] >= h]
\]

\[= 1 - F(h - s),
\]

where \( s = s(0) + a.(s(0) - s(-1)) \). Secondly,

\[
P[S[1] \in R(r) | S[0] = s(0), S[-1] = s(-1)]
\]

\[= P[S[0] + a.(S[0] - S[-1]) + E[1] \in R(r) | S[0] = s(0), S[-1] = s(-1)]
\]

\[= P[S(0) + a.(S(0) - s(-1)) + E[1] \in R(r)]
\]

\[= P(r,w - s) - P((r-1),w - s).
\]

Our next step is to show that

\[
P[N = n | S[1] \in R(r), S[0] = s(0), S[-1] = s(-1)]
\]

\[= P[N = n-1 | S[0] \in R(r), S[-1] = s(0)].
\]
By definition, $N = n$ implies that

$S[n] \geq h, S[m] < h, m < n$.

Applying the recursion relationship between the cumulative sums to $S[n]$ and $S[m]$, we obtain

$$S[n] = S[1] + \sum_{i=1}^{n-1} a_i \cdot (S[1] - S[0]) + \sum_{j=2}^{n-1} \sum_{k=0}^{L} a_j \cdot b_i [j]$$

$$S[m] = S[1] + \sum_{i=1}^{m-1} a_i \cdot (S[1] - S[0]) + \sum_{j=2}^{m-1} \sum_{k=0}^{L} a_j \cdot b_i [j].$$

By the same reasoning, $N = n-1$ implies that

$S[n-1] \geq h, S[m-1] < h, m-1 < n-1$

and (using $j'$ instead of $j$ as a dummy variable)

$$S[n-1] = S[0] + \sum_{i=1}^{n-1} a_i \cdot (S[0] - S[-1]) + \sum_{j'=1}^{n-1} \sum_{k=0}^{L} a_j \cdot b_i [j']$$

$$S[m-1] = S[0] + \sum_{i=1}^{m-1} a_i \cdot (S[0] - S[-1]) + \sum_{j'=1}^{m-1} \sum_{k=0}^{L} a_j \cdot b_i [j'].$$

Let $x$ be an arbitrary element of $R(\mathcal{I})$. In the expressions for $S[n]$ and $S[m]$, substitute $S[1] = x.$
\[ S[0] = s(0), \] and in the expressions for \( S[n-1] \) and \( S[m-1] \) substitute \( S[0] = x, S[-1] = s(0), \) and put \( j = j' + 1. \) The resulting expressions are

\[
S[n] = x + \sum_{i=1}^{n} a_i (x - s(0)) + \sum_{j=2}^{n-1} \sum_{k=0}^{n-k} E[j] \\
S[m] = x + \sum_{i=1}^{m} a_i (x - s(0)) + \sum_{j=2}^{m} \sum_{k=0}^{m-j} E[j] \\
S[n-1] = x + \sum_{i=1}^{n-1} a_i (x - s(0)) + \sum_{j=2}^{n-1} \sum_{k=0}^{n-1-k} E[j] \\
S[m-1] = x + \sum_{i=1}^{m-1} a_i (x - s(0)) + \sum_{j=2}^{m-1} \sum_{k=0}^{m-1-j} E[j].
\]

We see that the expressions for \( S[n] \) and \( S[m], \) \( m < n, \) given \( S[1] = x, S[0] = s(0), S[-1] = s(-1), \) are identical to the expressions for \( S[n-1] \) and \( S[m-1], \) \( m-1 < n-1, \) given \( S[0] = x, S[-1] = s(0). \) Since this is true for each fixed \( x \in \mathbb{R}(r), \) the events must be equivalent, and the corresponding probabilities must therefore be equal.

Combining the above results we find that

\[
P[N = n | S[0] = s(0), S[-1] = s(-1)] = 1 - P(h - s)
\]

for \( n = 1 \) and

\[
= \sum_{r=1}^{m} [P(r, v - s) - P((r-1), v - s)]
\]
\[ P[N = n-1| S[0] \in R(r), S[-1] = s(0)] \]

for \( n > 1 \), where \( s = s(0) + a_1 (s(0) - s(-1)) \). Letting \( w \to 0, m \to \infty \) and \( m \cdot w \to h \), we find that

\[ P[N = n| S[0] = s(0), S[-1] = s(-1)] = 1 - F(h - s) \]

for \( n = 1 \) and

\[ h \int_0^h P[N = n-1| S[0] = u, S[-1] = s(0)] \cdot dF(u-s) \]

for \( n > 1 \).

We have therefore derived a recursion relationship which can be used to compute the run-length distribution of a CUSUM scheme with positive first-order autoregressive observations. Note that it is of the same form as the Page-Ewan-Kemp recursion relationship for the run-length distribution of a CUSUM scheme with independent random variables, except that our probabilities are conditioned on the events \( S[0] = s(0), S[-1] = s(-1) \), while their probabilities are conditioned on the event \( S[0] = s(0) \). Also, the term corresponding to the state \( R(0) \) is missing in our equation, since we have assumed \( E[i] > 0 \). It is clear that we could proceed in this manner to obtain a recursion relationship for an arbitrary positive
autoregressive process. For example, for a
second-order positive autoregressive process we would
condition on the events $S[0] = s(0)$, $S[-1] = s(-1)$,
$S[-2] = s(-2)$.

Having obtained an integral equation for the
run-length distribution, it is now possible to obtain
an integral equation for the moment generating
function, $M'(t; s(0), s(-1))$, and the average run-length,
$N'(s(0), s(-1))$, of the scheme. If
$P[N=n|S[0]=s(0), S[-1]=s(-1)]$ is multiplied by $\exp[n.t]$
and summed over $n$, the following integral equation for
$M'(t; s(0), s(-1))$ is obtained:

$$
M'(t; s(0), s(-1)) = \exp[t].[ (1 - P(h-s)) \\
+ \int_{0}^{h} M'(t; u, s(0)) . dP(u-s) ]
$$

If this integral equation is differentiated with
respect to $t$ and $t$ is set equal to zero, we find that
$N'(s(0), s(-1))$ satisfies the integral equation

$$
N'(s(0), s(-1)) = 1 + \int_{0}^{h} N'(u, s(0)) . dP(u-s).
$$

We now consider some of the assumptions in the
above analysis. In general, when the process is in
control, the observations will not be strictly positive
and the above-derived recursion relationship cannot be
expected to hold, even approximately. On the other
hand, when the process has gone out of control, the mean of the process will be positive and hence there will be a relatively high probability that the observations will be positive, so in this case it is reasonable to expect that the recursion relationship will generate probabilities which are close to the true values. Actually, as long as the sequence of cumulative sums remains positive (i.e., does not hit the reflecting barrier at zero) we expect that the above approximation will be fairly good.

The reason that we excluded the region \( R(0) \) from the range of the cumulative sums is that, without this restriction, we found that we could not uniquely recover the \( X \)-values, given the \( S \)-values. There are in general an infinite number of values of \( X[n] \) which map \( S[n] \) into zero (e.g. any \( X[n] \) such that \( S[n-1] + X[n] \leq 0 \)) so that even if we were given the entire sample path of the cumulative sums, we could not uniquely determine \( X[n] \) if \( S[n] = 0 \). Since the crux of the above derivation is the relationship between \( S[n], S[m] \) and \( S[m-1] \), and since \( S[m] \) and \( S[m-1] \) cannot be used to determine \( X[n] \) if \( S[n'] = 0 \) for some \( n \geq n' \geq m \), we therefore were forced to exclude the possibility that \( S[n'] \notin R(0) \) for all \( n' \). We restricted our considerations to purely autoregressive processes, rather than moving average or ARMA processes, because, with or without the restriction that \( S[n] > 0 \), the
cumulative sums of the latter processes are not Markovian, and hence we cannot expect to obtain a simple recursion relationship between the run-length probabilities (at least not in the above manner).
SECTION 4.3: SOME PRACTICAL PROBLEMS ASSOCIATED WITH THE COMPUTATION OF THE RUN-LENGTH DISTRIBUTION OF A CUSUM SCHEME

So far in this thesis we have considered the problem of determining the run-length distribution of a CUSUM control procedure from a purely theoretical point of view. We now consider some of the practical problems which arise in this context. Specifically, we look at the numerical and computational difficulties which are associated with the techniques we have discussed, and we perform a Monte Carlo study to gauge the goodness-of-fit of the Bagshaw-Johnson approximation.

Part 1: Numerical and Computational Difficulties

We have discovered (the hard way) that there are some rather severe numerical and computational difficulties associated with all the approximation techniques discussed in this thesis. Although some of these difficulties are obvious, others are more subtle and so we feel that it would be useful to practitioners if we discussed them.

We begin our discussion by analyzing the formula for the Wiener process approximation to the run-length distribution given by Bagshaw and Johnson (1975a). It
is unnecessary for our purposes to reproduce the formula here; rather, we simply point out that it is of the form:

\[ u(t;0) = \exp[c/2].f(q(1), q(2), \ldots) \]

where \( c = 2.W'.h/W'' \) and the values of the function \( f(.) \) are obtained by evaluating an infinite series. The coefficients of the infinite series, \( q(1), q(2), \ldots \) are the positive solutions of

\[ \tan(q.h) = -q.W''/W'. \]

It can be seen from the above equations that if \( W' > 0 \) and \( W'/W'' \) increases, \( \exp[c/2] \) increases as an exponential function of \( c \), and since \( u(t;0) < 1 \), \( f(.) \) must also decrease as an exponential function of \( c \). This is an unfortunate circumstance, as it leads to numerical instability, as we shall now show.

If we make the substitution \( x = q.h \), the non-linear equation for \( q \) reduces to

\[ \tan(x) = -x/c, \]

and the roots of this equation are the intersections of the straight line \( y = -x/c \) with the curves \( y = \tan(x) \). By inspecting the graph of the set
of roots to this equation it becomes clear that
the roots approach multiples of $\pi/2$, which are singular
points of $\tan(x)$. Also, a small change in the value of
c results in a large change in the solution set. In
other words, the solutions of the non-linear equation
$\tan(x) = -x/c$ are unstable. This implies that it is
not easy to evaluate the infinite series accurately,
even with a high precision computational routine.

So now suppose that, instead of computing $f(.)$, we
compute $f(.) + e$, where $e$ is an error which is not
necessarily small. Then instead of computing $u(t;0)$,
we actually compute

$$u'(t;0) = \exp[c/2].(f(.)+e) = u(t;0) + \exp[c/2].e.$$  

If $c$ is large (i.e., $W' > 0$ and $W'/W''$ large) then
even if $e$ is small, the product $\exp[c/2].e$ is not
necessarily small. In fact, suppose that $W' = .5$, $W''$
= .25 and $h = 15$ (we actually had to compute $u(t;0)$
with these parameter values). Then $c = 60$ and $\exp[c/2]$
is on the order of 10, so that in order to have
$u'(t;0) < 1$ it is necessary for $e$ to be on the order of
$10^{-4}$, which is close to the maximum attainable
precision in double-precision in Fortran. Needless to
say, with values of $W' > .5$ and the same values of $W''$
and $h$ the situation rapidly deteriorates.

An added difficulty is that the run-length
distributions produced by the Wiener process
approximation (and, incidentally, the other
approximations as well) have high-order contact at
zero, which implies that even for moderate values of $c$
the error caused by the extra term $\exp[c/2]e$ can be
appreciable.

Since the coefficients of the infinite series must
be computed accurately by an iterative technique, and
since in many cases the infinite series converges
slowly (as pointed out by Sweet and Harden, 1970), the
evaluation of the run-length distribution can become
computationally quite expensive. Bagshaw and Johnson
(1977) found it necessary to resort to an extended
Newton-Raphson method in order to evaluate the roots of
the non-linear equation, as even 50 iterations of the
unextended Newton-Raphson method did not yield enough
accuracy. We found that a simple fixed-point iteration
method obtained reasonable accuracy without too much
computational effort, but that it was inefficient in
obtaining high-order accuracy.

Unfortunately, the Brook-Evans matrix equation and
the Page-Ewan-Kemp integral equation (and even more
unfortunately, our integral equation) involve
similar numerical and computational problems.

Consider first the Brook-Evans matrix equation.
In order to approximate the probability
$P[N[i] = n]$ it is necessary to multiply an $(m+1 \times m+1)$
matrix \( n \) times, where \( m \) is the number of sub-intervals into which the interval \((0, h)\) is partitioned. Each matrix multiplication takes on the order of \((m+1)^3\) arithmetic operations, so that the computation of \(P[N[i] = n]\) takes on the order of \(n(m+1)^3\) operations.

If the state-space of the observations is continuous, then \( m \) will have to be fairly large in order to approximate the run-length probabilities with sufficient accuracy. As the number of arithmetic operations required to compute the probabilities increases as a cubic function of \( m \), it is therefore computationally expensive to achieve high-order accuracy. But if high-order accuracy is not achieved and maintained, then the round-off errors generated by the computations will quickly build up as \( n \) increases and lower the precision of the upper-tail probabilities. Moreover, since the initial probabilities are usually very small (due to high-order contact at zero), and since in most cases the order of magnitude of the probabilities changes very rapidly near the mode, small errors in the initial probabilities can lead to very large errors in subsequent probabilities. Thus, the computation of the probabilities is numerically unstable as well as computationally expensive.

The same statement is true of the recursive integral equation of Page-Ewan-Kemp and our own
equation. Since approximations to the solutions of the integral equations can be obtained by converting them to matrix equations, many of the computational and numerical difficulties involved with the solution of the matrix equations are also involved with the solution of the integral equations. In particular, the problems posed by the computational expense of achieving high-order accuracy, the build-up of round-off errors and the numerical instability of the solutions are all present.

One might expect that efficient numerical integration methods could be used to relieve some of the computational effort, and although this is probably true in many cases, it is not necessarily true in all cases. In one case, for example, we found that the trapezoidal rule out-performed Simpson's rule, which is usually more efficient, because a large portion of the area under the curves was concentrated near the end-points of \((0,h)\), and the trapezoidal rule assigns more weight to the end-points than does Simpson's rule.

In light of the above analysis it might seem that the best way to approximate the run-length distribution of a CUSUM scheme is by Monte Carlo simulation. This is in fact true in many cases, but it is false in many others. As we shall see in the next section, the run-length distributions of CUSUM control procedures often have long, heavy right-hand tails, and
it is computationally expensive to approximate run-length distributions of this type using Monte Carlo methods. Suppose, for example, that there is a probability of p percent that a run-length will exceed the positive integer n. At least n observations must be simulated in order to generate just one run-length which is greater than n, and this process must be repeated a fairly large number of times to accurately estimate the upper-tail probabilities. Let us assume that, in order to achieve the desired level of accuracy, we need to generate at least m run-lengths which exceed n. Then we require MORE THAN m.n observations to approximate the upper-tail probabilities, and this number will be only p percent of the total number of observations required to approximate the run-length distribution. For run-length distributions with long, heavy tails, p and m will be large even for fairly large values of n, and a considerable computational effort will therefore be required to adequately approximate the distribution.

Part 2: Goodness-of-Fit of the Bagshaw-Johnson Wiener Process Approximation

One major problem with the Wiener process approximation is that the first passage time distribution of the Wiener process is often
stochastically larger than the run-length distribution of the CUSUM scheme. The reason for this is that a Wiener process is a continuous process, and it will therefore tend to signal at the same time or before a discrete process with the same mean, variance and absorbing and reflecting barriers. Bagshaw and Johnson (1977) suggested that the critical value, \( h \), should be 'increased somewhat' to improve the approximation, but they did not suggest a method of determining how much \( h \) should be increased. We have not been able to find a general method of determining an appropriate value of \( h \). In fact, we have found that under certain circumstances the first passage time distribution will be stochastically smaller than the corresponding run-length distribution, so that an increase in the value of \( h \) would make the approximation worse, not better.

The fact that the variance of the Wiener process approximation to \( S_z[n] \) changes when the mean changes also presents a practical problem. Due to a simultaneous change in both the mean and variance of the Wiener process, it can occur that the expected first passage time actually increases rather than decreases when the process goes out of control (i.e., switches from model 1 to model 2). This is certainly an unappealing result, since we therefore have the anomaly that it takes
longer, on average, for the process to signal when it has gone out of control that when it has remained in control. Although this result is unappealing, it should not be too surprising, as the upper-bound that we derived for $U'(2)/U'(1)$ was of the form $W'(1)/W'(2)$, so that if $W'(1)/W'(2) > 1$, then there is a non-zero probability that $U'(2) > U'(1)$. Unfortunately, we do not see how this problem can be avoided, as the variance of an ARMA process necessarily changes if one of the autoregressive or moving average parameters change, as it is a function of these parameters. Bagshaw and Johnson (1977) suggested that it might be possible to exploit the dependence of the variance of an ARMA process on the autoregressive and moving average parameters by developing a scheme which signals whenever there is a significant increase or decrease in the variance of the process. This seems like an excellent idea, but we have not had a chance to investigate it thoroughly.

We will now discuss the results of our Monte Carlo goodness-of-fit study. For the sake of continuity we started with many of the same parameter values as used by Bagshaw and Johnson, and then we changed some of them to study the effect of the changes. We concentrated our efforts on the CUSUM schemes corresponding to the statistic $S[n]$. Since the Wiener process approximations to $S[n]$ and $Sz[n]$ differ by only
a change in the mean and variance parameters, we expect that the conclusions which we make concerning the $S[n]$ CUSUM schemes will hold for the $Sz[n]$ CUSUM schemes.

Figures 4.3.1 - 4.3.5 are graphs of the cumulative run-length distributions and Wiener process approximations of five CUSUM schemes with first-order autoregressive observations. The schemes have common autoregressive parameters, variance parameters and critical values of $a = 0.5$, $W' = 4.0$ and $h = 17.32$, respectively. The mean parameters range from 0.7 to -0.1 by increments of 0.2. The cumulative run-length distributions generated by Monte Carlo simulation are denoted by $M[N <= n]$, and plotted with the symbol '.', and the Wiener process approximations are denoted by $P[W <= n]$ and plotted with the symbol 'x'. The letters 'AARL' stand for Asymptotic Average Run-Length (i.e. expected first passage time), and the letters 'ARL' stand for Average Run-Length. Baghsaw and Johnson (1975a, figure 5), considered the case corresponding to figure 4.3.2.

It is clear that the Wiener process approximations are stochastically larger than the simulated run-length distributions. It is also clear that the approximations become progressively worse as $W'$ decreases, in the sense that the absolute errors incurred by using the percentage points of the first
passage time distributions to approximate the corresponding percentage points of the run-length distributions increase markedly as $W'$ decreases. For example, with $W' = 0.7$, the difference between the true and estimated medians is less than $25 - 17 = 8$, but with $W' = -0.1$, the difference is greater than $135 - 76 = 59$. Also, with $W' = 0.7$ the difference between the true and estimated average run-lengths is only $26.19 - 20.67 = 5.52$, but with $W' = -0.1$, the difference is $181.85 - 102.28 = 79.57$. We have found a similar relationship between the value of $W'$ and the fit of the approximations in all of our simulations.

One reason for this relationship between the value of $W'$ and the goodness-of-fit of the approximations is that as $W'$ increases, the first passage time distributions and the run-length distributions converge to peaked, short-tailed distributions with similar mean and variance parameters in the interval $[0, 1]$, so that the absolute differences in their percentage points necessarily decrease. Another reason is that as $W'$ decreases, the tails of the run-length distributions become much heavier than the tails of the first passage time distributions, which tends to inflate the values of the average run-lengths relative to the expected first passage times.

Since $S[n]$ converges to a Wiener process as $n$ increases, and since the average run-length of a
CUSUM scheme increases as \( W' \) decreases, the run-length distributions must converge to the first passage time distributions as \( W' \to -\infty \). However, the rate of convergence appears to be very slow.

We next considered the effect of the value of the critical value on the goodness-of-fit of the approximation. Figures 4.3.6 - 4.3.8 have the same values of \( a, W' \) and \( W'' \) as figure 4.3.2 (the case considered by Bagshaw and Johnson), but their critical values are 29, 54 and 156, respectively. Note that the Wiener process approximation becomes appreciably better, in the sense defined above, as \( h \) increases. Of course this is to be expected, as the approximation should improve as the average run-length increases, and the average run-length increases as \( h \) increases. The results of all our simulations support the conclusion that the goodness-of-fit of the Wiener process approximation improves as \( h \) increases.

It is interesting to note that there appears to be a fixed bias of about -8.0 in the asymptotic average run-lengths of figures 4.3.2, 4.3.6 and 4.3.7 (the bias in figure 4.3.8 is slightly less, which is understandable, as the bias must approach zero as \( h \to \infty \)). Even more interesting is the fact that the formula for the asymptotic average run-length of the
Wiener process approximation is of the form

\[ U' = \frac{W''}{(2W')^2} \left[ \exp[-c] - 1 + c \right], \]

and with \( W' = 0.5 \) and \( W'' = 4 \) we have

\[ \frac{W''}{(2W')^2} = 8.0. \]

Thus, if we ignore the term '-1' in the equation for \( U' \), we obtain almost an exact fit between the AARL's and the ARL's in figures 4.3.2, 4.3.6 and 4.3.7.

Figure 4.3.9 illustrates the effect of a change in the variance parameter on the goodness-of-fit of the approximation. The values of \( W' \) and \( h' \) are the same in figures 4.3.2 and 4.3.9, but the value of \( a \) is 0.5 in the former figure and -0.5 in the latter figure, which changes the corresponding variance parameter from 4.0 to 0.4444... Obviously, the decrease in the variance parameter from 4.0 to 0.4444... improves the goodness-of-fit of the approximation dramatically. When \( W' \) is greater than zero and is fairly large (say \( \geq 0.5 \)) we have found that this is always the case.

We suspect that the reason that this is true is that as \( W' \to \infty \) and \( W'' \to 0 \), both the first passage time distributions and the run-length
distributions rapidly converge to peaked, short-tailed distributions with similar mean and variance parameters in the interval \([h/W', 1 + h/W']\).

The empirical formula

\[
ARL = 1 + h/m' \quad (m' > 0),
\]

which was suggested by Ewan and Kemp (1960), seems to support this argument. Also, our heuristic justification of the Ewan-Kemp formula, given in section 1 of this chapter, and our proof that the expected first passage time of a Wiener process converges to \(h/W'\), seems to lend further support to this argument. In addition, the interval \([h/W', 1 + h/W']\) converges to the interval \([0, 1]\) as \(W' \to \infty\), and, as stated above, the first passage time distributions and run-length distributions converge to peaked, short-tailed distributions with the similar mean and variance parameters in this latter interval as \(W' \to \infty\).

As a final point, we present the following short list of AARL's and ARL's, with the corresponding values of \(h/W'\) and \(1 + h/W'\) (the values of \(a, W''\) and \(h\) are \(-0.5, 0.4444...\) and 17.32, respectively):
We feel that this evidence, taken together, strongly supports our argument.

The purpose of figure 4.3.10 is simply to show that under certain circumstances run-length distributions can be stochastically larger than first passage time distributions.

If we consider the results of the above simulations as a whole, we come to the conclusion that the goodness-of-fit of the Wiener process approximations improve as

\[ c \to \infty \]

and \( c \to -\infty \),

where

\[ c = 2W'.h/W'' \]

Note that \( c \) is the constant that appears in \( W' \).
CONCLUSIONS:

Based on our analysis we conclude:

1.) There is no such thing as an all-purpose sequential control procedure. The sequential control procedure which is appropriate for a particular application will depend on the unique characteristics associated with that application and may have to be specially designed.

2.) At present there is no single method of generating or approximating the run-length distribution of a CUSUM statistic which is superior to all other methods in all cases.

3.) The Wiener process approximation obtained by Bagshaw and Johnson works well in some cases (e.g. as \( c \to \infty \) and for large average run-length) but it can be quite misleading in many other cases and should be used with caution.

4.) More research is required to find methods of improving the Wiener process approximation, and to find methods of efficiently solving the matrix and integral equations.
CUMULATIVE RELATIVE FREQ.

POLYGONS

**FIGURE 4.3.1**

\[ M[N \leq n] \text{ (o)} \quad W' = 0.7 \quad \text{AARL} = 20.67 \quad \alpha = 0.5 \]

\[ p[w \leq n] \text{ (x)} \quad h = 17.32 \quad \text{ARL} = 26.19 \quad b = 0.0 \]

**RUN-LENGTH \((n)\)**
Figure 4.3.2

Cumulative Relative Freq.

\[ P[R \geq n] \quad \text{PLW}[n] \quad (x) \quad M[\text{Lx} \geq n] \quad (\cdot) \]

\[ w_1 = 0.5 \quad w_2 = 0.4 \quad h_1 = 17.32 \quad ARL = 34.72 \quad b = 0.0 \quad AARL = 26.745 \]

\[ P[R \geq n]\]
**Cumulative Relative Freq. Polygons**

**Figure 4.3.3**

- $M_{N=n}$ (•) $w' = 0.3$, $AARL = 32.165$, $a = 0.5$
- $P_{W=n}$ (x) $h = 17.32$, $ARL = 51.176$, $b = 0.0$

**Run-Length ($\eta$)**
Figure 4.3.4

- $M[N \leq n]$ (star) $w' = 0.1$ AARL = 57.32 $\alpha = 0.5$
- $\rho[N \leq n]$ (cross) $w'' = 4.0$ $h = 17.32$ ARL = 86.44 $b = 0.0$

**RUN-LENGTH ($n$)**
CUMULATIVE RELATIVE FREQ.

POLYGONS

FIGURE 4.3.5

$M[n \leq n]$ (●) $w' = -0.1$ $AARL = 102.28$ $q = 0.5$

$P[w \leq n]$ (x) $w'' = 4.0$ $h = 17.32$ $ARL = 181.85$ $b = 0.0$

RUN-LENGTH ($\eta$)
CUMULATIVE RELATIVE FREQ.

POLYGONS

FIGURE 4.3.6

\[ M[N\leq n] \quad (*) \quad w' = 0.5 \quad AARL = 50.0 \quad a = 0.5 \]
\[ P[N\leq n] \quad (x) \quad h = 29.0 \quad ARL = 58.02 \quad b = 0.0 \]
**CUMULATIVE RELATIVE FREQ.**

**POLYGONS**

**FIGURE 4.3.7**

\[ M_{N \leq n} (\ast) \quad W' = 0.5 \quad AARL = 100.0 \quad \alpha = 0.5 \]

\[ P_{W \leq n} (x) \quad h = 54.0 \quad ARL = 108.19 \quad b = 0.0 \]
CUMULATIVE RELATIVE FREQUENCY POLYGONS

FIGURE 4.3.8

\[ M[N \leq n] \ (\ast) \quad w' = 0.5 \quad AARL = 304.0 \quad \alpha = 0.5 \]

\[ P[W \leq n] \ (x) \quad w'' = 4.0 \quad ARL = 304.84 \quad b = 0.0 \]

RUN-LENGTH \ (n - 200)
**Cumulative Relative Frequency Polygons**

**Figure 4.3.9**

- \( w' = 0.5 \)
- \( AARL = 33.75 \)
- \( b = 0.5 \)
- \( w'' = 0.444 \)
- \( h = 17.32 \)
- \( ARL = 34.509 \)
- \( b = 0.0 \)
CUMULATIVE RELATIVE FREQUENCY POLYGONS

Figure 4.310

\[ M[N > n] \quad (\star) \quad w' = 0.1 \quad A_{ARL} = 155.2 \quad q = 0.0 \]

\[ p[W > n] \quad (x) \quad h = 16.77 \quad A_{ARL} = 148.86 \quad b = 0.5 \]
BIBLIOGRAPHY


Int. J. Control, 23, 647-656.


Kemp, K. W. (1967a). 'Formal expressions which can be used for the determination of the operating characteristic and average sample number of a simple sequential test.', J. R. S. S., B, 29, 248-262.


Page, E. S. (1955). 'A test for a change in a parameter occurring at an unknown point.', Biometrika, 42, 523-527.

Page, E. S. (1957). 'On problems in which a change in a parameter occurs at an unknown point.', Biometrika, 44, 248-252.


Page, E. S. (1962b). 'A modified control chart with warning lines.', Biometrika, 49, 171-175.


