Monitoring Markov Dependent Binary Observations with a Log-Likelihood Ratio Based CUSUM Control Chart

Shabnam Modarres-Mousavi

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Statistics

Committee Chair: Marion R. Reynolds, Jr.
Committee Member: Jeffrey B. Birch
Committee Member: Robert S. Schulman
Committee Member: G. Geoffrey Vining
Committee Member: William H. Woodall

January 17, 2006
Blacksburg, VA

Keywords: Monitoring a proportion, Markov chain models, CUSUM control chart, Correlated observations, Binary data.

Copyright 2006 Shabnam Modarres-Mousavi

Monitoring Markov Dependent Binary Observations with a Log-Likelihood Ratio Based CUSUM Control Chart

Shabnam Modarres-Mousavi

ABSTRACT

Our objective is to monitor the changes in a proportion with correlated binary observations. All of the published work on this subject used the first-order Markov chain model for the data. Increasing the order of dependence above one by extending a standard Markov chain model entails an exponential increase of both the number of parameters and the dimension of the transition probability matrix. In this dissertation, we develop a particular Markov chain structure, the Multilevel Model (MLM), to model the correlation between binary data. The basic idea is to assign a lower probability to observing a 1 when all previous correlated observations are 0's, and a higher probability to observing a 1 as the last observed 1 gets closer to the current observation. We refer to each of the distinct situations of observing a 1 as a "level". For a given order of dependence, t, at most $t+1$ different values of conditional probabilities of observing a 1 can be assigned. So the number of levels is always less than or equal to $t+1$. Compared to a direct extension of the first-order Markov model to higher orders, our model is considerably parsimonious. The number of parameters for the MLM is only one plus the number of levels, and the transition probability matrix is $(t+1)^2$.

We construct a CUSUM control chart for monitoring a proportion with correlated binary observations. First, we use the probability structure of a first-order Markov chain to derive a log-likelihood ratio based CUSUM control statistic. Then, we model this CUSUM statistic itself as a Markov chain, which in turn allows for designing a control chart with specified statistical properties: the Markov Binary CUSUM (MBCUSUM) chart. We generalize the MBCUSUM to account for any order of dependence between binary observations through implying MLM to the data and to our CUSUM control statistic. We verify that the MBCUSUM has a better performance than a curtailed Shewhart $p-$chart. Also, we show that except for extremely large changes in the proportion (of interest) the MBCUSUM control chart detects the changes faster than the Bernoulli CUSUM control chart, which is designed for independent observations.

Acknowledgements

I am greatly indebted to Prof. Reynolds, whose gracious guidance made this work possible. Many helpful suggestions from Professors Birch, Schulman, Vining, and Woodall enriched my research experience.

This achievement is dedicated to my beloved husband Reza. His unconditional support makes even impossible goals of mine approachable.

Shabnam Mousavi

January 2006

Table of Contents

Table of Tables

Table of Figures

1 Introduction

For all organizations, success in the competitive marketplace depends heavily on continuous improvement of quality. Quality is conceived as the fitness for use and is inversely proportional to unwanted variability. “Quality improvement is the reduction of variability in processes and products” (Montgomery, 2005). The main source of waste in money, time, and effort that is associated with lower than standard condition of products is unwanted variability, so improving quality could be achieved through reducing waste in every possible form. Statistical techniques can be used to monitor the performance of a given process, to record the quality of the product, and to warn in case of the violation of standards. The statistical methods developed to improve quality form the field of statistical quality control, and statistical process control (SPC) is one of the major areas of statistical quality control.

1.1 Control Charts for Monitoring a Proportion

If our interest is to monitor a process and to be warned when process quality has changed, we can use what is called a *control chart*. A control chart is the primary technique of SPC. A control chart is based on plots of statistics obtained from measurements of a quality characteristic in samples taken from the process against time (or sample number). Unusual variability results in the plotted statistic falling outside of the upper or lower control limits (UCL and LCL) of a control chart. Whereas in the absence of unusual variability the plotted statistic tends to fall within the control limits.

1.1.1 Binary Data

Data on quality characteristics are measured in two forms: Continuous measurements (frequently called *variables* data), and discrete or count data (frequently called *attributes* data). In some cases a continuous measurement, such as a dimension, is obtained, and the inspected items are classified into two categories of conforming to the standards (when the dimension is within specifications) or *non-defective*; and nonconforming (when the dimension is outside of specifications) or *defective*. Count data that take only two values, 0 and 1, are called *binary* data. As an example, if a light bulb is tested by plugging in to see if it works, then a binary observation is obtained.

The practice of recording the measurements of quality characteristics as binary data with the goal of improving and controlling the process is not limited to manufacturing. In health-care surveillance, the outcome can be naturally binary. For example, the results of a certain treatment are cured or not cured. Simple examples of binary data outside of manufacturing are the requests received by a customer service department that are/are not answered within the standard reply period, or the deliveries that are/are not sent to the correct address, etc. Also, binary data can be produced by coding a continuous measurement. For example, in a certain test, the reading time is recorded for the test takers. Those who completed the reading in less than 2 minutes get a zero and those who went over two minutes get a one. The transformation of observed data into binary data is referred to as *hard limiting* (Kedem, 1980). Hard limiting or "clipping" results in the loss of information, and so should only be used if justifiable by a gain. We can use a Bernoulli model for our data if we can assume that the probability of being defective for each unit is constant, that is if we monitor the production process after it has reached stability and have available to us a certain probability of observing a defective item at each inspection.

1.2 The State of the Art

Production (or otherwise) process observations have traditionally been assumed to be independent, but they may actually have autocorrelation. As an example, consider the process of monitoring a production process and labeling the inspected products as defective or nondefective. If an unusual cause emerges in the machinery or in the input material, the chances that it only affects individual items is much slimmer than the scenario in which after the cause arises a series of affected items will be produced. Also, mechanical defects usually worsen over time. Thus, observing a defective impacts the probability of observing another defective in the future. So it makes intuitive sense to account for correlation among observations. Moreover, it has been formally investigated and shown that the quality of items are often serially dependent (see Broadbent, 1958), and that the existence of correlation has an adverse effect on the performance of the monitoring tools (such as control charts) that are designed based on assuming independent data (see Deligonul and Mergen, 1987).

In the next chapter, we review the literature that clarifies the motive and sets the background for our work. There are many papers on autocorrelation in control charts, but most of these are for continuous variables. We will review the work on discrete data to see that almost all of the published work on attribute control charts for monitoring a proportion of interest, p, in a dependent process share two characteristics: using independent random variables for the control statistic; and modeling the dependence in the underlying process as a first-order Markov chain. Hidden Markov models (HMM), motivated by finding an underlying generating mechanism for data by assuming a Markov chain structure for the error, are widely used in computerized speech recognition practice. This line of work is not directly related to our current framework, because we intentionally develop a form that can be modeled as a Markov chain.

1.2.1 Control Charts for Monitoring a Proportion Based on Independent Data

The traditionally used Shewhart p-chart is based on plotting the proportion of defective items in samples of size n. What is plotted on the Shewhart chart depends only on the current sample. CUSUM charts are based on plotting a cumulative sum of past data, so CUSUM charts are advantageous over Shewhart-type charts in that they accumulate information over time. Another disadvantage of the Shewhart chart is that n must be quite large if p_0, the in-control value of p, is close to 0. Also, the number of possible values for the in-control ANSS (average number of samples to signal) is limited with the Shewhart chart, which in turn, limits the possible values for the false alarm rate.

There are two CUSUM control charts for independent binary data that are related to our work. Reynolds and Stoumbos (1999) constructed the Bernoulli CUSUM that plots the original data as they are obtained from the continuous inspection. Therefore, a decision about the state of the process can be made at any point without the need to wait until a summary statistic is computed based on the data from a complete sample. The Bernoulli CUSUM chart, which is designed for independent binary stream of observations, detects changes in p much faster than the Shewhart $p-$chart. Bourke (1991) has used run-length (number of nondefective items between successive defective items) as a basis for a CUSUM control chart to detect a shift in p in the case of 100% inspection. He assumed that the attribute measures are independent and demonstrated that his chart is more efficient than the $p-\text{chart}$. In general, when the individual inspection results are available, any control chart that needs to group items into samples is disadvantageous compared to Bernoulli CUSUM that uses individual items (n = 1). The CUSUM chart that we develop in this dissertation reduces to the Bernoulli CUSUM if the observations are independent.

1.2.2 Modeling Dependence

One of the mathematically well developed methods for modeling the dependence between data represents the dependent data as a Markov chain. In this framework, the dependence can be limited to a certain number of consecutive observations. This number is called the order of the Markov chain. For example, if the probability of a certain value for an observation is assumed to depend on three previous observations, but not the ones before that, then this is an instance of a third-order Markov chain. This assumption is referred to as the *Markov assumption*.

All past work has used a first-order Markov model for constructing control charts for correlated binary observations. Bhat and Lal (1990) determined the upper and lower control limits of a Shewhart control chart for a production process that is first-order Markov dependent. Their chart is based on the number of defective items in sequential samples taken far enough apart for the samples to be considered independent. For the case of 100% inspection, Blatterman and Champ (1992) assumed that the correlation in the data follows a first-order Markov model, and evaluate a Shewhart chart based on the number of nondefective items between defective items (Y). Champ, Blatterman, and Rigdon (1994) proposed an attribute CUSUM chart for monitoring the proportion defective, based on the same random variable, Y. They outline a Markov chain approach for determining the run- length distribution of one-sided and two-sided charts. Lai, Xie, and Govindaraju (2000) studied the problem of monitoring a high quality process (i.e., small p_0) with first-order Markov dependence. Their chart is based on a random variable $X\ (=Y+1)$ that is the number of units observed to get a defective unit.

1.3 The Objective, Scope, and Contribution of this Work

The main objective of this dissertation is to find an efficient method for monitoring changes in a proportion when the binary process observations are correlated. The data collected for monitoring changes in processes, often demonstrate dependence. Also, the presence of correlation affects the performance of monitoring tools (such as control charts) that are designed based on assuming independent data. Therefore, it is useful and desirable to develop charts that incorporate the correlation among observations.

A particular Markov chain model is constructed and used as the model for correlated binary data. When the binary data are the result of the inspection of an item, we assume that the results of the inspection are immediately available after inspection. We consider the case of a continuous stream of inspected items, which can arise when 100% inspection is used or in the case when inspection is done continuously but the production rate exceeds the inspection rate.

We develop a CUSUM control chart for monitoring correlated binary data that follows the Markov chain model. Our CUSUM chart is constructed using a CUSUM statistic derived from the log-likelihood ratio for the correlated binary observations. The CUSUM statistic can itself be modeled as a Markov chain, so this allows the CUSUM chart to be set up to have specified statistical properties.

Our model for correlated binary observations gives a simple probability structure for modeling any order of dependence among binary data as a certain Markov chain. Most of the published work on our subject has used the probability structure of a first-order Markov chain to model the dependence, and a geometric random variable to construct the control chart statistic. Our approach differs from the existing work on three accounts. The first difference is in the structure of the statistics constructed from the data. Instead of considering the number of nondefectives between two consecutive defectives, we use the individual binary observations. Second, we construct a CUSUM statistic that is directly derived from the log-likelihood ratio for Markov dependent data. Third, our model of correlated binary observations requires considerably fewer parameters for capturing a particular Markov-type dependence, compared to the direct extension of the standard Markov chain model (see Section 4.2).

In Chapter 2, we review the current literature and give definitions. In Chapter 3, a CUSUM statistic is constructed based on the log-likelihood ratio that uses the first-order Markov probability structure to model the correlated binary observations. Modeling this CUSUM chart as a Markov chain enables us to design a control chart with specific statistical properties. In Chapter 4, a general model that accounts for a special Markov dependence in the underlying process is developed and then a log-likelihood ratio based CUSUM chart is developed for this model. This CUSUM chart can itself be modeled as a Markov chain. This generalized Markov model requires only three parameters: the long term probability of observing a defective, p, the correlation coefficient, ρ, and the order of dependence, t. We discuss a method for estimating the parameters of this CUSUM, in which we employ the properties of hidden Markov models (HMM). Then we show how a general model for correlated binary observations can be developed by further extending the idea that was used to generalize our CUSUM statistic from accounting for only first-order dependence to $t^{th}-$order dependence for $t \geq 1$. In Chapter 5, we address the effect of correlation on different control charts that are designed to detect shifts in p, and compare their performance. Chapter 6 gives a summary of all previous chapters, conclusions and discussion of some future directions. In Appendix-A the relationship between our CUSUM increments and the reference value of a CUSUM chart that is based on a geometric random variable is derived.

2 Control Charts for Monitoring a Proportion, Data Structure, and Measures of Performance

2.1 Introduction

The basic methods of statistical process control (SPC) are based on the Shewhart control chart, named after their developer, Walter A. Shewhart. These charts only use the information in the last collected sample. This in turn results in a low level of sensitivity to small shifts in process parameters. Applying supplemental sensitizing rules adversely affects the main advantage of Shewhart-type charts, which is the simplicity of construction and interpretation. There are two main alternatives to the Shewhart control chart for monitoring small shifts in the process: the cumulative sum (CUSUM) and the exponentially weighted moving average (EWMA) control charts. There is not a significant difference in terms of statistical performance between these two control charts. The control charts we design for monitoring a proportion with autocorrelated binary data in the next chapters are CUSUM charts.

In this chapter, we present the concepts, definitions, and models that will be used in this dissertation. After a general introduction to each subject, we elaborate on the material that is of specific interest in the later chapters.

2.2 Data Structure and Sampling

To obtain data for constructing a control chart we take samples of items produced by the production process. Let $\mathbf{X}_k = (X_{k1}, X_{k2}, \ldots, X_{kn})$ represent sample number k, where the sample size is n. In the most general setting, there are inspection periods when observations are obtained from the process and non-inspection periods when no observation is obtained. Also, there is a positive time between individual inspections. Furthermore, the parameter of interest, say p, can change any time within either the inspection or non-inspection period. This framework allows that a shift can happen anywhere during the sampling period.

Reynolds and Stoumbos (2000) introduced a general model in which there is a positive time, say d_1, between individual items, and a time, say d_2, between samples. As an example to illustrate different sampling schemes, they consider a process where the inspection of one item takes one minute and the production rate is one item every six seconds. Thus while one item is inspected, nine items will be produced that are not included in the sample. If every eight hours (every shift), samples of $n = 100$ items are inspected, then the 100 minute inspection period is followed by a non-inspected period of $8 * 60 - 100 = 380$ minutes. The time between the start of two consecutive inspections of $n = 100$ is $c = nd_1 + d_2$, so if each item's inspection takes $d_1 = 1$ minute, then $c = 100 * 1 + 380 = 480$ minutes.

For a constant sampling rate per time unit, different sample sizes can be used. Some alternatives to $n = 100$ every 8 hours might be $n = 50$ every 4 hours or $n = 25$ every 2 hours. Using the $p-$chart for detecting small changes in p requires large n, which under a constant inspection rate results in long non-inspection periods (in our example, 380 minutes) during which large changes in p could go undetected. The performance of the $p-$chart is therefore highly influenced by the choice of n.

The design of CUSUM charts for monitoring p is not as easy as designing a CUSUM chart for the mean of a normal distribution. The choice of control limits for achieving a specified average false-alarm rate depends on the sample size (n), the in-control value of p (p_0), and the shift in p that has to be detected quickly (p_1). Hawkins and Olwell (1998) considered the problem of monitoring p in a simple model that assumes negligible sampling time. Thus, changes in p only happen between samples. They recommend choosing n based on convenience, whereas, Reynolds and Stoumbos (2000) made specific recommendations for practical applications in this case.

In many research studies it is assumed that the time between the individual observations in a sample is short enough to be neglected. That is, relative to the time between samples, the n observations in a sample are taken at the same time ($d_1 \approx 0$). If the sampling is nonstop, we have a case that is called continuous sampling. In continuous sampling, it is not necessary to inspect all items. It might be the case that inspecting an item takes longer than producing an item, but as long as there is no non-inspection period, we have continuous sampling. In the special case of 100% inspection, d_2 is equal to d_1, which is the case investigated by Reynolds and Stoumbos (1999).

In this dissertation, we are interested in a specific type of count data that takes only two values, 1 and 0. This kind of data can be generated by Bernoulli trials, however, in order to stress the fact that we allow for dependence between the observations, we use the term *binary data*. The sampling scheme we consider in our work is continuous sampling, which arises when there is 100% inspection, or when the inspection rate is slower than the production rate and the inspection is continuous. An example for the latter is as follows. Consider a case in which the production of each item takes six seconds, whereas, the inspection of an item takes one minute. In this scenario, nine items will go un-inspected while one is being inspected. In this context, what we mean by samples is groups of specified size from the data obtained by continuous inspection. The result for the i^{th} inspected item in the k^{th} sample is either defective or non-defective, and can be represented by a random variable as follows:

$$X_{ki} = \begin{cases} 1 & \text{if the } i^{th} \text{ item in sample } k \text{ is defective} \\ 0 & \text{otherwise} \end{cases}$$

for $i = 1, 2, \ldots, n$, and $k = 1, 2, \ldots$. The total number of defectives in the k^{th} sample is $T_k = \sum_{i=1}^{n} X_{ki}$.

2.3 Control Charts for Independent Observations

2.3.1 The General Shewhart Control Chart

To define a general Shewhart chart, consider the case in which X has a distribution indexed by a general parameter θ, where the objective is to detect a change from the in-control value θ_0. Recall that $\mathbf{X}_k$ denotes the sample of n items for the k^{th} sample. Consider a statistic $W_k = w(\mathbf{X}_k)$ that is an estimator of θ.

The Shewhart chart consists of a center line (CL) at the in-control value θ_0, and two control limits at a certain multiple (usually three) of the standard deviation of w_k from CL. Although a smaller multiple of the standard deviation of W_k, corresponding to narrower control limits, would increase the sensitivity to small shifts in θ, this is achieved only at the expense of an increase in the rate of false alarms. Too many false alarms must be avoided as it would destroy the confidence of personnel in the control chart program.

To operate the Shewhart control chart, W_k is computed after each sample and plotted on the chart. If a point plots outside the control limits, or if a nonrandom pattern is observed, it is most likely that θ has shifted to a new level and so the process is declared to be out-of-control.

2.3.2 The *p*-chart

Consider the problem of statistical monitoring of the parameter p when samples of n items are taken from the process. The underlying distribution of the number of defective units in the k^{th} sample, T_k, for independent Bernoulli (binary) observations has a binomial (n, p) distribution:

$$P(T_k = t) = \binom{n}{t} p^t (1-p)^{n-t}, \text{ for } t = 0,1,2,...,n,$$

where the mean and variance of T_k are np and $np(1-p)$, respectively.

The sample proportion defective is $\hat{p}_k = T_k/n$, where the distribution of $\hat{p}_k$ can be obtained from the binomial distribution. The mean and variance of $\hat{p}_k$ are $\mu_{\hat{p}} = p$ and $\sigma^2_{\hat{p}} = p(1-p)/n$, respectively. The Shewhart control chart that monitors p is called the $p-$chart. The control limits for the $p-$chart when the in-control value, p_0, is known are usually specified as

$$\text{UCL} = p_0 + 3\sqrt{\frac{p_0(1-p_0)}{n}}; \ \text{CL} = p_0; \ \text{LCL} = p_0 - 3\sqrt{\frac{p_0(1-p_0)}{n}}$$

This is called 3-sigma limits. This chart is operated by plotting $\hat{p}_k$ on the chart.

2.3.2.1 Dependence Bias in $p-$Charts

Deligonul and Mergen (1987) studied the variance of the sample proportion defective under a distribution with autocorrelation. They verified that violating the binomial assumption of independent trials may result in underestimating the width of the region between the control limits, which corresponds to a higher false alarm rate than what is warranted by the appropriate statistical model. Also, they showed that the amount of this bias is higher for a stronger serial correlation. They developed an approximation that can be used for correcting the dependence bias of $p-$charts.

2.3.3 CUSUM Control Charts

Wald (1947) introduced the sequential probability ratio tests (SPRT) in the context of hypothesis testing. Page (1954) proposed the CUSUM chart, and recognized that a one-sided CUSUM for a general parameter, say θ, is equivalent to a sequence of SPRTs. This structure has been developed, among others, by Reynolds and Stoumbos (1998) for the case of $\theta = p$ as discussed above. Consider an in-control value θ_0, and a value to be detected θ_1. For the k^{th} sample of independent observations $\mathbf{X}_k = \left(X_{k1}, X_{k2}, ..., X_{kn}\right)$, define

$$Z_k = \ln\left(f(\mathbf{X}_k \mid \theta_1) / f(\mathbf{X}_k \mid \theta_0)\right).$$

If z_{ki} is defined as $\ln f(X_{ki} \mid \theta_1) / \ln f(X_{ki} \mid \theta_0)$, then Z_k can be written as $Z_k = \sum_{i=1}^{n} z_{ki}$. On a CUSUM chart, we plot the values of $C_k = Z_k + \max\{0, C_{k-1}\}$, where $C_0 = 0$. This chart produces a signal when C_k exceeds a specified upper control limit. For practical purposes, the statistics defined above are equivalent to the traditional CUSUM statistics $C_k = \max(0, C_{k-1} + Z_k)$, $k = 1, 2, \ldots$.

2.3.3.1 The Binomial CUSUM Chart

Recall that the binomial random variable T_k denotes the total number of defectives in the k^{th} sample of n items. For detecting an increase in the proportion defective p, the general CUSUM chart based on the likelihood ratio above is equivalent to a CUSUM chart using the statistic $S_k = \max(0, S_{k-1}) + (T_k - n\gamma)$ for $k = 1, 2, ...$, where $n\gamma$ is called the reference value. The parameter γ is chosen as follows. For a given in-control value p_0, and a given out-of-control value $p_1 > p_0$ that should be detected quickly, define two constants as follows:

$$r_1 = -\ln\frac{1-p_1}{1-p_0} \quad \text{and} \quad r_2 = \ln\frac{p_1(1-p_0)}{p_0(1-p_1)} \tag{2.1}$$

Then, from the definition of the SPRT, $\gamma = r_1/r_2$. This CUSUM chart, called the binomial CUSUM control chart, signals when $S_k \geq h$, for given upper control limit h.

2.3.3.2 The Bernoulli CUSUM Chart

If the observations become available individually, then we can base our CUSUM chart directly on the individual observations $X_1, X_2, \ldots$, without using any grouping of the items into samples. Instead of waiting until a sample is complete to evaluate the control statistic, the Bernoulli CUSUM chart can use individual data to plot a point on the control chart. Thus, as soon as the result of one item's inspection is ready, a point will be plotted on the chart. For detecting an increase in p, Reynolds and Stoumbos (1999) gave the following definition for the Bernoulli CUSUM control statistic

$$B_k = \max(0, B_{k-1}) + (X_k - \gamma), \; k = 1, 2, \ldots. \tag{2.2}$$

Although the Bernoulli CUSUM chart was considered by others, Reynolds and Stoumbos were the first to provide the definition given in Equation (2.2) and model this CUSUM chart as a Markov chain, as explained below. The Bernoulli CUSUM signals when $B_k \geq h$, for a specific upper control limit, h. The reference value is $\gamma = r_1 / r_2$ (with r_1 and r_2 as defined in equation (2.1)). This chart signals an increase in p for $B_k \geq h$, where h is the upper control limit.

If $\gamma = 1/m$ for some integer m, then $(X_k - \gamma)$ will take on only two values, $-1/m$ and $(m-1)/m$, and B_k will either move down by $-1/m$ or up by $(m-1)/m$. Reynolds and Stoumbos (1999) suggested that achieving $\gamma = 1/m$ can be accomplished by a small adjustment of p_1 (a slight modification of p_1 does not have any practical consequence). For example, if $p_0 = 0.01$ and $p_1 = 0.025$, then $r_1/r_2 = 0.0153/0.9316$, which is equal to $1/61.02$. By adding approximately 0.00001 to the initial value of p_1, we achieve $r_1/r_2 = 1/61$. Now, $\sum_{i=1}^{k}\left(X_i - r_1/r_2\right)$ can be plotted in $1/61$ increments and therefore a Markov chain model can represent this control statistic.

In the next chapter, we will extend this approach to account for correlated data.

2.3.3.3 The Geometric CUSUM Chart

One way of representing the information in binary data, is to consider the number of nondefective items between the $(j-1)^{th}$ and the j^{th} defective items. Denote this random variable as Y_j. It can be shown that the statistics $Y_1, Y_2, \ldots$ are independent and follow a geometric distribution. Define the geometric CUSUM control statistic as

$$G_j = \min\left(0, G_{j-1}\right) + \left(Y_j - \gamma_G\right), \ j = 1, 2, \ldots. \qquad (2.3)$$

The geometric CUSUM signals at the j^{th} defective item if $G_j \le h$.

For any given set of binary data, the information contained in $\left\{Y_j\right\}$ matches that of $\left\{X_k\right\}$ (the set of binary random variables used in constructing the Bernoulli CUSUM). Thus the geometric CUSUM, defined in (2.3) is mathematically equivalent to the Bernoulli CUSUM defined in (2.2), with a head-start. The reference values of these two charts are the reciprocal of each other, i.e., $\gamma_G = r_2/r_1$, for r_1 and r_2 defined in (2.1). (Bourke (1991), Reynolds and Stoumbos (1999), and Sego et al. (2005))

Bourke (1991) provides a detailed investigation of the geometric CUSUM chart. We will see later that the geometric random variable has also been used to construct Shewhart and CUSUM charts for correlated observations. Bourke (2001) used the binomial CUSUM to monitor p, when it is desired to detect small to moderate shifts. He finds that the performance is considerably better when the sample size is set at one, that is, when the Bernoulli CUSUM is used.

2.3.4 CUSUM Chart vs. *p*-chart

Traditionally, Shewhart charts are used for detecting large shifts and CUSUM charts for detecting small shifts in a process parameter. The $p-\text{chart}$ needs to use a reasonably large value of n (depending on p_0) for good performance, while a CUSUM chart can use smaller values of n. Reynolds and Stoumbos (2000) generated several numerical examples to illustrate that for monitoring all sizes of sustained shifts in p, CUSUM charts perform better than the $p-\text{chart}$, because CUSUM charts can use a smaller sample size and can accumulate the information in past samples.

Recall that the Shewhart $p-\text{chart}$ plots the control statistic T_k/n and signals if T_k/n falls outside three standard deviations from p_0, that is ($p_0 \pm 3\sqrt{p_0(1-p_0)/n}$). Therefore, when the lower three-sigma control limit takes a negative value, it does not provide an effective lower control limit. However, it is not just a problem with three-sigma limits, since, it may not be feasible to use a lower probability limit because the false alarm rate may be too high (depending on p_0 and n).

Note that a point is plotted on the Bernoulli CUSUM chart after each individual observation, while it is only plotted at the end of the sample on the binomial CUSUM chart. Nonetheless, for the sample size of $n = 1$ these two charts are exactly the same. In general, Reynolds and Stoumbos (2000) showed that there is a little difference between Bernoulli and binomial CUSUM charts in expected time to detect small and moderate shifts in p. The Bernoulli CUSUM is better for larger shifts.

2.4 Measures of Control Chart Performance

The performance of a control chart is reflected in the level of efficiency of the chart in detecting changes in a process. The faster a change is detected the better is the performance of the chart. In order to attain a valid comparison of performance when the sampling scheme is changed, we need to keep the sampling rate and the false alarm rate constant. The sampling rate per unit time is defined as the sample size divided by the sampling interval. Thus, if for some reason we need a larger sample size, then we must also increase the sampling interval, in order to maintain a constant sampling rate.

A measure that is usually used for the performance of control charts is the *Average Run Length* (ARL), which is defined as the expected number of samples (or sometimes as expected number of individual observations) required to signal. According to Woodall (2006) "run-length performance is typically used for comparison purposes in Phase II [and] corresponds to the number of points plotted on the chart." However, in the case of 100% inspection, the plotted points on the chart may correspond only to defective observations. So it is useful to have other performance measures that distinguish observations from samples. Reynolds and Stoumbos (2000) defined the *Average Number of Samples to Signal* (ANSS) as "the expected number of samples of n observations taken from a specified time point to the time that the chart signals"; the *Average Number of Observations to Signal* (ANOS) as "the expected number of individual observations taken from a specified time point to the time that the chart signals"; and the *Average Time to Signal* (ATS) as "the expected length of time from the start of monitoring to the time that the chart signals."

If a renewal process can be used to model the control charts, then the in-control ATS measures the average false-alarm rate. In other words, a signal is a *false-alarm* when $p = p_0$. The long-run average number of false signals per unit time that a chart produces is the reciprocal of the in-control ATS. For example, consider a control chart with ATS=500 hours at $p = p_0$, this chart produces $1/500 = 0.002$ false-alarms per hour.

For the $p-$chart, the ANSS, ANOS, and ATS are easily computed as follows:

$$\text{ANSS} = 1/P[(T_j/n) \text{ falls outside the control limits}]$$

$$\text{ANOS} = n\,\text{ANSS}$$

$$\text{ATS} = d_0 + c\ \text{ANSS} - d_2, \tag{2.4}$$

where d_0 is the time period before taking the first sample. The ANSS for the binomial CUSUM chart can be approximated closely by using Markov chain methods (see Gan, 1993). Then the ANOS and ATS can be derived from (2.4). But for the Bernoulli CUSUM, none of these relations hold. Reynolds and Stoumbos (1999) give expressions for the Bernoulli CUSUM ANOS that do not depend on d_0, d_1, or d_2. Reynolds and Stoumbos (2000) presented an intuitive relation for approximating the ATS of a Bernoulli CUSUM as

$$\text{ATS} \approx \text{ANOS}\ \frac{c}{n} = \text{ANOS}\ \frac{nd_1 + d_2}{n}, \tag{2.5}$$

where $\text{ATS} \gg d_0, d_2$ and $(\text{ANSS} - 1) \approx \text{ANOS}/n$. For the example in which samples of $n = 100$ are taken every 8 hours, the average time between inspections is approximately $c/n = 480/100 = 4.8$ minutes. If at $p = p_0$ we have ANOS=10000, then ATS $\approx$ (10000)(4.8) $= 48000$ minutes or 800 hours.

In Chapters 3 and 4, we develop a CUSUM control chart for monitoring proportions with correlated observations. To evaluate the performance of our CUSUM and to compare its performance to the Bernoulli CUSUM, we will use the ANOS.

2.4.1 Steady State Measures

In most cases, it is likely that the special cause will occur after monitoring has started, so the control statistic may not be at its starting value. An approach to measuring the performance of a CUSUM chart is to compute the ATS after the control statistic has reached a stationary distribution. In particular, we assume that a shift in p occurs after the process is in steady-state, that the probability that shift happens in a particular interval between items or between samples is proportional to the length of the interval, and that when the shift happens in an interval, the position of the shift within the interval is uniformly distributed over the interval. Reynolds and Stoumbos (2000) provided methods for calculating steady state ATS (SSATS) as well as steady state ANOS (SSANOS) and steady state ANSS (SSANSS) for the $p-\text{chart}$, the binomial CUSUM chart, and the Bernoulli CUSUM chart.

2.4.2 Design of Control Charts with Specific Performance

Designing a CUSUM chart for monitoring p requires specifying a sampling scheme and the choice of a sampling scheme can be limited by the corresponding sampling costs. If at every 8 hours we take samples of $n = 100$, then the sampling rate per unit time is $n/c = 100/8 = 12.5$ items per hour. For a given sampling scheme, the CUSUM parameters we must specify are p_1, representing a shift in the value of p that has to be detected quickly, and a control limit, h. The process parameter p_0 is the value of p that is attainable by the process, which in practice is estimated in Phase I. The parameter p_1 can be considered as a tuning parameter (Reynolds and Stoumbos, 1999) that can be used to tune the CUSUM chart to detect small or large shifts. The reference value γ is a function of p_0 and p_1. The value of p_1 can be adjusted slightly so that $\gamma = 1/m$ for an integer m. This slight adjustment of γ is one method that allows for the value of the control limit h to be chosen to achieve a desired in-control ANOS. The control limit determines the average false-alarm rate as well as the speed of detecting an increase in p. Notice that a low average false-alarm rate corresponds to a large h, whereas, fast detection of shifts in p requires a small h. Thus a tradeoff between these two desired characteristics of a control chart is imposed on the design, which is usually decided depending on practical considerations.

2.4.3 General Issues

Woodall (1997) provided an overview of control charts that have been designed for monitoring a proportion. He lists p and np charts (based on the binomial distribution) for monitoring the proportion of defective items. Then he mentions that when p_0 is small, recent text books recommend basing control charts on the number of nondefective items between defective items (assumed to follow a geometric distribution). These charts are all Shewhart-type charts the design of which generally involves determining the sample size n and constants (such as 3) that determine the control limits. (We do not have n in a case where we count the number of nondefectives between defectives.) Notice that the 3-sigma limits for the Shewhart chart may not be appropriate when the control statistic has a right skewed distribution (see Ryan and Schwertman (1997) for a discussion of determining control limits by using probability-based methods).

Besides the Shewhart-type charts, other control charts that are used with the count data are the exponentially weighted moving average (EWMA) and the cumulative sum charts. In this dissertation, we work with CUSUM charts. Reynolds and Stoumbos (2000) showed that the Bernoulli and the Binomial CUSUM charts are better than the Shewhart $p-$chart, which is expected because CUSUM can be based on a smaller sample size and thus is faster in detecting changes in p.

2.5 Modeling Autocorrelated Binary Data As a First-order Markov Process

To describe the serial correlation in a production process, a first-order Markov model have been used by Bhat and Lal (1988, 1990), Deligonul and Mergen (1987), and Champ, Blatterman, and Rigdon (1994). Here, we adopt the way this model was presented by Bhat and Lal (1999).

The transition probability matrix for the first-order Markov chain can be identified by two parameters: the probability that an item is defective in the long run, p, and the serial correlation, ρ. Suppose that $X_k,\ k \geq 0$ is a first-order Markov process with the general distribution $p_{ij} = P(X_k = j \mid X_{k-1} = i)$ for $x, y \in \{0,1\}$. From Bhat and Lal (1988), the probability transition matrix of the first-order Markov model for the stable process is

$$P = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} = \begin{bmatrix} 1-p(1-\rho) & p(1-\rho) \\ (1-p)(1-\rho) & p+\rho(1-p) \end{bmatrix} \qquad (2.6)$$

where p and ρ must satisfy $1-\min\{1/p, 1/(1-p)\} < \rho < 1$ or equivalently, $\max\{0, -\rho/(1-\rho)\} < p < \min\{1, 1/(1-\rho)\}$.

Bhat and Lal (1989) used two related sets of parameters interchangeably in their work. Their model can be described by two parameters, $a = p_{01}$ and $b = p_{10}$, instead of p and ρ, where these parameters are related by the following relationships

$$p = a/a+b \text{ and } \rho = 1-(a+b). \qquad (2.7)$$

While the models that we introduce in the next section could have used the parameters a and b, we found it more useful to represent our work by using the parameters p and ρ. Champ et al. (1994) used only the notation a and b for presenting the first-order Markov model. For comparability, in Appendix-A, we have shown the relationship between the parameters of Champ et al. (1994) and our parameters.

2.6 *Control Charts for Correlated Data*

In this section, we will review control charts that are designed for monitoring a proportion with autocorrelated binary data. All of this work is based on modeling the correlation by using a first-order Markov chain model.

2.6.1 A Shewhart Chart with Independent Samples

Bhat and Lal (1988) used the first-order Markov chain model presented in Section 2.5 to derive the distribution of the number of successes in a sequence of Markov trials. Bhat and Lal (1990) used this distribution to design a Shewhart chart based on the number of defective items in samples of fixed size n that are inspected with enough spacing between samples such that the serial correlation between samples becomes negligible, but there is correlation between the observations in a sample. Thus, similar to Shewhart sampling for independent production, the spacing between samples is enough to assume that samples are independent. They obtain the three-sigma control limits for an attribute control chart for Markov dependent production processes as

$$\left(\text{LCL, UCL}\right)=\left(p_0-3\sqrt{\text{var}(T_k)/n},\ p_0+3\sqrt{\text{var}(T_k)/n}\right), \tag{2.8}$$

where the variance of defectives in a sample of size n for given p_0 and ρ is given by

$$\text{var}(T_k)=np_0\left(1-p_0\right)+2p_0\left(1-p_0\right)\frac{\rho}{1-\rho}\left(n-\frac{1-\rho^n}{1-\rho}\right). \tag{2.9}$$

2.6.2 A Shewhart Chart with 100% Sampling

When samples are not far enough apart to be considered practically independent, the Shewhart chart designed by Bhat and Lal (1991) does not apply. For 100% sampling, Blatterman and Champ (1992) provide a method for designing a Shewhart chart by modeling the correlation in the data as a first-order Markov chain. Their control chart is based on the number of nondefective units between defective units $\left\{Y_j\right\}$ and plots the statistic used by Bourke (1991). They derive the pdf of the run length N as in (2.10) and (2.11), where N is defined as "the number of defective items observed until a signal is given."

$$P\left(N=t\right)=\begin{cases} u_1 & t=1 \\ \left(1-u_1\right)\left(1-u_2\right)^{t-2}u_2 & t=2,3,\dots \end{cases}$$

$$P\left(N\le t\right)=\begin{cases} u_1 & t=1 \\ u_1+\left(1-u_1\right)\left[1-\left(1-u_2\right)^{t+1}\right] & t=2,3,\dots \end{cases} \tag{2.10}$$

where u_1 and u_2 , which are functions of p and ρ represent the probability that Y_1 and Y_n $\left(n\ge 2\right)$, respectively, fall outside the control limits. They derived the ARL, which is the mean of N , as

$$E\left(N\right)=\frac{1-\left(u_1-u_2\right)}{u_2}. \tag{2.11}$$

They showed that the ARL, increases as the correlation decreases and that there is a significant change in the ARL as a function of the correlation. Thus, the correlation must be considered in designing a Shewhart chart for monitoring a dependent process.

2.6.3 A CUSUM Control Chart with 100% Sampling

For the case of 100% inspection, Champ, Blatterman, and Rigdon (1994) outlined a Markov chain approach for determining the run- length distribution of one-sided and two-sided CUSUM charts for monitoring p, based on Y_j the number of nondefective items before the j^{th} defective item. If the observations are independent, then their one-sided CUSUM chart reduces to that of Bourke (1991). The two-sided CUSUM chart for p, is based on the upper and lower CUSUM statistics

$$L_n = \max\{0, L_{n-1} + \gamma_L - Y_n\} \text{ and } U_n = \max\{0, U_{n-1} - \gamma_U + Y_n\}. \qquad (2.12)$$

They use a sequential probability ratio (SPRT) approach to evaluate the reference values, and thereafter convert them to integers. The reference values, in terms of our preferred notation can be expressed as follows:

$$\gamma_L = \left\lceil \frac{\ln\left(\dfrac{p_{10_L} p_{01_L}}{p_{10_0} p_{01_0}}\right)}{\ln\left(\dfrac{1-p_{10_L}}{1-p_{10_0}}\right)} - 1 \right\rceil \text{ and } \gamma_U = \left\lceil 1 - \frac{\ln\left(\dfrac{p_{01_U} p_{10_U}}{p_{01_0} p_{10_0}}\right)}{\ln\left(\dfrac{1-p_{10_U}}{1-p_{10_0}}\right)} \right\rceil \qquad (2.13)$$

where the subscripts represent the values of the current and past observations, and the subscripts of the subscripts are values of p at which the conditional probability is calculated. For example, p_{10_0} represents the conditional probability of observing a 0 when the previous observation has been 1, calculated at the in-control value of p. For the details of how (2.12) relates to our CUSUM, in terms of the parameters and notation, see Appendix-A.

2.6.4 Monitoring a High Quality Dependent Process

Lai, Xie, and Govindaraju (2000) studied the problem of monitoring a high quality process (i.e., small p_0) with serial dependence, when there is 100% sampling. Their Shewhart chart is based on a random variable X_j that is the number of units observed to get a defective unit. This is equivalent to the number of 0's between two 1's plus one. Notice that $X_j = Y_j + 1$, where Y_j is the number of nondefectives between two defectives, used in Champ et al. (1991, 1994). Lai et al (2000) discuss that for independent observations, the distribution function of X_j's reduces to the geometric distribution. In general, the mean of X is equal to $1/p$, and is not a function of ρ. Whereas, the variance of X, which can be expressed as $\left((1-p)/p^2\right)\left((1+\rho)/(1-\rho)\right)$, is a function of ρ. If the correlation is neglected and a geometric distribution is used for X, then the variance of X is underestimated when $\rho > 0$ and overestimated if $\rho < 0$. Because the variance of X is larger than that of the geometric in the presence of positive correlation, it is imperative to account for the correlation in our monitoring process.

Lai et al. (2000) showed that correlation affects the false alarm rate of a standard Shewhart chart that is constructed based on X_1, X_2, …, and gave numerical results indicating that when the control chart is designed assuming independence "even a small positive serial correlation increases the false alarm probability." Also, they numerically verify that the ARL decreases when ρ increases. Therefore, the traditional Shewhart-type control charts should not be used without modifications for autocorrelated observations. They conclude that for a small correlation the control limits can be revised, as in (2.14) and (2.15) below, otherwise different control schemes should be used. In the presence of autocorrelation, the upper and lower control limits that (approximately) achieve the false alarm probability α are

$$\mathrm{LCL} = 2 + \frac{\log\left(1-\alpha_{\mathrm{L}}\right) - \log\left\{(1-p)(1-\rho)\right\}}{\log\left\{1-p(1-\rho)\right\}} \tag{2.14}$$

$$\mathrm{UCL} = 1 + \frac{\log\left(1-\alpha_{\mathrm{U}}\right) - \log\left\{(1-p)(1-\rho)\right\}}{\log\left\{1-p(1-\rho)\right\}} \tag{2.15}$$

where $\alpha = \alpha_L + \alpha_U$.

Their concluding remarks and recommendations reinforce our motivation. Lai et al. (2000) pointed out that eliminating correlation is very hard. Based on investigating practical issues with monitoring procedures with correlated observations, they recommend using control schemes that take into account the presence of correlation.

In all of the work mentioned in Sections 2.6.1, 2.6.2, 2.6.3, and 2.6.4, the dependence in the data is assumed to follow a first-order Markov model. However, the control charts are based on control statistics that are functions of independent random variables that are derived from the dependent data. In the next chapters, we define CUSUM statistics that are directly derived based on the log-likelihood ratio for Markov dependent observations.

2.7 Markov Chains

Consider a stochastic process $(X_0, X_1, X_2, ...)$, where the following holds

$$P(X_{k+1} = j \mid X_k = i, X_{k-1} = i_{k-1}, ..., X_0 = i_0) = P(X_{k+1} = j \mid X_k = i)\,, \tag{2.16}$$

for all j, $i, i_{k-1}, \ldots, i_0$. This condition is referred to as the Markovian assumption. If the state space is discrete, a stochastic process that satisfies the Markovian assumption is called a discrete time *Markov chain* because the time parameter is discrete. A *time homogeneous* Markov chain is one for which the probability of going from state i to state j is independent from the location of the time period on the time axis. Therefore $p_{ij} =$ $P(X_{k+1} = j \mid X_k = i)$, the *transition* probability associated with this move, does not depend on k. $P = [p_{ij}]$ is called the transition matrix.

2.7.1 Stationary and Transition probabilities

Two states j and i *communicate* if it is possible to go from one to the other. That is if the transition probabilities $p_{ij}^{(k)}$ and $p_{ji}^{(k)}$ are positive for some k, where $p_{ij}^{(k)}$ = $P(X_{l+k} = j \mid X_l = i)$. The states that communicate with each other form a *class*. If all states communicate with each other (there is only one class), then we have an *irreducible* Markov chain. If the probability of returning to the current state is one, that state is called *recurrent*. And if this probability is less than one then the state is *transient*. A state is said to be *positive recurrent* if the expected time to return is finite.

All states are positive recurrent if the Markov chain is finite and irreducible. In this case, it can be shown that a unique stationary distribution, $\boldsymbol{\pi} = [\pi_j]$, always exists. The stationary probability π_j is the long run proportion of time that the Markov chain is in state j, which is independent of the initial state i.

State i is *absorbing* if $p_{ii} = 1$. Consider t transient states and a absorbing states, so that the total number of states is $s = t + a$. For convenience, number the states such that the first t are the transient states and the remaining a are the absorbing ones. Now, partition the *transition probability matrix* as follows:

$$\mathbf{P} = \begin{pmatrix} \mathbf{Q}_{t\times t} & \mathbf{U}_{t\times a} \\ \mathbf{0}_{a\times t} & \mathbf{I}_{a\times a} \end{pmatrix} \tag{2.17}$$

Define $v_{ij}^{(k)} = P(\text{absorbed in state } j \text{ in exactly } k \text{ steps}|\text{start in state } i)$, for $k = 1, 2, ...$; and $i = 1, 2, ..., t$; and $j = t+1, t+2, ..., t+a$. If absorption occurs at step k, then the process must have been in one of the transient steps at step $k-1$. Thus, it follows that $v_{ij}^{(k)} = \sum_{l=1}^{t} p_{il}^{(k-1)} p_{lk}$, or in matrix notation $\mathbf{V}^{(k)} = [v_{ij}^{(k)}] = \mathbf{Q}^{k-1}\mathbf{U}$. Let $v_{ij} = P(\text{absorbed in state } j|\text{ start in state } i) = \sum_{k=1}^{\infty} v_{ij}^{(k)}$, then we can write

$$\mathbf{V} = [v_{ij}] = \sum_{k=1}^{\infty} \mathbf{V}^{(k)} = \sum_{k=1}^{\infty} \mathbf{Q}^{k-1}\,\mathbf{U} = (\mathbf{I} - \mathbf{Q})^{-1}\,\mathbf{U}\,. \tag{2.18}$$

2.7.2 The ANOS Vector

Denote m_{ij} as the expected number of times that the process is in transient state j before absorption, given that the starting state is i, for $i, j = 1, 2, ..., t$. Define

$N_i =$ expected time to absorption, given that the starting state is i

$= \sum_{j=1}^{t}$(expected number of times in state j before absorption, given that the starting state is i)

$= \sum_{j=1}^{t} m_{ij}$.

Thus, for $\mathbf{N} = (N_1, N_2, ..., N_t)^T$, we can see that $\mathbf{N} = \mathbf{M1} = (\mathbf{I} - \mathbf{Q})^{-1}\,\mathbf{1}$, where $\mathbf{M} = [m_{ij}]$.

In Chapter 5, we use this vector to calculate the exact ANOS for a control chart, when the control statistic is represented by a Markov model.

2.8 Binary Time Series

Kedem (1980) defined "clipping" or "hard limiting" as transforming an original series of data into a binary series, which has the benefit of making the statistical analysis simple and fast. The case of monitoring processes with original non-binary data can potentially be converted to a situation in which we want to monitor a binary series. Binary series have desirable properties, such as remaining almost unharmed when the original data are truncated, and eliminating the problem of outliers. Theoretically, the pair-wise correlation will reduce when a binary series is produced from a Gaussian series, however, transforming continuous data to binary data entails information loss. Finally, finite conditional dependence that can be represented by a specific order of dependence in a binary series facilitates useful estimation. Thus, it may be desirable to convert observed data into binary series and model this series as a Markov chain with finite dependence. That is if the gain from the desirable properties of a binary series exceeds the unavoidable loss of information that results from clipping.

2.9 Summary

In this chapter, we defined the binary data, which is the data that we will use throughout our work, and outlined the notation for a general sampling scheme. In Section 2.3, we gave the general definitions for Shewhart charts and CUSUM control charts, and reviewed the control charts that have been applied to the problem of monitoring a proportion, such as the $p-$chart, as well as the Bernoulli, the binomial, and the geometric CUSUM charts. Section 2.4 presented the definitions for the usual measure of performance, the ARL, and for other measures of performance, the ANOS, ANSS, and ATS, that enable us to compare control charts with different sample sizes. Thereafter, as our interest is in monitoring dependent production processes, we reviewed the state of research in this area, including different control charts that have been designed to account for the autocorrelation in the binary data as well as the estimation of model parameters and the order of the Markov chain. In Section 2.7, we provided an introduction to the properties and terminologies of Markov chains. We will use the Markov chains in two ways, first to model the correlation among observation binary data, and second to model the CUSUM statistic that we will develop.

3 A Log-Likelihood Based CUSUM Control Chart for Monitoring a Proportion in the Presence of Autocorrelation

3.1 Introduction

The general form of a CUSUM chart cannot be written as a sum of log-likelihood ratios for individual observations when the observations are correlated. In this chapter, we derive the conditional probabilities for correlated data and use this probability structure to construct a CUSUM statistic that accounts for the correlation between binary data by modeling it as first-order Markov dependence. A desirable property of a control chart is the capability of conforming to certain statistical specifications. To achieve this property, we model our CUSUM statistic as a Markov chain, and call it the Markov Binary CUSUM (MBCUSUM). This enables us to use Markov chain properties to set up the MBCUSUM control chart with desired specifications, such as a specified false alarm rate. The MBCUSUM can be applied when the results of the inspection of individual items are available immediately after the inspection of each item.

In most applications, there is a period (usually called phase I) when the in-control parameters of the process are estimated. Then in Phase II, we monitor the process. Here, we are going to assume that we have estimated the in-control parameters, p_0 and ρ, without error, and are considering phase II.

3.2 Log-likelihood Ratio for Markov Dependent Data

A first-order Markov process is specified by two parameters: the long-term fraction defective, p, and the serial correlation, ρ. Assume that the out-of-control situation corresponds to a change in p, but not a change in ρ, so the in-control and out-of-control parameters are (p_0, ρ) and (p_1, ρ), respectively. Here, $p_1 > p_0$ represents a specific out-of-control value of p that we want to detect quickly. Remember that as discussed in Section 2.5, the first-order Markov chain can be represented by $p_{01} =$

$P\left(X_k = 1 \mid X_{k-1} = 0\right)$ and $p_{10} = P\left(X_k = 0 \mid X_{k-1} = 1\right)$, where $p = p_{01} / \left(p_{01} + p_{10}\right)$ and $\rho = 1 - p_{01} - p_{10}$.

Now, consider binary observations $X_k \in \{0,1\}$, for which the stationary distribution is given by π_i, where $i = 1,2$ for a first-order Markov chain. For each observation X_k, consider a control chart that plots $C_k = \max(0, C_{k-1}) + z_k$ and signals when $C_k \geq h$, where the variable z_k is defined as

$$z_k = \ln\left(f(X_k \mid p_1, \rho) / f(X_k \mid p_0, \rho)\right). \tag{3.1}$$

Before deriving an expression for z_k when the data have nonzero correlation, let us review the case of independent data. When the observations are independent, i.e., $\rho = 0$, the log of the likelihood ratio of the observations up to the current point can be written as in (3.1). Consider a stable process, and consider X_0 with distribution Bernoulli$\left(p\right)$ as a device to start the correlated process. The log-likelihood ratio for X_0 is $\ln\left(f\left(X_0 \mid p_1\right) / f\left(X_0 \mid p_0\right)\right)$. After the first observation, we have

$$Z_1 = \ln \frac{f(X_0, X_1 \mid p_1)}{f(X_0, X_1 \mid p_0)} = \ln \frac{f(X_0 \mid p_1)}{f(X_0 \mid p_0)} + \ln \frac{f(X_1 \mid p_1)}{f(X_1 \mid p_0)},$$

after the second observation we have

$$Z_2 = \ln \frac{f(X_0, X_1, X_2 \mid p_1)}{f(X_0, X_1, X_2 \mid p_0)} = \ln \frac{f(X_0 \mid p_1)}{f(X_0 \mid p_0)} + \ln \frac{f(X_1 \mid p_1)}{f(X_1 \mid p_0)} + \ln \frac{f(X_2 \mid p_1)}{f(X_2 \mid p_0)},$$

and after k observation we have

$$Z_k = \ln \frac{f(X_0, \ldots, X_k \mid p_1)}{f(X_0, \ldots, X_k \mid p_0)} = \sum\nolimits_{i=0}^{k} \ln \frac{f(X_i \mid p_1)}{f(X_i \mid p_0)} = \sum\nolimits_{i=0}^{k} z_i \,. \tag{3.2}$$

But when X_k's are dependent, i.e., $\rho \neq 0$, this simple additive relationship does not hold. Nevertheless, we can always write $f(X_1, X_2 \mid p, \rho) = f(X_1 \mid p, \rho) f(X_2 \mid X_1, p, \rho)$, and thus for a Markov dependent process, where X_k depends only on X_{k-1}, by Markovian assumption the following equality holds:

$f(\mathbf{X}_k \mid p, \rho)$

$$= f(X_0, X_1, ..., X_k \mid p, \rho)$$
$$= f(X_0 \mid p, \rho) f(X_1 \mid X_0, p, \rho) f(X_2 \mid X_1, p, \rho) ... f(X_k \mid X_{k-1}, p, \rho) . \quad (3.3)$$

In order to use this relation for constructing our CUSUM statistic, in what follows, we derive an expression for the general term $f(X_k \mid X_{k-1}, p, \rho)$.

3.2.1 Conditional Probabilities for Markov Dependent Data

To derive the conditional probability for first-order Markov dependent binary observations that can appear only in one of the four following sequences, $(X_{k-1}, X_k) \in \{(0,0),(0,1),(1,0),(1,1)\}$, we write

$$f(X_k \mid X_{k-1}) = P(X_k = 0 \mid X_{k-1} = 0)^{(1-X_{k-1})(1-X_k)} \times P(X_k = 1 \mid X_{k-1} = 0)^{(1-X_{k-1})X_k} \times$$
$$P(X_k = 0 \mid X_{k-1} = 1)^{X_{k-1}(1-X_k)} \times P(X_k = 1 \mid X_{k-1} = 1)^{X_{k-1}X_k} . \quad (3.4)$$

Recall from Section 2.5, that the conditional probabilities for a stable first-order Markov chain in terms of the parameters p and ρ are as follows:

$$\begin{aligned} p_{00} &= 1 - p(1-\rho) \\ p_{01} &= p(1-\rho) \\ p_{10} &= (1-p)(1-\rho) \\ p_{11} &= p + \rho(1-p). \end{aligned} \quad (3.5)$$

By replacing from (3.5) in (3.4), we get

$$f(X_k \mid X_{k-1}, p, \rho) =$$
$$[1-p(1-\rho)]^{(1-X_{k-1})(1-X_k)} [p(1-\rho)]^{(1-X_{k-1})X_k} [(1-p)(1-\rho)]^{X_{k-1}(1-X_k)} [1-(1-p)(1-\rho)]^{X_{k-1}X_k} , \quad (3.6)$$

which is the conditional probability structure for autocorrelated binary data, where the dependence has been modeled as a first-order Markov chain.

3.3 The Log-likelihood Ratio Based CUSUM for Correlated Binary Observations

After k observations, the log-likelihood ratio for Markov dependent binary observations is

$$
\begin{aligned}
Z_k &= \ln \frac{f(X_0,\ldots,X_k \mid p_1,\rho)}{f(X_0,\ldots,X_k \mid p_0,\rho)} \\
&= \sum_{i=0}^{k} \ln \frac{f(X_i \mid p_1,\rho)}{f(X_i \mid p_0,\rho)} \\
&= \ln \frac{f(X_0 \mid p_1)}{f(X_0 \mid p_0)} + \sum_{i=1}^{k} \ln \frac{f(X_i \mid X_{i-1}, p_1,\rho)}{f(X_i \mid X_{i-1}, p_0,\rho)}\,, \qquad (3.7)
\end{aligned}
$$

where it is assumed that X_0 follows a Bernoulli (p) model, and that p changes, but, ρ does not change. (However, p_{01} and p_{10} both change with changes in p .)

Equation (3.7) can be used to express the log-likelihood ratio $\ln\left(f(X_k \mid X_{k-1}, p_1,\rho) / f(X_k \mid X_{k-1}, p_0,\rho)\right)$ in terms of the parameters p_0 , p_1 , and ρ . But first, to create a simpler expression, we rewrite (3.4) as

$$
\begin{aligned}
&f(X_k \mid X_{k-1}, p,\rho) = \\
&p^{(1-X_{k-1})X_k} (1-p)^{X_{k-1}(1-X_k)} (1-\rho)\left(\frac{1}{1-\rho} - p\right)^{(1-X_{k-1})(1-X_k)} \left(\frac{1}{1-\rho} - (1-p)\right)^{X_{k-1}X_k}, \qquad (3.8)
\end{aligned}
$$

then, define a new parameter $s = 1/(1-\rho)$, which gives $1-\rho = 1/s$, and substitute in (3.4) to get

$$
f(X_k \mid X_{k-1}, p,\rho) = \frac{p^{(1-X_{k-1})X_k} (1-p)^{X_{k-1}(1-X_k)} (s-p)^{(1-X_{k-1})(1-X_k)} \left(s-(1-p)\right)^{X_{k-1}X_k}}{s}. \qquad (3.9)
$$

Taking logarithms of both sides of (3.9) gives

$$
\begin{aligned}
\ln f(X_k \mid X_{k-1}, p,\ \rho) = X_k \ln \frac{p}{s-p} + X_{k-1} \ln \frac{1-p}{s-p} + \\
X_{k-1}X_k \ln \frac{(s-p)(s-(1-p))}{p(1-p)} + \ln\left(s\,(s-p)\right). \qquad (3.10)
\end{aligned}
$$

Now, using (3.10) in the numerator and denominator of the log-likelihood ratio gives

$$\ln\frac{f(X_k\mid X_{k-1},p_1,\rho)}{f(X_k\mid X_{k-1},p_0,\rho)}=\ln\frac{s.\,p_1^{(1-X_{k-1})X_k}.(1-p_1)^{X_{k-1}(1-X_k)}.(s-p_1)^{(1-X_{k-1})(1-X_k)}.(s-(1-p_1))^{X_{k-1}X_k}}{s.\,p_0^{(1-X_{k-1})X_k}.(1-p_0)^{X_{k-1}(1-X_k)}.(s-p_0)^{(1-X_{k-1})(1-X_k)}.(s-(1-p_0))^{X_{k-1}X_k}}$$

$$=X_{k-1}.\ln\left(\frac{1-p_1}{s-p_1}\frac{s-p_0}{1-p_0}\right)+X_k.\ln\left(\frac{p_1}{p_0}\frac{s-p_0}{s-p_1}\right)$$

$$+X_{k-1}X_k\left(\ln\frac{s-(1-p_1)}{s-(1-p_0)}+\ln\frac{s-p_1}{s-p_0}+\ln\frac{1-p_0}{1-p_1}+\ln\frac{p_0}{p_1}\right)-\ln\frac{s-p_0}{s-p_1}. \tag{3.11}$$

Define the following four parameters that will be useful in constructing the Markov model for the CUSUM control statistic:

$$q_1=\ln\frac{s-p_1}{s-p_0};\ q_2=\ln\frac{p_1}{p_0};\ q_3=\ln\frac{1-p_1}{1-p_0};\ q_4=\ln\frac{s-(1-p_1)}{s-(1-p_0)}. \tag{3.12}$$

Rewriting (3.11) in terms of q_1 to q_4 gives

$$\ln\frac{f(X_k\mid X_{k-1},p_1,\rho)}{f(X_k\mid X_{k-1},p_0,\rho)}=X_{k-1}(-q_1+q_3)+X_k(-q_1+q_2)-X_{k-1}X_k(-q_1+q_2+q_3-q_4)+q_1.$$

Defining

$$\Delta_k=X_{k-1}(-q_1+q_3)+X_k(-q_1+q_2)-X_{k-1}X_k(-q_1+q_2+q_3-q_4)+q_1, \tag{3.13}$$

the exact log-likelihood ratio based CUSUM statistic for a first-order Markov process is

$$C_k=\max(0,C_{k-1})+\ln\frac{f(X_0\mid p_1)}{f(X_0\mid p_0)}+\Delta_k. \tag{3.14}$$

where $k=1,2,\ldots$, and $C_0=0$. The possible values of the random variable Δ_k (call it the CUSUM *increment*) for each value of two consecutive observations are shown in Table 3-1.

Table 3-1 The Values of the CUSUM Increment

(X_{k-1}, X_k)	Δ_k
$(0,0)$	$q_1 = \ln \dfrac{s - p_1}{s - p_0}$
$(0,1)$	$q_2 = \ln \dfrac{p_1}{p_0}$
$(1,0)$	$q_3 = \ln \dfrac{1 - p_1}{1 - p_0}$
$(1,1)$	$q_4 = \ln \dfrac{s - (1 - p_1)}{s - (1 - p_0)}$

3.4 The Markov Binary CUSUM Control Chart

In what follows, our goal is to find a form that would approximate the values of our increment (see Table 3.1) by rounding to integer multiples of $1/m$, where m is an integer. When this is achieved, we use Z_k from Equation (3.7) to construct a CUSUM statistic that can itself be modeled as a Markov chain: the MBCUSUM (the Markov Binary CUSUM) statistic denoted by C_k. The MBCUSUM statistic is constructed such that from any transient state to any other transient state in the transition probability matrix, the changes can be expressed in terms of only four multiples of $1/m$. Thus, the possible values of C_k are 0, $1/m$, $2/m$, etc. For a given value of the upper control limit, say h, the MBCUSUM control chart is designed to signal at $C_k \geq h$, where $h > 1$ is expressed as an integer multiple of $1/m$. It will be shown later that the number of transient states is $2H$, where $H = mh$. Notice that there are H values of C_k, each corresponding to two transient states. This happens because we account for first-order dependence, where the current value of the observation depends on the previous value. To find the properties of the control chart we only need the matrix of transition probabilities, $\mathbf{Q}$, that has $2H$ states.

Before presenting the general method of constructing the MBCUSUM, we illustrate the steps in the following example.

3.4.1 Example

Consider the case in which $p_0 = 0.01$, $p_1 = 0.025$, and $\rho = 0.05$. Then the value of s is 1.0526 and $q_1 = -0.0145$. Then we want the increment values to be rounded to integer multiples of $1/m$, where $m = [1/|q_1|] = 69$. Thus the first value of the increment is $-1/69$ corresponding to two consecutive zeros. Denote the value of Δ_k rounded to a multiple integer of $1/m$ as D_k. The increment values, corresponding to the four possible sequences are shown in Table 3-2.

Table 3-2 The Exact and Rounded Values of the CUSUM Increment

(X_{k-1}, X_k)	Δ_k	D_k
$(0,1)$	$q_1 = -0.0145$	$-1/69 = -0.0145$
$(0,1)$	$q_2 = 0.9163$	$63/69 = 0.9130$
$(0,1)$	$q_3 = -0.0153$	$-1/69 = -0.0145$
$(0,1)$	$q_4 = 0.2147$	$15/69 = 0.2174$

For $h = 100/69$ (i.e., $H = 100$), the matrix $\mathbf{Q}$ has 200 states. The in-control values ($p = p_0$) of the conditional probabilities for the correlated binary observations modeled as a first-order Markov chain are as follows:

$$p_{00} = 1 - p(1-\rho) = 1 - 0.010(1-0.05) = 0.09905$$

$$p_{01} = 1 - p_{00} = 0.0095$$

$$p_{10} = (1-p)(1-\rho) = (1-0.010)(1-0.05) = .9405$$

$$p_{11} = 1 - p_{10} = 0.0595 .$$

Each row of the matrix $\mathbf{Q}$ has at most two nonzero elements. When there are two nonzero elements the row sums to one, otherwise there is one smaller than one nonzero element as is shown in Figure 3-1, below.

i	X_{k-1}	C_{k-1}																
		j:	1	2	3	$\cdots$	31	32	$\cdots$	127	128	$\cdots$	197	198	199	200
		X_k:	0	1	0	$\cdots$	0	1	$\cdots$	0	1	$\cdots$	0	1	0	1
		C_k:	0	0	1/69	$\cdots$	15/69	15/69	$\cdots$	63/69	63/69	$\cdots$	98/69	98/69	99/69	99/69
1	0	0	p_{00}	0	0	$\cdots$	0	0	$\cdots$	0	p_{01}	$\cdots$	0	0	0	0
2	1	0	p_{10}	0	0	$\cdots$	0	p_{11}	$\cdots$	0	0	$\cdots$	0	0	0	0
3	0	1/69	p_{00}	0	0	$\cdots$	0	0	$\cdots$	0	0	$\cdots$	0	0	0	0
4	1	1/69	p_{10}	0	0	$\cdots$	0	0	$\cdots$	0	0	$\cdots$	0	0	0	0
5	0	2/69	0	0	p_{00}	$\cdots$	0	0	$\cdots$	0	0	$\cdots$	0	0	0	0
6	1	2/69	0	0	p_{10}	$\cdots$	0	0	$\cdots$	0	0	$\cdots$	0	0	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
73	0	36/69	0	0	0	0	0	0	$\cdots$	0	0	$\cdots$	0	0	0	p_{01}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
170	1	84/69	0	0	0	0	0	0	$\cdots$	0	0	$\cdots$	0	0	0	p_{11}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
199	0	99/69	0	0	0	0	0	0	$\cdots$	0	0	$\cdots$	p_{00}	0	0	0
200	1	99/69	0	0	0	0	0	0	$\cdots$	0	0	$\cdots$	p_{10}	0	0	0

Figure 3-1 The $\mathbf{Q}$ Matrix for the MBCUSUM with $p_0 = 0.010$, $p_1 = 0.025$, and $\rho = 0.05$

3.4.2 The Markov Binary CUSUM Statistic

Converting the increment values to integer multiples of $1/m$, where m is an integer, allows for modeling our CUSUM statistic as a Markov chain. In general, we could convert the CUSUM increments to integer multiples of some rational number, say m_1/m_2, where m_1 and m_2 are integers. Using integer multiples of m_1/m_2 should allow for arbitrarily close approximations to the exact CUSUM. However, this would increase the size of $\mathbf{Q}$, so there would be a limit to what we could compute. Here, for simplicity, we consider only $m_1 = 1$.

Denote the values of Δ_k that are converted to multiples of $1/m$ as D_k, and define the Markov Binary CUSUM statistic as

$$C_k = \max(0, C_k) + \ln\frac{f(X_0 \mid p_1)}{f(X_0 \mid p_0)} + D_k \tag{3.15}$$

The possible transitions and associated transition probabilities for the MBCUSUM, using a first-order Markov model for the process observations, are summarized in Table 3-3. The way in which we arrive at these values is explained later.

Table 3-3 Transitions and Transition Probabilities for the MBCUSUM

X_{k-1}	X_k	D_k	Transition	Probability	i
0	0	$-1/m$	$1 \to 1$	p_{00}	
			$i \to i-2$		$3,5,\ldots,2H-1$
0	1	d_{01}/m	$i \to i+2d_{01}+1$	p_{01}	$1,3,\ldots,2H-2d_{01}-1$
1	0	d_{10}/m	$2 \to 1$	p_{10}	
			$i \to i-2d_{10}-1$		$2d_{10}+2, 2d_{10}+4,\ldots,2H$
1	1	d_{11}/m	$i \to i+2d_{11}$	p_{11}	$2,4,\ldots,2H-2d_{11}$

The possible values of the increment D_k are obtained from q_1, q_2, q_3, and q_4 (see (3.12)) as follows. $m = \text{int}(|1/q_1|)$ is a positive integer, where $\text{int}(\cdot)$ indicates the closest integer value, and $q_1 = \ln\left((s-p_1)/(s-p_0)\right)$. We use $1/m$ as the basic step for the values of the increment, and for plotting the control chart. For $\rho > 0$ and $p_1 > p_0$, we define:

$$d_{01} = \text{int}(q_2 m), \quad q_2 = \ln\frac{p_1}{p_0}$$

$$d_{10} = |\text{int}(q_3 m)|, \quad q_3 = \ln\frac{1-p_1}{1-p_0}$$

$$d_{11} = \text{int}(q_4 m), \quad q_4 = \ln\frac{s-(1-p_1)}{s-(1-p_0)} \tag{3.16}$$

Notice, that $s = 1/(1-\rho)$ is larger than 1 for $\rho > 0$. Also, for $p_1 > p_0$, q_1 and q_3 are always negative, whereas, q_2 and q_4 are always positive. Therefore, the signs of the four values of the CUSUM increment do not depend on the values of p_0 and p_1 as long as $p_0 < p_1$.

The elements of the transition matrix are derived based on the following considerations. There are a total of $2H$ transient states, corresponding to H values of the CUSUM statistic. We denote the states by i, where $i = 1,2,3,\ldots,2H$. Then C_k takes the values that are integer multiples of $1/m$, i.e., 0, $1/m$, $2/m$, through $(H-1)/m = h - 1/m$. The largest value of the statistic for which a signal is not generated is equal to

$(H-1)/m$. Each value of the statistic corresponds to two transient states as expressed by the following relationship

$$C_k = \frac{\text{int}\left(\frac{i-1}{2}\right)}{m} \Leftrightarrow \begin{cases} i = 2mC_k + 1 \\ \\ i = 2mC_k + 2 \end{cases} \tag{3.17}$$

for $i = 1, 2, \ldots, H$. Remember that C_k is defined such that if C_{k-1} takes a negative value, the statistic will be reset to zero. Also, remember that the value of the increment for consecutive observations 00 and 10 is negative. So the transition probabilities corresponding to these sequences of observations and to a value of zero for both C_k and C_{k-1}, turn out to be p_{00} and p_{10}. These probabilities correspond to q_1 and q_3, respectively. Thereafter, $-1/m$ is added to C_{k-1} at any transition from state i to state $j = i-2$ when $(X_{k-1}, X_k) = (0,0)$, with probability p_{00}, until C_{k-1} reaches its maximum value of $(H-1)/m$, i.e., at state $i = 2mC_k + 1 = 2m((H-1)/m) + 1 = 2H - 1$.

When the second value of the increment, d_{01}/m, is added first to the minimum value of $C_{k-1} = 0$, the transition is to state j, where C_k is equal to d_{01}/m, that is $j = i + 2d_{01} + 1$ and $(X_{k-1}, X_k) = (0,1)$, with probability p_{01}. The last time that the second increment can be added to the CUSUM statistic without creating a signal is when C_k takes its maximum value of $(H-1)/m$, thus, $C_{k-1} = \max(C_k) - d_{01}/m = (H-1)/m - d_{01}/m$, which corresponds to state $i = 2mC_{k-1} + 1 = 2H - 2d_{01} - 1$.

For the sequence of $(X_{k-1}, X_k) = (1,0)$ the second increment d_{10}/m will be reduced from C_{k-1} to make C_k. The first time this reduction can happen is when C_{k-1} is equal to d_{10}/m and thus C_k will come out to be zero. This situation occurs at the state where $i = 2mC_{k-1} + 2 = 2d_{10} + 2$ and $j = 1$, and continues at every other row i, and column $j = i - 2d_{10} - 1$ until C_{k-1} reaches its maximum value of $(H-1)/m$, at state $i = 2H$ and $j = 2H - 2d_{10} - 1$.

The last value of the increment d_{11}/m is added to $C_{k-1} = 0$ for the first time, resulting in a transition from state $i = 2$ to the state where $C_k = d_{11}/m$, and for which $j = 2m(d_{11}/m) + 2 = 2d_{11} + 2$. Thereafter, this transition corresponds to $(X_{k-1}, X_k) = (1,1)$ and even numbered states until the last time where d_{11}/m can be added to C_{k-1}, which is when C_k takes its maximum value of $(H-1)/m$. Thus $C_{k-1} = (H-1)/m - d_{11}/m$, which corresponds to state $i = 2mC_{k-1} + 2 = 2H - 2d_{11}$. Denoting the transition probabilities from state i to state j by $p_{i\to j}$, we can define the $\mathbf{Q}$ matrix for the first-order Markov binary CUSUM as follows:

$$
\begin{aligned}
&p_{1\to 1} = p_{00} \\
&p_{2\to 1} = p_{10} \\
&p_{i\to i-2} = P(X_k = 0 \mid X_{k-1} = 0) = p_{00} && i = 3,5,\ldots,2H-1 \\
&p_{i\to i+2d_{01}+1} = P(X_k = 1 \mid X_{k-1} = 0) = p_{01} && i = 1,3,\ldots,2H-2d_{01}-1 \\
&p_{i\to i-2d_{10}-1} = P(X_k = 0 \mid X_{k-1} = 1) = p_{10} && i = 2d_{10}+2, 2d_{10}+4,\ldots,2H \\
&p_{i\to i+2d_{11}} = P(X_k = 1 \mid X_{k-1} = 1) = p_{11} && i = 2,4,\ldots,2H-2d_{11}.
\end{aligned}
\tag{3.18}
$$

3.4.2.1 ANOS and SSANOS

Denote the vector of ANOS values for the transient states as $\mathbf{N} = (N_1, N_2, \ldots, N_{2H})'$. N_i is the ANOS if the control statistic and the first-order process correspond to state i when we start. Thus, in the in-control case, N_i is the in-control ANOS when we start in state i; and in the out-of-control case, N_i is the out-of-control ANOS when we start with the shift already present, i.e., when $p = p_1$. The ANOS vector for our Markov chain can be expressed as

$$\mathbf{N} = (\mathbf{I}-\mathbf{Q})^{-1}\mathbf{1}, \tag{3.19}$$

where $\mathbf{Q} = [p_{i\to j}]$ is the transition matrix presented in (3.18), $(\mathbf{I}-\mathbf{Q})^{-1}$ is the fundamental matrix of the Markov chain and $\mathbf{1}$ is a column vector of 1's (for details see Section 2.5). If the CUSUM starts at 0 and X_0~Bernoulli (p), then the ANOS of interest is $pN_1 + (1-p)N_2$.

In practice, the shift in p happens at a random time, when the value of the control statistic may be different from 0. To account for this case, we use the steady state ANOS. The SSANOS can be calculated as $\boldsymbol{\pi}\mathbf{N}$, where $\boldsymbol{\pi}$ is the normalized left eigenvector of $\mathbf{Q}$ that corresponds to the largest eigenvalue.

3.5 Summary

We constructed a CUSUM control chart for monitoring the proportion defective in a first-order Markov dependent binary process based on a CUSUM statistic that is modeled as a Markov chain. We called it the Markov Binary CUSUM (MBCUSUM). Formulating our CUSUM statistic as a Markov chain allows us to construct a transition probability matrix, which in turn enables us to obtain specific statistical properties for our control chart. We defined an approximation that gives expressions for the CUSUM increment in terms of integer multiples of $1/m$, where m is an integer. In Chapter 5, we numerically verify that the performance of the MBCUSUM is superior to that of the traditional Shewhart chart and the Bernoulli CUSUM control chart.

4 A General Approach to Modeling and Monitoring Processes with Correlated Binary Observations

4.1 Introduction

In this chapter, we develop a general model that maps a binary series with arbitrary order (t) Markov dependence to two levels with a conditional probability of observing a defective at each level. Compared to the direct extension of the standard formulation of Markov chains, this simpler model substantially reduces the transition matrix dimensions, and requires only three parameters. The three-parameter model assigns a higher probability to observing a 1 when there has been another 1 in the previous t observations; and a lower probability to observing a 1 after t consecutive zeros. By using this model we obtain a general form for the CUSUM statistic developed in Chapter 3. Finally, we introduce a series of rules that will give a generalization of the three-parameter model and derive the increments for the corresponding CUSUM statistic.

4.2 Generalization of the Standard Markov Model

Define t as the number of previous observations that have Markov dependence with the current observation. This is referred to as the order of a Markov chain. For example, if X_k has Markov dependence with X_{k-1}, X_{k-2}, X_{k-3}, but does not depend on X_{k-3}, X_{k-4}, $\ldots$, then we have a third-order Markov chain.

Start from the simplest case of $t=1$. Recall the transition matrix for a first-order Markov chain is as presented in Figure 4-1,

$$\begin{array}{cc} & X_k \\ X_{k-1} & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} \end{array}$$

Figure 4-1 Transition Matrix for the First-Order Markov Chain

where the rows sum to one. Now, consider the case where the probability of observing a 1 depends on two previous observations, i.e., $t = 2$. For this case the transition probability matrix for the standard Markov model can be constructed as in Figure 4-2.

$$\begin{array}{cc} & X_{k-1}X_k \\ X_{k-2}X_{k-1} & \begin{array}{cccc} 00 & 01 & 10 & 11 \end{array} \\ \begin{array}{c} 00 \\ 01 \\ 10 \\ 11 \end{array} & \begin{bmatrix} p_{000} & p_{001} & 0 & 0 \\ 0 & 0 & p_{010} & p_{011} \\ p_{100} & p_{101} & 0 & 0 \\ 0 & 0 & p_{110} & p_{111} \end{bmatrix} \end{array}$$

Figure 4-2 The Full Transition Matrix for the Second-Order Markov Chain

Although we see eight probability values in Figure 4-2, only four parameters are needed to express this model because each row of this matrix sums up to one. Let us take one step further and consider the standard Markov model for the case of $t = 3$. The transition probability matrix can be expressed as in Figure 4-3.

	$X_{k-2}X_{k-1}X_k$							
$X_{k-3}X_{k-2}X_{k-1}$	000	001	010	011	100	101	110	111
000	p_{0000}	p_{0001}	0	0	0	0	0	0
001	0	0	p_{0010}	p_{0011}	0	0	0	0
010	0	0	0	0	p_{0100}	p_{0101}	0	0
011	0	0	0	0	0	0	p_{0110}	p_{0111}
100	p_{1000}	p_{1001}	0	0	0	0	0	0
101	0	0	p_{1010}	p_{1011}	0	0	0	0
110	0	0	0	0	p_{1100}	p_{1101}	0	0
111	0	0	0	0	0	0	p_{1110}	p_{1111}

Figure 4-3 The Full Transition Matrix for the Third-Order Markov Chain

Each row sums up to one, and so $8 = 2^3$ parameters are needed for this case. Generalizing this pattern would lead to requiring 2^t parameters, for modeling a Markov chain for which the dependence goes back through t observations. This exponential increase in the number of parameters that are needed to model the Markov dependence, imposes a high level of complexity on modeling the correlation structure of observed data. In the next section, we develop a simpler model with a transition probability matrix that uses a number of parameters which is considerably smaller than 2^t, and that does not increase exponentially with t.

4.3 The Model with Three Parameters

To construct a model for correlated binary data consider the following conditional probability structure for a defective:

(1) Assign a higher value α_2 to $P(\text{defective})$, if a 1 (another defective) has been observed in the past t observations, denote this condition as *one* ; and

(2) assign a lower value α_1 to $P(\text{defective})$, if all previous t observations are 0 (nondefective), denote this condition as *zero*.

The states of the Markov chain correspond to the previous t observations as follows. State 1 corresponds to *zero* (i.e., all previous t observations are 0). For $1 < i < t+2$, state i corresponds to $X_{k-t+i-2} = 1$, $X_{k-t+i-1}=\ldots= X_{k-1} = 0$, and $X_{k-t+i-3}$, …, X_{k-t} either 0 or 1. For example, at state t, X_{k-2}= 1 and X_{k-1}= 0, while rest of the observations could be 0 or 1.

To compare this approach with the direct extension of the standard Markov chain models, consider the case of $t = 3$. We saw in Section 4.2 that a third-order Markov chain requires 8 parameters and that its transition probability matrix has 8 states. Applying rules (1) and (2) to a third-order Markov chain results in assigning the conditional probabilities shown in Table 4-1 such that any set of values of X_{k-1}, X_{k-2}, and X_{k-3} that has at least one 1 is equivalent in terms of the probability for X_k. Table 4-1 can be summarized as Figure 4-4, which shows how our three-parameter model reduces the size of full transition matrix for a third-order Markov chain from eight states to four states.

Table 4-1 Assigned Conditional Probabilities to Binary Observations with Markov Dependence of Order Three

sequence	X_{k-3}	X_{k-2}	X_{k-1}	$P(X_k = 1 \mid X_{k-1}, X_{k-2}, X_{k-3})$
zero,1	0	0	0	α_1
one,1	0	0	1	α_2
one,1	0	1	0	α_2
one,1	0	1	1	α_2
one,1	1	0	0	α_2
one,1	1	0	1	α_2
one,1	1	1	0	α_2
one,1	1	1	1	α_2

$$
\begin{array}{c}
\\
\underline{X_{k-3}\,X_{k-2}\,X_{k-1}} \\
000 \\ 100 \\ {*10} \\ {**1}
\end{array}
\quad
\begin{array}{cccc}
\multicolumn{4}{c}{\underline{X_{k-2}\,X_{k-1}\,X_k}} \\
000 & 100 & {*10} & {**1}
\end{array}
$$

$$
\begin{bmatrix}
1-\alpha_1 & 0 & 0 & \alpha_1 \\
1-\alpha_2 & 0 & 0 & \alpha_2 \\
0 & 1-\alpha_2 & 0 & \alpha_2 \\
0 & 0 & 1-\alpha_2 & \alpha_2
\end{bmatrix}
$$

Figure 4-4 The Reduced Transition Matrix for the Third-Order Markov Chain (* can be 0 or 1)

4.3.1 The Long Term Probability Structure

For the three-parameter model the long-term probability for a defective item can be expressed as:

$$p = P(X=1) = \sum_{i=1}^{\tau} \pi_i P(X=1 \mid \text{in state } i)$$

$$= \pi_1\alpha_1 + \sum_{i=2}^{\tau} \pi_i \alpha_2 , \tag{4.1}$$

where π_i is the stationary probability of the Markov chain, and $\tau = t+1$. The vector $\boldsymbol{\pi}$ of stationary probabilities must satisfy:

$$\begin{bmatrix}\pi_1 & \pi_2 & \pi_3 & \cdots & \pi_\tau\end{bmatrix} = \begin{bmatrix}\pi_1 & \pi_2 & \pi_3 & \cdots & \pi_\tau\end{bmatrix} \begin{bmatrix} 1-\alpha_1 & 0 & 0 & \cdots & \alpha_2 \\ 1-\alpha_2 & 0 & 0 & \cdots & \alpha_2 \\ 0 & 1-\alpha_2 & 0 & \cdots & \alpha_2 \\ 0 & 0 & 1-\alpha_2 & \cdots & \alpha_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \alpha_2 \end{bmatrix}$$

and this gives the following set of equations:

$$\begin{aligned} \pi_1 &= \pi_1(1-\alpha_1) + \pi_2(1-\alpha_2) \\ \pi_2 &= \pi_3(1-\alpha_2) \\ \pi_3 &= \pi_4(1-\alpha_2) \\ &\vdots \\ \pi_\tau &= \pi_1\alpha_1 + \pi_2\alpha_2 + \pi_3\alpha_2 + \cdots + \pi_\tau\alpha_2. \end{aligned} \tag{4.2}$$

Notice that the last equation in (4.2) shows that the stationary probability corresponding to the last state is equal to p.

To find the values of π_i in terms of α_1, α_2, and t, we take the following steps. Solving the first equation for π_2 gives

$$\pi_2 = \pi_1 \frac{\alpha_1}{1-\alpha_2}. \tag{4.3}$$

Replacing from (4.3) in the second equation gives π_3 in terms of π_1

$$\pi_3 = \pi_1 \frac{\alpha_1}{(1-\alpha_2)^2}. \tag{4.4}$$

Continuing to replace π_{i-1} (in terms of π_1) in π_i gives

$$\pi_i = \pi_1 \frac{\alpha_1}{(1-\alpha_2)^{i-1}}, \tag{4.5}$$

for $i = 2,3,\ldots,\tau$. Now, use the fact that $\sum_{i=1}^{\tau} \pi_i = 1$ and write

$$\pi_1 = 1 - \sum_{i=2}^{\tau} \pi_i = 1 - \sum_{i=2}^{\tau} \pi_1 \frac{\alpha_1}{(1-\alpha_2)^{i-1}}. \tag{4.6}$$

Solving (4.6) for π_1 gives π_1 in terms of α_1, α_2, and t as

$$\pi_1 = \left(1 + \alpha_1 \sum_{i=2}^{\tau} \frac{1}{(1-\alpha_2)^{i-1}}\right)^{-1}. \tag{4.7}$$

Thus, equations (4.5) and (4.7) give all the stationary probabilities for a $t^{th}-$order Markov chain in terms of the parameters of our three-parameter model.

4.3.2 The Serial Correlation

In the last chapter, we saw that the stationary serial correlation for a first-order Markov process is expressed as $\rho = 1 - p_{01} - p_{10} = p_{11} - p_{01}$ and the long-term probability of observing a defective is $p = p_{01}/(p_{01} + p_{10})$. Here, we derive the correlation structure for our three-parameter model.

To find an expression for the correlation, recall that

$$\text{corr}(X_{k-1}, X_k) = \frac{\text{cov}(X_{k-1}, X_k)}{\left[\text{var}(X_{k-1})\,\text{var}(X_k)\right]^{1/2}}, \tag{4.8}$$

where $\text{cov}(X_{k-1}, X_k) = E(X_{k-1}, X_k) - E(X_{k-1})E(X_k)$.

From the definition of the three-parameter model, for states $1,2,\ldots,t$, we have $X_{k-1} = 0$, so

$$E(X_{k-1}X_k \mid X_{k-1} \text{ corresponds to one of the states } 1,2,\ldots,t) = E(0) = 0 .$$

Only for state τ , $X_{k-1} = 1$, so

$$E(X_{k-1}X_k \mid X_{k-1} \text{ corresponds to state } \tau) = E(X_k \mid X_{k-1} = 1) = \alpha_2 .$$

Therefore,

$$\begin{aligned}
&E(X_{k-1}X_k) \\
&= \sum_{i=1}^{\tau} \pi_i \, E(X_{k-1}X_k \mid X_{k-1} \text{ corresponds to state } i) \\
&= \sum_{i=1}^{t} \pi_i \left(0\right) + \pi_\tau \alpha_2 \\
&= \pi_\tau \alpha_2 \\
&= p\,\alpha_2 .
\end{aligned}$$

And so we can write

$$\begin{aligned}
\operatorname{cov}(X_{k-1}, X_k) &= E(X_{k-1}, X_k) - E(X_{k-1})E(X_k) \\
&= p\,\alpha_2 - p^2 .
\end{aligned}$$

To evaluate the denominator of (4.8), notice that $E(X_k^2) = E(X_k = 1) = p$, and $E\left(X_k\right) = p$. Substituting in $\operatorname{var}(X_k) = E(X_k^2) - E(X_k)^2$, gives $\operatorname{var}\left(X_k\right) = p - p^2$. So $\left[\operatorname{var}(X_{k-1}).\operatorname{var}(X_k)\right]^{1/2} = p - p^2$. Consequently, we have

$$corr(X_{k-1}, X_k) = \frac{p\,\alpha_2 - p^2}{p - p^2} ,$$

or

$$\rho = \frac{\alpha_2 - p}{1 - p} , \qquad (4.9)$$

where

$$p = \pi_\tau = \frac{\alpha_1}{\left(1-\alpha_2\right)^t} \pi_1 = \frac{\alpha_1}{\left(1-\alpha_2\right)^t} \left(1 - \alpha_1 \sum_{i=1}^{\tau} \frac{1}{\left(1-\alpha_2\right)^{i-1}} \right)^{-1} . \qquad (4.10)$$

So, for a t^{th} – order Markov chain, we have the expressions for p and ρ in terms of α_1 and α_2. Also, we can express α_1 and α_2 in terms of p and ρ.

$$\alpha_1 = \left(\frac{1}{\left(p\left(1-\alpha_2\right)^t\right)} - \sum_{i=2}^{\tau} \frac{1}{\left(1-\alpha_2\right)^{i-1}} \right)^{-1}, \tag{4.11}$$

where $\alpha_2 = p + \rho(1-p)$. Notice that α_2 is not a function of t, so it only changes with changes in the values of p and/or ρ.

4.3.3 Estimation of the Model Parameters

To use our model in practice, we need to estimate its parameters, α_1, α_2, and t. To find the maximum likelihood estimator for α_1 and α_2, remember that our three-parameter model has a $(t+1)\times(t+1)$ transition matrix. According to Bhat and Lal (1999) "a standard procedure to estimate the elements of the transition probability matrix is by observing transition counts." For estimating the parameters, after the process has reached stability, we take a sample of n binary observations. For the stable Markov process, denote the number of transitions from state i to state j by n_{ij}. Recall that in the three-parameter model, there are four possible probabilities associated with all the sequences of binary data. If t consecutive zeros have been observed, then the probability that the next observation is a zero is $\left(1-\alpha_1\right)$, and the number of times this happens equals to n_{11}. Let us define $n_1 = n_{11}$. If a one is observed after t zeros, the probability that the next observation is one is α_1, and there are $n_{1,t+1}$ of such transitions. Define $n_2 = n_{1,t+1}$. When at least one 1 has been observed among the t previous observations, the probability assigned to observing a zero is $1-\alpha_2$, and to observing a one is α_2. Denote the number of such transitions by $n_3 = \sum_{i=2}^{t+1} n_{i,i-1}$, and $n_4 = \sum_{i=2}^{t+1} n_{i,t+1}$, respectively. Then the likelihood function is

$$L_t = \left(1-\alpha_1\right)^{n_1} \alpha_1^{n_2} \left(1-\alpha_2\right)^{n_3} \alpha_2^{n_4} \tag{4.12}$$

and the maximum likelihood estimators, calculated from the derivatives of the log-likelihood function are $\widehat{\alpha}_1 = n_2/\left(n_1+n_2\right)$, and $\widehat{\alpha}_2 = n_4/\left(n_3+n_4\right)$.

To estimate the value of t, for $t = 1, 2, \ldots$ we find $\widehat{\alpha}_1$, $\widehat{\alpha}_2$, and L_t. Then we choose the value of t that maximizes L_t.

As an example consider the following data set with $n = 60$ observations:

000000000010000101000000000000000000100000000000000000000010

Suppose that $\hat{t} = 2$. In this case the probability of the first two observations cannot be determined. The probability (likelihood) for the last 58 observations from our three-parameter model is

$$(1-\alpha_1)^8 \alpha_1 (1-\alpha_2)^2 (1-\alpha_1)^2 \alpha_1 (1-\alpha_2) \alpha_2 (1-\alpha_2)^2 (1-\alpha_1)^{16} \alpha_1 (1-\alpha_2)^2 (1-\alpha_1)^{19} \alpha_1 (1-\alpha_2)$$

and we have $n_1 = 45$, $n_2 = 2$, $n_3 = 10$, and $n_4 = 1$. So $\hat{\alpha}_1 = 2/47$ and $\hat{\alpha}_2 = 1/11$.

Bhat and Lal (1989) also use ML estimators for their Markov model. In their work, by using the maximum likelihood method, p and ρ are estimated as

$$\hat{p} = \frac{\hat{p}_{01}}{\hat{p}_{01} + \hat{p}_{10}} \text{ and } \hat{\rho} = 1 - \left(\hat{p}_{01} + \hat{p}_{10}\right),$$

where $\hat{p}_{ij} = n_{ij} / n_{i0} + n_{i1}$, n_{ij} is the number of times the process moves from state i to state j for $i, j \in \{0,1\}$. This is a special case of our MLE, for $t = 1$, where we had $\alpha_1 = p_{01}$ and $\alpha_2 = p_{11}$.

MacDonald and Zucchini (1997, p. 144) suggested that the fit of different Markov chain models can be compared based on the unconditional log likelihoods (denoted by L), and AIC and BIC. Recall that these measures of fitness of a model are defined as $\text{AIC} = -2L + 2K$ and $\text{BIC} = -2L + K \log T$, where K is the number of parameters estimated in order to fit a model, and T represents the length of the observation sequence. We leave this subject for future work as an investigation of tests for goodness of fit is beyond the scope of this dissertation.

4.3.4 The Generalized Markov Binary CUSUM

We use the three-parameter model for Markov dependent binary data, developed in the previous section, to establish a CUSUM statistic for such observations that is itself represented as a Markov chain. We have only two values for the conditional probability of a defective (α_1 and α_2), and so we can use the log-likelihood ratio expression that was derived in Chapter 3. We distinguish only the following cases: all of the last t

observations were 0, or there was a 1 observed among the past t observations. To represent the past t observations, define

$$O_k = \begin{cases} 0 & \text{if } t \text{ consecutive zeros are observed,} \\ 1 & \text{otherwise.} \end{cases} \tag{4.13}$$

That is if $X_{k-t} = X_{k-t+1} = \cdots = X_{k-1} = 0$, then $O_k = 0$; otherwise $O_k = 1$. We interchange freely the terms "$O_k = 0$" with "*zero*", and "$O_k = 1$" with "*one*". The current observation, X_k, can take two values, 0 or 1. The CUSUM statistic can be constructed by using the same method as in Section 3.2. Recall the exact log-likelihood based CUSUM statistic for correlated binary data, for a first-order Markov chain, i.e., when $t = 1$, was derived as (see equation (3.10))

$$C_k = \max(0, C_{k-1}) + \ln \frac{f(X_0 \mid p_1)}{f(X_0 \mid p_0)} + \Delta_k, \qquad k = 1, 2, ...$$

where $\Delta_k\big|_{t=1} = X_{k-1}(-q_1 + q_3) + X_k(-q_1 + q_2) - X_{k-1}X_k(-q_1 + q_2 + q_3 - q_4) + q_1$. The values of Δ_k in terms of p_{01} and p_{11} are shown in Table 4-2 below, where the subscript of the subscript refers to the in- or out-of-control values. For example, p_{01_0} is the value of p_{01} for the in-control value of p, i.e., p_0.

Table 4-2 The CUSUM Increment Values in Terms of the Conditional Probabilities

(X_{k-1}, X_k)	Δ_k
$(0,0)$	$q_1 = \ln \frac{1 - p_{01_1}}{1 - p_{01_0}}$
$(0,1)$	$q_2 = \ln \frac{p_{01_1}}{p_{01_0}}$
$(1,0)$	$q_3 = \ln \frac{p_{10_1}}{p_{10_0}}$
$(1,1)$	$q_4 = \ln \frac{1 - p_{10_1}}{1 - p_{10_0}}$

To find the increment for the generalized MBCUSUM, replace X_{k-1} by O_k, p_{11} by α_2, and p_{01} by α_1 (when $t = 1$, we have $p_{01} = \alpha_1$, and $p_{11} = \alpha_2$,). Recall that in the three-parameter model the conditional probabilities of observing a 1 are α_1, after t zeros, and

α_2, when at least another 1 was observed among the previous t observations. The four values of the increment are calculated as a function of parameters α_1 and α_2, in Table 4-3, where the second subscript indicates that the value of the conditional probability has been calculated for the in-control (0), or out-of-control (1) value of p. For example, α_{20} indicates the value of α_2 based on p_0.

Table 4-3 Three-parameter Probabilities and Generalized Increments

O_k	X_k	$P(X_k \mid O_k)$	Δ_k
0	0	$1-\alpha_1$	$\delta_1 = \ln \frac{1-\alpha_{11}}{1-\alpha_{10}}$
0	1	α_1	$\delta_2 = \ln \frac{\alpha_{11}}{\alpha_{10}}$
1	0	$1-\alpha_2$	$\delta_3 = \ln \frac{1-\alpha_{21}}{1-\alpha_{20}}$
1	1	α_2	$\delta_4 = \ln \frac{\alpha_{21}}{\alpha_{20}}$

The increment for the generalized MBCUSUM thus can be expressed as follows:

$$\Delta_k = O_k(-\delta_1 + \delta_3) + X_k(-\delta_1 + \delta_2) - O_k X_k(-\delta_1 + \delta_2 + \delta_3 - \delta_4) + \delta_1 . \qquad (4.14)$$

For $p_1 > p_0$ and constant $\rho > 0$, as verified later in this chapter, the increments do not change signs as a function of p, ρ, and t (see Section 4.4.2.1).

4.3.4.1 Modeling The CUSUM as a Markov Chain

To approximate the CUSUM statistic as a Markov chain, we express the increments in terms of integer multiples of $1/m$, where m is a positive integer. We use the same approach that was developed in Chapter 3 for the first-order Markov chain. In the next section we explain the details of the required calculations for the general case of a t^{th} − order Markov chain in the context of our three-parameter model.

First, we investigate the sign of the increments, which will be useful in establishing the transitions in the $\mathbf{Q}$ matrix. For our range of interest, that is for $0 < p_0 < p_1 < 1$ and $0 < \rho < 1$, it is shown below that δ_1, δ_2, δ_3, and δ_4 have the same sign.

Recall that α_2 as a function of p and ρ was expressed as $\alpha_2 = p(1-\rho)+\rho$, and that α_2 does not depend on t. It can be seen that α_2 increases with p, because

$$\frac{\partial \alpha_2}{\partial p} = 1-\rho > 0, \text{ for } \rho \in (0,1).$$

For the in- and out-of-control values of p, and for the constant and positive serial correlation, we have

$$\alpha_{20} = p_0(1-\rho)+\rho,$$

$$\alpha_{21} = p_1(1-\rho)+\rho, \tag{4.15}$$

and so it holds that $\alpha_{21} > \alpha_{20}$ for $p_1 > p_0$.

In Section 4.3.2, an expression for α_1 , equation (4.11), was derived, which will be useful later in this chapter. Here, we will derive another expression for α_1 that is easier to use for the purposes of this section. Recall that $\pi_\tau = \pi_1\alpha_1/(1-\alpha_2)^t$. Also, notice that the last equation of $\pi = \pi\mathbf{P}$, i.e., $\pi_\tau = \pi_1\alpha_1 + \pi_2\alpha_2 + \pi_3\alpha_2 + \cdots + \pi_\tau\alpha_2$, can be written as $\pi_\tau = \pi_1\alpha_1 + (1-\pi_1)\alpha_2$. So $\pi_1\alpha_1/(1-\alpha_2)^t = \pi_1\alpha_1 + (1-\pi_1)\alpha_2$, which gives

$$\pi_1 = \frac{\alpha_2(1-\alpha_2)^t}{\alpha_1 + (\alpha_2-\alpha_1)(1-\alpha_2)^t}. \tag{4.16}$$

Solving (4.16) for α_1 gives

$$\alpha_1 = \frac{p\alpha_2(1-\alpha_2)^t}{\alpha_2 - p + p(1-\alpha_2)^t}. \tag{4.17}$$

Factoring out the term $p(1-\alpha_2)^t$ from both nominator and denominator of the right hand side and using $\alpha_2 - p = \rho(1-p)$ gives

$$\alpha_1 = \frac{\alpha_2}{1+\dfrac{\rho}{p}\dfrac{1-p}{(1-\alpha_2)^t}}. \tag{4.18}$$

Now, notice that because $\partial\alpha_2/\partial p > 0$, the numerator of (4.18) increases with p. In the denominator, ρ/p decreases as p increases. When p increases $(1-\alpha_2)$ decreases with a slower rate than $(1-p)$. This holds because

$$\frac{\partial \alpha_2}{\partial p} = 1-\rho < 1 \text{, for } 0<\rho<1$$

$$\Rightarrow \frac{\partial \alpha_2}{\partial p} < \frac{\partial p}{\partial p} = 1$$

$$\Rightarrow \frac{\partial(1-\alpha_2)}{\partial p} < \frac{\partial(1-p)}{\partial p} \qquad \text{, for } 0<\alpha_2,\ p<1$$

$$\Rightarrow \frac{\partial(1-\alpha_2)^t}{\partial p} < \frac{\partial(1-p)}{\partial p} \qquad \text{, for } 0<(1-\alpha_2)<1 .$$

So $(1-p)/(1-\alpha_2)^t$ is decreasing as p increases and therefore, the denominator of (4.18) is decreasing with an increase in p, resulting in α_1 increasing with p. Thus $\alpha_{11} > \alpha_{10}$ for $\rho > 0$.

Now, the signs for the four CUSUM increments can be derived as follows:

$$0<\alpha_{10}<\alpha_{11}<1 \Rightarrow \begin{cases} 0<(1-\alpha_{11})<(1-\alpha_{10})<1 \Rightarrow 0<\dfrac{1-\alpha_{11}}{1-\alpha_{10}}<1 \Rightarrow \delta_1 = \ln\dfrac{1-\alpha_{11}}{1-\alpha_{10}}<0 \\ \dfrac{\alpha_{11}}{\alpha_{10}}>1 \Rightarrow \delta_2 = \ln\dfrac{\alpha_{11}}{\alpha_{10}}>0 \end{cases}$$

$$0<\alpha_{20}<\alpha_{21}<1 \Rightarrow \begin{cases} 0<(1-\alpha_{21})<(1-\alpha_{20})<1 \Rightarrow 0<\dfrac{1-\alpha_{21}}{1-\alpha_{20}}<1 \Rightarrow \delta_3 = \ln\dfrac{1-\alpha_{21}}{1-\alpha_{20}}<0 \\ \dfrac{\alpha_{21}}{\alpha_{20}}>1 \Rightarrow \delta_4 = \ln\dfrac{\alpha_{21}}{\alpha_{20}}>0. \end{cases}$$

Next, we want to express the CUSUM statistic in terms of integer multiples of $1/m$. To do so, we need to convert the values of the increment as follows. Define $m = \text{int}(|\delta_1^{-1}|)$; $d_{01} = \text{int}(\delta_2 \times m)$; $d_{10} = \text{int}(|\delta_3| \times m)$; $d_{01} = \text{int}(\delta_4 \times m)$, where $\text{int}(\cdot)$ gives the closest integer value to the argument inside the parenthesis. Expressing the increments in terms of integer multiples of $1/m$ can be now achieved, and so the Markov model form for our CUSUM statistic is shown in Table 4-4 below.

The generalized Markov binary CUSUM is denoted by C_k with increment D_k, and will reduce to the MBCUSUM developed in Chapter 3 for the special case of $t=1$, as shown in the next subsection.

Table 4-4 Increment Values for the Generalized MBCUSUM

O_k	X_k	$P(X_k \mid O_k)$	D_k
0	0	$1-\alpha_1$	$-1/m$
0	1	α_1	d_{01}/m
1	0	$1-\alpha_2$	$-d_{10}/m$
1	1	α_2	d_{11}/m

4.3.5 The MBCUSUM as a Special Case of the Generalized MBCUSUM

In this section, we verify that the generalized Markov binary reduces to the MBCUSUM that was introduced in the previous chapter. Remember that α_2 is a function of p and ρ only, and does not depend on the order of the Markov chain, t. For $t=1$, the value of the second (lower) level probability is

$$\alpha_1 = \left.\frac{p\alpha_2\left(1-\alpha_2\right)^t}{\alpha_2 - p + p\left(1-\alpha_2\right)^t}\right|_{t=1} = \frac{p\alpha_2\left(1-\alpha_2\right)}{\alpha_2 - p + p\left(1-\alpha_2\right)} = \frac{p\alpha_2\left(1-\alpha_2\right)}{\alpha_2 - p\alpha_2} = p\frac{1-\alpha_2}{1-p} = p\left(1-\rho\right) = p_{01}.$$

Table 4-5 The MBCUSUM ($t=1$) as a Special Case

X_{k-1}, X_k	Δ_k	$P(X_k \mid X_{k-1})$	transition	i
0,0	$\ln\frac{1-p_{01_1}}{1-p_{01_0}}$	p_{00}	$1 \to 1$	
			$i \to i-2$	$3,5,\ldots,2H-1$
0,1	$\ln\frac{p_{01_1}}{p_{01_0}}$	p_{01}	$i \to i+2d_{01}+2$	$1,3,\ldots,2(H-d_{11})-1$
1,0	$\ln\frac{1-p_{11_1}}{1-p_{11_0}}$	p_{10}	$2 \to 1$	
			$i \to i-2d_{10}-1$	$2d_{10}+2, 2d_{10}+4,\ldots,2H$
1,1	$\ln\frac{p_{11_1}}{p_{11_0}}$	p_{11}	$i \to i+2d_{11}$	$2,4,\ldots,2(H-d_{11})$

The long term probability of observing a defective p, can be derived as follows:

$$\begin{aligned} p &= \pi_1\alpha_1 + \pi_2\alpha_2 \\ &= \frac{\alpha_2(1-\alpha_2)^1}{\alpha_1 + (\alpha_1-\alpha_2)(1-\alpha_2)^1}\alpha_1 + \frac{\alpha_1\alpha_2(1-\alpha_2)^0}{\alpha_1 + (\alpha_1-\alpha_2)(1-\alpha_2)^1}\alpha_2 \\ &= \frac{\alpha_1}{1-\alpha_2+\alpha_1} \\ &= \frac{p_{01}}{p_{01}+p_{10}}, \end{aligned}$$

where $p_{10} = 1 - p_{11} = 1 - \alpha_2$.

From substituting p_{01} for α_1, and p_{11} for α_2, and for $t = 1$ the generalized BMCUSUM increments, δ_i, are equal to q_i for $i = 1,2,3,4$, where q_i's are the increment values for the simple BMCUSUM derived in Chapter 3. The increments, conditional probabilities and transitions are shown in Table 4-5.

4.4 The Transition Probability Matrix

In this section, we first present a numerical example of the CUSUM and transition matrix for the case of t = 3 with specific values of the parameters, followed by another example with t = 2 in terms of nonspecific values for the other parameters, and finally we give the general case of $t \geq 1$ in terms of nonspecific values of all the model parameters.

4.4.1.1 A Numerical Example for the Generalized MBCUSUM

Consider a case in which $p_0 = 0.010$, $p_1 = 0.025$, $\rho = 0.05$, and $t = 3$. The higher and lower values of in-control probabilities are

$$\alpha_{20} = p_0 + \rho(1-p_0) = 0.0595,$$

$$\alpha_{10} = \left.\frac{p_0\alpha_{20}(1-\alpha_{20})^t}{\alpha_{20} - p_0 + p_0(1-\alpha_{20})^t}\right|_{t=3} = 0.0086.$$

For the out-of-control value of p, we obtain $\alpha_{21} = 0.0738$ and $\alpha_{11} = 0.0214$. From Table 4-3, the first increment value is $\delta_1 = \ln\left((1-\alpha_{11})/(1-\alpha_{10})\right) = -0.0130$, and its reciprocal equals -76.9231, which we convert to the closest integer and take the absolute value to get $m = 77$. The second value of the CUSUM increment is $\delta_2 = \ln\left(\alpha_{11}/\alpha_{10}\right) =$ 0.9140, which times 77 is equal to 70.378, and rounded to 70. The third increment value is calculated as $\delta_3 = \ln\left((1-\alpha_{21})/(1-\alpha_{20})\right) = -0.0153$, then multiplied by 77 gives -1.1781, and rounded to -1. The last value of the CUSUM increment is $\delta_4 = \ln\left(\alpha_{21}/\alpha_{20}\right) =$ 0.2147, multiplied by 77 equals 16.5319, and the closest integer value to that is 17. So the values of the CUSUM increment converted to integer multiples of $1/77$ corresponding to the values of (O_k, X_k) are as follows:

$$\{(zero,0),(zero,1),(one,0),(one,1)\} \rightarrow \{-1/77, 70/77, -1/77, 17/77\}.$$

The transition probabilities and values of increment for this Markov chain are summarized in Table 4-6.

Table 4-6 The Probabilities and Increments for the Generalized MBCUSUM ($t = 3$ and $m = 77$)

O_k	X_k	$P(X_k \mid O_k)$	Δ_k	D_k
0	0	0.9914	-0.0130	$\frac{-1}{77} = -0.0129$
0	1	0.0086	0.9140	$\frac{70}{77} = 0.9091$
1	0	0.9405	-0.0153	$\frac{-1}{77} = -0.0129$
1	1	0.0595	0.2147	$\frac{17}{77} = 0.2208$

The possible values of C_k are integer multiples of $1/77$. For $H = 100$, the control limit h is $400/77$, where H is equal to mh as was defined in Chapter 3. Then the values of C_k that are below h are $0, 1/77, 2/77, \ldots, 399/77$. The definition of C_{k-1} allows for a possible value of -1/77, but we can combine this state with the state for 0. Thus, the number of transient states for our control statistic that is expressed as a Markov chain is $\tau mh = 400$, where $\tau = t+1$. This is obtained from counting the steps of size $1/77$ starting at 0 and ending at $h - 1/77$. Since there are only three possible values for the increment (for 00 and 10, the increment takes the same value in this example, as shown in Table 4-6), after each observation of X_k, $k \ge 1$, the change in the value of C_k will be one of the three followings: (1) down 1 $\tau-$state ($-1/77$); (2) up 70 $\tau-$states ($+70/77$); and (3) up 17 $\tau-$state ($+17/77$), where by $\tau-$state we mean all the states that correspond to each value of C_k and for this example $\tau = 4$.

The transition matrix consists of 100^2 four by four blocks, each corresponding to a value of the CUSUM statistic. In the first 4×4 block of the transition matrix, both the previous and the current values of the CUSUM statistic are equal to zero. This can occur when the actual value of C_k was originally negative, but then we have reset it to zero. This corresponds to two situations, for which the increment takes a negative value, i.e., $(zero, 0)$ corresponding to probability $1-\alpha_1$, and $(one, 0)$ corresponding to probability $1-\alpha_2$. The latter corresponds to three instances as was seen in Table 4-6.

The first increment $-1/77$ can be added to C_{k-1} when it has the value $1/77$ and produce the value of zero for C_k. This would result in another 4×4 block under the first one that looks exactly the same as the previous one. Thereafter, this increment will be added to every value of C_{k-1} and gives a value for C_k such that $C_{k-1} - C_k = 1/77$. This stops when C_{k-1} reaches its maximum value of $99/77$ and $C_k = 98/77$, which creates the last block of this sort.

The second increment $70/77$ is first added to $C_{k-1} = 0$ at state $i = 1$, and gives $C_k = 70/77$ at state $j = 4 \times 77 \times (70/77) + 4 = 284$, when the observations are $(zero, 1)$ with probability α_1. Whereas, for the rest of observations at states $i = 2, 3, 4$, the forth increment value $17/77$ is added to $C_{k-1} = 0$. This creates a transition to states $j = 4 \times 77 \times (17/77) + i$, with probability α_2. These additions are possible until C_{k-1} reaches its maximum value of $99/77$. So for the second increment $70/77$, we have $C_{k-1} = 99/77 - 70/77 = 29/77$. The transition is from state $i = 4 \times 77 \times (29/77) + 1 = 117$ to state $j = 400$, with probability α_1. Finally, the second increment $17/77$, is last added to $C_{k-1} = 99/77 - 17/77 = 82/77$ at corresponding states 330, 331, and 332 and goes to the last state, with probability α_2. These results are summarized below.

$$p_{1\to 1} = 1-\alpha_1$$

$$p_{2\to 1} = 1-\alpha_2$$

$$p_{3\to 2} = 1-\alpha_2$$

$$p_{4\to 2} = 1-\alpha_2$$

$$p_{i\to i-4} = P(zero,\ 0) = 1-\alpha_1 \qquad i = 5,9,...,397$$

$$p_{i\to i+283} = P(zero,1) = \alpha_1 \qquad i = 1,5,...,117$$

$$p_{i\to i-5} = P(one,0) = 1-\alpha_2 \qquad i = 6,10,...,398;\ 7,11,...,399;\ 8,12,...,400$$

$$p_{i\to j} = P(one,1) = \alpha_2 \qquad i = \begin{cases} 2,6,...,330 & ,\text{for } j = i+70; \\ 3,7,...,331 & ,\text{for } j = i+69; \\ 4,8,...,332 & ,\text{for } j = i+68. \end{cases} \tag{4.19}$$

		j	1	2	3	4	⋯	69	70	71	72	⋯	281	282	283	284	⋯	393	394	395	396	397	398	399	400
		X_{k-2}	0	1	*	*	⋯	0	1	*	*	⋯	0	1	*	*	⋯	0	1	*	*	0	1	*	*
		X_{k-1}	0	0	1	*	⋯	0	0	1	*	⋯	0	0	1	*	⋯	0	0	1	*	0	0	1	*
		X_k	0	0	0	1	⋯	0	0	0	1	⋯	0	0	0	1	⋯	0	0	0	1	0	0	0	1
i	$X_{k-3}X_{k-2}X_{k-1}$	$C_{k-1} \backslash C_k$	0	0	0	0	⋯	[	17/77		]	⋯	[	70/77		]	⋯	[	98/77		]	[	99/77		]
1	000	0	$1-\alpha_1$	0	0	0	⋯	0	0	0	0	⋯	0	0	0	α_1	⋯	0	0	0	0	0	0	0	0
2	100	0	$1-\alpha_2$	0	0	0	⋯	0	0	0	α_2	⋯	0	0	0	0	⋯	0	0	0	0	0	0	0	0
3	*10	0	0	$1-\alpha_2$	0	0	⋯	0	0	0	α_2	⋯	0	0	0	0	⋯	0	0	0	0	0	0	0	0
4	**1	0	0	0	$1-\alpha_2$	0	⋯	0	0	0	α_2	⋯	0	0	0	0	⋯	0	0	0	0	0	0	0	0
5	000	1/77	$1-\alpha_1$	0	0	0	⋯	0	0	0	0	⋯	0	0	0	0	⋯	0	0	0	0	0	0	0	0
6	100	1/77	$1-\alpha_2$	0	0	0	⋯	0	0	0	0	⋯	0	0	0	0	⋯	0	0	0	0	0	0	0	0
7	*10	1/77	0	$1-\alpha_2$	0	0	⋯	0	0	0	0	⋯	0	0	0	0	⋯	0	0	0	0	0	0	0	0
8	**1	1/77	0	0	$1-\alpha_2$	0	⋯	0	0	0	0	⋯	0	0	0	0	⋯	0	0	0	0	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
117	000	29/77	0	0	0	0	⋯	0	0	0	0	⋯	0	0	0	0	⋯	0	0	0	0	0	0	0	α_1
118	100	29/77	0	0	0	0	⋯	0	0	0	0	⋯	0	0	0	0	⋯	0	0	0	0	0	0	0	0
119	*10	29/77	0	0	0	0	⋯	0	0	0	0	⋯	0	0	0	0	⋯	0	0	0	0	0	0	0	0
120	**1	29/77	0	0	0	0	⋯	0	0	0	0	⋯	0	0	0	0	⋯	0	0	0	0	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
329	000	82/77	0	0	0	0	⋯	0	0	0	0	⋯	0	0	0	0	⋯	0	0	0	0	0	0	0	0
330	100	82/77	0	0	0	0	⋯	0	0	0	0	⋯	0	0	0	0	⋯	0	0	0	0	0	0	0	α_2
331	*10	82/77	0	0	0	0	⋯	0	0	0	0	⋯	0	0	0	0	⋯	0	0	0	0	0	0	0	α_2
332	**1	82/77	0	0	0	0	⋯	0	0	0	0	⋯	0	0	0	0	⋯	0	0	0	0	0	0	0	α_2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
397	000	99/77	0	0	0	0	⋯	0	0	0	0	⋯	0	0	0	0	⋯	$1-\alpha_1$	0	0	0	0	0	0	0
398	100	99/77	0	0	0	0	⋯	0	0	0	0	⋯	0	0	0	0	⋯	$1-\alpha_2$	0	0	0	0	0	0	0
399	*10	99/77	0	0	0	0	⋯	0	0	0	0	⋯	0	0	0	0	⋯	0	$1-\alpha_2$	0	0	0	0	0	0
400	**1	99/77	0	0	0	0	⋯	0	0	0	0	⋯	0	0	0	0	⋯	0	0	$1-\alpha_2$	0	0	0	0	0

Figure 4-5 The Transition Matrix for the Generalized MBCUSUM with third-Order Dependence

For $H = 100$, a signal is produced by the Markov binary CUSUM control chart, when $C_k \geq 100/77$. The transition probability matrix has 400 transient states as shown in Figure 4-5.

4.4.1.2 The Generalized MBCUSUM when $t = 2$ and $0 < p, \rho < 1$

To demonstrate the way in which our general formulation works, here, we consider the Markov process with second order dependence, and derive the transition probability matrix from the generalized model for the binary CUSUM control statistic. The conditional probability of observing a 1 for second-order Markov dependent binary observations are shown in Table 4-7.

Table 4-7 Assigned Probabilities to the Full Form for $t = 2$

sequence	X_{k-2}	X_{k-1}	$P(X_k = 1 \mid X_{k-1}, X_{k-2})$
$zero,1$	0	0	α_1
$one,1$	0	1	α_2
$one,1$	1	0	α_2
$one,1$	1	1	α_2

$$\begin{array}{c|ccc} X_{k-2}X_{k-1} \backslash X_{k-1}X_k & 00 & 10 & *1 \\ \hline 00 & 1-\alpha_1 & 0 & \alpha_1 \\ 10 & 1-\alpha_2 & 0 & \alpha_2 \\ *1 & 0 & 1-\alpha_2 & \alpha_2 \end{array}$$

Figure 4-6 The Reduced Transition Matrix for a Second-Order Markov Chain

Figure 4-7 gives the transition matrix for the CUSUM statistic modeled as a Markov chain (the MBCUSUM) and is summarized in Table 4-9.

Table 4-8 Transitions for the Generalized MBCUSUM with $t = 2$ and arbitrary m

obs	increment	prob	transition	i
$zero, 0$	$-1/m$	$1-\alpha_1$	$1 \to 1$	
			$i \to i-3$	$4, 7, \ldots, 3H-2$
$zero, 1$	d_{01}/m	α_1	$i \to i+3d_{01}+2$	$1, 5, \ldots, 3(H-d_{01}-1)+1$
$one, 0$	$-d_{10}/m$	$1-\alpha_2$	$i \to i-1$	$2, 3$
			$i \to i-3d_{10}-1$	$3(d_{10}+k-1)+2,\ 3(d_{10}+k-1)+3, \ldots, 3(d_{10}+k)$ $k = 1, 2, \ldots, H-d_{10}$
$one, 1$	d_{11}/m	α_2	$i+(k-1) \to i+3d_{11}+1$	$2, 5, \ldots, 3(H-d_{11}-1)+2$ $k = 1, 2$

		j	1	2	3	⋯	[	$3d_{11}$	]	⋯	[	$3d_{01}$	]	⋯	[	$3(H-d_{10})$	]⋯	[			$3H$		]	
							$\cdots j+1$	$j+2$	$j+3 \cdots j+1$			$j+2$	$j+3 \cdots j-2$			$j-1$	$j \cdots j-5$			$j-4$	$j-3$	$j-2$	$j-1$	$3H$
		$X_{k-1}X_k$	00	10	*1	⋯	00	10	*1	⋯	00	10	*1	⋯	00	10	*1	⋯	00	10	*1	00	10	*1
		C_k	0	0	0	⋯	[	d_{11}/m	]	⋯	[	d_{01}/m	]	⋯	[	$(H-1-d_{10})/m$	]⋯		[	$(H-2)/m$	]	[	$(H-1)/m$	]
i	$X_{k-2}X_{k-1}$	C_{k-1}																						
1	00	0	A	0	0	⋯	0	0	0	⋯	0	0	α_1	⋯	0	0	0	⋯	0	0	0	0	0	0
2	10	0	B	0	0	⋯	0	0	α_2	⋯	0	0	0	⋯	0	0	0	⋯	0	0	0	0	0	0
3	*1	0	0	B	0	⋯	0	0	α_2	⋯	0	0	0	⋯	0	0	0	⋯	0	0	0	0	0	0
4	00	$1/m$	A	0	0	⋯	0	0	0	⋯	0	0	0	⋯	0	0	0	⋯	0	0	0	0	0	0
5	10	$1/m$	0	0	0	⋯	0	0	0	⋯	0	0	0	⋯	0	0	0	⋯	0	0	0	0	0	0
6	*1	$1/m$	0	0	0	⋯	0	0	0	⋯	0	0	0	⋯	0	0	0	⋯	0	0	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋯	⋮	⋮	⋮	⋯	⋮	⋮	⋮	⋯	⋮	⋮	⋮	⋯	⋮	⋮	⋮	⋮	⋮	⋮
$3d_{10}+1$	00	d_{10}/m	0	0	0	⋯	0	0	0	⋯	0	0	0	⋯	0	0	0	⋯	0	0	0	0	0	0
$3d_{10}+2$	10	d_{10}/m	B	0	0	⋯	0	0	0	⋯	0	0	0	⋯	0	0	0	⋯	0	0	0	0	0	0
$3d_{10}+3$	*1	d_{10}/m	0	B	0	⋯	0	0	0	⋯	0	0	0	⋯	0	0	0	⋯	0	0	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋯	⋮	⋮	⋮	⋯	⋮	⋮	⋮	⋯	⋮	⋮	⋮	⋯	⋮	⋮	⋮	⋮	⋮	⋮
$3(H-d_{01})-2$	00	$\frac{H-1-d_{01}}{m}$	0	0	0	⋯	0	0	0	⋯	0	0	0	⋯	0	0	0	⋯	0	0	0	0	0	α_1
$3(H-d_{01})-1$	10		0	0	0	⋯	0	0	0	⋯	0	0	0	⋯	0	0	0	⋯	0	0	0	0	0	0
$3(H-d_{01})$	*1		0	0	0	⋯	0	0	0	⋯	0	0	0	⋯	0	0	0	⋯	0	0	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋯	⋮	⋮	⋮	⋯	⋮	⋮	⋮	⋯	⋮	⋮	⋮	⋯	⋮	⋮	⋮	⋮	⋮	⋮
$3(H-d_{11})-2$	00	$\frac{H-1-d_{11}}{m}$	0	0	0	⋯	0	0	0	⋯	0	0	0	⋯	0	0	0	⋯	0	0	0	0	0	0
$3(H-d_{11})-1$	10		0	0	0	⋯	0	0	0	⋯	0	0	0	⋯	0	0	0	⋯	0	0	0	0	0	α_2
$3(H-d_{11})$	*1		0	0	0	⋯	0	0	0	⋯	0	0	0	⋯	0	0	0	⋯	0	0	0	0	0	α_2
⋮	⋮	⋮	⋮	⋮	⋮	⋯	⋮	⋮	⋮	⋯	⋮	⋮	⋮	⋯	⋮	⋮	⋮	⋯	⋮	⋮	⋮	⋮	⋮	⋮
$3H-2$	00	$\frac{H-1}{m}$	0	0	0	⋯	0	0	0	⋯	0	0	0	⋯	0	0	0	⋯	A	0	0	0	0	0
$3H-1$	10		0	0	0	⋯	0	0	0	⋯	0	0	0	⋯	B	0	0	⋯	0	0	0	0	0	0
$3H$	*1		0	0	0	⋯	0	0	0	⋯	0	0	0	⋯	0	B	0	⋯	0	0	0	0	0	0

Figure 4-7 Transition Matrix for the Generalized MBCUSUM with $t = 2$

($\alpha_1 = P(1 \mid zero)$, $\mathrm{A} = 1-\alpha_1$, $\alpha_2 = P(1 \mid one)$, $\mathrm{B} = 1-\alpha_2$)

4.4.1.3 The Transition Probability Matrix for the Generalized MBCUSUM

The possible values of C_k are integer multiples of $1/m$. So the control limit h can be taken to be an integer multiple of $1/m$, when m is an integer. Then the values of C_k that are below h are $0, 1/m, 2/m, \ldots, h-1/m$. The definition of C_k allows for a possible value of $-1/m$, but we can combine this state with the state for $C_k = 0$. Thus, the number of transient states for our approximated control statistic that is expressed as a Markov chain is τmh, where $\tau = t+1$ (this is obtained from counting the steps of size $1/m$ starting at 0 and ending at $h-1/m$). The factor τ comes from the fact that we need more states to account for the dependence between X_k and the previous t observations, which requires exactly τ states for each value of the CUSUM statistic. Since there are only four possible values for the increment (see Table 4-4), after each observation of X_k, $k \geq 1$, the change in the value of C_k will be one of the four followings: (1) down 1 $\tau -$state; (2) up d_{01} $\tau -$states ; (3) down d_{10} $\tau -$states; and (4) up d_{11} $\tau -$states.

The transition matrix is a square matrix of dimension τH, where $H = mh$. This matrix can be pictured as $\tau \times \tau$ blocks with nonzero elements that correspond to the different values of the increment. (Notice the 4×4 blocks in the transition matrix in Section 4.4.2.4). For each value of C_k, the τ corresponding states can be derived from the following equation

$$C_k = \frac{\operatorname{int}\left(\dfrac{i-1}{\tau}\right)}{m}, \tag{4.20}$$

and this yields the state numbers $i = \tau mC_k + 1, \tau mC_k + 2, \ldots, \tau mC_k + \tau.$

The top-left $\tau \times \tau$ block of transition probabilities corresponds to $C_{k-1} = 0$ and $C_k = 0$, and to states $1 \leq i, j \leq \tau$. This block reflects the fact that the term '$\max\left(0, C_{k-1}\right)$' resets the negative value of the CUSUM statistic to zero. Element (1,1) of this block corresponds to $O_k = 0$, which assigns $1-\alpha_1$ to the conditional probability of next observation being also 0, i.e., $X_k = 0$. The other nonzero elements of this block correspond to $O_k = 1$ that assigns probability $1-\alpha_2$ to observing $X_k = 0$. Thus, the first $\tau \times \tau$ block of the $\mathbf{Q}$ matrix is formed by

$$p_{1\to 1} = 1-\alpha_1$$

$$p_{i\to i-1} = 1-\alpha_2\ ,\ i = 2,3,\ldots,\tau\ . \tag{4.21}$$

The $\tau\times\tau$ block corresponding to states $\tau+1\le i\le 2\tau$ and $1\le j\le\tau$, also corresponds to $C_{k-1} = 1/m$ and $C_k = 0$. These value of the CUSUM increment that links these CUSUM statistic values equals $-1/m$. This block has only one nonzero element corresponding to observing (*zero*, 0). The same block corresponds to all values of the CUSUM statistic that satisfy $C_k - C_{k-1} = -1/m$. That is it will appear until the last row $(i = \tau H)$, and the column for which $j = \tau H - \tau - 1$. These blocks can be created through the following formulas:

$$p_{i\to i-\tau} = 1-\alpha_1\ ,\ i = \tau+1, 2\tau+1,\ldots,\tau H - t\ . \tag{4.22}$$

Now consider the $\tau\times\tau$ blocks corresponding to $C_k - C_{k-1} = d_{01}/m$. For example, consider this block for the states $1\le i\le\tau$ and $\tau d_{01}+1\le j\le \tau d_{01}+\tau$. At the first state $(i=1)$, when d_{01}/m is added to $C_{k-1}=0$ a transition is made to where $C_k = d_{01}/m$ for which $j = \tau d_{01}+\tau$. The maximum value for C_k is $h-1/m$, so the last state at which d_{01}/m can be added to C_{k-1} is where $C_{k-1} = \max(C_k) - d_{01}/m$, that is $i = \tau H - \tau d_{01} - t$. These $\tau\times\tau$ blocks have only one nonzero element corresponding to the conditional probability of observing a 1 after t zeros. We have

$$p_{i\to i+\tau d_{01}+t} = \alpha_1\ ,\ i = 1,\ \tau+1,\ 2\tau+1,\ \ldots,\ \tau(H-d_{01})-t\ . \tag{4.23}$$

The $\tau\times\tau$ blocks that correspond to $C_{k-1} - d_{10}/m = C_k$ have three nonzero elements equal to $1-\alpha_2$. For example consider $C_{k-1} = d_{10}/m$ and $C_k = 0$. These CUSUM statistic values correspond to the following states: $(i,j) = (\tau d_{10}+2,1),(\tau d_{10}+3,2),\ldots,(\tau d_{10}+\tau,t)$. The last time d_{10}/m can be subtracted from C_{k-1} is when C_{k-1} takes its maximum value of $h-1/m$. This results in the CUSUM value of $C_k = \max(C_{k-1}) - d_{10}/m = h - 1/m - d_{10}/m$. These CUSUM statistic values correspond to the last t states of the **Q** matrix, i.e., $i = \tau H - t + 1,\ \tau H - t + 2,\ \ldots,\tau H$, and $j = \tau H - \tau d_{10} - t, \tau H - \tau d_{10} - t + 1, \ldots, \tau H - \tau d_{10}$, respectively. The corresponding transition probabilities are

$$p_{i\to i-\tau d_{10}-1} = 1-\alpha_2\ ,$$

$$i = \tau d_{10} + (k-1)\tau + 2,\ \tau d_{10} + (k-1)\tau + 3,\ldots,\tau d_{10} + (k-1)\tau + \tau\ ,$$

$$k = 1,2,\ldots,H - d_{10}\ . \tag{4.24}$$

For $C_{k-1} + d_{11}/m = C_k$, corresponding $\tau \times \tau$ blocks take value α_2 in all rows of the last column except for the first row. For example consider states $i = 2, 3, \ldots, \tau$, then the value of C_{k-1} is zero and $C_k = d_{11}/m$, which corresponds to state $j = \tau d_{11} + \tau$. The last block of this sort corresponds to $C_{k-1} = \max(C_k) - d_{11}/m$. The transition probabilities for these blocks are

$$p_{i+(k-1)\to i+2d_{11}} = \alpha_2,\ i = 2, \tau+2, \ldots, (H-d_{11})\tau - t - 1,\ \ k = 1, 2, \ldots, t-1. \qquad (4.25)$$

These results are summarized in Table 4-10. Notice that if the increment takes equal values for different sequences of observations, say (*zero*,0) and (*one*,0), then the blocks created by these values will overlap. (In Section 4.4.1.1 for $t = 3$, we saw an instance of a case for which the first and third increment values were the same.)

Table 4-9 Transitions for the Generalized Markov Binary CUSUM

obs	increment	prob	transition	i
$zero, 0$	$-1/m$	$1-\alpha_1$	$1 \to 1$	
			$i \to i - \tau$	$\tau+1, 2\tau+1, \ldots, \tau H - t$
$zero, 1$	d_{01}/m	α_1	$i \to i + \tau d_{01} + t$	$1,\ \tau+1,\ 2\tau+1,\ \ldots,\ \tau(H-d_{01}) - t$
$one, 0$	$-d_{10}/m$	$1-\alpha_2$	$i \to i - 1$	$2, 3, \ldots, \tau$
			$i \to i - \tau d_{10} - 1$	$\tau(d_{10}+k-1)+2, \tau(d_{10}+k-1)+3, \ldots, \tau(d_{10}+k)$ $k = 1 : H - d_{10}$
$one, 1$	d_{11}/m	α_2	$i + (k-1) \to i + \tau d_{11} + t - 1$	$2, \tau+2, \ldots, \tau(H - d_{11} - 1) + 2$ $k = 1 : t$

4.5 Generalization of the Three-Parameter Model

Recall that for a t^{th} – order Markov chain every $t+1$ consecutive observations are correlated, and we denoted $t+1$ by τ. To generalize the three-parameter model, for each value of $t \geq 1$, consider a model with the following properties:

1. It maps all the 2^t sequences of binary data into η levels, where $2 \leq \eta \leq \tau$;

2. It assigns a value α_l to the conditional probability of observing a defective at level l, for $1 \le l \le \eta$, such that $\alpha_1 < \alpha_2 < \cdots < \alpha_\eta$; and
3. It generates a reduced transition matrix with τ states.

The assignment of α_l is based on the idea that observing a defective increases the probability of observing another defective. Thus if there is no defective in the last t observations, assign the lowest value of α_l to $P(X_k = 1 \,|\, \text{at level } l)$, and call this level one (denote by $l = 1$). If $\eta = 2$, that is if we want only two levels, then the conditional probability of observing a defective in all remaining cases is $\alpha_2 > \alpha_1$, and the assignment is complete. Otherwise, if $t-1$ previous observations are nondefectives that come after one defective, assign $\alpha_2 > \alpha_1$ to $P(X_k = 1 \,|\, l = 2)$. If $\eta = 3$, then the assigned value to the conditional probability of observing a defective in all remaining cases is $\alpha_3 > \alpha_2$. The probability assignment is complete. This algorithm can be continued until the case in which the previous observation is equal to one, which corresponds to the case of $\eta = \tau$, and the highest probability is assigned to that last level, α_η. For example, a third-order Markov chain can be modeled at 2, 3, or 4 levels as illustrated in Table 4-11.

Now, we have a general model for correlated binary data that requires $\eta + 1$ parameters for a t^{th} – order Markov chain, where η is the number of different conditional probabilities that are assigned to observing a defective and $2 \le \eta \le \tau$. We name this model the *multi-level model* and denote it by MLM. When the number of levels is two, the multi-level model reduces to the three-parameter model developed in Section 4.3. The MLM gives a transition matrix of dimension $\tau \times \tau$ for a t^{th} – order Markov chain. The transition matrix for a third-order Markov chain is 4×4 and is shown in Figure 4-8 for $\eta = 4$, where $1 \le i \le \tau$ represents the state in the Markov chain.

		$X_{k-2}X_{k-1}X_k$			
i	$X_{k-3}X_{k-2}X_{k-1}$	000	100	*10	**1
1	000	$1-\alpha_1$	0	0	α_1
2	100	$1-\alpha_2$	0	0	α_2
3	*10	0	$1-\alpha_3$	0	α_3
4	**1	0	0	$1-\alpha_4$	α_4

Figure 4-8 The Reduced Transition Matrix for a Third–Order Markov Chain (* can be 0 or 1)

The MLM with four levels ($\eta = 4$) reduces to the MLM with three levels by replacing α_4 with α_3, and that the three-parameter model is generated by further replacing α_3 with α_2. Also, notice that to obtain the MLM with three levels for the second-order Markov chain, α_2 and α_1 should collapse into α_1, α_3 should be replaced with α_2, and α_4 with α_3. Table 4-11 illustrates these relationships.

Table 4-10 All Possible Conditional Probabilities by MLM for Binary Observations with Markov Dependence of Order Three

X_{k-3}	X_{k-2}	X_{k-1}	X_k	$P(X_k \mid l)$	$P(X_k \mid l)$	$P(X_k \mid l)$
				$\eta = 4$	$\eta = 3$	$\eta = 2$
0	0	0	1	α_1	α_1	α_1
0	0	1	1	α_4	α_3	α_2
0	1	0	1	α_3	α_3	α_2
0	1	1	1	α_4	α_3	α_2
1	0	0	1	α_2	α_2	α_2
1	0	1	1	α_4	α_3	α_2
1	1	0	1	α_3	α_3	α_2
1	1	1	1	α_4	α_3	α_2

4.5.1 The Long Term Probability Structure

For the multi-level setting, the long-term probability for a defective item can be expressed as:

$$p = P(X = 1) = \sum_{i,l} \pi_i^l P(X = 1 \mid \text{at level } l), \tag{4.26}$$

where π_i^l is the stationary probability of the Markov chain corresponding to state i, and level l. The states of the Markov chain are defined in the same way as the three-parameter model as follows. State 1 corresponds to *zero* (i.e., all previous t observations are 0). For $1 < i < t+2$, state i corresponds to $X_{k-t+i-2} = 1$, $X_{k-t+i-1} = \ldots = X_{k-1} = 0$, and $X_{k-t+i-3}$, …, X_{k-t} either 0 or 1. There are $\tau = t+1$ states and η levels. For example, when $t = 3$ and $\eta = 3$, we have $p = P(X = 1) = \pi_1^1\alpha_1 + \pi_2^2\alpha_2 + \pi_3^2\alpha_2 + \pi_4^3\alpha_3$. The vector of stationary probabilities must satisfy $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$. First, consider the case of $\eta = \tau$. We have

$$\begin{bmatrix} \pi_1^1 & \pi_2^2 & \pi_3^3 & \cdots & \pi_\tau^\tau \end{bmatrix} = \begin{bmatrix} \pi_1^1 & \pi_2^2 & \pi_3^3 & \cdots & \pi_\tau^\tau \end{bmatrix} \begin{bmatrix} 1-\alpha_1 & 0 & \cdots & 0 & \alpha_1 \\ 1-\alpha_2 & 0 & \cdots & 0 & \alpha_2 \\ 0 & 1-\alpha_3 & \cdots & 0 & \alpha_3 \\ 0 & 0 & \cdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 1-\alpha_\tau & \alpha_\tau \end{bmatrix}$$

and this gives the set of equations:

$$\begin{aligned} \pi_1^1 &= \pi_1^1(1-\alpha_1) + \pi_2^2(1-\alpha_2) \\ \pi_2^2 &= \qquad\qquad \pi_3^3(1-\alpha_3) \\ \pi_4^4 &= \qquad\qquad\qquad \pi_4^4(1-\alpha_4) \\ &\vdots \\ \pi_\tau^\tau &= \pi_1^1\alpha_1 + \pi_2^2\alpha_2 + \pi_3^3\alpha_3 + \cdots + \pi_\tau^\tau\alpha_\tau \end{aligned} \tag{4.27}$$

To find the values of π_i^l in terms of t and α_l, we take the following steps. Solving the first equation for π_2^2 gives

$$\pi_2^2 = \pi_1^1 \frac{\alpha_1}{1-\alpha_2}. \tag{4.28}$$

Replacing for π_2^2 from (4.27) in the second equation gives π_3^3 in terms of π_1^1 as

$$\pi_3^3 = \pi_1^1\alpha_1 \left(\frac{1}{1-\alpha_2} \frac{1}{1-\alpha_3} \right). \tag{4.29}$$

Continuing to replace π_{i-1}^{l-1} (in terms of π_1^1) in π_i^l gives

$$\pi_i^l = \pi_1^1\alpha_1 \left(\prod_{l=2}^{\tau} (1-\alpha_l)^{-1} \right), \tag{4.30}$$

for $l, i = 2,3,\ldots,\tau$. Now, use the fact that $\sum_{i=1}^{\tau} \pi_i^l = 1$ and write

$$\pi_1^1 = 1 - \sum_{i=2}^{\tau} \pi_i^l = 1 - \sum_{i=2}^{\tau} \left(\pi_1^1\alpha_1 \left(\prod_{l=2}^{\tau} (1-\alpha_l)^{-1} \right) \right) \tag{4.31}$$

Solving for π_1^1 gives

$$\pi_1^1 = \left[1 + \alpha_1 \sum_{i=2}^{\tau} \left(\prod_{l=2}^{\tau} (1-\alpha_l) \right)^{-1} \right]^{-1}, \tag{4.32}$$

which together with the expression for π_i^l in terms of π_1^1 gives all the stationary probabilities for the case of $\eta = \tau$. Notice that in this case π_i^l is only nonzero when $i = l$.

Now consider the general case of $2 \le \eta \le \tau$. Then from solving the equations for the stationary probabilities satisfying $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$, and following the same steps as above, we get

$$\pi_i^l = \frac{\alpha_1}{\prod_{j=2}^{l}(1-\alpha_j)\prod_{j=l+1}^{i}(1-\alpha_\eta)} \pi_1^1 , \tag{4.33}$$

for $2 \le i \le \tau$, and $2 \le l \le \eta$. The last equation in $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$ has π_t^η equal to the long-term probability of observing a defective, that is $p = E(X_k) = \pi_\tau^\eta$. From $\pi_1^1 = 1 - \sum \pi_i^l$, we have

$$\pi_1^1 = \left[1 + \alpha_1 \sum_{i=2}^{\tau} \left(\prod_{j=2}^{\min(i,\eta)} (1-\alpha_j) \prod_{j=\eta+1}^{i} (1-\alpha_\eta) \right)^{-1} \right]^{-1}, \tag{4.34}$$

for $2 \le i \le \tau$, and $2 \le l \le \eta$, which will reduce to the simpler form shown in (4.32) if we are using the maximum number of levels, i.e., if $\eta = \tau$. These derivations are illustrated in a special case below.

4.5.1.1 The Stationary Probabilities when $t = 3$ and $\eta = 3$

Consider the case of modeling a third-order Markov chain by assigning three different values to the conditional probability of observing a 1, i.e., choose η to be 3. The probability assignment for this case can be seen in Table 4-11. To obtain the stationary probabilities we solve the following:

$$\begin{bmatrix} \pi_1^1 & \pi_2^2 & \pi_3^3 & \pi_4^3 \end{bmatrix} = \begin{bmatrix} \pi_1^1 & \pi_2^2 & \pi_3^3 & \pi_4^3 \end{bmatrix} \begin{bmatrix} 1-\alpha_1 & 0 & 0 & \alpha_1 \\ 1-\alpha_2 & 0 & 0 & \alpha_2 \\ 0 & 1-\alpha_3 & 0 & \alpha_3 \\ 0 & 0 & 1-\alpha_3 & \alpha_3 \end{bmatrix},$$

which gives the set of equations:

$$\begin{aligned}
\pi_1^1 &= \pi_1^1(1-\alpha_1) + \pi_2^2(1-\alpha_2) \\
\pi_2^2 &= \pi_3^3(1-\alpha_3) \\
\pi_3^3 &= \pi_4^3(1-\alpha_3) \\
\pi_4^2 &= \pi_1^1\alpha_1 + \pi_2^2\alpha_2 + \pi_3^3\alpha_3 + \pi_4^3\alpha_3
\end{aligned}$$

and solves as

$$\pi_1^1 = \left[1 + \frac{\alpha_1}{1-\alpha_2} + \frac{\alpha_1}{1-\alpha_2}\frac{1}{1-\alpha_3} + \frac{\alpha_1}{1-\alpha_2}\frac{1}{(1-\alpha_3)^2}\right]^{-1}$$

$$\pi_2^2 = \pi_1^1\frac{\alpha_1}{1-\alpha_2}$$

$$\pi_3^3 = \pi_2^2\frac{1}{1-\alpha_3} = \pi_1^1\frac{\alpha_1}{1-\alpha_2}\frac{1}{1-\alpha_3}$$

$$\pi_4^3 = \pi_3^3\frac{1}{1-\alpha_3} = \pi_1^1\frac{\alpha_1}{1-\alpha_2}\frac{1}{(1-\alpha_3)^2}$$

The long term probability of observing a defective is $p = P(X=1) = E(X_k) = \pi_4^3$. It can be seen that by setting $\alpha_3 = \alpha_2$, that is by choosing η equal to two, the three-parameter model for third–order Markov chain will be produced as a special case.

4.6 The MBCUSUM for the Multi-level Model

The multi-level model (MLM) can be used to further generalize our MBCUSUM control chart. To do so, we need to distinguish between sequences of the previous t observations with different numbers of consecutive zeros. Let us denote the number of consecutive nondefectives in the items observed immediately before item k as R_k, where R_k can take any positive value. We consider the values 0, 1, …, t for R_k, which correspond to the conditional probabilities of observing a defective, α_l as follows. If we have observed at least t consecutive zeros, that is if $R_k \geq t$, then we are at level 1 and we assign α_1 to $P(X_k \mid R_k)$. If $X_{k-t} = 1$ and $X_{k-t-1} = X_{k-t-2} = \cdots = X_{k-1} = 0$, then $R_k = t-1$, and we are moving to a higher level of probability for observing a defective, say α_2. By continuing this method, we assign the highest probability level of α_τ to observing a defective when $R_k = 0$, which corresponds to $X_{k-1} = 1$ and X_{k-2}, X_{k-3}, …, X_{k-t} either 0 or 1. In general, we are at level l and use α_l, when $R_k = \tau - l$ for $1 \le l \le \tau$. Recall that the generalized CUSUM increment is the following:

$$\Delta_k = O_k(-\delta_1 + \delta_3) + X_k(-\delta_1 + \delta_2) - O_k X_k(-\delta_1 + \delta_2 + \delta_3 - \delta_4) + \delta_1$$

Now, we derive the values of the increment that reflect the conditional probability of X_k for given R_k. When $R_k = t$, that is for $l = 1$, we use $P(X_k \mid R_k) = \alpha_1$, and the values of δ_1 and δ_2 are equal to those for the generalized MBCUSUM. This means that there is no change in the value of the increment for the case of $O_k = 0$. The reason is that δ_1 and δ_2 are functions of α_1 and do not depend on α_2, $\alpha_3, \ldots, \alpha_\tau$. For $l = 2, 3, \ldots, \tau$, the values of δ_3 and δ_4 are derived as a function of α_l, where α_l is the probability of observing a defective, given R_k. The second subscript shows the corresponding value of p that is used for calculating α_l, i.e., α_{l0} and α_{l1} are obtained as functions of p_0 and p_1, respectively. These results are summarized in Table 4-12.

Table 4-11 The CUSUM Increment for the General form of the MLM

O_k	X_k	R_k	l	$P(X_k \mid R_k)$	Δ_k
0	0	t	1	$1-\alpha_1$	$\delta_1 = \ln \frac{1-\alpha_{11}}{1-\alpha_{10}}$
0	1	t	1	α_1	$\delta_2 = \ln \frac{\alpha_{11}}{\alpha_{10}}$
1	0	$\tau - l$	$2,3,\ldots,\tau$	$1-\alpha_l$	$\delta_3 = \ln \frac{1-\alpha_{l1}}{1-\alpha_{l0}}$
1	1	$\tau - l$	$2,3,\ldots,\tau$	α_l	$\delta_4 = \ln \frac{\alpha_{l1}}{\alpha_{l0}}$

For an MLM with η levels, the MBCUSUM increment has 2η possible values, where $2 \le \eta \le \tau$.

4.7 Summary

In this chapter, we first reviewed how the dimensions of the direct generalization of the first-order Markov chain model increases exponentially with the order of the chain. Based on a simplifying idea, a probability structure for a $t^{th}-$order Markov model was developed for correlated binary observations, which required only three parameters. We derived the model parameters in terms of p and ρ for $t \geq 1$, and discussed the estimation of parameters. Expressions for the long-term probability of observing a defective and the serial correlation were derived. Next, a CUSUM chart was developed for the three-parameter model. This CUSUM chart was approximated by converting the increments to the desirable forms. Thus, a general method was developed for designing CUSUM control charts for monitoring a proportion in correlated processes (of any order of dependence) with specific desirable statistical properties. Illustrative examples were provided.

In the last section, we defined certain rules that give a generalization of the three-parameter model. The MLM assigns increasing values to the conditional probability of observing a defective as the number of nondefectives previously observed decreases. In terms of these assigned probabilities, we derived the stationary probabilities of a Markov chain and the corresponding CUSUM increments.

5 Numerical Results

5.1 Introduction

In Chapter 3, we first constructed a CUSUM statistic based on the log-likelihood ratio for Markov dependent binary observations, and then modeled it as a Markov chain and named it the Markov Binary CUSUM. The MBCUSUM reduces to the Bernoulli CUSUM if the data are independent. In this chapter, we first discuss how comparisons will be done by showing how the Bernoulli CUSUM will be evaluated in the presence of correlation. Next, we numerically verify the known fact that CUSUM charts have better overall performance than the traditional Shewhart chart, and afterwards, we focus on the comparison of the MBCUSUM that models the correlation among observations as a Markov chain, with the Bernoulli CUSUM that is designed assuming independent data. In Chapter 4, a generalization of MBCUSUM was developed, which can account for any order of dependence among observed data. Here, we report the in-control ANOS values for several settings of the MBCUSUM. Also, under changing correlation and increasing order of dependence, we investigate the sensitivity of the generalized MBCUSUM chart to the model parameters. In the end, we provide a table that gives the values of the upper control limit which correspond to some desired values of the in-control ANOS.

5.2 Comparisons involving the Bernoulli CUSUM and the MBCUSUM

In order to compare the performance of the MBCUSUM with the Bernoulli CUSUM, we take the following approach. First, we consider a desired value of ANOS that corresponds to a desired false alarm rate. Then, we find the upper control limits that give the closest value to this desired value of ANOS for the Bernoulli CUSUM and the MBCUSUM charts. Afterwards, we find the SSANOS for values of $p > p_0$ to compare the relative performance of these two CUSUM charts. These calculations are based on the Markov model of the CUSUM statistics. Remember that the Bernoulli CUSUM of Reynolds and Stoumbos (1999) is designed for independent data. The transition matrix that they provide is also for independent observations. However, to investigate the performance of the Bernoulli CUSUM when observations are correlated, we need to find the appropriate probability transition matrix.

5.2.1 Transition Matrix for The Bernoulli CUSUM with Correlated Observations

In Section 2.3.3.2, we reviewed the Bernoulli CUSUM formulation as a Markov chain (Reynolds and Stoumbos, 1999). Recall from equation (2.2) that the control statistic was defined as $B_k = \max(0, B_{k-1}) + (X_k - \gamma)$, for $k = 1, 2, \ldots$, where γ was converted to $1/m$, and m is an integer. As a result, the changes in the consecutive values of the Bernoulli CUSUM statistic took only two values: $-1/m$, when $X_k = 1$, and $(m-1)/m$, when $X_k = 0$. For a given upper control limit h, the transition matrix for the Bernoulli CUSUM statistic modeled as a Markov chain has dimension $mh \times mh$. But, this is only the case for independent observation data. When the observations are correlated, as was discussed in Chapter 4, by using the three-parameter model for the binary observation data, the dimensions of this matrix increases to $\tau H \times \tau H$, where τ equals one plus the order of dependence. This happens because for each value of B_k, there are τ corresponding states. The relationship between the state number (i) and the corresponding values of the CUSUM statistics is as shown below:

$$B_k = \frac{\operatorname{int}\left(\frac{i-1}{\tau}\right)}{m} \Leftrightarrow i = \begin{cases} \tau m B_k + 1 \\ \tau m B_k + 2 \\ \vdots \\ \tau m B_k + \tau \end{cases} \tag{5.1}$$

The possible transitions and associated transition probabilities for the Bernoulli CUSUM when the observations are modeled as a Markov chain by using the three-parameter model (see Chapter 4) are shown in Table 5-1. The transition matrix for $t = 2$ is presented in Figure 5.1 as an example.

Table 5-1 Transition Probabilities for the Bernoulli CUSUM

O_k	X_k	$B_k - \max(0, B_{k-1})$	Transition	Probability	i
zero	0	$-1/m$	$1 \to 1$	$1-\alpha_1$	
			$i \to i-\tau$		$\tau+1, 2\tau+1, ..., \tau H - t$
one	0	$-1/m$	$i \to i-1$	$1-\alpha_2$	$2, 3, ..., \tau$
			$i \to i-\tau-1$		$k\tau+2, k\tau+3, ..., k\tau+\tau$
zero	1	$(m-1)/m$	$i \to i+\tau m-1$	α_1	$1, \tau+1, ..., \tau(H-m)+1$
one	1	$(m-1)/m$	$i+l \to k\tau$	α_2	$2, \tau+2, ..., \tau(H-m)+2$
					$k = m, m+1, \ldots, H$
					$l = 0, 1, \ldots, t-1$

		j	1	2	3	$\cdots$	$3m-2$	$3m-1$	$3m$	$\cdots$	$3H-5$	$3H-4$	$3H-3$	$3H-2$	$3H-1$	$3H$
		$X_{k-1}X_k$	00	10	*1	$\cdots$	00	10	*1	$\cdots$	00	10	*1	00	10	*1
		B_k	0	0	0	$\cdots$	$1-1/m$	$1-1/m$	$1-1/m$	$\cdots$	$h-2/m$	$h-2/m$	$h-2/m$	$h-1/m$	$h-1/m$	$h-1/m$
i	$X_{k-2}X_{k-1}$	B_{k-1}														
1	00	0	$1-\alpha_1$	0	0	$\cdots$	0	0	α_1	$\cdots$	0	0	0	0	0	0
2	10	0	$1-\alpha_2$	0	0	$\cdots$	0	0	α_2	$\cdots$	0	0	0	0	0	0
3	*1	0	0	$1-\alpha_2$	0	$\cdots$	0	0	α_2	$\cdots$	0	0	0	0	0	0
4	00	$1/m$	$1-\alpha_1$	0	0	$\cdots$	0	0	0	$\cdots$	0	0	0	0	0	0
5	10	$1/m$	$1-\alpha_2$	0	0	$\cdots$	0	0	0	$\cdots$	0	0	0	0	0	0
6	*1	$1/m$	0	$1-\alpha_2$	0	$\cdots$	0	0	0	$\cdots$	0	0	0	0	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$3(H-m)+1$	00	$h-1$	0	0	0	$\cdots$	0	0	0	$\cdots$	0	0	0	0	0	α_1
$3(H-m)+2$	10	$h-1$	0	0	0	$\cdots$	0	0	0	$\cdots$	0	0	0	0	0	α_2
$3(H-m)+2$	*1	$h-1$	0	0	0	$\cdots$	0	0	0	$\cdots$	0	0	0	0	0	α_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$3H-2$	00	$h-1/m$	0	0	0	$\cdots$	0	0	0	$\cdots$	$1-\alpha_1$	0	0	0	0	0
$3H-1$	10	$h-1/m$	0	0	0	$\cdots$	0	0	0	$\cdots$	$1-\alpha_2$	0	0	0	0	0
$3H$	*1	$h-1/m$	0	0	0	$\cdots$	0	0	0	$\cdots$	0	$1-\alpha_2$	0	0	0	0

Figure 5-1 Transition Matrix for the Bernoulli CUSUM with Correlated Observations, *t* =2

5.3 MBCUSUM vs. Shewhart Chart vs. Bernoulli CUSUM

Table 5-2 provides the ANOS and SSANOS values for Shewhart chart as well as for the Bernoulli and MBCUSUM charts for an example with the in-control value of $p_0 = 0.01$. The upper control limits of the CUSUM charts were determined to correspond to a specific value for the in-control ANOS around 17000, that corresponds to an upper control limit of 5 for a Shewhart chart with $n = 100$. The value of p that has to be detected quickly by the CUSUM chart, namely, the out-of-control values is $p_1 = 0.04$. The SSANOS values for a range of out-of-control values of $p > p_0$ were calculated for the Shewhart chart with curtailed sampling, and for the Bernoulli and Markov Binary CUSUM charts at two values of t, 1 and 4. In Table 5.2, the row corresponding to $p_0 = 0.01$ gives the in-control ANOS values, while the other rows corresponding to $p > p_0$ give out-of-control SSANOS values. For the Shewhart chart, the shift may occur in the middle of the n observations being taken, so the number of observations to the point where a statistic is plotted may be less than n.

For correlated observations, when $\rho =$ 0.05, the out-of-control SSANOS values in Table 5-2 show that the MBCUSUM has a better performance than the Bernoulli CUSUM chart and that both CUSUM charts perform more efficiently than the Shewhart chart. While the MBCUSUM performs clearly better that the Shewhart chart and Bernoulli CUSUM for a wide range of shifts, for very large shifts (larger than twenty times the in-control value in Table 5-2) the Bernoulli CUSUM and the Shewhart chart become advantageous over the MBCUSUM chart. This is an artifact of the model, imposed by the assumption that the value of ρ remains constant. Another fact that is shown in Table 5-2 is that the number of states in the Markov chain is much lower for CUSUM charts than for the Shewhart chart, and also a little lower for MBCUSUM than for the Bernoulli CUSUM.

Table 5-2 ANOS and SSANOS values for Curtailed Shewhart chart, MBCUSUM and Bernoulli CUSUM, when $p_0 = 0.01$

	ρ =0.05, t=1			ρ =0.05, t=4		
	n=100 UCL=5	p_1 = 0.04		n=100 UCL=6	p_1 = 0.04	
	Curtailed	m=34	m=46	Curtailed	m=41	m=46
p	Shewhart	MB	Bernoulli	Shewhart	MB	Bernoulli
0.010	16935.5	16914.3	17046.1	19736.1	19545.6	19841.2
0.015	4210.2	2876.8	3155.0	6087.2	3354.6	4298.4
0.020	1658.3	1004.6	1102.0	2624.6	1188.3	1544.8
0.025	848.3	515.7	559.9	1386.5	619.9	781.9
0.030	512.8	327.9	353.0	840.3	399.9	489.6
0.040	254.9	183.1	195.1	404.3	228.7	268.4
0.070	91.7	77.9	81.4	129.2	102.6	111.3
0.100	56.6	50.1	51.3	75.6	68.9	70.2
0.200	25.4	23.7	23.3	33.1	37.5	31.7
0.300	16.3	16.3	15.3	21.5	28.6	20.5
0.400	12.0	13.0	11.4	16.2	24.1	15.5
0.500	9.5	11.1	9.1	13.4	21.2	12.9
H	1200	174	189	3500	217	253
(t+1)H		348	378		1085	1265

5.4 The Effect of Higher Orders of Dependence

Table 5-3 gives the ANOS and SSANOS values for several values of the order of dependence t for the Bernoulli and Markov Binary CUSUM charts when $\rho = 0.05$, $p_0 = 0.01$, and $p_1 = 0.04$. When there is no correlation, these two charts are equivalent. It can be seen that when the observations are correlated (even at a small value such as 0.05), the performance of MBCUSUM is increasingly superior to that of the Bernoulli CUSUM as the order of dependence increases. In producing these results, the upper control limits of the CUSUM charts have been adjusted to give an in-control ANOS of about 29000. This choice is made to make these results comparable to the Shewhart chart used in Reynolds and Stoumbos (1999).

Table 5-3 The Performance of CUSUM charts as *t* increases, when $p_0 = 0.01$ and $p_1 = 0.04$

	ρ =0	ρ =0.05							
	t=0	t=1		t=4		t=8		t=12	
p	CUSUM	MB	Bernoulli	MB	Bernoulli	MB	Bernoulli	MB	Bernoulli
0.010	29050.8	29132.1	28994.5	29399.0	29209.4	29373.6	29356.9	29109.2	29112.7
0.015	3875.3	3988.6	4380.1	4278.6	5477.1	5172.4	6412.4	5329.5	7036.8
0.020	1201.2	1240.5	1361.2	1390.7	1807.5	1800.5	2279.3	1970.5	2672.9
0.025	587.4	603.0	654.0	697.4	877.0	919.1	1141.0	1064.0	1387.3
0.030	366.6	373.9	402.2	442.1	538.8	583.9	709.9	705.7	879.1
0.040	202.6	204.9	218.4	249.6	291.4	329.5	387.3	420.7	487.2
0.070	85.4	86.3	90.3	111.2	119.9	149.6	160.8	203.9	206.2
0.100	54.2	55.4	56.9	74.7	75.5	103.0	102.1	143.4	133.5
0.200	25.1	26.5	25.3	40.8	34.3	58.1	48.6	84.7	72.4
0.300	16.7	18.2	16.1	31.1	22.7	44.3	35.9	85.1	77.1
0.400	12.5	14.4	11.9	26.3	17.3	41.6	35.2	167.9	162.5
0.500	10.0	12.4	9.4	23.1	14.4	56.0	51.1	687.8	683.7
H	186	192	209	234	273	265	358	392	447
(t+1)H	186	384	418	1170	1365	2385	3222	5096	5811

For a certain range of changes in p, the out-of-control SSANOS values decrease consistently as p increases. But, for large values of t and extremely high values of p, the SSANOS starts to increase. This phenomenon can be explained by noticing the implication of the changes that happen in the values of the conditional probabilities of observing a defective as p takes large values. Tables 5-4, 5-5, and 5-6 are presented to illustrate this artifact of our model.

Table 5-4 gives the values that the three-parameter model assigns to the lower (α_1) and higher (α_2) conditional probabilities of observing a 1 for a first-order Markov chain, for different values of p and ρ. Notice that α_1 and α_2 both increase as p increases, however, the rate of increase is much higher for α_1. That is, for fixed ρ and t, when the long term probability of observing a 1 increases the probability of observing a 1 after a 0 increases faster than the probability of observing a 1 after another1. For higher orders of

Table 5-4 The Values of α_1 and α_2 for Different Values of p and ρ, when t =1

p	ρ	.010	.025	0.050	0.075	0.100	0.150	0.200	0.250	.300	.400
0.001	α_1	.00099	.00098	.00095	.00093	.00090	.00085	.00080	.00075	.00070	.00060
	α_2	.01099	.02598	.05095	.07593	.10090	.15085	.20080	.25075	.30070	.40060
0.010	α_1	.00990	.00975	.00950	.00925	.00900	.00850	.00800	.00750	.00700	.00600
	α_2	.01990	.03475	.05950	.08425	.10900	.15850	.20800	.25750	.30700	.40600
0.040	α_1	.03960	.03900	.03800	.03700	.03600	.03400	.03200	.03000	.02800	.02400
	α_2	.04960	.06400	.08800	.11200	.13600	.18400	.02320	.28000	.32800	.42400
0.100	α_1	.09900	.09750	.09500	.09250	.09000	.08500	.08000	.07500	.07000	.06000
	α_2	.10900	.12250	.14500	.16750	.19000	.23500	.28000	.32500	.37000	.46000
0.250	α_1	.24750	.24375	.23750	.23125	.22500	.21250	.20000	.18750	.17500	.15000
	α_2	.25750	.26875	.28750	.30625	.32500	.36250	.40000	.43750	.47500	.55000
0.500	α_1	.49500	.48750	.47500	.46250	.45000	.42500	.40000	.37500	.35000	.30000
	α_2	.50500	.51250	.52500	.53750	.55000	.57500	.60000	.62500	.65000	.70000

dependence, however, the behavior of α_1 changes in a manner that implies unrealistic structure to the sequences of observations. Table 5-5 investigates the effect of changes in the order of dependence at fixed values of ρ on α_1 and α_2, as p increases. Recall that α_1 is a function of p, ρ, and t, whereas, α_2 is only a function of p and ρ, but not of t. Thus, for each value of p and ρ, changes in the value of t does not affect the value of α_2. Table 5-5 gives the values of α_1 and α_2, for two values of ρ, 0.05 and 0.20, and four orders of dependence, 1, 4, 8, and 12, and six values of p. For $\rho = 0.05$ (low

Table 5-5 The Values of α_1 and α_2 for Different Values of p and ρ, when t Increases

		$\rho =$ 0.05				$\rho =$ 0.20			
p	t	1	4	8	12	1	4	8	12
0.001	α_1	.00095	.00081	.00066	.00054	.00080	.00041	.00017	.00007
	α_2	.05095	.05095	.05095	.05095	.20080	.20080	.20080	.20080
0.010	α_1	.00950	.00812	.00655	.00525	.00800	.00405	.00161	.00064
	α_2	.05950	.05950	.05950	.05950	.20800	.20800	.20800	.20800
0.040	α_1	.03800	.03218	.02509	.01902	.03200	.01568	.00571	.00202
	α_2	.08800	.08800	.08800	.08800	.23200	.23200	.23200	.23200
0.100	α_1	.09500	.07872	.05629	.03672	.08000	.03637	.01080	.00299
	α_2	.14500	.14500	.14500	.14500	.28000	.28000	.28000	.28000
0.250	α_1	.23750	.18173	.08823	.02945	.20000	.07105	.01089	.00145
	α_2	.28750	.28750	.28750	.28750	.40000	.40000	.40000	.40000
0.500	α_1	.47500	.26486	.02587	.00138	.40000	.06809	.00195	.00005
	α_2	.52500	.52500	.52500	.52500	.60000	.60000	.60000	.60000

correlation), at each value of p, α_1 decreases as t increases, whereas α_2 remains unchanged. Also, for t equal 1 and 4 when p increases both α_1 and α_2 take increasing values, which is consistent with our findings in Table 5-4. But when t takes larger values, such as 8 and 12, α_1 first increases with increase in p and then as p takes larger values, the value of α_1 starts to decrease. This phenomenon is stronger when the correlation is higher (i.e., $\rho = 0.20$).

This phenomenon suggests that large values of p and t result in certain types of sequences of observations. It is a situation in which the long term probability of observing a 1 increases, and at the same time, the probability of observing a 1 after t zeros, will decrease for high values of t. This implies very long sequences of nondefectives for a process with a high long term probability of defectives. It means that assuming fixed ρ makes our model unrealistic for very large values of p and t. Table 5-6 shows that even for a moderate value of $p = 0.05$, when the correlation is very high, say 0.50, and the order of dependence is high too, say 20, the very low value of α_1 that is below 0.000001, implies extremely long sequences of nondefectives that would result in a large value of SSANOS.

Table 5-6 The Values of α_1 and α_2 for p=0.05, when t and ρ Increase

	T	1	2	4	8	12	20
ρ							
0.01	α_1	.49500	.48519	.43289	.13379	.01069	.00004
	α_2	.50500	.50500	.50500	.50500	.50500	.50500
0.05	α_1	.47500	.42976	.26486	.02587	.00138	.00000
	α_2	.52500	.52500	.52500	.52500	.52500	.52500
0.20	α_1	.40000	.26667	.06809	.00196	.00005	.00000
	α_2	.60000	.60000	.60000	.60000	.60000	.60000
0.50	α_1	.08333	.00581	.00581	.00002	.00000	.00000
	α_2	.75000	.75000	.75000	.75000	.75000	.75000

5.5 The Performance for a High Quality Process

The previous comparisons between the Bernoulli CUSUM and the MBCUSUM chart were for the case of $p_0 = 0.01$. In this section, we consider smaller values of p_0 that represent high quality production processes. Table 5-7 gives ANOS values for the

MBCUSUM and for the Bernoulli CUSUM, when the in-control value of p is $p_0 = 0.001$ and the our-of-control value of p to be detected is $p_1 = 0.005$. To correspond to Table 5-3, the upper control limit is adjusted to give an in-control ANOS as close as possible to 29000. As the value of the correlation increases to 0.05 and then to 0.20, MBCUSUM performs increasingly better than the Bernoulli CUSUM. Moreover, when the order of dependence is higher (here, fourth-order), for the same amount of correlation, the MBCUSUM shows an even more advantageous performance compared to the Bernoulli CUSUM. So both the value of correlation and the order of dependence have an adverse effect on the performance of the Bernoulli CUSUM. The MBCUSUM is more efficient than the Bernoulli CUSUM in the presence of correlation, and its efficiency increases as the order of dependence grows higher. However, for large values of p, even when the correlation takes higher values, the Bernoulli CUSUM has a better performance than the MBCUSUM. This feature might not be of any practical consequence, because the values of p for which the Bernoulli CUSUM performs better than the MBCUSUM may be much higher than the out-of-control values of p that need to be detected by the control chart.

Table 5-7 ANOS and SSANOS Values for MBCUSUM and Bernoulli CUSUM at $p_0 = 0.001$ and $p_1 = 0.005$

	t=0	t=1				t=4			
	ρ =0	ρ = 0.05		ρ = 0.20		ρ = 0.05		ρ = 0.20	
p	CUSUM	MB	Bernoulli	MB	Bernoulli	MB	Bernoulli	MB	Bernoulli
0.001	29081.3	29117.1	29159.1	29123.3	29123.1	29119.9	29118.9	29125.2	29093.6
0.002	4746.6	4729.6	5267.6	5075.2	6568.7	5051.4	6386.3	6900.1	9699.0
0.003	2002.3	1986.0	2199.5	2195.3	2877.8	2184.0	2785.4	3329.6	5038.5
0.005	837.4	822.8	897.0	928.6	1184.3	925.9	1146.7	1513.3	2306.4
0.007	519.7	508.3	549.5	578.6	724.1	578.7	702.2	963.5	1446.6
0.010	330.6	322.7	345.4	370.6	454.9	372.4	442.1	622.6	919.4
0.015	206.9	202.0	212.0	234.4	280.6	237.7	274.0	391.5	571.1
0.020	151.7	148.3	152.3	173.0	203.3	177.1	199.5	285.0	414.9
0.025	120.5	117.7	118.8	137.7	159.8	142.3	157.4	223.9	326.4
0.030	100.1	97.8	97.4	114.6	131.9	119.7	130.3	184.6	269.5
0.040	75.0	73.4	72.0	86.1	98.1	91.8	97.1	137.6	200.5
0.070	42.9	42.5	40.7	49.8	55.6	56.7	55.3	81.9	114.8
0.100	30.0	30.3	28.4	35.5	38.8	43.2	38.7	61.9	81.2
0.200	15.0	16.3	14.2	19.1	19.4	28.6	19.6	43.1	42.7
H	948	984	1053	1110	1328	1104	1288	1860	2405
(t+1)H	948	1968	2106	2220	2656	5520	6440	9300	12025

Table 5-8 ANOS and SSANOS Values for MBCUSUM and Bernoulli CUSUM at $p_0 = 0.004$ and $p_1 = 0.020$

	t=0	t = 1				t = 4			
	ρ = 0	ρ = 0.05		ρ = 0.2		ρ = 0.05		ρ = 0.2	
p	CUSUM	MB	Bernoulli	MB	Bernoulli	MB	Bernoulli	MB	Bernoulli
0.004	28901.7	29021.3	29140.7	29172.9	29202.5	29192.1	29199.4	29227.4	29136.4
0.008	2338.4	2359.1	2702.0	2592.6	3708.5	2584.3	3583.7	3865.4	6092.8
0.012	780.7	785.6	879.5	889.2	1239.5	894.4	1194.9	1526.6	2460.4
0.016	428.3	429.4	472.6	490.8	662.3	498.4	639.3	894.1	1393.1
0.020	290.1	290.2	316.2	332.3	440.7	344.2	425.9	626.1	945.2
0.024	218.5	218.4	236.4	251.3	328.4	259.5	317.6	482.3	709.7
0.030	159.4	159.4	171.2	184.0	237.1	192.0	229.4	360.7	514.9
0.040	110.1	110.4	117.2	128.0	162.0	135.8	156.9	257.2	352.9
0.070	58.0	58.9	59.8	69.1	83.5	76.2	80.8	145.6	182.2
0.100	40.1	41.1	39.8	48.5	56.5	55.4	54.3	108.5	123.6
0.200	20.0	21.3	19.1	25.4	26.6	33.9	25.7	79.7	61.7
H	315	333	356	379	481	390	466	665	996
(t+1)H	315	666	712	758	962	1950	2330	3325	4980

Table 5-8 demonstrates the same results, as in Table 5-7 for a higher value of in-control probability of $p_0 = 0.004$ and an out-of-control value of p to be detected that is five times bigger, i.e., $p_1 = 0.020$. The upper control limit for the CUSUM charts was adjusted such that the in-control value of ANOS is as close as possible to 29000. Thus, the results of Table 5-8 correspond to the results obtained in Tables 5-7 and 5-3.

Table 5-9 gives ANOS values for six sets of the parameters (p_0, p_1, ρ), when the in-control ANOS was desired to be at 29000, same as in Tables 5-3, 5-4, and 5-5. All corresponding values for m, number of states H, UCL h, as well as the transition probabilities p_{01} and p_{11} are reported toward a more complete portrayal of the characteristics of the MBCUSUM. For instance, it can be seen that the number of states are realistically manageable. Now, we can verify the practical usefulness of accounting for higher orders of dependence in the design of the generalized MBCUSUM control chart. If we want an in-control ANOS of approximately 29000, then for $p_0 = 0.01$, $p_1 = 0.05$, and $\rho = 0.01$, the upper control limit for the simple MBCUSUM with first order dependence has to be set to 6.0833 (see Table 5-9). At this setting the value of the integer that specifies our steps for plotting the control chart is $m = 24$.

Table 5-9 ANOS and SSANOS values for the MBCUSUM with Different Parameters

				p_0=0.01	and	p_1=0.05				
	ρ	0.01			0.05			0.2		
	m	24			26			30		
	H	146			160			174		
	h	6.0833			6.1538			5.8000		
p		α_1	α_2	ANOS	α_1	α_2	ANOS	α_1	α_2	ANOS
0.0100		0.0099	0.0199	29595	0.0095	0.0595	29477	0.0080	0.2080	29946
0.0150		0.0149	0.0249	4417	0.0142	0.0643	4356	0.0120	0.2120	4991
0.0200		0.0198	0.0298	1344	0.0190	0.0690	1336	0.0160	0.2160	1634
0.0300		0.0297	0.0397	357	0.0285	0.0785	364	0.0240	0.2240	473
0.0400		0.0396	0.0496	179	0.0380	0.0880	186	0.0320	0.2320	247
0.0500		0.0495	0.0595	116	0.0475	0.0975	122	0.0400	0.2400	164
0.0600		0.0594	0.0694	85	0.0570	0.1070	91	0.0480	0.2480	123
0.0700		0.0693	0.0793	68	0.0665	0.1165	72	0.0560	0.2560	98
0.0800		0.0792	0.0892	56	0.0760	0.1260	60	0.0640	0.2640	82
0.0900		0.0891	0.0991	48	0.0855	0.1355	52	0.0720	0.2720	71
0.1000		0.0990	0.1090	42	0.0950	0.1450	45	0.0800	0.2800	62
0.5000		0.4950	0.5050	7	0.4750	0.5250	9	0.4000	0.6000	15
				p_0=0.04	and	p_1=0.10				
	ρ	0.01			0.05			0.2		
	m	16			16			20		
	H	109			103			125		
	h	6.8125			6.4375			6.25		
p		α_1	α_2	ANOS	α_1	α_2	ANOS	α_1	α_2	ANOS
0.0400		0.0396	0.0496	28923	0.0380	0.0880	29462	0.0320	0.2320	29619
0.0500		0.0495	0.0595	4045	0.0475	0.0975	4540	0.0400	0.2400	4758
0.0600		0.0594	0.0694	1125	0.0570	0.1070	1286	0.0480	0.2480	1408
0.0700		0.0693	0.0793	506	0.0665	0.1165	568	0.0560	0.2560	643
0.0800		0.0792	0.0892	301	0.0760	0.1260	331	0.0640	0.2640	384
0.0900		0.0891	0.0991	210	0.0855	0.1355	227	0.0720	0.2720	267
0.1000		0.0990	0.1090	160	0.0950	0.1450	171	0.0800	0.2800	204
0.1200		0.1182	0.1288	108	0.1140	0.1640	115	0.0960	0.2960	138
0.1500		0.1485	0.1585	73	0.1425	0.1925	77	0.1200	0.3200	94
0.2000		0.198	0.2080	47	0.1900	0.2400	50	0.1600	0.3600	62
0.2500		0.2475	0.2575	35	0.2375	0.2875	38	0.2000	0.4000	48
0.5000		0.4950	0.5050	16	0.4750	0.5250	18	0.4000	0.6000	26

In constructing Table 5-10, for $p_0 = 0.01$, $p_1 = 0.05$, and $\rho = 0.01$, we kept the upper control limit constant at h=6.0833, then we explored the effect on the generalized MBCUSUM of changes in the order of dependence and changes in the value of ρ. A lower in-control ANOS value indicates a higher false alarm thus the adverse effect of increase in t and ρ on the performance of the control chart is illustrated. These results suggest that it is important to account for the order of dependence in the set up of control charts that are designed for monitoring processes with correlated observations. It also demonstrates that an MBCUSUM, when it is designed for a first order Markov chain, would perform acceptably well even if the actual order of dependence is as high as third order, and while the actual correlation is more than twice of what it has been originally designed for (at $\rho = 0.025$, when designed for $\rho = 0.01$).

Table 5-10 The Change in the Performance of the MBCUSUM with Increase in *t* and *ρ*

ANOS (*m*)	*t*	1	2	3
ρ				
0.01		29174 (24)	23453 (25)	24339 (25)
0.025		24180 (25)	16368 (26)	17596 (26)
0.05		17141 (26)	16657 (27)	10874 (29)
0.075		19275 (26)	9640 (29)	7746 (31)
0.1		16999 (27)	9892 (30)	5519 (34)
0.125		12934 (28)	7794 (32)	4017 (37)
0.15		10321 (29)	5347 (34)	3037 (41)

Table 5-11 The Change in the Performance of the MBCUSUM with Decrease in *t* and *ρ*

ANOS (*m*)	*t*	3	2	1
ρ				
0.15		29361 (41)	84033 (34)	259500 (29)
0.125		50089 (37)	164502 (32)	394616 (28)
0.1		88239 (34)	239682 (30)	651149 (27)
0.075		158863 (29)	230100 (27)	8072837746 (26)
0.05		300614 (29)	626053 (27)	637569 (26)
0.025		692345 (26)	603038 (26)	1159618 (25)
0.01		1193442 (25)	1120258 (25)	1633937 (24)

The importance of accounting for the actual values of correlation and the order of dependence is illustrated in Table 5-11. For $p_0 = 0.01$, $p_1 = 0.05$, and $\rho = 0.01$ we found the upper control limit of $h = 242/41 = 5.9024$ that corresponds to an in-control ANOS as close as possible to 29000. Then we reduced the value of correlation as well as the order of dependence to investigate the change in the performance of an MBCUSUM chart that is designed for a high value of correlation when the actual correlation is low. It can be seen that the performance deteriorates quickly as the correlation decreases. The same is true for orders of dependence that are lower than the design value.

5.6 UCL for Some Desired Values of ANOS

Table 5-12 Values of H Which Will Approximately Give a Desired In-Control ANOS for Markov Chains with First to Sixth Order Dependence. (Exact In-Control ANOS Values Appear in Parenthesis)

p_1=0.05	**ρ**	0.01					0.05	
p_0=0.01	**α_2**	0.0199					0.0595	
			Desired	**ANOS(p_0)**			**Desired**	**ANOS(p_0)**
t	**α_1**	**m**	10000	30000	**α_1**	**m**	10000	30000
1	0.0099	24	121	147	0.0095	26	132	160
			(9962)	(30776)			(9928)	(29069)
2	0.0098	25	125	152	0.0090	27	133	161
			(9902)	(30102)			(10269)	(29477)
3	0.0097	25	125	151	0.0086	29	143	175
			(9985)	(29732)			9794	(30316)
4	0.0096	25	124	151	0.0081	31	155	189
			(9835)	(30287)			(10074)	(29797)
5	0.0095	26	133	162	0.0077	33	166	204
			10017	30185			10026	29772
p_0=0.04	**ρ**	0.01					0.05	
p_1=0.10	**α_2**	0.0496					0.0880	
			Desired	**ANOS(p_0)**			**Desired**	**ANOS(p_0)**
t	**α_1**	**m**	10000	30000	**α_1**	**m**	10000	30000
1	0.0396	16	91	110	0.0380	16	86	103
			(10272)	(30482)			(10232)	(29345)
2	0.0392	16	90	109	0.0360	17	91	110
			(10125)	(30459)			(9977)	(29838)
3	0.0388	16	89	108	0.0341	19	101	123
			(9896)	(30120)			(10084)	(30683)
4	0.0383	16	89	107	0.0322	20	117	139
			(10186)	(29502)			(10170)	(29986)
5	0.0379	17	89	107	0.0303	22	130	160
			(10427)	(30395)			(10088)	(29892)

The main advantage of modeling the CUSUM statistic as a Markov chain is that we can achieve specific statistical properties. In Table 5-12, we have derived the values of H that will approximately give a specific value of the in-control ANOS. The corresponding values of m are reported. From $h = H / m$, the upper control limits can be calculated.

5.7 Summary

In this Chapter, we reconfirmed this fact in the context of correlated binary observations. Also, we saw that as the value of the correlation and/or the order of dependence increases the MBCUSUM performs considerably and consistently better than the Bernoulli CUSUM. The exception arises for very large shifts, for which the Bernoulli CUSUM show better performance. A table giving general information for the MBCUSUM at six sets of parameter values was reported. The change in the SSANOS with increasing t and ρ reconfirms the necessity of accounting for correlation in observed data when monitoring processes. Finally, we provided values of the upper control limit and the integer values that approximately give a desired in-control ANOS for different orders of dependence and for two values of ρ. Thus, we saw that both the amount of correlation between observations and the order of dependence influence the performance of control charts. Thus, these factors have to be taken into account in the design of the charts.

6 Overview and Perspective

In this chapter, we first review the main thrust of previous chapters, then summarize our conclusions, and close with an outline of possible future directions of the current work.

6.1 Definitions and Assumptions

The subject of our study is monitoring correlated binary observation data. Such data arise naturally from situations where only binary outcomes are possible. An instance of such situation, in the context of quality control is monitoring a proportion in a production process, where the changes in the proportion defective is detected.

In a quality control context, an item is said to be *defective* if the item does not conform to required standards on one or more of several inspected quality characteristics. Consider the objective of detecting a change in the proportion defective, p, of the items produced by a process, when the inspected results are binary (defective/nondefective). Process monitoring aims at detecting any cause that shifts p from an in-control value p_0 to an out-of-control value p_1. In many cases it is desired to detect an increase in p, which corresponds to a deterioration in process quality. The primary tools for monitoring processes are control charts. To construct a chart we need to collect data from the process. We assumed that sampling starts at time 0 with an initial distribution that immediately gives the stationary distribution, and that observations are generated by a Markov chain, for which the probabilities are derived as a function of the correlation, ρ, and the long term probability of observing a defective item, p, and the order of the chain, t. In practice, for designing a control chart, it may be necessary to estimate the in-control process parameter, p_0 and correlation coefficient ρ.

The performance of control charts is usually measured by ARL which is defined based on samples. It would be appropriate to compare the performance of different control charts based on ARL, only if the sample sizes are the same. To compare control charts with different sample sizes, what we need are separate measures for the number of samples and the number of observations required to signal. We need to distinguish between the number of samples and the number of observations required to signal, and therefore, we used ANOS and SSANOS measures that.

6.2 Our Approach

In this dissertation, our main goal was to find a practical way to design a control chart for monitoring changes in processes that are represented by correlated binary data. Our focus was to incorporate the dependence of observations into the structure of the chart by using a Markov model for the dependence. We achieved these goals by taking the steps that are outlined below.

6.2.1 Core Derivations

We derived the conditional probability of $P(X_k = y \mid X_{k-1} = x, p, \rho)$ for the first-order Markov dependent observations in terms of two parameters, p and ρ, as

$$[p(1-\rho)]^{(1-x)y}\,[(1-p)(1-\rho)]^{x(1-y)}\,[1-p(1-\rho)]^{(1-x)(1-y)}\,[1-(1-p)(1-\rho)]^{xy}, \qquad (6.1)$$

where $x, y \in \{0, 1\}$. By using the conditional probability of a first-order Markov chain, we constructed an exact log-likelihood ratio CUSUM control statistic, $\ln\left(f(\mathbf{X} \mid p_1, \rho)/f(\mathbf{X} \mid p_0, \rho)\right)$, where $\mathbf{X}$ represents a vector of binary observations, and p_0 and p_1 are the in-control and out-of-control values of p, respectively.

The next step was developing a form for our CUSUM statistic which can be represented as a Markov chain. To provide a concise form of the primary statistic, we rewrote this log-likelihood ratio by defining and using a term s (a function of ρ), and four terms q_1 to q_4 (as functions of s, p_0, and p_1). Then we defined a basic step of $1/m$, where m is the integer value of q_1^{-1}. We derived the increments of the CUSUM statistic as integer multiples of this basic step and functions of q_1 to q_4 and expressed the increment as (see equation (3.9))

$$\Delta_k = X_{k-1}(-q_1 + q_3) + X_k(-q_1 + q_2) - X_{k-1}X_k(-q_1 + q_2 + q_3 - q_4) + q_1.$$

Table 6-1 provides all four values of the increment in two ways: in terms of our parameters, p and ρ, and as functions of conditional probabilities of observing a defective in a first-order Markov chain, p_{01} and p_{10}. These functions will give the increment values for given in-control (subscript 0), and out-of-control (subscript 1) values of p. (The relationship between our parameters and these conditional probabilities are given in Appendix-A.)

Table 6-1 The Increment Values for the Markov Binary CUSUM

(X_{k-1}, X_k)	Δ_k	
$(0,0)$	$q_1 = \ln \frac{s - p_1}{s - p_0}$	$\equiv \ln \frac{1 - p_{01_1}}{1 - p_{01_0}}$
$(0,1)$	$q_2 = \ln \frac{p_1}{p_0}$	$\equiv \ln \frac{p_{01_1}}{p_{01_0}}$
$(1,0)$	$q_3 = \ln \frac{1 - p_1}{1 - p_0}$	$\equiv \ln \frac{p_{10_1}}{p_{10_0}}$
$(1,1)$	$q_4 = \ln \frac{s - (1 - p_1)}{s - (1 - p_0)}$	$\equiv \ln \frac{1 - p_{10_1}}{1 - p_{10_0}}$

Notice, that if observations are independent (ρ = 0), then $q_1 = q_3$ and $q_2 = q_4$, and the increment reduces to $X_k(-q_1 + q_2) + q_1$. But $-q_1 + q_2 = r_2$ and $q_1 = -r_1$, where r_1 and r_2 are the parameters defined in Reynolds and Stoumbos (1999), which was reviewed in Chapter 2. This means that our control statistic reduces to that of the Bernoulli CUSUM for monitoring a proportion when the observations are independent.

6.2.2 The Markov Binary CUSUM

The four values of the increment correspond to the four possible sequences of the two consecutive binary observations (i.e., 00, 01, 10, and 11). This structure enabled us to model the CUSUM statistic as a Markov chain, for which the transition probability matrix for the transient states, $\mathbf{Q}$, was derived. The values that the CUSUM statistic, C_k, takes are integer multiples of $1/m$. For a given value of the upper control limit, say h, the CUSUM control chart is designed to signal at $C_k \geq h$, where $h > 1$ is expressed as an integer multiple of $1/m$. So the largest value of the CUSUM statistic that does not create a signal is $h-1/m$. To find the properties of the chart we only need the matrix of transient states, $\mathbf{Q}$, which is of dimension $2H \times 2H$, where $H = mh$. There are two states for each value of C_k, whereas, in the independent case we have one state for each value of the CUSUM statistic. We called the control chart that is based on this control statistic, the *Markov Binary CUSUM*.

6.2.3 The Multilevel Model for Correlated Binary Observations

We showed that a direct extension of our method to incorporate higher orders of dependence in the Markov model is not desirable, because of the exponential growth of the number of parameters needed for the Markov model. We introduced an idea that reduces the complexity of this problem and so provides a manageable extended framework for modeling Markov dependence of order $t \geq 1$.

We developed a method that provides a Markov model for any order of dependence among binary data. The idea is that observing a defective increases the probability of observing another defective. We associate this increase in probability with moving from a lower to a higher *level.* A $t^{th}-$order Markov chain can be represented by up to $t+1$ levels. The multi-level model requires only $t+1$ parameters and reduces the transition probability matrix to $(t+1)\times(t+1)$. We used a special case of this model to design a CUSUM control chart that can be applied to correlated observations and which can satisfy specific statistical properties. In doing so, we limited the probability levels assigned to observing a defective to only two values, and this gives a three-parameter model. Using the three-parameter model for correlated data resulted in the number of parameters increasing from two to only three. The only parameter we added to the long term probability of defective, p, and the correlation, ρ, was the order of dependence, t. The idea was to divide all possible sequences of binary observations into two categories, then assign a lower probability to one category, and a higher probability to the other category.

6.2.4 The Generalized MBCUSUM

To generalize the MBCUSUM to apply to data from the three-parameter model, we generalized the increment structures derived for a first-order Markov chain (see Table 6-1), by replacing the first-order conditional probabilities with these two probability values. The lower and higher probability values are functions of our three model parameters. We called the CUSUM statistic that uses this generalized increment structure the *generalized binary CUSUM*, modeled it as a Markov chain to derive the generalized MBCUSUM, and derived the general form of the transition probabilities for transitions from state i to state j, for any Markov chain with dependence of order $t \geq 1$ (see Chapter 4). The generalized MBCUSUM that uses the three-parameter model for $t^{th}-$order Markov chain and has a control limit of magnitude H, corresponds to a square matrix of dimension $(t+1) \times H$.

6.3 Control Charts for Monitoring Changes in a Proportion

The Shewhart chart for monitoring p is called a $p-$chart. To design the $p-$chart for the proportion defective, we need to specify three parameters: the sample size, the sampling interval, and the width of the control limits. The $p-$chart requires a large sample size for detecting small changes in p. Other control charts that are more responsive to small shifts are EWMA and CUSUM charts. The choice between EWMA and CUSUM is a matter of personal preference. The chart that we developed, extended, and discussed in the present work was a CUSUM chart, which as its very special case reduces to the Bernoulli CUSUM control chart developed by Reynolds and Stoumbos (1999).

6.4 Conclusions

We designed a CUSUM chart (MBCUSUM) that accounts for the correlation between binary observations by modeling the data as a particular Markov chain. We showed that in the presence of correlation, the overall performance of the MBCUSUM chart is better than the traditional Shewhart $p-$chart and the Bernoulli CUSUM chart.

However, for very large shifts, as an artifact of the model the relative performance of the MBCUSUM deteriorates. We derived a form for our CUSUM chart that enabled us to model it as a Markov chain. This in turn allows for designing an MBCUSUM control chart that corresponds to specific statistical properties.

For correlated binary observations, we developed a general probability model. This model considerably reduces the number of parameters needed for specifying a Markov chain. For practical use, we showed the way in which the parameters are estimated for a special case of our general model. Thus the significant extension of this work over past work is in providing a model for correlated observations that goes beyond the first-order dependence. Being extremely parsimonious in the number of required parameters, our generalized model offers a manageable framework for capturing the dependence of any order. Whereas in the existing literature, the correlation of observations is modeled only as a first-order Markov model.

We used our specific Markov model to derive a general form for the MBCUSUM. We noticed that a precise estimation of the model parameters is important in that it directly affects the performance of the MBCUSUM. Also, the usefulness of our MBCUSUM chart is limited by the range of shifts that it can detect effectively, because assuming fixed ρ makes our model unrealistic for high orders of dependence when the shift is very large.

6.5 Future Directions

The ideas and models that were developed in the current work can be extended and used in the following directions and areas. In the field of public-health, attribute CUSUM control charts are widely used. Moreover, according to Woodall (2006), the data collected for health-related studies, in many cases is binary by nature. For example consider successful or unsuccessful treatment with a certain medication, death rate for a specific procedure, etc. These characteristics of the problems and methods related to control charts in health-care studies suggests that the MBCUSUM can be applied to such problems.

In this dissertation, we considered the case of continuous sampling. Thus, a potential extension is toward incorporating non-continuous sampling schemes. One advantage that our approach has over the other works on the same subject is in that they either assume independent samples, or 100% sampling. Whereas our control chart can be used for sampling schemes with non-inspection periods, where samples are not independent from each other. The reason is that our CUSUM statistic itself is modeled as a Markov chain, and so by using the Markov model the amount of correlation that is carried from the end of one sample to the beginning of the next sample can be calculated.

Appendix A: The Relationship between Two Approaches

In Chapter 2, we reviewed the CUSUM control chart developed by Champ, Blatterman, and Rigdon (1994). In this section, we demonstrate the relationship between their CUSUM statistic and the MBCUSUM statistic. First, we derive our four parameters in terms of $a\,(=p_{01})$ and $b\,(=p_{10})$.

The long term probability of observing a defective ($X=1$), for a first-order Markov chain is $p = a/(a+b)$ and for a stationary process, the serial correlation can be expressed as $\rho = 1-a-b$.

Thus, it can be shown that

$$a = p(1-\rho) = \frac{p}{s},$$

$$b = 1-p(1-\rho)-\rho = (1-p)(1-\rho) = \frac{1-p}{s},$$

where $s = 1/(1-\rho)$, that is $1-\rho = 1/s$.

We assume that ρ is a known constant (or it has been approximated accurately). Therefore, we essentially are imposing that the sum of $a+b$ remains constant. Below, we derive q_1 through q_4 as functions of a and b.

$$q_1 = \ln\frac{s-p_1}{s-p_0} = \ln\frac{s-\dfrac{a_1}{a_1+b_1}}{s-\dfrac{a_0}{a_0+b_0}} = \ln\left(\frac{s(a_1+b_1)-a_1}{s(a_0+b_0)-a_0}\times\frac{a_0+b_0}{a_1+b_1}\right) \equiv \ln\frac{1-a_1}{1-a_0},$$

$$q_2 = \ln\frac{p_1}{p_0} = \ln\frac{\dfrac{a_1}{a_1+b_1}}{\dfrac{a_0}{a_0+b_0}} = \ln\left(\frac{a_1}{a_0}\times\frac{a_0+b_0}{a_1+b_1}\right) \equiv \ln\frac{a_1}{a_0},$$

$$q_3 = \ln\frac{1-p_1}{1-p_0} = \ln\frac{1-\dfrac{a_1}{a_1+b_1}}{1-\dfrac{a_0}{a_0+b_0}} = \ln\left(\frac{(a_1+b_1)-a_1}{(a_0+b_0)-a_0}\times\frac{a_0+b_0}{a_1+b_1}\right) \equiv \ln\frac{b_1}{b_0},$$

$$q_4 = \ln\frac{s-(1-p_1)}{s-(1-p_0)} = \ln\frac{s-sb_1}{s-sb_0} = \ln\frac{s(1-b_1)}{s(1-b_0)} \equiv \ln\frac{1-b_1}{1-b_0}.$$

Now, recall that based on the random variable Y_n, defined as the number of 0's between the $(n-1)^{th}$ and the n^{th} defective item for $n \geq 2$, the CUSUM statistic was:

$$U_n = \max\{0, U_{n-1} + Y_n - k_u\}.$$

By using an SPRT method the value of k_U, converted to integers, was

$$k_U = \text{int}\left(1 - \frac{\ln\left(\frac{a_U b_U}{a_0 b_0}\right)}{\ln\left(\frac{1-b_U}{1-b_0}\right)}\right).$$

The CUSUM chart signals at the N^{th} defective observation if $U_N \geq h_U$, for $h_U \geq 0$. Then, we can rewrite the reference value as

$$k_U = \text{int}\left(1 - \frac{\ln\frac{a_U}{a_0} + \ln\frac{b_U}{b_0}}{\ln\frac{1-b_u}{1-b_0}}\right) = \text{int}\left(1 - \frac{q_2 + q_3}{q_4}\right) = \text{int}\left(1 - \frac{\frac{q_2 \times m}{m} + \frac{q_3 \times m}{m}}{\frac{q_4 \times m}{m}}\right).$$

Now, recall the definition of the increments for the MBCUSUM, $d_{01} = \text{int}(q_2 * m)$; $d_{10} = |\text{int}(q_3 * m)|$; and $d_{11} = \text{int}(q_4 * m)$. It can be seen that Champ et al. (1994) CUSUM, in terms of the increments of the MBCUSUM can be written as the following:

$$U_n = \max\{0, U_{n-1}\} + (Y_n - 1) + \frac{d_{01} + d_{10}}{d_{11}}.$$

References

Bhat, N., and Lal, R. (1988), "Number of Successes in Markov Trials," *Advances in Applied Probability* **20**, pp. 677-680.

Bhat, N., and Lal, R. (1990), "Attribute Control Charts for Markov Dependent Production Processes," *IIE Transactions* **22**, pp. 181-188.

Blatterman, D. K., and Champ, C.W. (1992), "A Shewhart Control Chart Under 100% Inspection for Markov Dependent Attribute Data," *Proceedings of the 23rd Annual Modeling and Simulation Conference*, Pittsburg, PA, pp. 1769-1774.

Bourke, P. D. (Jul. 1991), "Detecting a Shift in Fraction Nonconforming Using Run-Length Control Charts with 100% Inspection," *Journal of Quality Technology* **23 (**3), pp. 225-238.

________(2001), "Sample Size and the Binomial CUSUM Control Chart: The Case of 100% Inspection," *Metrika* **53**, pp. 51-70.

Broadbent, S. R. (1958), " The Inspection of a Markov Process," *Journal of the Royal Statistical Society*, Series B-Statistical Methodology **20**, 111-119.

Champ, C. W.; Blatterman, D. K., and Rigdon, S. E. (1994), "A CUSUM Quality Control Chart Under 100% Inspection for Markov Dependent Attribute Data," *Proceedings of the Third Industrial Engineering Research Conference*, Atlanta, GA. pp. 34-39.

Deligonul, Z.S., and Mergen, A.E. (1987), "Dependence Bias in Conventional p Charts and its correction with an Approximate Lot Quality Distribution," *Journal of Applied Statistics* **14**, pp. 75-81.

Gan, F.F. (1993), "An Optimal Design of CUSUM Control Charts for Binomial Counts". *Journal of Applied Statistics* **20**, pp. 445-460.

Hawkins, D. M. (1992). "Evaluation of Average Run Lengths of Cumulative Sum Charts for an Arbitrary Data Distribution". *Communications in Statistics – Simulation and Computation* **21**, pp. 1001-1020.

Hawkins, D. M. and Olwell, D. H. (1998), *Cumulative Sum Charts and Charting for Quality Improvement*. Springer-Verlag. New York, NY.

Kedem, Benjamin (1980), *Binary Time Series*. Marcel Dekker, Inc., NY 10016.

Kemeny, John G., and Snell, J. Laurie (1960), *Finite Markov Chains*. Van Nostrand Reinhold Company, NY 10001.

Lai, C. D., Govindaraju, K., and Xie, M. (Aug. 1998), "Effects of Correlation on Fraction Non-Conforming Statistical Process Control Procedures," *Journal of Applied Statistics* **25** (4), pp. 535-543.

Lai, C. D., Xie, M., and Govindaraju, K. (May 2000), "Study of a Markov Model for a High-Quality Dependent Process," *Journal of Applied Statistics* **27** (4), pp. 461-473.

MacDonald, Iain L., and Zucchini, Walter (1997), *Hidden Markov and Other Models for Discrete-valued Time Series*. Chapman & Hall, UK.

Montgomery, D. C. (2005), *Introduction to Statistical Quality Control*. John Wiley and Sons, Inc., 5th ed.

Montgomery, D. C., and Woodall, W. H. (1997), "A Discussion on Statistically-Based Process Monitoring and Control," *Journal of Quality Technology* **29** (2), pp. 121-132.

Page, E. S. (1954), "Continuous Inspection Schemes," *Biometrika* **41**, pp. 100-114.

Reynolds, M. R., Jr., and Stoumbos, Z. G. (1998), "A General Approach to Modeling CUSUM Charts for a Proportion," *IIE Transactions* **30**, pp. 545-561.

_________ (1999), "A CUSUM Chart for Monitoring a Proportion When Inspecting Continuously," *Journal of Quality Technology* **31**, 87-108.

_________ (2000), "A General Approach to Modeling CUSUM Charts for a Proportion," *IIE Transactions on Quality and Reliability Engineering*, Special Issue on Emerging Trends in Quality Engineering **32 (**6), pp. 515-535.

_________ (2001), "Monitoring a Proportion Using CUSUM and SPRT Control Charts," in *Frontiers in Statistical Quality Control* **6**, eds. H.-J. Lenz and P.-Th. Wilrich, Heidelberg, Germany: Springer-Verlag, pp. 155-175.

_________ (2004), "Control Charts and Efficient Allocation of Sampling Resources," *Technometrics* **46 (**2), pp. 200-214.

Sego, L.; Woodall, W. H.; and Reynolds, M. R. Jr. (2005). " A Comparison of Methods for the Surveillance of Congenital Malformations". Presented at the 2005 Joint Statistical Meetings in Minneapolis, Minnesota.

Spitzner, Dan J., and Boucher, Thomas R. (2005), "Markov-Chain Asymptotic Variance Calculation for High Order Partial Sums," working paper. Department of Statistics, Virginia Tech.

Turin, William (1998), *Digital Transmission Systems: Performance Analysis and Modeling*. McGraw-Hill, New York, NY 10011.

Wald, A. (1947) *Sequential Analysis*. Dover Publications, Inc. New York, NY.

Warakagoda, Naryan (1996), http://jedlik.phy.bme.hu/~gerjanos/HMM/node4.html.

Woodall, W. H. (1997), "Control Charts Based on Attribute Data: Bibliography and Review," *Journal of Quality Technology* **29** (2), pp. 172-183.

________ (2006) "The Use of Control Charts in Health-Care and Public-Health Surveillance," *Journal of Quality Technology* **38** (2).

SHABNAM MODARRES-MOUSAVI
Research Associate and Assistant Professor
Department of Statistics, Statistical Consulting Center
The Pennsylvania State University, University Park, PA 16802
Office: (814) 863-8127, Cell: (540) 250-6554, E-mail: sxm70@psu.edu
Webpage: http://www.stat.psu.edu/people/faculty/smousavi.html

Education
Ph.D., Statistics, Virginia Tech, Spring 2006.
Dissertation Title: "Monitoring Markov Dependent Binary Observations with a Log-Likelihood Based CUSUM Control Chart," Advisor: Marion R. Reynolds.
Ph.D., Economics, Virginia Tech, Fall 2002.
Dissertation Title: "Methodological Foundations for Bounded Rationality as a Primary Framework," Advisor: Nicolaus Tideman.
B.Sc., Electrical Engineering, Sharif University of Technology, Tehran, Iran, 1995.

Research Interests
In Statistics: Statistical Process Control, Markov Chain Models, Subjective Probability.
In Economics: Nature of Uncertainty, Decision-Making, Axiomatization of Bounded Rationality, Methodology of Modeling Choice Behavior.

Teaching Experience
Schools: Virginia Tech, 2001, 2002, Roanoke College, 2004-5, Penn State University 2005- present.
Courses Taught: Statistical Consulting Practicum I and II, Elementary Statistics, Problem Solving and Communication in Applied Statistics, Marketing Management, International Marketing, Consumer Behavior, Promotion Management, Statistics for Social Sciences, Principles of Economics (Macroeconomics and Microeconomics).

Statistical Consulting
Faculty Consultant, Statistical Consulting Center, Department of Statistics, The Pennsylvania State University, August 2005 - present.
Team member/leader, Consulting Center,Department of Statistics, Virginia Tech, 2003-4.

Publications
Papers in Peer-Reviewed Journal
"A Classroom Exercise: Voting by Ballots and Feet," with Charles Holt et al, *Southern Economics Journal* 2005, 72(1), 253-263.

"Toward a Transactional Theory of Decision Making: Creative Rationality as Functional Coordination in Context," with Jim Garrison, *Journal of Economic Methodology* 10(2), 131-156, June 2003.

Contributed Chapter
"Household Welfare and Income Distribution, Poverty Line, and Characteristics of the Poor Families 1989-1996," with Mohammad Tabibian et al., in *Poverty and Income Distribution in Iran*, Institute for Research in Planning and Development, 2000.

Conference Proceedings

"CUSUM Control Charts for Autocorrelated Binary Data," with Marion R. Reynolds, *Papers and Proceedings, 5th Annual Hawaii International Conference on Statistics, Mathematics, and Related Fields. Honolulu, Hawaii.* January 16-18, 2006.

"Rationality, Bounded Rationality and Actual Decision Making," *Papers and Proceedings, 23rd Annual Conference of Society for the Advancement of Behavioral Economics*, July 2003.

"Role of Rural Women in Development in Korea, Indonesia, and Bangladesh: A Comparative Study," with Reza Kheirandish, *Papers and Proceedings, The Conference on Rural Women and Development, Tehran, Iran, July 1997.*

Under Review:

"Uncertainty Improves the Second Best," with Hans H. Haller. Submitted to *FUR VI Volume* (Foundations and Applications of Utility, Risk, and Decision Theory).

"An Outside The Box Approach to Bounded Rationality," Submitted to the *Atlantic Economic Journal*.

"Bounded Rationality and Full Rationality Do Not Fit in the Same Paradigm– A Note," Submitted to the *Journal of Economic Psychology*.

Research Reports

"Drug Abuse among College Students: A Survey for a National Plan to Confront Drug Abuse in Iran," with Amir H. Mehryar and Reza Kheirandish. Research Report, Institute for Research in Planning and Development, 1997.

"Project on Development Planning in Iran." I studied the two latest five-year development plans for Iran and had primary responsibility for writing six of thirteen TV programs on the evaluation of those plans and the future perspectives for several fields, including: *Primary and Higher Education, Health, Income Distribution, Women, Social Participation, Industrial Structure.* The project goal was to provide insight into social situations by educating citizens in an easily comprehended medium. Three TV channels broadcast these programs. 1998.

Research in Progress

"A General Model for Correlated Binary Observations," with Marion R. Reynolds, Department of Statistics, Virginia Tech.

"A Simple Method for Approximating the Variance In Contingency Tables," with George Terrell, Department of Statistics, Virginia Tech.

Principal Professional Presentations

"CUSUM Control Charts for Correlated Binary Data," INFROMS Annual Meeting, San Francisco, CA, November 2005.

"How to Develop Bounded Rationality as a Primary Framework," Society for Judgment and Decision Making **(SJDM)**, Minneapolis, MN, November 2004.

"Rationality, Bounded Rationality, and Actual Decision Making," Society for the Advancement of Behavioral Economics **(SABE)**, Lake Tahoe, Nevada, July 2003.

"Toward a Transactional Theory of Decision Making: Creative Rationality as Functional Coordination in Context," International Network of Economic Methodology, Allied Social Science Associations **(ASSA-INEM)**, Atlanta, GA, January 2002.

"Theory of Choice vs. Theory of Inquiry: Can John Dewey Solve a Puzzle?" 2nd Annual Symposium on the Foundations of the Behavioral Sciences: Behavioral Economics and Neoclassical, American Institute for Economic Research- Behavioral Research Council **(AIER- BRC)**, Great Barrington, MA, July 2002.

Organization of a Conference Session

Rationality and Mathematical Logic, Society for the Advancement of Behavioral Economics, Lake Tahoe, Nevada, July 2003.

Workshops Attended

"Public Choice Outreach Conference", George Mason University, VA, May/June 2002.

"American Pragmatism—Dewey and Bentley Transactional Epistemology," American Institute for Economic Research (AIER), MA, June 2002.

"NSF Sponsored Workshop on Classroom Experiments in Economics," College of William and Mary, May 2002.

"Experimental Economics," George Mason University, August 2001.

"Summer Fellowship Program," AIER, June- July, 2000 and 2001.

Honors and Awards

Research Fellowship, (visiting and in-absentia award) American Institute for Economic Research, Great Barrington, MA, 2000 and 2001.

Young Professional Program Candidate, UNESCO, Paris, France, Fall 1997.

Ranked 81 in special courses and 8 in general courses, Nation-Wide Entrance Exam for Colleges and Universities, Iran, 1987 (approximately 200,000 test takers).

Campus Leadership Activities

Faculty Advisor, Iranian Club, Virginia Tech. 2004-5.

Elected Member at Large, Council for International Student Organization, Spring, 1999.

Acting President, then President, Iranian Club at Virginia Tech, 1998-2000. Cooperated with other associations at Virginia Tech in organizing various activities. Organized art exhibition and lecture for the renowned Persian surrealist painter, Iran Darroudi.

Languages: Persian, English, familiar with French and Arabic