Modified cumulative sum procedures for count data with application to early detection of morbidity in radio frequency-monitored animals

by Frederick Kweku Annan Holdbrook

A dissertation submitted in partial fulfillment , of the requirements for the degree of Doctor of Philosophy in Statistics

Montana State University



Abstract:

Radio frequency (RF) technology is used in electronic monitoring and data acquisition devices currently available to the commercial animal feeding industry for continuously monitoring the feeding and watering behaviors of feedlot animals.

There is therefore the need for statistical process control (SPC) procedures that can be used on-line, in conjunction with these electronic data collection systems, to achieve a cost effective system that can quickly detect animal morbidity.

In this dissertation, a modified cumulative sum (modified CUSUM) procedure is proposed based on modifying the traditional CUSUM scheme for count data. Additional CUSUM design parameters and additional out-of-control conditions are introduced to give the procedure several enhanced features, improve its detection capability, and reduce the average run length (ARL).

The method is evaluated using simulated Poisson data, and recommendations for the choice of the extra design parameters are discussed. An outline is provided to show how the modified procedure can be implemented for a generic cattle feedlot data, which can be acquired from a digital monitoring and data collection system based on radio frequency (RF) ear-tag technology.

Results demonstrate that the modified CUSUM scheme can indeed achieve higher sensitivity and performance than the traditional CUSUM scheme. The proposed modified CUSUM schemes can also be designed to be relatively robust to isolated outliers that may be present in the data. It is further demonstrated that an optimal design of this modified CUSUM scheme can indeed be very useful in the early detection of morbidity among group-fed animals.

# MODIFIED CUMULATIVE SUM PROCEDURES FOR COUNT DATA WITH APPLICATION TO EARLY DETECTION OF MORBIDITY IN RADIO FREQUENCY-MONITORED ANIMALS

by

Frederick Kweku Annan Holdbrook

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Statistics

MONTANA STATE UNIVERSITY
Bozeman, Montana

February 2001

APPROVAL

of a dissertation submitted by

Frederick Kweku Annan Holdbrook

This dissertation has been read by each member of the dissertation committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the College of Graduate Studies.

William F. Quimby William F. Quimby 2/16/01
(Signature) Date

Approved for the Department of Statistics

John Lund John Lund 2/16/01
(Signature) Date

Approved for the College of Graduate Studies

Bruce McLeod Bruce R. McLeod 2-26-01
(Signature) Date

## STATEMENT OF PERMISSION TO USE

In presenting this dissertation in partial fulfillment of the requirements for a doctoral degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library. I further agree that copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U. S. Copyright Law. Requests for extensive copying or reproduction of this dissertation should be referred to Bell & Howell Information and Learning, 300 North Zeeb Road, Ann Arbor, Michigan 48106, to whom I have granted "the exclusive right to reproduce and distribute my dissertation in and from microform along with the non-exclusive right to reproduce and distribute my abstract in any format in whole or in part."

Signature 

Date 2/16/04

I dedicate this manuscript to the loving memory of my dear mother Dinah. She had always encouraged and supported me wholeheartedly in my life decisions. Thank you Mom for having faith in my dreams.

Dinah Maame Gyamansa Akyempon
1929–1999

# ACKNOWLEDGEMENTS

First of all, I thank my Lord Jesus Christ who has continually satisfied my deepest needs even when study fatigue and discouragement often set in.

I sincerely thank my advisor, Dr. William Quimby for his understanding, patience, guidance, assistance and suggestions during the preparation of this manuscript.

I am very grateful to Dr. John Borkowski who provided invaluable ideas, comments and suggestions in the fields of quality control and stochastic processes.

I especially thank my reading committee made up of professors W. Quimby, J. Borkowski, and Robert Boik as well as other members of my committee including professors James Robison-Cox, Warren Esty, and Lea Acord for their time and helpful comments and suggestions.

My thanks also go to James Jacklitch for helping me with LaTeX, and Lois Carbaugh for her immense help during my studies in Montana.

Most of all, I wish to express my gratitude to my entire family, brothers and sisters for their everlasting support and encouragement. I am most indebted to my brother Edward Delke Holdbrook without whose assistance none of these would have been possible. To my wonderful children Angela, Linda and Jesse Holdbrook, I say thank you very much for being understanding, loving, and cheerful all these times.

TABLE OF CONTENTS

LIST OF TABLES ..... ix

LIST OF FIGURES ..... xi

1. INTRODUCTION AND BACKGROUND INFORMATION ..... 1

Definitions and Terminology ..... 3
Background and Overview of Relevant SPC Literature ..... 9
General Control Charting Procedures ..... 23
Motivating Example: Cattle Feedlot Study ..... 29
Description of a Generic Feedlot Data ..... 29
Purpose and Initial Problems ..... 30
Goals and Objectives of Current Research ..... 31
Outline of Related Problems ..... 32

2. PROPOSED UNDERLYING MODEL ..... 36

Basic Assumptions ..... 36
Stochastic Process ..... 37
Time Series Process ..... 40
Markov Process ..... 42
Markov Property ..... 43
Transition Probability ..... 44
Lumped and expanded Markov chains ..... 47
Markov Model for a Generic Feedlot Data ..... 48
Poisson Process ..... 50
Poisson Distribution ..... 50
Assumptions of the Poisson Process ..... 51
Negative Binomial Distribution ..... 52
The Need For a Modified CUSUM ..... 55

3. DESIGN OF TRADITIONAL CUSUM PROCEDURES ..... 60

Notation and Conventions ..... 60
Design of Count Data CUSUMS ..... 62
Determining $K$ and $H$ ..... 64
Formulation For General CUSUMS ..... 65
Markov Chain Approach ..... 69
Run Length Distribution and Average Run Length ..... 71

4. PROPOSED MODIFIED CUSUM PROCEDURE ..... 77

Additional Notations and Terminology ..... 79
General Description of the Procedure ..... 85
Modified CUSUM Algorithm ..... 89
Generalization: $n_B$ Buffer States with *M-in-a-row* ..... 91
Simple Illustrative Cases ..... 99
Possible Transitions in the Modified CUSUM Scheme ..... 102
Construction of Extended Transition Probability Matrix ..... 106
Computing The Probability of Extremeness ..... 108
Run Length Distribution and Average Run Length ..... 110
Choosing $K, H,\ W,$ and $\pi_\alpha$ ..... 111

5. NUMERICAL EXAMPLES ..... 116

Examples: Based On Sample Data ..... 116
With No Buffer State ..... 117
With One Buffer State ..... 122
With Two Buffer States ..... 129
Example: Changing $\pi_\alpha$ Values ..... 133

6. COMPUTATIONAL RESULTS ..... 137

Evaluation and Summary of Results ..... 137
Example: Using The FIR Feature ..... 139
Design Implementation ..... 143
Design Example ..... 145

7. APPLICATION ..... 147

Computing the Group Transition Matrix ..... 147

8. CONCLUSIONS, DISCUSSIONS AND FUTURE RESEARCH ..... 153

Conclusions ..... 153
Discussions and Future Research ..... 154

REFERENCES CITED ..... 159

APPENDICES ..... 169

APPENDIX A – Figures (Plots) of Average Run Lengths (ARLs) For Modified Upper CUSUM (With No FIR) Using In-Control Means $\mu_0 = 2.0, 4.0, 6.0, 8.0$ and $10.0$. ..... 170
APPENDIX B – Average Run Length (ARL) Tables For Modified Upper CUSUM (With No FIR) Using In-Control Means $\mu_0 = 2.0, 4.0, 6.0, 8.0$ and $10.0$. ..... 186

APPENDIX C – Average Run Length (ARL) Tables For Modified Upper CUSUM (With FIR: Headstart $S_0^+ = H/2$) Using In-Control Means $\mu_0 = 2.0, 4.0, 6.0, 8.0$ and 10.0. ..... 194
APPENDIX D – S-PLUS Programs For Implementing Modified CUSUM . 205

# LIST OF TABLES

Table Page

1. Sample output for a modified upper CUSUM design with $K = 4$, $H = 6$, and $W = 5$, based on Poisson data with in-control mean $\mu = 3.8$. This output for Example 5.1 illustrates the case of no buffer state in region $B$. ..... 120

2. Extended Transition Matrix for the modified upper CUSUM design in Example 5.1. The design has $K = 4$, $W = 5$, and $H = 6$, based on Poisson data with in-control mean $\mu = 3.8$. This matrix is identical to a transition matrix for a regular Poisson CUSUM with $K = 4$ and $H = 6$ (because there is no buffer state in region $B$). 121

3. Sample output for a modified upper CUSUM design with $K = 4$, $H = 6$, $W = 4$, and $\pi_\alpha = 0.05$, based on Poisson data with in-control mean $\mu = 3.8$. This is sample output for Example 5.2, illustrating the case with one buffer state in region $B$. The out-of-control signal is given at sample 14, where $\pi_{5,3} \leq 0.05$. ..... 126

4. Extended Transition Matrix for the modified upper CUSUM design with $K = 4$, $H = 6$, $W = 4$, and $\pi_\alpha = 0.05$ for Poisson data with in-control mean $\mu = 3.8$ in Example 5.2. This design has one buffer state (state $E_5$) in region $B$..... 127

5. Modified Transition Matrix for the modified upper CUSUM design with $K = 4$, $H = 6$, $W = 4$, and $\pi_\alpha = 0.05$ for Poisson data with in-control mean $\mu = 3.8$ in Example 5.2. This design has one buffer state (state $E_5$) in region $B$..... 127

6. Output showing probabilities of extremeness for the design in Example 5.3, with $\mu_0 = 3.8$, $K = 4$, $H = 6$ and $W = 3$. ..... 130

7. Sample output for a modified upper CUSUM design with $K = 4$, $H = 6$, $W = 3$, and $\pi_\alpha = 0.05$, based on Poisson data with in-control mean $\mu = 3.8$. This output is part of Example 5.3, demonstrating the case with two buffer states. Out-of-control signal of type C-ABS occurs at sample 15. ..... 131

8. Transition table for the modified upper CUSUM design with $K = 4$, $H = 6$, $W = 3$ and $\pi_\alpha = 0.05$, for Poisson data with in-control mean $\mu_0 = 3.8$. This is part of Example 5.3 and it demonstrates the case of two buffer states. ..... 132

9. Comparison of Probabilities of Extremeness and Selected Alpha-Expanded States for Example 5.4, using $K = 7$, $H = 5$, $\pi_\alpha = 0.05$, and $W = 4, 3, 2, 1$..... 134

10. ARLs at in-control mean ($\mu = 3.5$) and several shifts for Example 5.4, using a modified upper CUSUM with $K = 7$, $H = 5$, $\pi_\alpha = 0.05$, and $W = 4, 3, 2, 1$..... 134

11. Average run lengths for modified CUSUM (No FIR), for five in-control means $\mu_0 = 2.0$, 4.0, 6.0, 8.0, and 10.0; $H = 7.0, 10.0$; $W = 3, 6$; $K = 1$,2, 3, 5, 7; and several shift levels..... 138

12. Percentage reductions in ARLs (with NO FIR) of a modified upper CUSUM scheme with in-control mean $\mu_0 = 4.0$, $H = 7$, $K = 7$, $\pi_\alpha = 0.05$, $W = 6, 5, 4, 3$, and several shift levels for Example 6.1. The first column of ARLs are the in-control ARLs values..... 141

13. ARLs (With and Without) the FIR Feature for Poisson modified upper CUSUM. Italicized values correspond to the use of FIR with $S_0^+ = H/2 = 3.5$. The in-control mean is $\mu = 4.0$; with several shift values (for Example 6.1). Parameters are $H = 7$, $K = 7$, $\pi_\alpha = 0.05$, and various $W$ values. ..... 143

# LIST OF FIGURES

Figure Page

1. A graphical representation of the modified CUSUM chart, showing a typical sample path with the two regions $A$ and $B$. Here $K$ is the reference value, $\mu_0$ is the target, $W$ is the warning limit, and $H$ is the decision interval. $S^+_{n,c_n}$ is a one-sided upward modified CUSUM statistic. ..... 86

2. A schematic representation (flowchart) of the modified CUSUM procedure, based on the four-in-a-row warning runs rule. ..... 92

3. A Modified (Upper) CUSUM chart for simulated Poisson data with an in-control mean of $\mu = 3.8$ for Example 5.1. The design parameters are $K = 4$, $H = 6$, and $W = 5$ (illustrating the case where there is no buffer state in region $B$). Hence, this design is equivalent to a regular Poisson CUSUM chart with parameters $K = 4$, and $H = 6$. Out-of-control signal occurs at sample 17 with type H-ABS. ..... 118

4. A Modified (Upper) CUSUM chart for simulated Poisson data with in-control mean $\mu = 3.8$. The design parameters are $K = 4$, $H = 6$, $W = 4$ and $\pi_\alpha = 0.05$. This is part of Example 5.2 and it illustrates the case of one buffer state in region $B$. Out-of-control signal occurs at sample 14, with type A-ABS. ..... 125

5. A Modified (Upper) CUSUM chart for simulated Poisson data with in-control mean $\mu = 3.8$. The design parameters are $K = 4$, $H = 6$, and $W = 3$ and it illustrates Example 5.3 with a four-in-a-row warning runs rule. The process signals out of control (with type C-ABS) at sample 15. ..... 130

6. Decreasing trend of ARLs (with no FIR) for Modified (Upper) CUSUM scheme used in Example 6.1. ..... 141

7. ARLs for Modified (Upper) CUSUM charts for various values of $H$, $K = 1, 2$, and $W$ (determined by $n_B$), when the process experiences a step shift of $\mu$ (given in multiples of the in-control mean $\mu_0 = 2.0$). 171

8. ARLs for Modified (Upper) CUSUM charts for various values of $H$, $K = 3, 5$, and $W$ (determined by $n_B$), when the process experiences a step shift of $\mu$ (given in multiples of the in-control mean $\mu_0 = 2.0$). 172

9. ARLs for Modified (Upper) CUSUM charts for various values of $H$, $K = 5, 7$, and $W$ (determined by $n_B$), when the process experiences a step shift of $\mu$ (given in multiples of the in-control mean $\mu_0 = 2.0$). 173

10. ARLs for Modified (Upper) CUSUM charts for various values of $H$, $K = 1, 2$, and $W$ (determined by $n_B$), when the process experiences a step shift of $\mu$ (given in multiples of the in-control mean $\mu_0 = 4.0$). 174

11. ARLs for Modified (Upper) CUSUM charts for various values of $H$, $K = 3, 5$, and $W$ (determined by $n_B$), when the process experiences a step shift of $\mu$ (given in multiples of the in-control mean $\mu_0 = 4.0$). 175

12. ARLs for Modified (Upper) CUSUM charts for various values of $H$, $K = 7$, and $W$ (determined by $n_B$), when the process experiences a step shift of $\mu$ (given in multiples of the in-control mean $\mu_0 = 4.0$). 176

13. ARLs for Modified (Upper) CUSUM charts for various values of $H$, $K = 1, 2$, and $W$ (determined by $n_B$), when the process experiences a step shift of $\mu$ (given in multiples of the in-control mean $\mu_0 = 6.0$). 177

14. ARLs for Modified (Upper) CUSUM charts for various values of $H$, $K = 3, 5$, and $W$ (determined by $n_B$), when the process experiences a step shift of $\mu$ (given in multiples of the in-control mean $\mu_0 = 6.0$). 178

15. ARLs for Modified (Upper) CUSUM charts for various values of $H$, $K = 7$, and $W$ (determined by $n_B$), when the process experiences a step shift of $\mu$ (given in multiples of the in-control mean $\mu_0 = 6.0$). 179

16. ARLs for Modified (Upper) CUSUM charts for various values of $H$, $K = 1, 2$, and $W$ (determined by $n_B$), when the process experiences a step shift of $\mu$ (given in multiples of the in-control mean $\mu_0 = 8.0$). 180

17. ARLs for Modified (Upper) CUSUM charts for various values of $H$, $K = 3, 5$, and $W$ (determined by $n_B$), when the process experiences a step shift of $\mu$ (given in multiples of the in-control mean $\mu_0 = 8.0$). 181

18. ARLs for Modified (Upper) CUSUM charts for various values of $H$, $K = 7$, and $W$ (determined by $n_B$), when the process experiences a step shift of $\mu$ (given in multiples of the in-control mean $\mu_0 = 8.0$). 182

19. ARLs for Modified (Upper) CUSUM charts for various values of $H$, $K = 1, 2$, and $W$ (determined by $n_B$), when the process experiences a step shift of $\mu$ (given in multiples of the in-control mean $\mu_0 = 10.0$). 183

20. ARLs for Modified (Upper) CUSUM charts for various values of $H$, $K = 3, 5$, and $W$ (determined by $n_B$), when the process experiences a step shift of $\mu$ (given in multiples of the in-control mean $\mu_0 = 10.0$). 184

21. ARLs for Modified (Upper) CUSUM charts for various values of $H$, $K = 7$, and $W$ (determined by $n_B$), when the process experiences a step shift of $\mu$ (given in multiples of the in-control mean $\mu_0 = 10.0$). 185

# ABSTRACT

Radio frequency (RF) technology is used in electronic monitoring and data acquisition devices currently available to the commercial animal feeding industry for continuously monitoring the feeding and watering behaviors of feedlot animals.

There is therefore the need for statistical process control (SPC) procedures that can be used on-line, in conjunction with these electronic data collection systems, to achieve a cost effective system that can quickly detect animal morbidity.

In this dissertation, a modified cumulative sum (modified CUSUM) procedure is proposed based on modifying the traditional CUSUM scheme for count data. Additional CUSUM design parameters and additional out-of-control conditions are introduced to give the procedure several enhanced features, improve its detection capability, and reduce the average run length (ARL).

The method is evaluated using simulated Poisson data, and recommendations for the choice of the extra design parameters are discussed. An outline is provided to show how the modified procedure can be implemented for a generic cattle feedlot data, which can be acquired from a digital monitoring and data collection system based on radio frequency (RF) ear-tag technology.

Results demonstrate that the modified CUSUM scheme can indeed achieve higher sensitivity and performance than the traditional CUSUM scheme. The proposed modified CUSUM schemes can also be designed to be relatively robust to isolated outliers that may be present in the data. It is further demonstrated that an optimal design of this modified CUSUM scheme can indeed be very useful in the early detection of morbidity among group-fed animals.

# CHAPTER 1

## INTRODUCTION AND BACKGROUND INFORMATION

The rapid evolution of electronic monitoring and control systems has stimulated the search for more advanced statistical quality control (SQC) procedures that can be linked on-line to operate in conjunction with these automatic data acquisition and monitoring systems. There appears to be two main reasons for this rapid development. First are the tremendous advances in computer and information technology, which have made digital monitoring and data collection systems cheaper and more effective (in terms of their capability to continuously monitor a process and obtain measurements on several variables for every item produced). The second reason is the theoretical interest and the inherent challenge posed by the usually non-normal, autocorrelated nature of the data generated by these electronic data collection systems.

The incredible potential of automated systems to improve productivity and quality as well as their other related issues, have recently been discussed by Keats [60], among several others. Rather than a system that analyzes the sample data after the product is produced, many manufacturing and process industries are looking for a cost effective system that can continuously monitor the process during production and that identifies and reports trouble spots before bad products are made (Keats

[60]). Many industries, such as the automobile, chemical, semi-conductor, and food and beverage industries are already implementing sophisticated versions of such digital control systems (Palm, Rodriguez, Spiring, and Wheeler [85]), and many more are set to follow, including the commercial cattle-feeding industry. SPC related problems in the cattle-feeding industry discussed in this dissertation can be summarized as follows: we have discrete data from a cattle feedlot that are potentially overdispersed, and could be autocorrelated because they were obtained from a high-speed data acquisition system. A simple easily-implemented statistical process control procedure is desired, that can be used to predict morbidity in cows by monitoring the mean of this process.

In this dissertation, a modified cumulative sum (modified CUSUM) control chart technique is proposed for use in the cattle-feeding industry. The method involves modifying the traditional CUSUM scheme by introducing additional parameters and additional out-of-control conditions to give the procedure several enhanced features. Determination of the parameter settings and the average run length are discussed, and simulated Poisson data are used to evaluate the procedure. Several examples are provided to illustrate the method. Finally, an outline is provided for implementing the modified procedure for a generic cattle feedlot data.

In the remainder of this chapter, I will first provide and discuss several general definitions and terminology from the field of statistical quality control that are used

frequently in this dissertation. Next will be a brief discussion on background information and overview of the relevant statistical process control literature. This is followed by a review of general control charting procedures, and a discussion on the main motivation behind this research (which also outlines the main problem being considered in this dissertation).

## Definitions and Terminology

This section outlines and discusses several quality control definitions and terminology conventions that are used throughout this dissertation. The definitions have been adopted mainly from texts by Besterfield [7]; Burr [19] and Montgomery [77].

1. A product refers to all categories of manufactured goods such as computers and clothing, processed goods including meat products and other food items, and services such as banking and health care. Every product is considered to be the result of a process.

2. Quality is defined formally as the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs (*International Standards Organizations*, ISO 8402).

   Implicit in the above definition of quality are the following meanings discussed in Montgomery [77, page 4]:

   (a) quality means fitness for use.

(b) quality is inversely proportional to variability. Thus, a decrease in the variability of an on-aim process will, by implication, increase the quality of the output (or product) from that process.

3. Quality improvement is the reduction of variability in products and processes.

4. A quality product is a good or excellent product that fulfills customers' expectation based on the intended use and selling price.

5. Control is the process of regulating or directing an activity to verify its conformance to a standard and to take corrective action if required (Besterfield [7, page 1]).

6. Quality control is the regulatory process for those activities which measure a product's performance, compare the performance with established standards, and pursue corrective action regardless of where those activities occur (Besterfield [7]).

7. Statistical quality control (SQC) is the branch of quality control that uses statistical methods to collect data in a sound, unbiased manner and to analyze and interpret the results appropriately so as to obtain satisfactory, dependable and economic quality (Besterfield [7]).

8. Statistical process control (SPC) is the part of SQC consisting of a collection

of powerful problem-solving tools useful in achieving process stability and improving capability through the reduction of variability.

9. Data collected from any service or production process have one thing in common: they vary. They can vary from time to time, customer to customer, operator to operator, piece to piece, and sample to sample. Burr [19] puts it more succinctly: "whenever we have variation we have a statistical problem, whether we like it or not, whether we know it or not." For data collected by on-line data acquisition systems, the sampling intervals tend to be very short, and large amount of data are generated within that short sample interval. Consequently, data obtained from such automatic data collection systems often tends to be correlated. The two main sources of variation discussed in the SPC environment are defined below.

   (a) Chance causes (or common causes) of variation refer to the natural, purely random variability or background noise that are an inherent part of the process. This represents the "stable" pattern of variation in the process.

   (b) Assignable causes (or special causes) of variation are the systematic non-random patterns of variation that are not part of the chance cause pattern, but which may occasionally be present in the output of a process. Variations of this type are generally larger than the background noise and can result in an unacceptable level of process performance. Removing special

causes usually improves the process (and the quality of the product).

In the above definitions, it should be mentioned that the terms chance and assignable causes were introduced by Shewhart, whereas W. Edwards Deming [25], who pioneered the philosophy of total and continuous quality improvements, suggested the corresponding alternative terms common and special causes.

10. The process is said to be in a state of statistical control when the assignable causes have been eliminated and the process is operating with only chance causes of variation present. At this stage, Besterfield [7, page 73] says that "no higher degree of uniformity can be attained with the existing process, ... except through a change in the basic process itself."

11. An out-of-control process is a process that is operating in the presence of special or assignable causes of variation.

12. The control chart is a very powerful SPC tool that can be used on-line to monitor and quickly detect the presence of assignable causes of process shifts so that investigative and corrective actions can be taken.

13. Shewhart charts are the conventional (or classical) control charts that follow the general construction principles developed by Dr. Walter Shewhart [101]. Examples of these are the $X$-chart (individuals chart), $R$-chart, $\overline{X}$-chart and $s$-chart for variables data, and $p$- chart, $np$-chart, $c$-chart and $u$-charts for attribute

data.

14. A CUSUM control chart (or CUSUM) is an abbreviation for the cumulative sums control chart.

15. Traditional (or standard or regular) CUSUM refers to the classical CUSUM design without any of the recent modifications or enhancements.

16. EWMA control chart (or EWMA) is an abbreviation for the exponentially weighted moving average control chart.

17. ARL is the abbreviation for the average run length (which is the mean of the run length distribution). In other words, the ARL is the average number of observations (or samples) before the process signals out-of-control. An essential distinction is made between the in-control ARL and the out-of-control ARL. The in-control ARL is the expected time until the process yields a false out-of-control signal, while it is in control. A signal occurring while the process is deemed to be in-control, is also referred to as a false alarm. The out-of-control ARL is the expected time until the process signals an alarm once it shifts to an out-of-control state. Ideally, it is desired that the in-control ARL will be high (meaning low false alarm rate) while the out-of-control ARL will be low.

18. The optimality property of a control scheme can be stated as follows: among all procedures with the same in-control ARL, the optimal procedure is the one

that has the smallest (or quickest) expected time until it signals a change, once the process shifts to the out-of-control state (Moustakides [81]).

19. An item is said to be nonconforming if it fails to conform to product specifications in some respect. Otherwise, it is said to be conforming. Specific instances of failure in a particular item are called nonconformances (or nonconformities).

20. Discrete data refer to observations made on a discrete variable. For example, the result of counting nonconformities or nonconforming pieces in a sample so that only whole numbers or integers may occur is discrete. The term "attribute" data originally meant counts of the number of conforming and nonconforming items in a sample (Hawkins and Olwell [47, page 105]). In recent usage however, attribute data is typically used when the emphasis is on the classification of items into conforming and nonconforming. This is an attempt to distinguish attribute data from "count data", a term often used when the emphasis is on the frequency of occurrence of nonconformances or other discrete events. Most authors think that this technical distinction is not essential, since both count and attribute data are discrete. Subsequently, the term count data is used for all discrete data because it has wider appeal (Lucas [68]), and this convention will be adopted here in this dissertation.

21. Variables data refers to the numerical value of a quality characteristic that is measured on a continuous scale.

## Background and Overview of Relevant SPC Literature

The purpose of this section is to provide background information and discuss relevant statistical quality control literature pertaining to the aims of this dissertation.

Quality control is a very broad field that deals with the regulatory process for those activities which measure a product's performance, compare the performance with established standards, and pursue corrective action regardless of where those activities occur. The formal beginning of statistical quality control (SQC) can be traced back to 1924, when Dr. Walter A. Shewhart of Bell Telephone Laboratories developed the first control chart. Later in 1931, Shewhart's [101] classic book *Economic Control of Quality of Manufactured Product* was published. It contained a complete exposition of the theory, practical application, and economics of the control charts. Burr [19, page 2] remarks in his 1979 book *Elementary Statistical Quality Control* that "seldom has a whole field of knowledge been so well explored and its application so well pointed in the first publication."

Statistical quality control charts made a great impact on manufacturing processes during World War II. The American Society for Quality Control which was subsequently formed in 1946, has been instrumental in promoting the use of quality control for all types of production and service. Today, control charts have become a popular and powerful statistical process control (SPC) tool, and they have found useful application in many fields including the manufacturing industry, management, government, crime control, health care and environmental surveillance. Volumes have

been written about the construction, applications, and properties of statistical control procedures. Texts such as the 1988 book by Grant and Leavenworth [39] titled *Statistical Quality Control* and Besterfield's [7] 1979 book *Quality Control* together with Burr [19], provide insightful and straight forward introduction to the Shewhart quality control charts. *Introduction To Statistical Quality Control* by Montgomery [77] is a more accessible textbook, and it provides excellent discussions on Shewhart charts, as well as the CUSUM and EWMA control schemes with interesting applications.

While Shewhart charts are widely used, easy to construct and interpret, and very effective for detecting large shifts in the parameter of the process, they are not so useful for detecting smaller shifts. In addition, there are violations of the underlying assumptions and the occurrence of trends and unusual patterns that needed to be studied. Goldsmith and Whitfield [38] had this to say about the Shewhart charts in 1961:

> The Shewhart quality control chart with fixed control lines suffers from the disadvantage that the observations are viewed independently, no account being taken of runs of observations all higher or all lower than the long-term mean of the process. This results in a relative insensitivity to moderate changes in the short-term mean value.

In response to these deficiencies, several adaptations and modifications of the Shewhart charts have been considered. These include schemes devised with both warning lines and action limits based on zone rules. Many of these modified Shewhart charts are discussed in Hill [49], Page [87, 88], and Mandel [75] who examined control charts that exhibit certain types of trends. The Western Electric [119] *Statistical Quality*

*Control Handbook* of 1956 contains several of these supplementary runs tests. The use of these supplementary runs tests (or *sensitizing rules* as they are often called) were advocated in an effort to improve the sensitivity of the Shewhart charts in detecting shifts of relatively smaller magnitude. Because the implementation of Shewhart control charts with these supplementary runs rules is often cumbersome and difficult to maintain, the search for other improvements and alternatives continued.

E. S. Page [86], in 1954, introduced the cumulative sum (CUSUM) control chart, and Roberts [95] proposed the exponentially weighted moving average (EWMA) control scheme in 1959, both as alternatives to the Shewhart charts. Duncan [26] also proposed models in 1956 that deal with the economic aspects in the design of control charts.

In what follows, a greater emphasis will be placed on the CUSUM procedure because it will be used to address the aims of this dissertation. Early papers that deal with the mathematical foundations of CUSUMs include Page [86], Ewan and Kemp [27], Kemp [62] and a series of three illuminating articles by Johnson and Leone [55, 56, 57] who also developed optimal CUSUM schemes for several distributions in the exponential family. Of particular interest is the realization that the CUSUM is based on the likelihood ratio strategy, and it can be considered as a sequence of Wald sequential probability ratio tests (SPRTs).

Two forms of the CUSUMs are generally discussed in the literature: tabular (or decision interval) CUSUM and the graphical V-Mask. The tabular form of the

CUSUM is most preferred because it has decision intervals which serve as control limits. The original CUSUM scheme proposed by Page plots the cumulative score

$$S_n = \sum_{i=1}^{n} x_i$$

against $n$, where $S_0 = 0$, and $x_i$ is a score assigned to the $i$th observation. Based on Page's procedure, a one-sided control scheme would take action after the $n$th observation if

$$S_n - \min_{0 \le i < n} S_i \ge h,$$

where $h$ is a positive constant called the decision interval (Goldsmith and Whitfield [38]). This initial procedure, however, looked more complicated when applied to the two-sided control case. G. A. Barnard [4], in a very stimulating article published in 1959, proposed a simpler method by shifting the origin of measurements to a target value $\mu$, and plotting the cumulative sums of deviations

$$S_n = \sum_{i=1}^{n} (x_i - \mu).$$

In this way the CUSUM can be considered as a random walk model. The symmetric V-shaped mask was advocated for a two-sided control scheme based on this cumulative deviation chart. In 1960, Ewan and Kemp [27] also conducted a detailed study of the CUSUM procedure, and they showed that instead of plotting the cumulative sum of deviations, it is better to plot

$$S_n = \sum_{i=1}^{n} (x_i - k)$$

where $k$ is a reference value. They also showed that for a $k$ value near $(\mu_0 + \mu_1)/2$, the ARL at the acceptable mean level $\mu_0$ is near its maximum for a given ARL at an unacceptable process level $\mu_1$. In a CUSUM design therefore, the values of $k$ and $h$ need to be determined to achieve desirable ARL properties.

An extensive body of literature now exists on the design of CUSUM control charts. Bissell [9], Lucas [69], and Woodall [124] (on general designs of CUSUM techniques); Hawkins and Olwell [46] (on CUSUMs for location and shape parameters based on the inverse Gaussian distribution); Bourke [14] and Lucas [70] (on CUSUMs for low count-level processes); and Taylor [110] (on economic design of CUSUM charts) are very informative. Van Dobben de Bruyn's [113] 1968 monograph titled *Cumulative Sum Tests: Theory and Practice*, is one of the earliest texts published about CUSUMs. A more recent text is the *Cumulative Sum Charts and Charting for Quality Improvement* by Hawkins and Olwell [47] which is an extremely useful, compact book devoted entirely to CUSUMs. Hawkins and Olwell discuss virtually all the essential aspects and techniques in the design and application of CUSUMs, and they also provide an extensive bibliographical index. In Chapter 6, they provide concise summary of the theoretical foundations of the CUSUM, and the derivation of the optimal CUSUM designs for several distributions belonging to the exponential family (several of them originally discussed by Johnson and Leone [55, 56, 57]). Hawkins and Olwell [47] also provide interesting discussions on the various methods currently available for computing the ARLs of the CUSUM. What is remarkable is that the CUSUMs for count

distributions are generally defined in exactly the same way as CUSUMs for continuous variables, and that the CUSUM can be easily designed for other distributional parameters apart from the mean.

Both CUSUM and EWMA schemes are superior to the Shewhart charts in many ways. Both make use of past data, and are able to react quickly to small persistent shifts in the parameter of the process. Although the performance of the Shewhart control scheme with supplementary runs rules improves greatly, it does not equal the performance level of the CUSUM (Champ and Woodall [20]). Page [88] showed that many of the amended Shewhart rules are in fact equivalent to a restricted type of the CUSUM scheme. In particular, the classical Shewhart chart with fixed control limits is known to be equivalent to the a CUSUM scheme where the parameters $h = 0$, and $k$ is equal to the control limit (see for example, Lucas [74]). It has also been shown by Bissell [10] that the CUSUM performs much better than the Shewhart chart when the process has a linear trend rather than an abrupt shift. Further discussions on control charts under a linear trend can be found in Gan [32], among others.

Optimality properties of the CUSUM design have also been examined by several authors. Lorden [65] studied the asymptotic optimality of CUSUM procedures for detecting a step change in distribution. Moustakides [81] used Lorden's criteria and showed that among all tests with the same false alarm rate, the CUSUM has the smallest reaction time for detecting the shift from an in-control distribution to a single specific out-of-control distribution. Ritov [94] (using a decision theory approach),

Yashchin [133], Gan [31] and Buckley [18] have also studied optimality properties of the CUSUM. Optimal CUSUM designs have been discussed for binomial counts (Gan [33]), and Gan [34] has discussed optimal design of exponential CUSUM control charts. The EWMA control chart has also been studied extensively and its optimality properties and ARL have been evaluated (see for example, Crowder [22, 23]; Lucas and Saccucci [73]; Srivastava and Wu [107]).

Concerning CUSUM and EWMA, most of the literature I have seen so far seems to suggest that, while they may be nearly close in performance and efficiency, the CUSUM is more widely used and preferable. Even when forecast errors from correlated data are being monitored, properly designed CUSUMs can outperform the EWMAs (Vander Wiel [114]). Most implementation of statistical control charts are concerned with the monitoring of the mean of a process data, often under Gaussian assumptions. While both the CUSUM and EWMA schemes are efficient for detecting small shifts in the mean, they perform poorly against large shifts which, on the contrary, can be promptly detected by the Shewhart individuals charts.

To address this and other inadequacies, several adaptations, extensions and modifications of the CUSUM and EWMA schemes have been proposed. This includes the combined Shewhart-CUSUM schemes for the process mean (Lucas [67]; Yashchin [129]) which utilizes the nice features of both procedures to rapidly detect small as well as large shifts in the process mean. CUSUM procedures for variability and scale parameters (Hawkins [41]; Srivastava [106]) and combined Shewhart-CUSUM schemes

for the process variance (Gan [30]) have also been discussed. In fact as far back as 1961, Goldsmith and Whitfield [38] had suggested that: "it may sometimes be better to plot a Shewhart chart and a cumulative deviation chart simultaneously to obtain a picture of the process variation."

Hawkins [42] described another interesting extension of the CUSUM, called the self-starting CUSUM scheme, which seeks to avoid the problem of parameter estimation by standardizing individual, successive observations using the mean and standard deviation of all observations accumulated to date. Implementing this scheme for a count data CUSUM can, however, be complicated, and the results can be greatly impacted by the presence of outliers in the data. Adaptive and Bayesian procedures have also been discussed for the CUSUM (e.g. Healy [48]; Joseph and Bowen [59]) and for the EWMA (e.g. Hubele and Chang [50]). CUSUM schemes based on variable sample size have also been studied (e.g. Hughes, Reynolds and Arnold [51]). Other authors like Bissell [12], and Yashchin [130] have studied weighted CUSUMs, which can be used to monitor process data when they are paired with measurements on another variable. Hawkins and Olwell [47, chapter 3] provides a detailed discussion of these procedures.

Two exceptionally useful enhancements to the CUSUM control schemes are the Fast Initial Response (FIR) CUSUMs (Lucas and Crosier [71]), where the CUSUM is set to an initial positive head start value to give a faster signal if the process is out of control, and the robust CUSUMs of Lucas and Crosier [72] and Hawkins [45]. Lucas

and Crosier [72] use a method based on a two-in-a-row outlier rule, while Hawkins [45] uses a windsorization procedure whereby an observation exceeding a specified threshold value would be replaced by that threshold. This "windsorized" value would then be used in updating the CUSUM. Lucas and Saccucci [73] also discussed EWMA control schemes with the fast initial response feature, combined Shewhart-EWMA and robust EWMA control charts.

The performance of a control chart is usually measured by the average run length (ARL) which, for the CUSUM and EWMA, is generally very difficult to calculate except through the use of numerical and analytical approximations (e.g. Page [86, 88]; Ewan and Kemp [27]; Goel and Wu [37], and Bissel [9]). In 1972, Brook and Evans [17] developed a Markov chain model for analyzing CUSUMs, which allowed simpler calculations of the run length distribution and its moments. Their method has since been extended by many, including Woodall [122], Lucas [67], Lucas and Crosier [71, 72], and Lucas and Saccucci [73] who used the Markov chain technique to compute the ARLs of several modified CUSUM schemes. The Markov chain approach will be discussed further in Chapter 3.

Although most of the earlier work dealing with CUSUMs involved continuous variable data, there seem to be a growing number of recent applications involving CUSUMs for count data. A comprehensive bibliography and review of control charts based on attributes can be found in Woodall [125]. Brook and Evans [17] were the first to consider CUSUMs based on the Poisson distribution, and they demonstrated how

the ARL and other moments of the run length distribution can be obtained via the Markov chain approach. Lucas [68] provided a more detailed discussion of count data CUSUMs, with particular emphasis on Poisson CUSUMs and time-between-events CUSUMs.

One very important fact about the design of CUSUMs is that a CUSUM scheme requires the precise statistical distribution for the data to be specified together with the exact parameter values of the model. This can be very problematic since in practice the exact model is never known, and parameter values must be estimated. Munford [82], in 1980, developed the cumulative score (CUSCORE) technique which discretizes the observations by assigning a score of -1, +1, or 0 to the sample values according to whether they are "extreme negative", "extreme positive" or otherwise. Even though his procedure is attractive in the sense that it avoids the distributional problems with the data and provides simple and compact expressions for the calculation of the ARL, it is not very efficient for detecting larger deviations in the process mean. Munford's procedure was later refined by Ncube and Woodall [84] who used a combined Shewhart-CUSCORE scheme to achieve better reaction to both large and small shifts. More recently, Radaelli [92] extended Munford's procedure to the CUSUM surveillance of rare health events, and Radaelli [93] considered an extension to the Poisson and negative binomial dynamics. Related procedures for grouped-data, based on gauging, have been discussed (e.g. Xiao [128]; Steiner, Geyer and

Wesolowsky [109]; Steiner [108]) where an observable continuous variable is measured, but the data values are simply classified into one of several classes or intervals of values of the variable.

The following discussion concerns the treatment of autocorrelated data in SPC, and it is very important because it has direct bearing on the justification of assumptions that are made in the last section of chapter 2 of this dissertation. Conventional control charts were initially introduced as monitoring tools for detecting the presence of out-of-control situations, which are typically caused by systematic nonrandom patterns of variation in the process. Identifying and eliminating (or drastically reducing) those causes of variation is very essential. The control chart relies on the assumption that if the process is operating in the state of statistical control, then the observations from the process can be regarded as independent and identically distributed random variables. Within the environment of SQC, the assessment of the state of a process is usually considered in terms of hypothesis testing, where the null hypothesis is often stated as

$H_0$ : The process is in the state of statistical control

and the alternative hypothesis is

$H_1$ : The process is out of control.

The emphasis on hypothesis testing when the data are serially correlated and SPC is used to monitor the residuals from a time series model, has generated some vigorous debate among statisticians. The presence of autocorrelation makes the assumption of

independence inappropriate, and can invalidate the Markov property of the control schemes (Yashchin [132]) and lead to poor control chart performance. Goldsmith and Whitfield [38], Johnson and Bagshaw [54], and Bagshaw and Johnson [5] were among the first to carry out extensive study about the effects of autocorrelation on the CUSUM and other control schemes. They showed that positive autocorrelation can reduce the in-control ARL of the process while negative correlation leads to higher ARLs. However for larger shifts or deviations in the current mean, the serial correlation has little effect on the ARL.

Two general approaches to dealing with inherently autocorrelated process data have been discussed in the literature (e.g. Adams, Woodall and Superville [1]; Vander Wiel [114]). One approach calls for widening the control limits (usually for positively correlated observations) to achieve an acceptable rate of false alarms. This method, while capable of reducing the number of false alarms, can also render the control chart useless for the purpose of signaling unusual behavior that could result in process improvement. For negatively correlated observations the control limits are usually narrowed to achieve similar effect. The second approach involves modeling the autocorrelation and it appears to be more appealing to researchers, partly because of the academic challenge and interest posed by the nature of the data. Most of these procedures still emphasize hypothesis testing. Barnard [4] first touched on this issue in his 1959 classic article where he suggested that:

> Now that the theory of stochastic processes has grown into a well-rounded

> body of theory, it would seem appropriate to consider industrial processes whose natural, intrinsic variability is best described as a run of dependent random variables, that is, as a stochastic process.

Barnard's proposal that the object of control charts should be to model and estimate the process mean drew both widespread support and concern from many statisticians at the time.

Recently these views have resurfaced in an article by Alwan and Roberts [2] who believe that "if a process can be modeled, then the traditional objectives of quality control–or surveillance–can be better served." Expressing concern that the concentration on hypotheses testing runs the danger of narrowing the perspectives of SPC procedures, Alwan and Roberts argued that the dichotomy of a "state of statistical control" versus "out of control" is too sharp, and they wrote that " a state of statistical control is often hard to attain; indeed, in many applications it appears that this state is never achieved, except possibly as a crude approximation." Also, in a four-panel discussion on the subject of "Process Control and Statistical Inference" (Crowder, Hawkins, Reynolds, and Yashchin [24]), three of the discussants affirmed the dominant view that any problem involving detection of changes in a process is inherently related to hypothesis testing, while Crowder thinks there should be more emphasis on estimation and engineering control, arguing that the emphasis on hypothesis represents a major deficiency in most control charts with autocorrelation.

Indeed, several authors, including Alwan and Roberts [2], Montgomery and Friedman [78], Montgomery and Mastrangelo [79], Box and Kramer [15], Harris and Ross

[40], Wardell, Moskowitz, and Plante [116], and have also recently suggested that autocorrelation should be modeled by fitting an appropriate time series effects, and then applying standard control charting procedures to the residuals (or the one-step forecast errors). However, most of these papers still emphasize hypothesis testing. The Box-Jenkins family of time series models are often employed in this endeavor. Montgomery and Mastrangelo and Alwan and Roberts have also recommended that the EWMA be used as an approximation to the underlying time series model. Others authors such as Fellner [29] and Lucas [74] have expressed great concern about this modeling approach. During the discussion of Wardell et al. [116], Lucas [74] argued that "it is essential for the data to be correlated (or for there to be a restriction on sample size) for a CUSUM or EWMA control scheme to work effectively", and he went further to suggest variance-component modeling as an alternative approach for dealing with correlated process data.

The above discussions also relate directly to data obtained from on-line electronic control systems of which the generic feedlot data, discussed later in the next section, is an obvious example. Runger and Willemain [97] and Keats [60] have discussed the prevalent trends in many industries to employ digital control and on-line data collection systems and their impact on SPC. Two direct impacts on SPC are mentioned in the literature. First, the measurement interval is so small, often leading to serial correlation among the observations. Second is the tremendous amount of data generated by these high-speed data collection systems for analysis. It is strongly desired to have

a SPC scheme to work in conjunction with these on-line systems during production rather than analyzing the sample data after production.

The lack of consensus in these discussions is evident, suggesting that much less guidance is available for choosing and designing monitoring schemes for potentially autocorrelated data.

The final extensions and adaptations I would like to mention regarding statistical control procedures is in the area of multivariate quality control. This presents far more challenging problems of detection and diagnosis of persistent shifts. Excellent discussion on this subject can be found in Hawkins and Olwell [47, chapter 8] who remark in page 190 of their book that "multivariate control may be much more sensitive to shifts than is the collection of univariate control", and that it "can be more specific in diagnosing causes than is a collection of univariate charts."

## General Control Charting Procedures

In this section, the three general types of control charts commonly discussed in the quality control literature will be reviewed. These are the Shewhart, EWMA and the CUSUM control charts. The same general design principles for these schemes can be applicable for both attribute and variable data types.

1. **Shewhart Charts:**

Suppose some quality characteristic $\theta$ (such as the process mean) is to be monitored, and let $\widehat{\theta}$ be an estimate of $\theta$ based on a random sample of $n$ units. Typically this will be a sample drawn from a process that is operating in statistical control, where the output or the observations are assumed to be independent identically distributed random variables. The values of $\widehat{\theta}$ for each successive sample are then plotted on the control chart. Let $\mu_{\widehat{\theta}}$ and $\sigma_{\widehat{\theta}}$ be the mean and the standard deviation of the sampling distribution of $\widehat{\theta}$. Then the center line (CL), upper control limit (UCL) and lower control limit (LCL) of the Shewhart control chart are given by the general model

$$\begin{aligned} \text{UCL} &= \mu_{\widehat{\theta}} + k_1\sigma_{\widehat{\theta}} \\ \text{CL} &= \mu_{\widehat{\theta}} \\ \text{LCL} &= \mu_{\widehat{\theta}} - k_1\sigma_{\widehat{\theta}} \end{aligned}$$

where $k_1$ is a constant representing the number of standard deviations a particular value of $\widehat{\theta}$ is allowed to vary from $\mu_{\widehat{\theta}}$ without triggering an alarm (or an out-of-control process). The value of $k_1$ is based on the distribution of $\widehat{\theta}$; but in practice, it is customary to choose $k_1 = 3$ (that is, "3-sigma" limits).

Oftentimes, warning limits are also constructed between the center line and the control limits, according to the formulas

$$\begin{aligned} \text{UWL} &= \mu_{\widehat{\theta}} + k_2\sigma_{\widehat{\theta}} \\ \text{LWL} &= \mu_{\widehat{\theta}} - k_2\sigma_{\widehat{\theta}} \end{aligned}$$

where UWL and LWL are the upper and lower warning limits, and where $k_2$ ($k_2 < k_1$) is the number of standard deviations a particular value of $\widehat{\theta}$ is allowed to vary from $\mu_{\widehat{\theta}}$ without triggering a warning alarm. In practice, it is common to choose $k_2 = 2$.

2. **EWMA Charts:**

The EWMA control chart is based on the control statistic

$$Z_i = \lambda X_i + (1-\lambda)Z_{i-1}, \qquad 0 < \lambda \leq 1, \tag{1.1}$$

where $\lambda$ is a specified constant weighting factor, and $X_i$ are a sequence of quality measurements assumed to be independent and identically distributed random variables with mean $\mu$ and variance $\sigma$. Here again, the $X_i$ can be individual observations or some empirical estimate of the process parameter. When the process is in control, the target value is $\mu = \mu_0$. The starting value $Z_0$ is usually taken to be $Z_0 = \mu_0$, otherwise it is computed as the average of the observations. A graphical display of the chart is obtained by plotting $Z_i$ against the time order (or the sample number) $i$.

After successive substitutions, the recursion in (1.1) can be written out as

$$\begin{aligned} Z_i &= \lambda X_i + \lambda(1-\lambda)X_{i-1} + \lambda(1-\lambda)^2 X_{i-2} + \cdots + \lambda(1-\lambda)^{i-1} + (1-\lambda)^i Z_0 \\ &= \lambda \sum_{j=0}^{i-1} (1-\lambda)^j X_{i-j} + (1-\lambda)^i Z_0, \end{aligned}$$

which is an exponentially weighted average of all past observations, where the weights $\lambda \sum_{j=0}^{i-1}(1-\lambda)^j$ and $(1-\lambda)^i$ sum up to unity. It can also be shown that

the variance of the control statistic $Z_i$ is

$$\begin{aligned} \sigma^2_{z_i} &= \text{Var}\left(\lambda\sum_{j=0}^{i-1}(1-\lambda)^j X_{i-j} + (1-\lambda)^i Z_0\right) \\ &= \frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2i}]\sigma^2. \end{aligned} \tag{1.2}$$

The control limits for sample $i$ are obtained by the formulas

$$\begin{aligned} \text{UCL}_i &= \mu_0 + L\sigma\sqrt{\frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2i}]} \\ \text{CL}_i &= \mu_0 \\ \text{UCL}_i &= \mu_0 - L\sigma\sqrt{\frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2i}]} \end{aligned}$$

where $L$ is the multiple of standard deviations a particular value of $Z_i$ is allowed to vary from $\mu_0$ without triggering an alarm. The EWMA control chart triggers an alarm when $Z_i > \text{UCL}_i$ or $Z_i < \text{LCL}_i$. Appropriate values of the parameters $\lambda$ and $L$ have to be determined in the design. As $i \to \infty$, the variance in (1.2) $\sigma^2_{z_i} \to \lambda/(2-\lambda)$. Thus, the control limits can also be based on the asymptotic variance, using the formulas

$$\begin{aligned} \text{UCL} &= \mu_0 + L\sigma\sqrt{\frac{\lambda}{2-\lambda}} \\ \text{UCL} &= \mu_0 - L\sigma\sqrt{\frac{\lambda}{2-\lambda}}. \end{aligned}$$

This represents the situation where the process has been running continuously for a long time and has reached steady-state values.

The EWMA statistic in (1.1) can also be written in the form

$$Z_i = Z_{i-1} + \lambda(X_i - Z_{i-1}),$$

in which case the EWMA can be considered as a one-step ahead forecast for the process (Lucas and Saccucci [73]). It has recently been pointed out by Woodall [126] that because the EWMA control chart limits are determined based on the assumption of independence of the data over time, the EWMA chart is not more robust than other control charts, and that "the fact that the EWMA chart is based on a statistic which could be used (in another way) with autocorrelated data is of no help in this regard."

3. **CUSUM Charts:**

CUSUMs also work by directly accumulating information across successive observations to gain greater sensitivity towards small persistent shifts in the process parameter. The Shewhart and the EWMA control charts defined earlier are inherently two-sided (for monitoring both increases and decreases in the parameter of the process). The tabular CUSUM, on the other hand, is inherently a one-sided control procedure. Maintaining a two-sided control will therefore require two separate, one-sided tabular CUSUM schemes. We will describe in this section the most common CUSUM scheme, which is the normal CUSUM for monitoring the mean of a process (based on Gaussian assumptions). Within this framework, an upper (one-sided) tabular CUSUM scheme $S_i^+$ is used for detecting an increase in the process mean from an *acceptable level* or target $\mu_0$ to an *unacceptable level* $\mu_1$ ($\mu_1 > \mu_0$), where $\mu_1$ is the shift for which maximum sensitivity is required. If the purpose is to detect a decrease in the mean

($\mu_1 < \mu_0$), then the lower (one-sided) CUSUM scheme $S_i^-$ will be required. These two separate schemes can be used together to monitor both increases and decreases from the target value.

Suppose $X_1, X_2, \ldots$, is a sequence of measurements on a process characteristic. The $X_i$'s are assumed to be independent and identically distributed. For monitoring the mean of a normal process, $X_i$ are often sample averages, but they can also be individual observations. Let $\mu$ be the process mean that is to be monitored. Then the classical tabular normal CUSUM for a two-sided control of the process mean is defined by the recursions

$$S_0^+ = 0 \tag{1.3a}$$

$$S_0^- = 0 \tag{1.3b}$$

$$S_i^+ = \max[\,0,\ \ X_i - (\mu_0 + K) + S_{i-1}^+] \tag{1.3c}$$

$$S_i^- = \max[\,0,\ \ (\mu_0 - K) - X_i + S_{i-1}^-] \tag{1.3d}$$

where $S_i^+$ and $S_i^-$ are respectively the upper and lower CUSUM statistics at sample $i$, and where $\max[a,\ b]$ denotes the maximum of $a$ and $b$. The design parameter $K$ is the reference level, and it represents a level of indifference that prevents the scheme from drifting toward the signal level $H$ when the process is on target (Yashchin [132]). The signal level $H$ is commonly called the decision interval. The process will be declared out-of-control if the CUSUM either signals an upward shift in the process mean ($S_i^+ > H$) or signals a lower shift in the

process mean ($S_i^- < H$). It is common in practice to standardize the design parameters $K$ and $H$ for detecting a maximum specified shift of $\Delta$ as multiples of the process standard deviation $\sigma$, where $K = k\sigma$, $H = h\sigma$, and $\Delta = \delta\sigma$. In applications, typical values for the design parameters are $h = 4$ or 5 and $k = 0.5$.

## Motivating Example: Cattle Feedlot Study

Research for this dissertation was motivated by the challenges encountered in dealing with an extensive amount of information collected from a cattle feedlot study using the radio frequency (RF) technology. The radio frequency technology is based on the Growsafe® System, which is a new feeding behavior technology on the market for commercial feedlot (Sowell et al. [104]; Schwartzkopf-Genswein et al. [100]).

### Description of a Generic Feedlot Data

The Growsafe® System typically consists of a black mat that contains an antenna located in the back of the feedbunk, a radio frequency transponder molded into a plastic ear-tag, and a personal computer for the data collection. Each animal in the feedlot is fitted with an ear-tag containing an identification number. Whenever an animal wearing the transponder is close enough to the feedbunk (usually very close, with its head almost lowered into the feedbunk), the system records data such as the identification number, location of the animal in the feedlot, time the visit was made, and the duration of visit. These electronic recordings are sent to a reader panel every 5-6 seconds (an average of about 5.25 sec), and the data collected are sent to a

personal computer for storage and future analysis. This recording of the presence of the animal at the feedbunk is termed a **hit** (or a **feeding hit**).

A generic feedlot data description might look like this: Over 100 cattle will be kept in a feedlot for some prolonged period of time (say, three months), and up to 18 hours of feedlot activity from 6 AM to 12 midnight will be recorded continuously for each animal.

A suitable aggregation interval (or sampling interval) is chosen for the analysis. Typically, a 3-hour aggregation interval will be adequate for this purpose. This aggregation or sampling interval will be selected such that it is large enough in time, to minimize the effects of autocorrelation in the data while retaining enough details about the feeding patterns of the animals in the feedlot.

If a 3-hour sampling interval is chosen, then the 18 hours of feedlot activity will be divided into six 3-hour time periods: 6 AM - 9 AM, 9 AM - 12 PM, 12 PM - 3 PM, 3 PM - 6 PM, 6 PM - 9 PM, and 9 PM - 12 AM. The total number of feeding hits recorded within each 3-hour interval will be taken as one observation, giving a total of six observations per day for each animal. The data for all animals, taken together, will therefore indicate a strong positive skewness and large over-dispersion.

Purpose and Initial Problems

The Growsafe® System was originally developed to improve the diagnosis of sickness and reduce chick mortality in ostriches. The current study however, is being conducted to evaluate the potential for using this technology for a similar purpose in

beef cattle.

Sickness among feedlot cattle continues to have a major economic consequence for commercial cattle feeding industry. A sick animal often experiences loss of appetite, depression and elevation of temperature. If not detected and treated early, a sickness can result in the death of the animal or lead to the sick animal infecting others, which ultimately results in unneeded mass medications. One of the most common sickness is the bovine respiratory disease (BRD), which can also cause detectable lung lesions that result in decreased growth performance of the cattle.

Because medical treatment is more effective the earlier it is administered in the disease, early identification of sick animals is desirable. The current method of identifying sick animals at the feedlot employs pen riders who set out to look for visual signs of morbidity, such as nasal discharge, soft cough and rapid shallow breathing. Obviously, this process is subjective and difficult to monitor.

## Goals and Objectives of Current Research

A more objective criteria than the use of pen-riders, such as the amount of feed intake, is needed to more effectively assess the presence and severity of sickness within individual animals. Unfortunately, the Growsafe® System does not provide direct measurement of the feed intake or weight gain after bunk visits.

Animal science literature however, strongly suggests that feeding and watering patterns are directly related to the status of health of the animals in the feedlot

(Putnam et al. [90]); and (Basarab et al. [6]). In particular, feeding frequency (or the number of feeding hits recorded in the duration of bunk visits) is a good indicator of an animal's feed intake while in the stall (Putnam et al. [90]; Schwartzkopf-Genswein et al. [100]). Pijpers et al. [89] found that in general, feed intake decreases with the onset of sickness, and that healthy animals tend to feed more and spend longer time at the feedlot than morbid ones. Longer time at the feedlot translates into higher number of feeding hits.

The main goal therefore, is to develop a simple, on-line statistical process control (SPC) procedure that could be linked with the automatic acquisition of feeding behavior data, to objectively monitor and detect morbid individual animals earlier than conventional methods. I refer to this as the 'broad objectives' or 'our purpose'. More specifically, the application of a cumulative sum (CUSUM) procedure will be explored.

## Outline of Related Problems

SPC procedures can also be used for modeling which will involve estimation, hypothesis testing, and prediction. In terms of hypothesis testing we may, for our present purpose, consider the null hypothesis associated with a specific animal

$H_{0i}$: Animal $i$ is healthy

against the alternative hypothesis

$H_{ai}$: Animal $i$ is sick.

One can envisage several problems and difficulties in a CUSUM analysis of this problem, such as those listed below:

1. The health of the animal, in this case, is not a variable that is directly observable or measurable by the system. It is latent.

2. Animal's *health* could lie within a continuum of values. Thus, there may not be a clear, distinct threshold or cut-off point for classifying an animal as either "healthy" or "sick". Such a threshold value, even if it exists, may be difficult to find.

3. Data on only one variable (*the number of feeding hits*) is available. A procedure that can best utilize the observations made on this single variable will be most desired.

4. We could not find a suitable transformation that would make these data approximately normally distributed. It is essential however, not to transform the count data, but to leave the data in the original units in order to ensure simplicity of use, evaluation, and interpretation of the results. Fortunately, the CUSUM procedure can be applied to count data.

5. Implementation of a SPC procedure such as the CUSUM scheme requires an explicit and precise statistical model for the observations; see for example Hawkins and Olwell ([47, page 14]). Specifying such a model may be difficult, and would

require extensive historical data. Additional problems may be encountered if the data are over-dispersed.

6. An important assumption underlying most SPC methods is that the observations are statistically independent and identically distributed. In practice however, the assumption of independence is not exactly satisfied by data collected over time.

7. Determination of the model parameters and optimal settings of the design parameters for the SPC or CUSUM procedure may be difficult.

This dissertation will be structured as follows. In Chapter 2, proposed underlying models will be discussed, beginning from the point of view of a stochastic process. Time series and Markov processes will be discussed as special types of stochastic processes. The Poisson process will be discussed briefly. The negative binomial distribution will also be discussed because it is one of the simplest and common models for overdispersed count data. The need for a modified CUSUM scheme will also be discussed in Chapter 2. In Chapter 3, the derivation of the general CUSUM for exponential family of distributions will be given. Count data CUSUMs will also be reviewed in some detail, and a summary of results on count data CUSUMs such as the computation of the ARL using the Markov chain approach of Brook and Evans [17] will be presented. The proposed modified CUSUM methods are presented in Chapter 4. In Chapter 5, numerical examples of the modified CUSUM scheme based

on the Poisson model are presented. Computational results involving summary and evaluation of the procedure, and design implementation are discussed in Chapter 6. An outline of how the procedure may be applied to the generic Feedlot data is given in Chapter 7, and finally conclusions, discussions and suggestions for future research are given in Chapter 8.

# CHAPTER 2

## PROPOSED UNDERLYING MODEL

The purpose of this chapter is to outline the basic assumptions and to suggest and discuss suitable underlying models (or class of models) for analyzing the type of generic feedlot data described in Chapter 1, keeping in view both the broad objectives and the specific goals that were also stated in Chapter 1. The unpredictable nature of future observations (for example, the number of feeding hits per 3-hr period) recorded on the animals in the generic feedlot, even when each animal is perfectly healthy, suggests a general stochastic model will be a good starting point in the analysis. After a brief general introduction, this discussion will move on to time series and Markov processes as special types of stochastic processes that could be reasonable and relevant models for this problem. The Poisson process will be discussed next because the Poisson distribution is the most commonly used model for count data. The negative binomial distribution will also be considered as one of the several generalizations of the Poisson distribution, and also as a useful, bona fide count data model in its own right. This chapter will also establish the need for a modified cumulative sum procedure, which will be presented in Chapter 4.

## Basic Assumptions

Based on the discussions in the previous chapter, the following five basic assumptions are essential throughout this dissertation. Additional assumptions will be made as and when they become necessary.

1. Prior to being received in the feedlot, all cows are checked to make sure they are 'perfectly healthy'.

2. The health status of an animal in the feedlot at any time period $t$ can be "directly assessed" by the cumulative amount of feed it has consumed up to that time.

3. The amount of feed consumed by the cow can be "estimated" by the number of feeding hits.

4. Using a "wide enough" sampling interval can greatly reduce potential serial correlation in the data, and diminish its impact on the analysis.

5. No animal in the feedlot has any influence on, or is influenced by, the feeding behavior of other animals in the feedlot.

Although assumptions 4 and 5 may not sound very reasonable, they are useful in this case for developing simple statistical ideas.

## Stochastic Process

Stochastic processes cover a very broad field and have several classes, each with extensive literature base and numerous areas of application. The literature includes texts such as Bhat [8], Taylor and Karlin [111], Ross [96], and also Krishnan [63] which contains a more technical viewpoint. The following definition of a stochastic process is given in Krishnan [63, page 39].

DEFINITION 2.1. A stochastic process $\{X_t, t \in T\}$ is a familiy (or a collection) of random variables defined on the probability space $(\Omega, \mathcal{F}, P)$ and taking values in the measurable space $(\Re, \mathcal{R})$, where:

$\Omega$ = nonempty set called the sample space;

$\mathcal{F}$ = a $\sigma$-field of subsets of $\Omega$;

$P$ = a probability measure defined on the measureable space $(\Omega, \mathcal{F})$;

$\Re$ = the real line; and

$\mathcal{R}$ = the $\sigma$-field of Borel sets on the real line.

The probability space $(\Omega, \mathcal{F}, P)$ is called the *base space* and the random variable $X_t$ is called the *state* of the process at time $t$ (for each $t \in T$). The measurable space $(\Re, \mathcal{R})$ is called the *state space* of the stochastic process and it is the set of all possible values that $X_t$ can assume (Ross [96, page 72]). The time set $T$ is called the *index set* (or *parameter space*). If $T$ is the countable set $\mathbb{Z}$, the stochastic process is often represented as $\{X_n, n \in \mathbb{Z}\}$ and is ascribed the name a *discrete-time process*. When

T is an interval of the real line ($T = \Re$), the stochastic process is called a *continuous-time process.* Thus, a stochastic process is simply a familiy of random variables that describes the evolution through time of some (physical) process (Ross [96, page 72]).

Consider a stochastic process $\{X_t, t \in T\}$ which satisfies $\mathbb{E}[|X_t|^2] < \infty$ for every $t \in T$, where the symbol $\mathbb{E}[\ .\ ]$ denotes the mathematical expectation. Assuming the components of the model are additive, a stochastic representation of the process (i.e. a stochastic model for the process) can be given as

$$X_t \ = \ \mu_t + \epsilon_t, \tag{2.1}$$

where $X_t$ is a random variable representing the state of the process at time $t$, and where $\mu_t$ is a non-random component of the process, and $\{\epsilon_t\}$ are independent identically distributed random variables with a finite variance. The process $\epsilon_t$ is commonly called the *innovation* because $\epsilon_t$ need not have a zero mean; rather $\epsilon_t$ is new or innovative at time $t$ (Joe [53, page 260]).

Analyzing the data via stochastic model (2.1) requires fitting a 'satisfactory' model, which involves parameter estimation and model checking or diagnostics. If an appropriate model can be obtained, it can be used to detect sick animals in the feedlot, and it can also be used to greatly enhance our understanding of the underlying process mechanism that generated the data. The immediate problem however is that the stochastic model given in (2.1) can involve a large number of parameters, and it can also exhibit some very complicated dependence structure such that even the mean function $\mathbb{E}[X_t]$ and the variance function $\sigma_t^2 = \mathbb{E}[X_t - \mathbb{E}(X_t)]^2$ can vary

with time. In view of this, certain "stationarity" conditions will be imposed. In this dissertation, a stochastic process $\{X_t, t \in \mathbb{Z}\}$ where $\{\mathbb{Z} = 0, \pm 1, \pm 2 \ldots\}$ will be called *stationary* (more specifically, weakly stationary or second-order stationary) if its mean and variance functions exist and are both constants (not dependent on the time parameter $t$), and if its covariance function $\gamma(t,s) = \mathbb{E}[(X_t - \mu_t)(X_s - \mu_s)]$, for all $s, t \in \mathbb{Z}$, is a function only of the time difference $|t - s|$. A stochastic process that does not satisfy these conditions is said to be *nonstationary*. Generally, depending on the nature of the data, it is possible to examine the dependence structure of model (2.1) and determine what kind of time series dependence (dependence decreasing with lag) is involved.

## Time Series Process

Time series modeling and analysis have been around for a long time, and have found applications in a variety of areas in scientific, commercial, industrial, and socio-economic fields. Some of the more accessible texts are Wei [117], and Brockwell and Davis [16] both of which discuss in-depth theory and applications. The traditional analysis of time series assumes an underlying continuous variable (usually a Gaussian random variable). The series is then decomposed into different components such as the trend component, seasonal component, and random noise component. Box and Jenkins developed an effective iterative procedure for analyzing time series models. Their method also provides techniques for handling certain nonstationary time series

models.

DEFINITION 2.2. A time series process $\{X_t, t \in T\}$ is a stochastic process where the set of observations $\{x_1, x_2, \ldots\}$ are obtained sequentially in time.

A discrete-time series can then be defined as a discrete-time stochastic process where the time index set $T$ is a discrete set, $\mathbb{Z}$ (Brockwell and Davis [16]). Each observation $x_t$ is considered to be a realized value of a certain random variable $X_t$. The sequence of observations $\{x_t, t \in \mathbb{Z}\}$ is then a realization (a single sample-path) of the family of random variables $\{X_t, t \in \mathbb{Z}\}$. A general time series model such as the one given in representation (2.1) can involve a very large (possibly infinite) number of parameters. The estimation of all those parameters may either be impossible or it can be inefficient because it is based on a single realization (a fixed number of observations). Such parameter estimates are often unstable and can consequently invalidate the identification of the model. It is essential therefore to begin with parsimonious models—models which are simple, and can adequately describe the time series with relatively few parameters.

Two common classes of time series models are the Box-Jenkins mixed autoregressive, moving average process of order $(p, q)$ (that is, ARMA($p$, $q$) process), and the autoregressive integrated, moving average process of order ($p$, $d$, $q$), (that is, ARIMA($p$, $d$, $q$)) for certain nonstationary processes. Both the ARMA($p$, $q$) and ARIMA($p$, $d$, $q$) classes of models may be used as preliminary models for count (discrete) data if the values of the observations are large enough to justify the assumption

of normally distributed innovations as a reasonable approximation. The reality however, is that most count data often include a high proportion of very low count regimes, and therefore the classical general ARIMA methods discussed in this section may not be very suitable. Indeed, West and Harrison [118, page 7] believe that *"a time series model is essentially a confession of ignorance, generally describing situations statistically without relating them to explanatory variables"* and that *"when and only when, any significant structure has been modelled should the modeller resort to time series."* Also, because the data are discrete (counts), we need data-admissible models that explicitly take account of the fact that the observations must be non-negative integers. Several techniques have been proposed for dealing with time-sequenced count data. These include the Bayesian-based dynamic modeling approach for non-normal data discussed in West and Harrison [118, chapter 14]; the ARMA-type models for discrete data and the copula-based approach discussed in Joe [53, chapter 8], and other structural-based approaches discussed in the literature.

## Markov Process

The discussion in this section will be limited to stochastic processes that exhibit first-order Markov dependence and how that may be applied to the generic feedlot data problem. Markov processes form an important class of stochastic processes where the probability of future evolution of a process when its present state is known exactly, does not change with additional knowledge of the past behavior (Krishnan

[63, page 57]). The notations and equations used here are consistent with Taylor and Karlin [111], Ross [96], and Bhat [8]. The following definition of a Markov process is provided by Taylor and Karlin.

DEFINITION 2.3. A Markov process $\{X_t\}$ is a stochastic process with the property that, given the value of $X_t$, the values of $X_s$ for $s > t$ are not influenced by the values of $X_u$ for $u < t$.

Since our main problem involves count (discrete) data, we will be particularly interested in Markov processes where both the state space and the time index set are discrete. This is often called a Markov chain, and is defined as below.

DEFINITION 2.4. A Markov chain (or a discrete-time Markov process) is a Markov process whose state space is a finite or countable set, and whose time index set is $T = (0, 1, 2, \ldots)$.

Markov Property

In view of the above definition, let $\{X_n, n = 0, 1, 2, \ldots\}$ be a Markov chain where the state space (set of all possible values of the process) is the set of nonnegative integers $\mathbb{N} = \{0, 1, 2, \ldots\}$. Let $X_n = i$ denote the statement: 'the process is in state $i$ at time epoch $n$'. Then the Markov property commonly stated in the literature has the form

$$\begin{aligned} &\Pr\{X_{n+1} = j | X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}, X_n = i\} \\ &\qquad = \Pr\{X_{n+1} = j | X_n = i\} \end{aligned} \tag{2.2}$$

for all states $i_0, i_1, \ldots, i_{n-1}, i, j$ and all $n \geq 0$. A more precise form of the Markov property is given by Krishnan [63, page 57] in the following definition.

DEFINITION 2.5. Let $\{X_t, t \in \mathbb{Z}\}$ be a stochastic process defined on a probability space $(\Omega, \mathcal{F}, P)$. Let $\mathcal{F}_t$ be the $\sigma$-field generated by $\{X_u, u \leq t\}$. Let $\mathcal{F}_t^c$ be the $\sigma$-field generated by $\{X_s, s > t\}$, and let $A \in \mathcal{F}_t$ and $B \in \mathcal{F}_t^c$. Then the process $\{X_t, t \in T\}$ is a Markov process with respect to the family $\{\ \mathcal{F}_t, t \in T\}$ if

$$\Pr(A \cap B \,|\mathcal{F}_t) \ = \ \Pr(A \,|\mathcal{F}_t) \Pr(B \,|\mathcal{F}_t). \tag{2.3}$$

The Markov property can again be restated in terms of the conditional distribution by the following equivalent statements:

1. the conditional distribution of any future state $X_{n+1}$ given the past states $X_0, X_1, \ldots, X_{n-1}$ and the present state $X_n$, is independent of the past states and depends only on the present state (Ross [96, page 132]).

2. The future, conditioned on the past history up to the present, is equal to the future, given the present (Krishnan [63, page 60]).

Transition Probability

The movement of the process from one state to another is referred to as a transition. The conditional probability $P_{ij}^{n,n+1}$ of $X_{n+1}$ being in state $j$ given that $X_n$ is in state $i$ is called the *one-step transition probability*, and is given by

$$P_{ij}^{n,n+1} \ = \ \Pr\{X_{n+1} = j | X_n = i\} \tag{2.4}$$

where $P_{ij}^{n,n+1}$ is a function of the pre-transition state $i$, the post-transition state $j$, and the time variable $n$. If the transition probabilities are time-homogeneous (or independent of the time variable $n$, but depends only on the time difference), the Markov chain is said to have stationary transition probabilities. Thus, the one-step transition probabilities $P_{ij}$ (where the time difference is one) are given by

$$P_{ij}^{n,n+1} = P_{ij} \tag{2.5}$$

for $P_{ij} \geq 0, \quad i,j \geq 0; \quad \sum_{j=0}^{\infty} P_{ij} = 1, \quad i = 0,1,2,\ldots.$

Similarly, the $n$-step transition probabilities $P_{ij}^{n}$ defined as the probability that a process in state $i$ will be in state $j$ after $n$ additional transitions, is given by

$$P_{ij}^{(n)} = \Pr\{X_{m+n} = j | X_m = i\} \quad m = 0,1,\ldots. \tag{2.6}$$

Let $\mathbf{P}$ denote a matrix of the one-step stationary transition probabilities $P_{ij}$, where

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ \vdots \\ i \\ \vdots \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & \ldots \end{array} \\ \left\| \begin{array}{cccc} P_{00} & P_{01} & P_{02} & \ldots \\ P_{10} & P_{11} & P_{12} & \ldots \\ \vdots & \vdots & \vdots & \vdots \\ P_{i0} & P_{i1} & P_{i2} & \ldots \\ \vdots & \vdots & \vdots & \vdots \end{array} \right\| \end{array}. \tag{2.7}$$

For all future references to a Markov chain it is assumed that the transition probabilities are stationary. The analysis of a Markov chain requires the computation of the transition probabilities of the possible realizations of the process. For most practical applications the Markov chains that are encountered are finite Markov chains,

and therefore the transition probability matrix (2.7) will usually be a finite square matrix. Excellent details on Markov processes can be found in Ross [96], Taylor and Karlin [111], Bhat [8], Kemeny and Snell [61], and Shiryayev [102]. The following points are, however, worthy of mention because they will be useful in the next two chapters.

1. A Markov process is completely defined once its transition probability matrix $\mathbf{P}$ and the probability distribution of the initial state $X_0$ are specified.

2. State $j$ is said to be *accessible* from state $i$ if $j$ can be reached from $i$ in a finite number of steps. This is denoted by writing $(i \rightarrow j)$.

3. If the two states $i$ and $j$ are accessible to each other, then they are said to *communicate*, and this is denoted by $(i \leftrightarrow j)$.

4. A state $i$ is said to be *recurrent* if and only if, starting from state $i$, eventual return to this state is certain.

5. A state $i$ is said to be *transient* if and only if, starting from state $i$, there is a positive probability that the process may not eventually return to this state.

6. A state $i$ is said to be an *absorbing* if and only if the transition probability $P_{ii} = 1$. Thus, once entered, it is never left.

7. The analysis of a Markov chain requires the computation of the transition probabilities of the possible realizations of the process.

Lumped and expanded Markov chains

The following concepts of lumped and expanded processes discussed in Kemeny and Snell [61, chapter 6] and Bhat [8, chapter 5] will be useful in Chapter 4 of this dissertation.

*Lumpable Markov chains* are Markov chains in which the states can be lumped (or pooled or combined) together without losing the basic Markovian property. This process, also called the *lumped process*, can be used to reduce an elaborate model with a very large number of states to a process with a smaller number of states by lumping certain states together when the identification of some states are not necessary any further. A formal definition of lumpability can be stated as follows.

DEFINITION 2.6. A Markov chain $\{X_n, n = 0, 1, 2, \ldots\}$ is lumpable with respect to a partition $\mathbf{G} = \{\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_m\}$ of the state space if for every initial state probability vector $\mathbf{v} : \{v_1, v_2, \ldots v_l\}$ the resulting lumped process is a Markov chain and the transition probabilities do not depend on the choice of $\mathbf{v}$.

THEOREM 2.7. *A necessary and sufficient condition for a Markov chain to be lumpable with respect to a partition* $\mathbf{G} = \{\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_m\}$ *is that for every pair of sets* $\mathbf{G}_i$ *and* $\mathbf{G}_j$, *the elements of the one-step transition probability matrix of the lumped chain* $\tilde{p}_{ij}$ *have the same value for every state* $k$ *in* $\mathbf{G}_i$. *That is,*

$$\tilde{p}_{ij} = \sum_{r \in \mathbf{G}_j} p_{kr} \qquad i, j = 1, 2, \ldots, l \quad \text{for all } k \in \mathbf{G}_i.$$

For proof of the lumpability theorem, see Kemeny and Snell [61, chapter 6].

DEFINITION 2.8. Expanded Markov chains (also called the *expanded process*) are

Markov chains that are obtained by enlarging the number of states in the original chain to give more detailed information about the process being studied.

Markov Model for a Generic Feedlot Data

In addition to its simplicity, a Markov model seems appropriate as an underlying model for the generic feedlot data problem for the following reason. Recalling that the observations are the numbers of feeding hits aggregated over 3-hr periods, it can be argued that for each animal, this 3-hr interval seems wide enough in time, to greatly reduce the effects of serial correlation in the data. If the 3-hr sampling interval is inadequate in diminishing the effects of autocorrelation in the data, we will assume that a "reasonably wide sampling interval" can be found to reduce the effects of the autocorrelation in successive observations. Thus, we will make the assumption that the 3-hr aggregated observations behave in a way that can be reasonably considered independent.

To illustrate how a Markov model might be used for the feedlot data problem, consider the following simple example. Let:

$\eta$ = maximum possible number of feeding hits in any given time period. (For example in a 3-hr interval, if individual feeding hits are recorded every 5.25 sec by the Growsafe® System, then $\eta = (3 \times 60 \times 60)/5.25 \approx 2057$).

$\xi_n$ = number of feeding hits recorded for a cow during time period $(n, n+1]$, where $0 \leq \xi_n \leq \eta$, $n = 0, 1, 2, \ldots$.

$C_n$ = cumulative feed intake $= \sum_{i=1}^{n} \xi_i = C_{n-1} + \xi_n$.

$V_n$ = state of the feed intake process at time $n$ for an animal in the feedlot. (From the basic assumptions, $V_n$ is a good indicator of the health status of the animal at time $n$, and it is a process).

Assume $\xi_0, \xi_1, \xi_2, \ldots$ are random variables and let the states of the process $\{V_n, n = 0, 1, 2, \ldots\}$ be given by the following cumulative feeding class intervals: state 0: *0*, state 1: *1–50* feeding hits, state 2: *51–100* feeding hits, ..., state 9: *401–450* feeding hits, state 10: *451* or more feeding hits. Then, the health status of the animal at time $n+1$ is a function of $C_n$ and $\xi_{n+1}$, where

$$V_{n+1} = \begin{cases} 0 & C_n + \xi_{n+1} = 0 \\ 1 & 1 \leq C_n + \xi_{n+1} \leq 50 \\ 2 & 51 \leq C_n + \xi_{n+1} \leq 100 \\ \vdots & \vdots \\ 9 & 401 \leq C_n + \xi_{n+1} \leq 450 \\ 10 & C_n + \xi_{n+1} \geq 451. \end{cases}$$

The process $\{V_n, n = 0, 1, 2, \ldots\}$ is therefore a Markov chain with the restricted state space $\{0, 1, 2, \ldots, 10\}$, and the transition probabilities are given by

$$P_{ij} = \Pr\{V_{n+1} = j | V_n = i\}, \quad i, j = 0, 1, 2, \ldots, 10.$$

The transition probability matrix $\mathbf{P} = \|P_{ij}\|$ can be computed when the distribution of the initial state $V_0$ is specified. In latter chapters which deal with CUSUM procedures,

the state corresponding to very low cumulative feeding hits will be designated as an absorbing state. Based on our model, very low cumulative feeding hits would naturally suggest that the animal is "sick".

## Poisson Process

### Poisson Distribution

DEFINITION 2.9. A discrete random variable $X$ is said to have a Poisson distribution if the probability mass function (pmf) of $X$ is given by

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad x = 0, 1, 2, \ldots, \tag{2.8}$$

where $x$ is the number of occurrences of an event, and $\lambda > 0$ is the parameter of the distribution. The mean $\mu$, and variance $\sigma^2$ of the Poisson distribution are given by $\mu = \lambda$ and $\sigma^2 = \lambda$.

In general, the Poisson distribution is skewed to the right, but the distribution approaches symmetry in appearance as $\lambda$ becomes larger. Montgomery [77], and Hawkins and Olwell [47], among others, discuss several applications of the Poisson distribution in statistical quality control. Many other applications of the Poisson distribution can be found in the literature on stochastic processes. Since the Poisson random variable assumes only nonnegative integers, it provides a realistic model for many random phenomena that involve the count of occurrences (or events) per unit time, length, region, space, area, or volume. Examples include the number of flaws per unit length of fabric, the number of defects per unit area of some material, the

number of traffic accidents (or industrial accidents) per week in a given location, the demands for telephone connection per hour, the number of emissions of radioactive material per unit time, and the number of organisms per unit volume of some fluid. One could derive the Poisson distribution either as a limiting form of the binomial distribution or by pure probability arguments (see, for example, Mood, Graybill, and Boes [80, Section 2.4]).

Assumptions of the Poisson Process

The Poisson distribution usually occurs as the outcome of some random process called the Poisson process. Following Mood, Graybill, and Boes [80] and Leon-Garcia [64], suppose that we are observing the occurrence of certain events such as the number of feeding hits (that is, the recording of the presence of an animal at the feedbunk) during a time period of length $T$. Assume $\nu$ ($\nu > 0$) is the mean rate at which those events occur per unit time; that is, $\nu$ is the *mean rate of feeding hits.* Assume that $\nu$ is the same for all healthy animals and it is constant for each day and for each time period. Suppose further that the interval $(0, T]$ is divided into $n$ non-overlapping subintervals, each of very short duration $\delta = \frac{T}{n}$, and assume the following conditions hold:

1. The probability of observing an outcome in a small time interval of length $\delta$ is approximately equal to $\nu\delta$. For the generic feedlot data, an outcome corresponds to a feeding hit in an interval of length $\delta = 5.25$ seconds.

2. The probability of occurrence of more than one outcome (eg. a feeding hit) in a small time interval of length $\delta$ is negligible compared to the probability of observing one or zero in the same time interval.

3. Whether or not an event occurs in a non-overlapping subinterval is independent of the outcomes in other non-overlapping subintervals.

The second assumption implies that we can view the outcome in each subinterval as a Bernoulli trial (Leon-Garcia [64]); that is either the subinterval has a hit, or it does not. Under these conditions then, the probability of observing exactly $x$ hits in a total length $T$ is Poisson with mean parameter $\lambda = \nu T$. The probability function of $X$ is then given by

$$p(x) = \frac{(\nu T)^x e^{-\nu T}}{x!} \quad x = 0, 1, 2, \ldots. \tag{2.9}$$

Notationally, $X \sim P(\nu T)$. That is, the distribution of $X$ is Poisson with parameter $\nu T$.

Negative Binomial Distribution

The negative binomial distribution can arise in many ways, and it is one of the most useful and widely-used count data models. The rigid conditions of the Poisson process cannot be exactly satisfied in most practical situations. For example, the probability of occurrence of feeding hits may not be constant. It may vary with time periods, or days, or animals. Montgomery [77], Lucas [68], Hawkins and Olwell [47]

among others have suggested that the Poisson model would usually work reasonably well as long as these departures from the assumptions of the Poisson process are not too severe. There are certain situations, however, in which the Poisson model for count data is completely inadequate and cannot be realistically modeled with a Poisson distribution. The generic feedlot data is a clear example. In such situations, a more reasonable approximation could be achieved using a discrete distribution other than the Poisson. An example is the case where the Poisson process has a parameter $\lambda$ that is not constant but is also a random variable, suggesting a mixture of distributions (or a contagious distribution). We can also have a Poisson process in which the occurrence of events in a fixed time interval takes the form of clusters. While cautioning that there is no single technique which will cover these mixture situations, Jackson [52] suggests that *"if no knowledge about the underlying mechanism is available, a mixture of Poissons or Poisson and various standard contagious distributions would be a reasonable thing to try."* For an extensive discussion of the various classes of discrete distributions and their mixtures and generalizations, including the *truncated* and *inflated* versions, the book by Johnson, Kotz and Kemp [58] is an excellent reference.

In the feedlot study for example, the recorded number of feeding hits in the time period of length $T$ could occur in *clusters*. It is conceivable, for instance, that an animal at the feedbunk is likely to remain there and feed continuously for some time. It may pause for a while, and then resume feeding again, resulting in several

clusters of recorded feeding hits. Thus, there could be at least two processes going in the feedlot: one giving rise to the number of clusters, and the other generating the number of feeding hits within each cluster. If the among-cluster distribution of $\lambda$ is gamma and the distribution within clusters given is $\lambda$ Poisson, then the result is the negative binomial. The resulting negative binomial distribution is said to have been derived as a *gamma mixture of Poissons* (see for example, Mood, Graybill and Boes [80, page 123] for a derivation).

DEFINITION 2.10. A discrete variable $X$ is said to have the negative binomial distribution if the probability mass function of $X$ is given by

$$p(x) = \frac{\Gamma(r+x)}{x!\Gamma(r)}\left(\frac{\beta}{\beta+1}\right)^r\left(\frac{1}{\beta+1}\right)^x, \quad x = 0,1,2,\ldots,\ \beta > 0. \tag{2.10}$$

The parameters are $r$ and $p = \beta/(\beta+1)$. In this form (2.10) of the negative binomial, $X$ is defined as the number of failures before $r$ successes of a Bernoulli process. Hawkins and Olwell [47] believe the parameterization (2.10) is more useful for modeling because it assumes all non-negative integer values, and it supports asymptotic convergence to the Poisson distribution. The mean and variance of (2.10) are

$$\mu = \frac{r}{\beta} \tag{2.11}$$

and

$$\sigma^2 = \frac{r}{\beta}\left(1+\frac{1}{\beta}\right) = \frac{r(\beta+1)}{\beta^2}. \tag{2.12}$$

The parameter $\beta$ determines over-dispersion. Thus, the negative binomial distribution (2.10) assumes the same set of non-negative integers as the Poisson, but it also

has an extra parameter, $\beta$, to model over-dispersion (Hawkins and Olwell [47]). If a Poisson process has parameter $\lambda$ that is not constant but varies randomly, then the parameter $\beta$ measures the variability in $\lambda$. A small or near zero value of $\beta$ corresponds to a highly variable $\lambda$, and therefore a large over-dispersion, leading to a negative binomial distribution for $X$ that is very different from the Poisson. Large values of $\beta$, on the other hand, mean there is very little over-dispersion and therefore a nearly constant $\lambda$, leading to a near-Poisson behavior in the negative binomial distribution. Thus, the negative binomial is to be preferred if over-dispersion is present. In the feedlot study, the Poisson parameter $\lambda$ is obviously not constant. At least it varies over the six 3-hour time periods during which the data were collected daily. This is likely to result in large over-dispersion. Jackson [52] observed that

> The negative binomial is related, at least asymptotically, to a number of other discrete distributions and, hence, often serves as a good approximation in situations where the underlying mechanism is unknown.

The negative binomial has also been suggested as a robust alternative to the Poisson (Gelman, Carlin, Stern, and Rubin [36, page 483]).

## The Need For a Modified CUSUM

It should be emphasized here that our goal is *not* to "fit the 'correct' model to the data". We seek a simple, easily-implemented CUSUM scheme for these data that could be used as an online process to help in the early detection of sick cows. In doing so, we want to base the CUSUM on a useful and interpretable model that can

reasonably describe the underlying process that generates the data.

CUSUM schemes are preferable here because they have several interesting features and advantages. They can be very effective in detecting small persistent shifts in the process mean. They are also very useful for continually attaining and maintaining control of quality during production, and they can be applied quickly and inexpensively to any measurable variable.

CUSUM methods can be readily adapted and modified to achieve various improvements in performance. They are also defined for both discrete and continuous standard distributional types used in charting, (Crowder, Hawkins, Reynolds and Yashchin [24]). In fact, CUSUM schemes have been implemented with count data models such as the Binomial, Poisson, and the Negative Binomial distributions. In addition, CUSUM is based on the likelihood ratio strategy (Johnson and Leone [55, 56, 57], Hawkins and Olwell [47]) and so it has several optimal properties (e.g. Moustakides [81]). Hawkins and Olwell [47, page 112] describes the effectiveness of the CUSUM procedures for count data distributions in the following words: *"for discrete data, the CUSUM remains the optimal test for detecting a persistent shift in the process parameters."* The CUSUM method therefore seems appropriate for our purpose which involves count data.

Although several adaptations of the CUSUM and its modifications exist, none of them seems to adequately meet the unique needs of our present problem. Quimby et al. [91] was the first to consider the application of CUSUMs to the cattle feedlot

problem. Their approach used the classical, two-sided standardized normal CUSUM to predict cattle morbidity in the feedlot, and then they compared their results with those of the pen riders, using a paired t-test. The main problem with this procedure is that the standardized normal CUSUM assumes that the observations from the process (when it is in-control) are independent and normally distributed. However, the nature of the data clearly suggests that non-normality is fundamentally intrinsic to the problem. Furthermore, there is a considerably high proportion of small counts in the data, making a normal approximation inapplicable. Under these conditions, using a count data CUSUM would be more appropriate. In addition, the observations from this process appear to be overdispersed, and any analysis should try to take this into consideration. This may call for the use of a CUSUM based on an overdispersed count data model, or alternatively, as is considered in this research, a new procedure that would not be very sensitive to the presence of individual isolated outliers.

The assessment of the presence and cause of potential autocorrelation is also very important, since (relatively strong) autocorrelation can distort the results of the analysis, and the cause of autocorrelation will also dictate what further action may be required. For this problem, the original raw data, made up of a very long sequence of feeding *hits* (1's) and *misses* (0's) occur in clusters and are inherently autocorrelated (due to the natural process dynamics). However, for the purposes of statistical analysis, an *observation* was recorded as the number of feeding hits over a 3-hr period. The 3-hr aggregation interval was chosen after preliminary empirical results

(based on historical data from an earlier feedlot study) indicated it was sufficiently spaced in time, to dilute the effect of the serial correlation while retaining some of the details of the daily feeding patterns in the feedlot. This can have the effect of making the data approximately uncorrelated (see for example, Runger and Willemain [97]; Faltin, Mastrangelo, Runger and Ryan [28]).

Therefore, based on the assumption of independence of the observations, we consider an original research (a modified CUSUM procedure) that modifies the traditional count data CUSUM control scheme so it can be used in conjunction with on-line data acquisition systems, and which can incorporate some of the following enhancements in CUSUM performance:

1. retain the CUSUM's known (well-acknowledged) sensitivity to small shifts;

2. be effective in detecting larger shifts more quickly than the traditional CUSUM scheme;

3. be relatively more robust (less sensitive) to the presence of isolated individual outliers than the traditional CUSUM schemes;

4. be able to respond more quickly to genuine out-of-control situations than the traditional CUSUM scheme.

My literature search shows that some of the existing work that come close to the goals outlined above include the combined Shewhart-CUSUM schemes (Lucas [67]) which allows the combined scheme to rapidly detect both small and large shifts.

The fast initial response (FIR CUSUM) by Lucas and Crosier [71] is another useful enhancement to the CUSUM procedure. This procedure, in which the CUSUM is reset to an initial nonzero value (usually taken as $S_0 = H/2$, where $H$ is the decision interval), allows the CUSUM to respond rapidly to out-of-control situations especially at start-up or after the CUSUM has given an out-of-control signal. The final paper that comes close to the goals listed above is the two-in-a-row robust CUSUM of Lucas and Crosier (1982b) in which a single outlier is ignored, but two consecutive outliers automatically trigger a signal. These papers however differ entirely in methodology and procedure from the methods proposed in this dissertation.

# CHAPTER 3

## DESIGN OF TRADITIONAL CUSUM PROCEDURES

This section presents a summary of the results on the count data CUSUM procedure (Brook and Evans [17]; Lucas [68]; Hawkins and Olwell [47]), and the derivation of the general tabular CUSUM scheme for distributions in the exponential family (Hawkins and Olwell [47, chapter 6]). Count data CUSUM schemes are similar in many ways to the continuous variables CUSUM schemes. They can be used in low, moderate, and in some cases, high count regimes to control counts involving rare events, the number non-conforming or the number conforming to certain specifications of a product. Only one-sided count data CUSUMs of the decision interval (tabular) form of scheme will be discussed here. Computation of the run length distribution and its properties will be based on the Markov chain approach of Brook and Evans [17].

### Notation and Conventions

Notations and conventions used in this section will be outlined first to achieve consistency with Brook and Evans [17]; Hawkins and Olwell [47], and with notational forms used in this dissertation. First, unless otherwise stated, the term upper (or upward) CUSUM refers to the upper one-sided tabular count data CUSUM. Similarly, the term lower (or downward) CUSUM will be used for the lower one-sided tabular

count data CUSUM. Second, the term traditional or regular count data CUSUM will refer to the standard cumulative sum procedure for discrete (or count) distributions, without any enhancements. Regular CUSUMs generally start from zero. Other items are listed below.

1. Random variables are denoted by italic upper-case Latin letters, commonly $X$, $Y$, $Z$, and $N$. These random variables will typically contain indexing subscripts.

2. The symbol $\mathbb{E}[.]$ denotes mathematical expectation, eg. $\mathbb{E}[N_i]$.

3. The CUSUM statistic at sample $n$ is denoted by $S_n$, usually with (+) or (-) superscripts corresponding to upper ($S_n^+$) or lower ($S_n^-$) CUSUM. In specific situations, $C_n$, $T_n$, and $D_n$ may also be used.

4. The maximum and minimum of $a$ and $b$ are denoted by $\max[a, b]$ and $\min[a, b]$, respectively.

5. The natural logarithm is denoted by ln, example $\ln(\lambda)$.

6. $E_i$ now denotes CUSUM state $i$. A transition from state $i$ to state $j$ is denoted by $E_i \rightarrow E_j$ (indicating state $j$ is accessible from state $i$).

7. Pr is the abbreviation for probability. For example, the probability of transition from state $E_i$ to state $E_j$ is denoted by $P_{ij} = \Pr(E_i \rightarrow E_j)$. The cumulative distribution function (cdf) is denoted by $F$ (usually with a subscript). For

example, $F_r = \Pr(X_n - K \leq r)$. Similarly, the probability mass function (pmf) is denoted by $p$ (usually with a subscript). For example, $p_r = \Pr(X_n - K \leq r)$.

8. A vector of ones is denoted by $\mathbf{1}$. The length will usually be specified.

9. Bold upper-case Latin letters are used for matrices. $\mathbf{I}$ is used for the identity matrix, $\mathbf{P}$ for the one-step transition probability matrix, and $\mathbf{R}$ for the matrix obtained from $\mathbf{P}$ by deleting the last row and column. Matrix (or vector) transpose is denoted by the "prime" symbol ($'$). For example, $\mathbf{0}'$ denotes the zero vector.

10. Bold lower-case Greek and bold italic upper-case Latin letters, are used for vectors. For example, $\boldsymbol{\mu}$ is a vector of ARLs, and $\boldsymbol{N}$ is a vector of run lengths.

## Design of Count Data CUSUMS

Let $\{X_n, n = 1, 2, \ldots\}$ be independent random variables representing the observed number of counts in sampling interval $n$. Then a regular upper CUSUM for detecting an increase in counts is defined by

$$S_0^+ = 0 \tag{3.1}$$

$$S_n^+ = \max(0, X_n - K + S_{n-1}^+) \qquad n > 0 \tag{3.2}$$

where $S_n^+$ is the upper CUSUM statistic at sample $n$, and $K$ is some reference value below which the CUSUM is not designed to react. The process starts from zero and so long as $X_n \leq K$, the process is considered to be in statistical control. If

$X_n > K$, the process is not considered to be in statistical control, and the CUSUM begins to accumulate the difference. If $S_n^+$ reverts to zero, the process is once again considered to be in control. If $S_n^+ \geq H$, where $H$ is known as the decision interval, the CUSUM will signal an alarm (or out-of-control). When this happens, the process will be stopped until an appropriate action is taken.

A regular lower (downward) CUSUM for detecting a decrease in counts is defined similarly (see for example, Hawkins and Olwell [47, chapter 5]) by

$$S_0^- = 0 \tag{3.3}$$

$$S_n^- = \min(0,\ X_n - K + S_{n-1}^-) \tag{3.4}$$

which can be turned into the upper CUSUM by writing

$$T_0^+ = 0 \tag{3.5}$$

$$T_n^+ = \max(0,\ K - X_n + T_{n-1}^+). \tag{3.6}$$

To design a CUSUM, one needs to choose suitable reference value ($K$) and decision interval ($H$) to so that the scheme will have desired ARL properties. Typically, the parameter $K$ is determined first, based on the in-control mean level $\mu_0$ and the out-of-control mean level $\mu_1$ that is to be detected quickly. After $K$ is determined, a value for the parameter $H$ is selected to give some desired in-control ARL. Hawkins and Olwell [47, page 108] explained that due to the discrete nature of the observations, if $K$ (and the initial CUSUM value $S_0^+$) is chosen to be an integer, then the CUSUM statistic (say, $S_n^+$, for all $n$), can only assume one of the limited set of integer values

$0, 1, 2, \ldots, H$. This restriction on the set of possible values of $S_n^+$ results in a similar limited set of possible ARLs, causing "graininess" in the ARL function. If either $K$ or the initial CUSUM value $S_0^+$ is a quarter-integer, then $S_n^+$ can take on integer, half-integer as well as quarter-integer values. This will also expand the attainable ARL values. The down side however, is that using fractional values can dramatically increase the number of states in the resulting Markov chain model (Hawkins and Olwell [47, page 108]).

Determining $K$ and $H$

To determine the reference value $K$, it is good to first choose the shift for which quickest detection (or maximum sensitivity) is desired. The parameter $K$ is then chosen to obtain the optimal response to a shift from the acceptable (in-control) mean level to the unacceptable (out-of-control) mean level. For example, if the counts $X_n$ follow a Poisson distribution (where $\mu = \lambda$), and we want to monitor an increase in the mean (i.e. $\lambda_1 > \lambda_0$ for a specific $\lambda_1$), then the reference value $K$ lies between $\lambda_0$ and $\lambda_1$, and it is given by

$$K = \frac{\lambda_1 - \lambda_0}{\ln(\lambda_1) - \ln(\lambda_0)}, \tag{3.7}$$

(see, for example, Lucas [68]). Usually $K$ is rounded to an integer value. However, as explained earlier, $K$ can be rounded to a half-integer or a quarter-integer value to gain some flexibility in the attainable ARLs. The derivation of (3.7) will be given later in this chapter. The value of $H$ is selected to give an appropriately large ARL

when the process is in control. The parameter $H$ should also be selected to give an appropriately small ARL when the process is running at an unacceptable mean level. This can be done using software or published nomograms or tables.

## Formulation For General CUSUMS

Following Hawkins and Olwell [47], let $Y$ be a random variable with a distribution which is a member of the exponential family of distributions with a single parameter $\theta$. Then the probability density (or the probability mass function) $f(y|\theta)$ can be written as

$$f(y \mid \theta) = \exp\{a(y)b(\theta) + c(y) + d(\theta)\} \tag{3.8}$$

where $\theta$ is the parameter of the distribution; $a(y)$ gives the minimum sufficient statistic; $c(y)$ is a function not involving $\theta$; and $b(\theta)$ and $d(\theta)$ are functions only of $\theta$, where $b(\theta)$ is called the "natural parameter".

The likelihood ratio (LR) $\Lambda_n$ for a sequential test of the simple hypotheses

$$H_0 : \theta = \theta_0 \qquad \text{vs.} \qquad H_1 : \theta = \theta_1$$

to determine whether the process has shifted from an acceptable (in-control) parameter level $\theta_0$ to an unacceptable (out-of-control) level $\theta_1$, based on independent observations $\{Y_i\}$, is given by

$$\Lambda_n = \prod_{i=1}^{n} \frac{f_1(Y_i)}{f_0(Y_i)}, \qquad n = 1, 2, \ldots. \tag{3.9}$$

The logarithms of the likelihood ratio is given by

$$\ln \Lambda_n = \sum_{i=1}^{n} \ln \left( \frac{f_1(Y_i)}{f_0(Y_i)} \right) = \sum_{i=1}^{n} Z_i \tag{3.10}$$

where $f_j(Y_i)$ is the probability density or mass function of $\{Y_i\}$ associated with hypothesis $j$ ($j = 0, 1$). The variable $Z_i$ can be interpreted as the contribution of the $i$th observation to the evidence against the assumption of in-control process (Yashchin [132]). For the exponential family (3.8), we have

$$\begin{aligned} Z_i &= \ln \left( \frac{f_1(Y_i)}{f_0(Y_i)} \right) \\ &= a(Y_i) \left[ b(\theta_1) - b(\theta_0) \right] + \left[ d(\theta_1) - d(\theta_0) \right]. \end{aligned} \tag{3.11}$$

A generic one-sided tabular CUSUM can be defined by the recursion

$$\begin{aligned} D_n &= \max \left[ 0,\ D_{n-1} + Z_n \right] \\ &= \max \left[ 0,\ D_{n-1} + a(Y_i) \left[ b(\theta_1) - b(\theta_0) \right] + d(\theta_1) - d(\theta_0) \right] \end{aligned} \tag{3.12}$$

which accumulates when LR $> 1$, and signals when $D_n \geq U$, where $U$ is some positive value. If $b(\theta_1) - b(\theta_0) > 0$, a general one-sided upper CUSUM $C_n^+$ is obtained by dividing through the generic CUSUM in (3.12) by $b(\theta_1) - b(\theta_0)$. This gives

$$\begin{aligned} C_n^+ &= \max \left[ 0,\ C_{n-1}^+ + a(Y_n) + \frac{d(\theta_1) - d(\theta_0)}{b(\theta_1) - b(\theta_0)} \right] \\ &= \max \left[ 0,\ C_{n-1}^+ + X_n - K \right] \end{aligned} \tag{3.13}$$

which signals if $C_n^+ \geq H$, where

$$\begin{aligned} C_n^+ &= D_n \,/\, (b(\theta_1) - b(\theta_0)) \\ X_n &= a(Y_n) \qquad \text{(gives the sufficient statistic)} \\ H &= U \,/\, (b(\theta_1) - b(\theta_0)) \qquad \text{(the decision interval)} \\ K &= -\frac{d(\theta_1) - d(\theta_0)}{b(\theta_1) - b(\theta_0)} \qquad \text{(the reference value).} \end{aligned} \tag{3.14}$$

Similarly, if $b(\theta_1) - b(\theta_0) < 0$, the general formulation for the one-sided lower tabular CUSUM $C_n^-$ is obtained by dividing through the generic CUSUM (3.12) by the negative $(b(\theta_1) - b(\theta_0))$ to give

$$C_n^- = \min\left[0,\ C_{n-1}^- + X_n - K\right], \tag{3.15}$$

which signals when $C_n^- \leq -H$, where $C_n^- = D_n/(b(\theta_1) - b(\theta_0))$, and $X_n$, $H$ and $K$ are as previously defined above. Thus the count data CUSUMS (3.1), (3.3) are defined in exactly the same way as the general CUSUM (3.13),(3.15).

EXAMPLE 3.1. *Derivation of Poisson CUSUM*

The Poisson distribution, defined by equation (2.8) in Chapter 2, belongs to the exponential family, and its pmf can be written as

$$f(y \,|\, \lambda) \;=\; \exp\{y \,\ln\lambda \,-\, \lambda \,-\, \ln(y\,!)\} \tag{3.16}$$

giving the components

$$\begin{aligned} a(y) &= y \\ b(\lambda) &= \ln \lambda \\ d(\lambda) &= -\lambda, \quad \text{and therefore} \\ K &= -\frac{d(\lambda_1) - d(\lambda_0)}{b(\lambda_1) - b(\lambda_0)} = \frac{\lambda_1 - \lambda_0}{\ln \lambda_1 - \ln \lambda_0}. \end{aligned} \tag{3.17}$$

Thus, the optimal Poisson CUSUM for detecting an increase from acceptable mean level $\lambda_0$ to an unacceptable level $\lambda_1$ ($\lambda_1 > \lambda_0$), is given by the upper CUSUM

$$\begin{aligned} C_0^+ &= 0 \\ C_n^+ &= \max\left[0,\ C_{n-1}^+ + X_n - K\right] \end{aligned} \tag{3.18}$$

which signals when $C_n^+ \geq H$, with $K$ as given in (3.17), The lower CUSUM for detecting a decrease in $\lambda$ can similarly be defined.

EXAMPLE 3.2. *CUSUM For The Negative Binomial Distribution*

The negative binomial pmf is also defined in (2.10) of Chapter 2. Considering $r$ as fixed, the pmf can be written as a member of the exponential family with

$$f(y \,|\, r, \beta) = \exp\left\{-y \ln(1+\beta) + \ln\left[\frac{\Gamma(r+y)}{y!\,\Gamma(r)}\right] + r \ln\left(\frac{\beta}{1+\beta}\right)\right\} \tag{3.19}$$

where

$$\begin{aligned} a(y) &= y \\ b(\beta) &= -\ln(1+\beta) \\ d(\beta) &= r\ln\left(\frac{\beta}{1+\beta}\right) \\ K &= -\frac{d(\beta_1)-d(\beta_0)}{b(\beta_1)-b(\beta_0)} = \frac{r\ln\left[\frac{\beta_0(1+\beta_1)}{\beta_1(1+\beta_0)}\right]}{\ln\left(\frac{1+\beta_0}{1+\beta_1}\right)}. \end{aligned} \tag{3.20}$$

A CUSUM can therefore be designed for monitoring changes in the parameter $\beta$ (which will reflect both changes in the mean of $Y$ and in the degree of over-dispersion). As $\beta$ increases the degree of over-dispersion decreases, and the negative binomial distribution becomes more Poisson-like. Because the mean of $Y$ is inversely related to $\beta$ ($\mu = r/\beta$), monitoring for an increase in the mean of $Y$ corresponds to detecting a downward shift in $\beta$, which can be accomplished with the upper CUSUM

$$\begin{aligned} C_0^+ &= 0 \\ C_n^+ &= \max\left[0,\ C_{n-1}^+ + X_n - K\right] \end{aligned} \tag{3.21}$$

which signals when $C_n^+ > H$, and where $K$ is obtained from (3.20). A lower CUSUM for detecting a decrease in the mean of $Y$ (and consequently monitoring for upward shift in $\beta$) can similarly be derived.

Markov Chain Approach

Brook and Evans [17] first proposed the Markov chain method to determine the exact and approximate run length distribution of the CUSUM, its moments, and its

percentage points. This was extended to the two-sided CUSUM scheme by Lucas and Crosier [71] and Woodall [122]. Further details on the procedure were provided by Lucas and Crosier [72] and Lucas and Saccucci [73]. Hawkins and Olwell [44] devised a procedure that combines the Markov chain approach with integral equations to obtain more accurate approximations of the run length distribution and its properties. The Markov chain procedure can give exact solution for discrete distribution CUSUMS such as the Poisson CUSUM with a rational $K$ value. We will consider only upper CUSUMs in the following discussion, since the lower one-sided CUSUM can also be turned into the upper CUSUM as in (3.5).

Assuming $S_0^+ = 0$, the state space of the count data CUSUM $\{S_n^+\}$ is divided into $H+1$ intervals or discrete points, given by the integer values $0, 1, 2, \ldots, H$. These values correspond to states $E_0, E_1, E_2, \ldots, E_H$ respectively, where $E_H$ is an absorbing state (corresponding to the out-of-control condition $S_n^+ \geq H$), and the rest are transient states. If $S_n^+ = i$, then the CUSUM said to be in state $E_i$. Because of the recursive definition of $S_n^+$ and the independence of sampled observations $X$ (see for example, Hawkins and Olwell [47, page 152]), the CUSUM has a random walk structure with the Markov property:

$$\begin{aligned} &\Pr\{S_n^+ = j | S_0^+ = i_0, S_1^+ = i_1, \ldots, S_{n-1}^+ = i\} \\ &\qquad = \Pr\{S_n^+ = j | S_{n-1}^+ = i\}. \end{aligned} \tag{3.22}$$

Thus, the possible values of the CUSUM statistic $\{S_n^+, n = 0, 1, 2, \ldots\}$ form a Markov chain.

Run Length Distribution and Average Run Length

The one-step transition probabilities from state $E_i$ $(i = 0, 1, 2, \ldots)$ to state $E_j$ are determined only by the probability distribution of $X_n$. Let $F_r = \Pr(X_n - K \leq r)$ and $p_r = \Pr(X_n - K = r)$ be the cumulative distribution function (cdf) and the probability mass functioin (pmf) respectively of $X_n - K$. Then the transition probabilities are given as follows (see Brook and Evans [17]):

$$\begin{aligned} P_{i0} &= \Pr(E_i \to E_0) = \Pr(X_n \leq K - i) \\ &= F_{-i} \\ P_{ij} &= \Pr(E_i \to E_j) = \Pr(X_n = K + j - i) \\ &= p_{j-i} \quad (j = 1, \ldots, H-1) \\ P_{iH} &= \Pr(E_i \to E_H) = \Pr(X_n \geq K + H - i) \\ &= 1 - F_{H-1-i}. \end{aligned}$$

The transition probability matrix $\mathbf{P}$ is given by:

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \vdots \\ i \\ \vdots \\ H-1 \\ H \end{array} \begin{array}{c} \begin{array}{ccccccccc} 0 & & 1 & & 2 & \cdots & j & \cdots & H-1 \end{array} \\ \begin{pmatrix} F_0 & p_1 & p_2 & \cdots & p_j & \cdots & p_{H-1} & 1-F_{H-1} \\ F_{-1} & p_0 & p_1 & \cdots & p_{j-1} & \cdots & p_{H-2} & 1-F_{H-2} \\ F_{-2} & p_{-1} & p_0 & \cdots & p_{j-2} & \cdots & p_{H-3} & 1-F_{H-3} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots \\ F_{-i} & p_{1-i} & p_{2-i} & \cdots & p_{j-i} & \cdots & p_{H-1-i} & 1-F_{H-1-i} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots \\ F_{1-H} & p_{2-H} & p_{3-H} & \cdots & p_{j-(H-1)} & \cdots & p_0 & 1-F_0 \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 1 \end{pmatrix} \end{array} \tag{3.23}$$

where $\mathbf{P}$ is a square $(H+1) \times (H+1)$ matrix. Because state $E_H$ is absorbing, the last row of $\mathbf{P}$ consists of zeros except for the last element ($p_{HH}$), which has a value of 1. Thus once the CUSUM enters state $E_H$, it can never leave, and so the process will have to be restarted. Matrix $\mathbf{P}$ can be partitioned as

$$\mathbf{P} = \begin{pmatrix} \mathbf{R} & (\mathbf{I}-\mathbf{R})\,\mathbf{1} \\ \mathbf{0}' & 1 \end{pmatrix} \tag{3.24}$$

where matrix $\mathbf{R}$ contains the probabilities of going from one transient state to another transient state, and it is obtained from $\mathbf{P}$ by deleting the last row and column. For integer $K$ values, the $\mathbf{R}$ matrix is a square $H \times H$, while for rational $K$ values ($K = a/b$), the $\mathbf{R}$ matrix is a square $bH \times bH$. Matrix $\mathbf{I}$ is an identity matrix, and vector $\mathbf{1}$ is a column of ones.

The run length vector of the control statistic is the number of steps required to reach the absorbing state $E_H$ for the first time, starting from state $E_i$ ($i = 0, 1, 2, \ldots, H-1$). Thus, the run length is a random variable and it can vary greatly. Let $\boldsymbol{N} = [N_0 \;\; N_1 \;\; \cdots \;\; N_{H-1}]'$ be the run length vector of the control statistic, where $N_i$ is defined by

$$N_i = \min_{n \geq 1}\{n : \; S_n^+ \geq H \mid S_0^+ = i\}. \tag{3.25}$$

The distribution of $N_i$ is called the run length distribution, and it can be obtained from the $t^{\text{th}}$-stage transition probability matrix $\mathbf{P}^t$:

$$\mathbf{P}^t = \prod_{j=0}^{t} \mathbf{P} = \begin{pmatrix} \mathbf{R}^t & (\mathbf{I} - \mathbf{R}^t)\,\mathbf{1} \\ \mathbf{0}' & 1 \end{pmatrix} \quad (t = 1, 2, \ldots). \tag{3.26}$$

The last column of matrix $\mathbf{P}^t$ gives the cumulative distribution function $\Pr(N_i \leq t)$ of the run length starting from state $E_i$ ($i = 0, 1, 2, \ldots, H-1$), together with unity as the last element. Define $\boldsymbol{F}_t$ to be the vector whose elements are the cumulative distribution functions of the run length starting from states $E_0, E_1, \ldots, E_{H-1}$. Then

$$\begin{aligned} \boldsymbol{F}_t &= \left[\Pr(N_0 \leq t) \;\; \Pr(N_1 \leq t) \;\; \cdots \;\; \Pr(N_{H-1} \leq t)\right]' \\ &= (\mathbf{I} - \mathbf{R}^t)\,\mathbf{1} \quad (t = 1, 2, \ldots). \end{aligned}$$

Similarly, if we define $\boldsymbol{L}_t$ as the vector whose elements are the probability mass

functions of the run lengths, then

$$\begin{aligned}
\boldsymbol{L}_t &= \begin{bmatrix}\Pr(N_0 = t) & \Pr(N_1 = t) & \cdots & \Pr(N_{H-1} = t)\end{bmatrix}' \\
&= \boldsymbol{F}_t - \boldsymbol{F}_{t-1} \\
&= \left(\mathbf{R}^{t-1} - \mathbf{R}^t\right) \mathbf{1} \\
&= \mathbf{R}^{t-1}\boldsymbol{L}_1, \qquad (3.27)
\end{aligned}$$

where $\boldsymbol{L}_1 = (\mathbf{I} - \mathbf{R})\,\mathbf{1}$ (see, for example, Lucas and Saccucci [73]; Brook and Evans [17]). If $e_i$ is a vector whose $i$th element is unity and zero elsewhere, then it follows that

$$\begin{aligned}
\Pr(N_i = t) &= \Pr(N_i \leq t) - \Pr(N_i \leq t-1) \\
&= e_i'(\mathbf{R}^{t-1} - \mathbf{R}^t)\,\mathbf{1} \\
&= e_i'\mathbf{R}^{t-1}(\mathbf{I} - \mathbf{R})\,\mathbf{1}
\end{aligned}$$

and

$$\mathbb{E}[N_i] = \sum_{t=1}^{\infty} t \Pr(N_i = t) = \sum_{t=1}^{\infty} t\, e_i'\mathbf{R}^{t-1}(\mathbf{I} - \mathbf{R})\,\mathbf{1}.$$

The ARL vector $\boldsymbol{\mu} = [\mu_0 \;\; \mu_1 \;\; \cdots \;\; \mu_{H-1}]'$ gives the mean of the run length distribution starting from states $E_0, E_1, \ldots, E_{H-1}$. Using the similarity of the matrix series with

the univariate series

$$(1-x)^{-1} = \sum_{j=0}^{\infty} x^j \tag{3.28}$$

$$(1-x)^{-2} = \sum_{j=1}^{\infty} j x^{j-1} \tag{3.29}$$

$$2(1-x)^{-3} = \sum_{j=2}^{\infty} j(j-1) x^{j-2} \tag{3.30}$$

$$2x(1-x)^{-3} = \sum_{j=2}^{\infty} j(j-1) x^{j-1} \tag{3.31}$$

provided that $|x| \leq 1$, it follows that

$$\begin{aligned} \boldsymbol{\mu} &= \mathbb{E}[\boldsymbol{N}] = \sum_{t=1}^{\infty} t\, \mathbf{R}^{t-1}\, \boldsymbol{L}_1 \\ &= (\mathbf{I}-\mathbf{R})^{-2}\, \boldsymbol{L}_1 \\ &= (\mathbf{I}-\mathbf{R})^{-1}\, \mathbf{1}, \end{aligned} \tag{3.32}$$

provided that each eigen-value of $\mathbf{R}$ has absolute value less than 1. The first element ($\mu_0$) of $\boldsymbol{\mu}$ is typically taken to be the ARL of the process, and it corresponds to the mean number of steps required to reach the absorbing state for the first time, starting from state $E_0$. If $\odot$ denotes the Hadamard product (or element-wise multiplication), then the second factorial moment is

$$\begin{aligned} \boldsymbol{\mu}^{(2)} &= \mathbb{E}[\boldsymbol{N} \odot (\boldsymbol{N}-1)] \\ &= \sum_{t=2}^{\infty} t\,(t-1)\, \mathbf{R}^{t-1}\, \boldsymbol{L}_1 \\ &= 2\,\mathbf{R}\,(\mathbf{I}-\mathbf{R})^{-2}\, \mathbf{1}. \end{aligned} \tag{3.33}$$

From the Woodbury binomial inverse theorem (Woodbury [127]), it follows that

$$(\mathbf{I}-\mathbf{R})^{-1} \;=\; \mathbf{I} \;+\; \mathbf{R}(\mathbf{I}-\mathbf{R})^{-1}$$

Noting further that $(\mathbf{I}-\mathbf{R})^{-1}\mathbf{R} \;=\; \mathbf{R}(\mathbf{I}-\mathbf{R})^{-1}$, it therefore follows that

$$\boldsymbol{\mu}^{(2)} = 2\left[(\mathbf{I}-\mathbf{R})^{-1}-\mathbf{I}\right](\mathbf{I}-\mathbf{R})^{-1}\mathbf{1}. \tag{3.34}$$

The vector of variances (or second central moments) of the run length distribution can also be easily computed from

$$\begin{aligned}
\mathrm{Var}[\boldsymbol{N}] &= \mathbb{E}[\boldsymbol{N}\odot(\boldsymbol{N}-1)] \;+\; \mathbb{E}[\boldsymbol{N}] \;-\; \{\mathbb{E}[\boldsymbol{N}]\odot\mathbb{E}[\boldsymbol{N}]\} \\
&= 2\,\mathbf{R}\,(\mathbf{I}-\mathbf{R})^{-2}\,\mathbf{1} \;+\; (\mathbf{I}-\mathbf{R})^{-1}\,\mathbf{1} \;-\; \{(\mathbf{I}-\mathbf{R})^{-1}\,\mathbf{1}\}^2 \\
&= 2\,(\mathbf{I}-\mathbf{R})^{-2}\,\mathbf{1} \;-\; (\mathbf{I}-\mathbf{R})^{-1}\,\mathbf{1} \;-\; \{(\mathbf{I}-\mathbf{R})^{-1}\,\mathbf{1}\}^2. \qquad (3.35)
\end{aligned}$$

Although the Markov chain procedure can give exact solution for a count data distribution such as the Poisson CUSUM with a rational $K$ value, in general the exact determination of the run length distribution and its moments is not possible and approximations are required (Woodall [122]; Lucas [68]).

# CHAPTER 4

## PROPOSED MODIFIED CUSUM PROCEDURE

The purpose of this chapter is to propose a one-sided tabular CUSUM procedure for some nonstandard count data, such as the generic feedlot data described in Chapter 1. The CUSUM control chart is a very useful and powerful monitoring tool. Appropriate choice of the design parameters $K$ and $H$ can be determined and used to design an optimal CUSUM to detect a specific shift (or step change) in the parameter of a distribution. The run length distribution and the average run length can be used to give an indication of the performance of the CUSUM scheme. It is also known that while the CUSUM chart is very effective in detecting small shifts, it is not as effective as the Shewhart chart in detecting large shifts. Several procedures aimed at improving the sensitivity and responsiveness of the CUSUM have been suggested. For example, the use of a combined CUSUM-Shewhart procedure for on-line control helps to improve the ability of the CUSUM to detect large shifts in the process parameter. If an out-of-control signal occurs on either (or both) charts, the process is considered out-of-control, and investigative and corrective actions will be implemented. The process will then be restarted at zero. This procedure has been discussed at length by Lucas [67]. Other useful enhancements to the CUSUM procedure are the robust CUSUMs (Lucas and Crosier [72]; Hawkins [45]) and the fast

initial response (FIR) CUSUMs or CUSUMs with a "head start" (Lucas and Crosier [71]). The FIR CUSUM is started at some nonzero value $S_0$, usually chosen to be one-half of the decision interval $H$ (that is, $S_0 = H/2$), and it is aimed at improving the sensitivity of the CUSUM for the case in which the process parameter has already shifted at the time the CUSUMs are reset. The improvement in performance is typically measured by the reduction in the average run length (ARL) of the process.

The modified CUSUM procedure described here can be viewed as another way that one could potentially improve the performance and sensitivity of the CUSUM chart. Although the procedure could also be used for continuous distributions, we will consider only the discrete (count data) case.

Four proposed modifications of the traditional CUSUM scheme will be introduced. These are

1. Additional design parameters, given by the warning level ($W$), and the pre-specified threshold value ($\pi_\alpha$) which serves as a "rejection level".

2. Expanded transition states, extended transition matrix, and the concept of a *probability of extremeness* $\pi_{j,c_n}$, where $j$ represents a specified state in a designated region (called region $B$; see Figure 1 on page 86) of the transition matrix, and $c_n$ is the value of some counter at sample $n$.

3. Additional out-of-control conditions *A-ABS* and *C-ABS*, in addition to the usual out-of-control condition *H-ABS*.

4. Warning runs rules such as the *M-in-a-row* and the $\pi_{j,c_n} \leq \pi_\alpha$ criteria which can be used in the CUSUM design to quickly detect shifts, detect trends, and to achieve some robustness against isolated outlying observations.

The following sections will discuss notations and terminology, describe the proposed modified CUSUM procedure, and provide a general algorithm and discuss a generalization of the method. Subsequent sections will illustrate the method with simple cases, discuss allowable transitions, the transition probability matrix and the computation of the ARL, and finally the determination of the design parameters.

## Additional Notations and Terminology

In this section, additional notations and terminology required for the modified cumulative sum procedure (which will now be called modified CUSUM) will be discussed. The reference value and the decision interval of the modified CUSUM scheme are denoted by $K$ and $H$ respectively, same as in the regular CUSUM scheme. Other items are outlined below.

1. Original states refer to state 0, state 1, state 2, ..., state $H-1$, state $H$ in the original CUSUM process (or the original Markov chain). These states are now denoted by $E_0$, $E_1$, $E_2$, ..., $E_{H-1}$, $E_H$ where $E_H$ is the usual absorbing state. If no confusion arises from these notations, then either "state $i$" or "$E_i$" can be used.

2. $W$ denotes the warning level of the modified CUSUM scheme, which is a parameter always chosen to be smaller than the decision interval $H$. The warning level divides the original states into two regions, labeled $A$ and $B$.

3. Region $A$ of the modified CUSUM process represents the "more stable" part of the process which suggests the process mean is near an aim value, and it involves the original transient states $E_0, E_1, E_2, \ldots, E_{W-1}$.

4. Region $B$ of the modified CUSUM consists of the original transient states $E_{W+1}$, $E_{W+2}$, $E_{W+3}, \ldots, E_{H-1}$. These original states in region $B$ may also be called buffer states. Region $B$ represents the "less stable" part of the process which suggests the process mean is larger than an aim value. Thus, the buffer states can be interpreted as 'warning states'.

5. $n_B$ denotes the total number of buffer states (in region $B$). For a CUSUM that moves in unit integer steps, $n_B$ is given by $n_B = (H - W) - 1$. If the CUSUM moves in either fractional steps $1/b$ (or rational steps $a/b$), then $n_B = b\,(H - W) - 1$.

6. $c_n$ denotes the current counter value. It represents the number of consecutive CUSUM values that fall in region $B$ of the modified CUSUM process. For example, counter value $c_n = 0$ means the current CUSUM value is in region $A$, $c_n = 1$ means the current CUSUM value is in region $B$, and $c_n = 2$ means that the two most current CUSUM values are in region $B$. The counter can

take values in the set $\{0, 1, 2, \ldots, M\}$. The value $c_n$ can be interpreted as the current number of consecutive warnings received. Thus, larger values of $c_n$ suggest the process mean may be off aim.

7. $M$ denotes the maximum threshold for counter value $c_n$. It denotes the maximum number of consecutive warnings before an out-of-control signal is given.

8. Set $\mathcal{A}$ consists of states in region $A$ as well as the warning level $W$. That is,

$$\mathcal{A} = \{E_0, E_1, E_2, \ldots, E_{W-1}, E_W\}. \tag{4.1}$$

9. $E_{(i,c_n)}$ represents a state formed by the indexed pair $(i, c_n)$, $i = 0, 1, 2, \ldots, H-1$ and $c_n = 0, 1, 2, \ldots, M$. Since all the states in set $\mathcal{A}$ (Equation (4.1)) have counter value $c_n = 0$, $\mathcal{A}$ can also be denoted by

$$\mathcal{A} = \left\{E_{(0,0)}, E_{(1,0)}, E_{(2,0)}, \ldots, E_{(W,0)}\right\}. \tag{4.2}$$

10. Expanded states in the modified CUSUM scheme are states created via an expanded process to gain more information about region $B$ of the original process. Expanded states are contained in the subsets

$$\mathcal{B}_{c_n} = \left\{E_{(W+s,c_n)},\ s = 1, 2, \ldots, n_B\right\} \quad c_n = 1, 2, \ldots, M$$

and they are formed by combining the buffer states $E_i$ $(i = W+1, W+2, \ldots, W+n_B)$ and counter values $c_n = 1, 2, \ldots, M$ in region $B$. That is, each state in $\mathcal{B}_{c_n}$ is represented by an indexed pair $(W+s, c_n)$. For example,

$E_{(W+1,1)}$ is an expanded state, and it refers to state $E_{W+1}$ with counter value $c_n = 1$. That is, the CUSUM is in state $E_{W+1}$ for the first time. In addition, all expanded states in the subset $\mathcal{B}_M = \{E_{(W+s,M)},\ s = 1, 2, \ldots, n_B\}$ are declared as absorbing states. Because the number of absorbing states has been expanded, the ARL will be affected using this modification.

11. Probability of extremeness (denoted by $\pi_{j,c_n},\ c_n = 2, 3, \ldots, M-1$) is the probability of having $c_n - 1$ consecutive points in region $B$ and ending up in a specified state $(j, c_n)$ (or $E_{(j,c_n)}$), given that the process is currently in region $B$ for the first time. From the definitions and terminology discussed earlier in this section, it implies that the $\pi_{j,c_n}$ rule is ignored by transitions from region $B$ to any state in set $\mathcal{A}$, transitions from set $\mathcal{A}$ to set $\mathcal{A}$, and all transitions to the states in set $\mathcal{B}_1$ and to the absorbing states.

12. $\pi_\alpha$ is a parameter which serves as a "rejection level" for $\pi_{j,c_n}$ in the modified CUSUM procedure. That is, $\pi_\alpha$ represents a pre-specified threshold or a cut-off value of the probability of extremeness $\pi_{j,c_n}$.

13. $\mathcal{B}^*$ is the set of $\pi_\alpha$–expanded states which consists of the expanded states for which the probability of extremeness is less than or equal to the threshold; that is, for which $\pi_{j,c_n} \leq \pi_\alpha$. Let $\mathcal{B}^*_{c_n}$ be the subset of the expanded states such that

$$\mathcal{B}^*_{c_n} = \left\{E_{(W+s,c_n)} :\ \pi_{W+s,c_n} \leq \pi_\alpha \text{ for } s = 1, 2, \ldots, n_B\right\},\ \ c_n = 2, 3, \ldots, M-1.$$

Then, for example, we may have $E_{(W+1,2)} \in \mathcal{B}^*$ but $E_{(W+1,1)} \notin \mathcal{B}^*$. All $\pi_\alpha$–expanded states are considered as absorbing states, and they are determined for a given in-control mean and $\pi_\alpha$.

14. The modified upper (or upward) CUSUM statistic at sample $n$ is denoted by $S^+_{n,c_n}$, where $c_n$ is the counter value. Similarly, the modified lower (or downward) CUSUM statistic at sample $n$ is denoted by $S^-_{n,c_n}$. For modified CUSUM values that fall in set $\mathcal{A}$, the counter value is $c_n = 0$, and so the control statistic can be written without the $c_n = 0$ value. That is, we can write $S^+_{n,0} = S^+_n$. The initial value of the control statistic can also be denoted by, for example, $S^+_{0,0}$ or $S^+_0$.

15. $E_{H-ABS}$ in the modified CUSUM scheme denotes the usual absorbing state $E_H$, and it represents the usual out-of-control condition where the CUSUM value goes beyond the decision boundary (eg. $S^+_{n,c_n} \geq H$). This out-of-control condition is termed *H-ABS* where "ABS" is an abbreviation for absorbing state.

16. $E_{C-ABS}$ denotes all the absorbing states in the subset of expanded states $\mathcal{B}_M = \{E_{(W+s,M)},\ s = 1, 2, \ldots, n_B\}$, and it represents the out-of-control condition due the maximum of $M$ consecutive CUSUM values in region $B$. This scenario is called *M-in-a-row*, and the type of out-of-control condition is termed *C-ABS*.

17. $E_{A-ABS}$ denotes all the absorbing states formed by the $\pi_\alpha$–expanded states in the set $\mathcal{B}^*$, and it represents the out-of-control condition due to the "rejection level" $\pi_\alpha$. This type of out-of-control condition is termed *A-ABS*.

18. $E_H^*$ denotes the single absorbing state formed by lumping (or combining, or pooling) together the three kinds of absorbing states $E_{A-ABS}, E_{C-ABS}$, and $E_{H-ABS}$, based on the lumpability theorem discussed in Chapter 2.

19. Extended transition matrix is the transition probability matrix for the expanded Markov chain in the modified CUSUM procedure. The extended transition probability matrix is denoted by $\mathbf{P_E}$ and it consists of all the original states in set $\mathcal{A}$, the expanded states in subsets $\mathcal{B}_{c_n}$, $c_n = 1, 2, \ldots, M-1$, and the two types of absorbing states $E_{C-ABS}$ and $E_{H-ABS}$ have been lumped together as $E_H^*$.

20. Modified transition matrix $\mathbf{P_M}$ is the transition matrix obtained by (possibly) reducing the extended matrix $\mathbf{P_E}$ by the number of $\pi_\alpha$–expanded states (states for which $\pi_{j,c_n} \leq \pi_\alpha$), and where the three types of absorbing states $E_{C-ABS}$, $E_{H-ABS}$ and $E_{A-ABS}$ have been lumped together as $E_H^*$. The number of $\pi_\alpha$–expanded states to be declared absorbing depends on the in-control process distribution and the threshold $\pi_\alpha$, and hence the modified matrix depends on a given in-control distribution and threshold $\pi_\alpha$. If the in-control distribution is Poisson($\mu_0$), then the number of $\pi_\alpha$-expanded states and the modified matrix depend on the process mean mean $\mu_0$ and $\pi_\alpha$.

21. Warning runs rule refers to decision rules based on the use of *C-ABS* (that is, *M-in-a-row*, or a run of M consecutive CUSUM points in region $B$) and *A-ABS*

(with the probability of extremeness $\pi_{j,c_n} \leq \pi_\alpha$) as out-of-control conditions.

General Description of the Procedure

Consider the discrete random variable $X$ to be the count (or number) of occurrences of some event. Suppose the independent random variables $X_1, X_2, \ldots$ are observed successively, with probability mass functions $f(x;\theta_i)$, $i = 1, 2, \ldots$, and cumulative distribution functions $F(x;\theta_i)$, $i = 1, 2, \ldots$, respectively. Typically, $\theta_i$, $(i = 1, 2, \ldots)$ would be location or scale parameters, but we would assume here that $\theta$ is a location parameter. It is usually of interest to monitor the condition $\theta_i = \theta_0$, $i = 1, 2, \ldots$. This is basically a test of the simple hypotheses $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$, where $\theta_1 > \theta_0$ if we want to monitor an increase in $\theta$, and $\theta_1 < \theta_0$ if a decrease in $\theta$ is to be monitored.

The CUSUM can be considered as a discrete state Markov process, as discussed in Chapter 3. In the modified scheme, a warning level $W$ is suitably chosen to classify the state space of the regular count data CUSUM into two regions labeled $A$ and $B$, as in Figure 1 (page 86).

Region $A$ is made up of the original CUSUM states $E_0, E_1, E_2, \ldots, E_{W-1}$, and is considered the "stable" part of the process, while region $B$ comprises states $E_{W+1}, E_{W+2}, \ldots, E_{H-1}$. The warning level then corresponds to state $E_W$, and together with the states in region $A$, they form a set $\mathcal{A} = \{E_0, E_1, E_2, \ldots, E_W\}$. Assuming the decision interval $H$ and the reference value $K$ of the CUSUM design are known, then

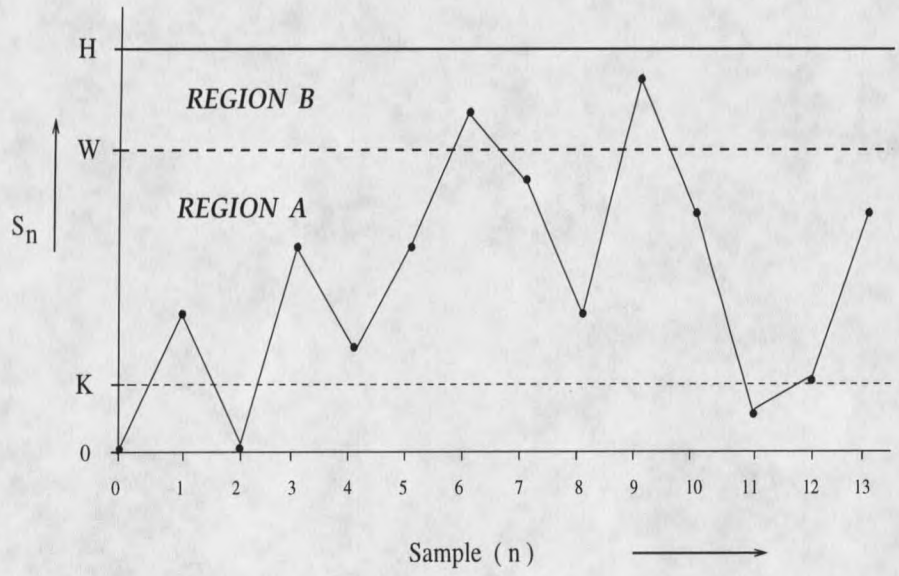


**Figure 1**: A graphical representation of the modified CUSUM chart, showing a typical sample path with the two regions $A$ and $B$. Here $K$ is the reference value, $\mu_0$ is the target, $W$ is the warning limit, and $H$ is the decision interval. $S^+_{n,\,c_n}$ is a one-sided upward modified CUSUM statistic.

$E_H$ represents the usual absorbing state of the Markov chain.

It is naturally important to closely monitor the behavior of the process in region $B$. Accordingly, a counter is kept with value $c_n = 0, 1, 2, \ldots, M$ corresponding to zero, one, two, ..., $M$ consecutive CUSUM values falling into region $B$. All states in set $\mathcal{A}$ have a counter value $c_n = 0$.

In order to gain more information about the process in region $B$, expanded states are created using the pairwise combinations of the states $(E)$ and counter values $(c_n)$ in region $B$. These expanded states are also designed to accommodate the consecutive CUSUM values that fall into region $B$.

DEFINITION 4.1. Let $K$ be the reference value, $\pi_\alpha$ the threshold value, $H$ the

decision interval, and $W$ the warning level where $K$, $W$, and $H$ are positive integer values with $W \leq H - 1$. If non-integer rational values are encountered, they could be rescaled, and that would not introduce any major changes in the procedure. An upper (or upward) modified (count data) CUSUM, after sample $n$, is defined by the recursive equations

$$S^+_{0,0} = S^+_0 = 0 \tag{4.3}$$

$$S^+_{n,c_n} = \max(0,\ X_n - K + S^+_{n-1,c_{n-1}}) \tag{4.4}$$

where $S^+_{n,c_n}$ is the modified CUSUM statistic at sample $n$, at which time there are $c_n$ consecutive points in region $B$. Note that $S^+_{n,c_n}$ is a Markov process because of its random walk structure, and therefore its run length distribution and the average run length (ARL) can be computed using the Markov chain approach of Brook and Evans [17], which was described in Chapter 3. The lower (or downward) modified CUSUM scheme can be defined similarly by

$$S^-_{0,0} = S^-_0 = 0 \tag{4.5}$$

$$S^-_{n,c_n} = \min(0,\ X_n - K + S^-_{n-1,c_{n-1}}) \tag{4.6}$$

which can be turned into an upward CUSUM as

$$D^+_{0,0} = D^+_0 = 0 \tag{4.7}$$

$$D^+_{n,c_n} = \max(0,\ K - X_n + D^+_{n-1,c_{n-1}}). \tag{4.8}$$

In view of this, we will consider just the upward CUSUM in describing the modified CUSUM method. If $S^+_{n,c_n} \leq W$ (that is, if the CUSUM is either in region $A$, or at the

warning level), the process is considered to be in-control, and sampling will continue unimpeded. If the CUSUM enters the usual absorbing state $E_{H-ABS}$ representing the condition $S^{+}_{n,c_n} \geq H$, it will signal out of control and sampling will stop. Since region $B$ is of great interest to us, certain restrictions are imposed on what can happen there.

Generally, if the process is in control, it is expected that the CUSUM values will be within $[0, W]$ most of the time. That is, with a large enough $W$, the CUSUM values should infrequently exceed the warning level, if the process is really in control. It would be even rarer for a process that is in-control to have $M$ consecutive CUSUM values falling inside region $B$ (which is close to $H$, with a large enough $W$). On the other hand, if the process is truly out-of-control, it is generally desired that the CUSUM would reach the decision interval $H$ quickly (to allow for early detection of out-of-control behavior). It is reasonable then to expect that $M$ consecutive CUSUM values falling inside region $B$ (and therefore closer to $H$, with a large enough $W$), would not be a rare event, if the process is substantially out-of-control. Consequently, whenever there is an occurrence of $M$ consecutive points falling into region $B$, the process is automatically declared out-of-control of type *C-ABS* (due to the *M-in-a-row* criterion).

If the CUSUM enters into region $B$ for the first time (with $c_n = 1$), this transition only triggers a warning and so no action is taken, since we are more interested in the occurrence of multiple consecutive CUSUM points in region $B$, while ignoring isolated, individual outlying observations. Accordingly, we will like to know "how unusual" it

is for an in-control process to have two, three, ..., $M-1$ consecutive CUSUM values cross the warning level and into region $B$ given that the process is currently in region $B$ for the first time. This is measured by the probability of extremeness, $\pi_{j,c_n}$, which would be compared with the pre-specified cut-off value $\pi_\alpha$. A typical sample path is shown in Figure 1. If the CUSUM value stays in region $B$ for the second consecutive sample, the counter will increment to $c_n = 2$, and $\pi_{j,2}$ will be calculated. If $\pi_{j,2}$ is less than or equal to $\pi_\alpha$, the process will be stopped and declared out-of-control with type *A-ABS*. If $\pi_{j,2}$ is greater than $\pi_\alpha$, the procedure would be continued.

Also, if the CUSUM crosses the decision boundary $H$ at any time, the process is stopped and declared out-of-control of type *H-ABS*. Alternatively, if the CUSUM returns to region $A$, (that is, if the CUSUM value is less than or equal to $W$) at any time, the counter value $c_n$ is reset to zero, and the process is once again said to be in-control. Sampling would then continue as usual. In Figure 1 (page 86), at $n = 6$ and $n = 9$, the counter $c_n = 1$. At all other points $c_n = 0$. Thus, with the modified CUSUM, there are three ways to trigger an out-of-control signal: (1) reaching the decision boundary $H$ (type H-ABS), (2) having $M$-in-a-row in region $B$ (type C-ABS), or (3) meets the probability of extremeness criterion (type A-ABS).

It should be noted that the additional out-of-control conditions (type H-ABS and type H-ABS) have the effect of reducing the average run length (ARL) of the modified CUSUM procedure in comparison to the traditional count data CUSUM schemes.

## Modified CUSUM Algorithm

The modified CUSUM algorithm will be described here in terms of the modified upper (or upward) CUSUM scheme ($S^+_{n,\,c_n}$) for the case of *M-in-a-row* warning runs rule. It will initially be assumed here that all the CUSUM design parameters are known. Later in the chapter, we will discuss how to choose appropriate values for these design parameters. A flowchart for implementing the *four-in-a-row* warning runs rule is provided in Figure 2.

A Modified CUSUM scheme at $\pi_\alpha$, with starting value $S^+_{0,\,0} = S^+_0 = 0$, reference value $K$, warning level $W$, and the decision interval $H$, can be designed by following steps **C1**—**C3** of the algorithm given below. The same algorithm can be used for the lower (or downward) CUSUM scheme.

*Algorithm*

**C1.** Specify the modified CUSUM design parameters: $K$, $W$, $H$, and $\pi_\alpha$. Set counter value to $c_n = 0$, and initialize the CUSUM at $S^+_{0,\,0} = S^+_0 = 0$.

**C2.** After the $n$th sample ($n = 1, 2, \ldots$), calculate the values of the CUSUM control statistic $S^+_{n,\,c_n}$.

**C3.** Do the following:

(A) If $S^{+}_{n,c_n} \geq H$ at any time, stop sampling and declare the process out-of-control (beyond the usual decision boundary $H$). This type of out-of-control is termed "H-ABS".

(B) If $S^{+}_{n,c_n} \leq W$, reset counter value to $c_n = 0$, and return to step **C2**.

(C) If $W + 1 \leq S^{+}_{n,c_n} \leq H - 1$, increment counter value by one unit and calculate probability of extremeness $\pi_{j,c_{n+1}}$ for the new counter value (at sample $n + 1$).

- If $\pi_{j,c_{n+1}} \leq \pi_\alpha$ (for $c_{n+1} = 2, 3, \ldots, M - 1$), stop sampling and declare the process out-of-control for counter value $c_{n+1}$ at sample $n+1$. This type of out-of-control condition is termed "A-ABS" (due to the probability of extremeness being less than or equal to the pre-specified threshold value).
- If $\pi_{j,c_{n+1}} > \pi_\alpha$, increment counter value by one unit. If $c_{n+1} \leq M - 1$, go back to step **C2**. Otherwise, if $\pi_{j,c_{n+1}} > \pi_\alpha$ and $c_{n+1} = M$ stop sampling, and declare the process as out-of-control. This type of out-of-control is "C-ABS" (due to the *M-in-a-row* criterion).

Generalization: $n_B$ Buffer States with *M-in-a-row*

For a generalization of this procedure, consider the state space of the regular CUSUM statistic, divided into $H + 1$ intervals or discrete points $0, 1, 2, \ldots, H$. This

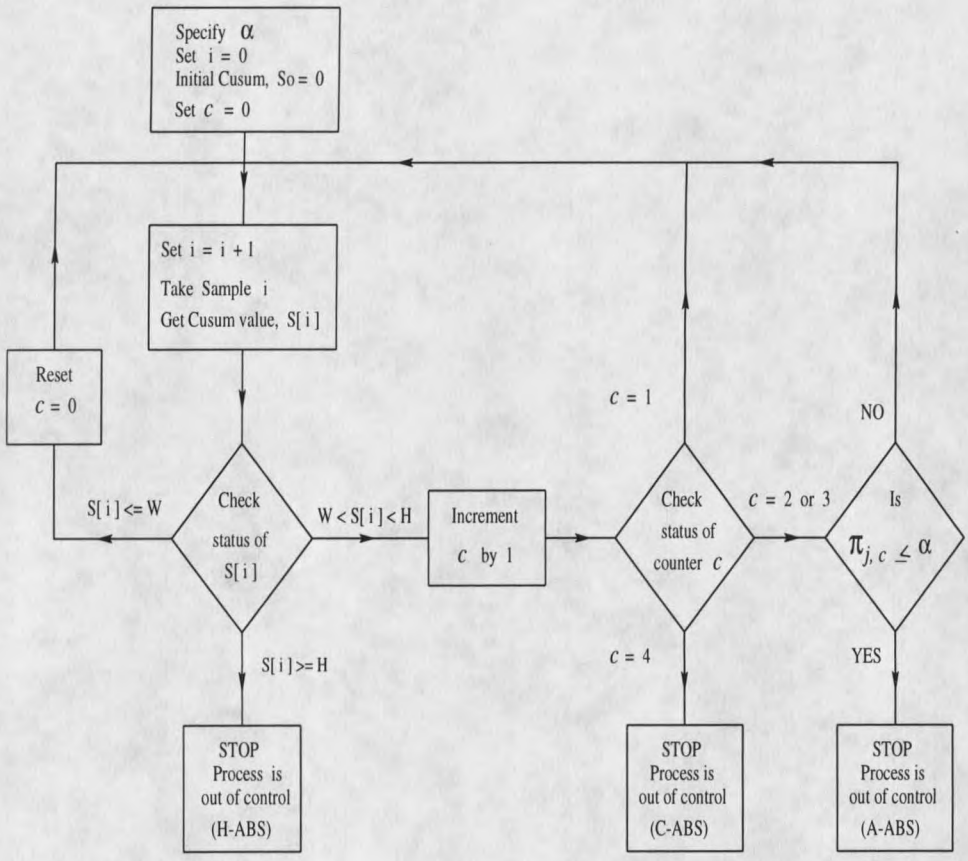


**Figure 2**: A schematic representation (flowchart) of the modified CUSUM procedure, based on the four-in-a-row warning runs rule.

forms a Markov chain with original states $E_0, E_1, E_2, \ldots, E_H$, where $E_H$ is an absorbing state and the rest are transient states. Suppose the decision interval $H$ and reference value $K$ are known. Choose a suitable warning level $W$ (with $W < H$) to separate the original transient states in the CUSUM space into two regions labeled $A$ and $B$, where

$$\text{Region } A = \{E_0, E_1, E_2, \ldots, E_{W-1}\} \quad \text{and} \tag{4.9}$$

$$\text{Region } B = \{E_{W+1}, E_{W+2}, E_{W+3}, \ldots, E_{H-1}\}. \tag{4.10}$$

The number of original states in region $B$ (that is, buffer states) is given by $n_B$, where

$$n_B = b(H-W)-1 \quad \text{(if CUSUM moves in fractional steps of } 1/b) \tag{4.11}$$

$$= (H-W)-1 \quad \text{(if CUSUM moves in unit integer steps).} \tag{4.12}$$

A counter value $c_n$ ($c_n = 0, 1, 2, \ldots, M$) keeps track of the number of consecutive CUSUM values that fall in region $B$, where $M$ is the maximum counter value. A cut-off value or threshold $\pi_\alpha$ is specified for the probability of extremeness $\pi_{j,c_n}$, where $c_n$ is the counter value at some sample number $n$ (at the CUSUM is in region $B$).

To obtain more information about the process, a new Markov chain (called the expanded process) is formed from the states given in sets $\mathcal{A}$ and $\mathcal{B}_{c_n}$ as well as state $E_{H-ABS}$.

(a) Set $\mathcal{A}$ contains states with counter value $c_n = 0$ (states with CUSUM value outside region $B$):

$$\mathcal{A} = \{E_0, E_1, E_2, \ldots, E_{W-1}, E_W\}. \tag{4.13}$$

(b) Subsets $\mathcal{B}_{c_n}$ contain expanded states formed from pairs of (*state, counter*) combinations. These are states containing $\{1, 2, \ldots, M\}$ consecutive CUSUM values in region $B$. Thus,

$$\mathcal{B}_{c_n} = \left\{E_{(W+s,\,c_n)},\ s = 1, 2, \ldots, n_B\right\}, \quad c_n = 1, 2, \ldots, M, \tag{4.14}$$

from which the following subsets can be obtained:

$$\begin{aligned}
\mathcal{B}_1 &= \left\{E_{(W+s,1)},\ s = 1, 2, \ldots, n_B\right\} && (4.15)\\
&= \left\{E_{(W+1,1)}, E_{(W+2,1)}, \cdots, E_{(W+n_B,1)}\right\}\\
\mathcal{B}_2 &= \left\{E_{(W+s,2)},\ s = 1, 2, \ldots, n_B\right\} && (4.16)\\
&= \left\{E_{(W+1,2)}, E_{(W+2,2)}, \cdots, E_{(W+n_B,2)}\right\}\\
&\vdots \qquad\qquad \vdots\\
\mathcal{B}_{M-1} &= \left\{E_{(W+s,M-1)},\ s = 1, 2, \ldots, n_B\right\} && (4.17)\\
&= \left\{E_{(W+1,M-1)}, E_{(W+2,M-1)}, \cdots, E_{(W+n_B,M-1)}\right\}\\
\mathcal{B}_M &= \left\{E_{(W+s,M)},\ s = 1, 2, \ldots, n_B\right\}\\
&= \left\{E_{(W+1,M)}, E_{(W+2,M)}, \cdots, E_{(W+n_B,M)}\right\}. && (4.18)
\end{aligned}$$

(c) Subsets $\mathcal{B}^*_{c_n}$ are the alpha–expanded states (with probability of extremeness $\pi_{j,c_n} \leq \pi_\alpha$). These are absorbing states, which are determined by the in-control mean and the threshold $\pi_\alpha$, result in out-of-control condition *A-ABS*. Alpha-expanded states belong to the subsets

$$\mathcal{B}^*_{c_n} = \left\{E_{(W+s,\,c_n)} :\ \pi_{W+s,c_n} \leq \pi_\alpha \text{ for } s = 1, 2, \ldots, n_B\right\},\ c_n = 2, 3, \ldots, M-1. \tag{4.19}$$

It should be noted that the transition probability matrix of this new Markov chain (extended transition matrix) can grow dramatically in size. The usual absorbing state $E_H$ (of the original chain) is now denoted by $E_{H-ABS}$ in the new chain, and it belongs

to the out-of-control condition *H-ABS*.

Warning runs rules are consequently implemented as follows. Since $M$ consecutive CUSUM points in region $B$ exhausts the maximum counter value (*M-in-a-row*), all the $n_B$ states in subset $\mathcal{B}_M$ (4.18) are automatically declared absorbing, due to out-of-control condition *C-ABS*. Hence it is not necessary to identify the individual expanded states in $\mathcal{B}_M$. Furthermore, it can be recognized that the states in $\mathcal{B}_M$ will form an $\mathbf{I}_{n_B \times n_B}$ identity matrix in the lower right side of extended transition matrix, and can thus be lumped (or pooled) together as $E_{C-ABS}$, by the lumpability theorem (discussed earlier in Chapter 2). A reduced matrix (modified matrix) can be obtained by removing the alpha-expanded states from the transient sets. These alpha-expanded states can also be lumped together as $E_{A-ABS}$, since the identification of individual states is necessary.

The extended transition matrix $\mathbf{P_E}$ has a total of $W + (M-1)\,n_B + 3$ states consisting of:

$$\left\{ \overbrace{E_0,\, E_1,\, E_2,\, \ldots,\, E_W}^{W+1 \text{ states in set } \mathcal{A}},\ \overbrace{E_{(W+1,1)},\, \ldots,\, E_{(W+n_B,M-1)}}^{n_B\,(M-1) \text{ expanded states}},\ \overbrace{E_{C-ABS},\, E_{H-ABS}}^{2 \text{ types of absorbing states } (E_H^*)} \right\} \tag{4.20}$$

where the $(M-1)\,n_B$ expanded states come from subsets $\{\mathcal{B}_{c_n},\ c_n = 1, 2, \ldots, M-1\}$ defined in equations (4.15)–(4.17). If $q^*$ of the expanded states violate the $\pi_{j,c_n} \leq \pi_\alpha$ rule for a fixed in-control mean, then the modified matrix $\mathbf{P_M}$ will consist of states

$$\left\{ \overbrace{E_0,\, E_1,\, E_2,\, \ldots,\, E_W}^{W+1 \text{ states in set } \mathcal{A}},\ \overbrace{E_{(W+1,1)},\, E_{(W+2,1)},\, \ldots,}^{n_B\,(M-1)-q^* \text{ expanded states}}\ \overbrace{E_{A-ABS},\, E_{C-ABS},\, E_{H-ABS}}^{3 \text{ types of absorbing states } (E_H^*)} \right\} \tag{4.21}$$

where again $E_H^*$ represents the lumping of absorbing states $E_{A-ABS}, E_{C-ABS}$ and $E_{H-ABS}$. Each of the $(M-1)\,n_B$ expanded states in equations (4.15)–(4.17) is associated with a column (or a row) in the extended transition matrix. The columns (or rows) are contained in corresponding subsets

$$\begin{aligned} \mathcal{C}_{c_n} &= \{(M-1)(s-1)+(W+1)+c_n,\ \ s=1,2,\ldots,n_B\},\quad c_n=1,2,\ldots,M-1 \\ &= \{W+M\,(s-1)+(2+c_n-s),\ \ s=1,2,\ldots,n_B\},\quad c_n=1,2,\ldots,M-1 \end{aligned}$$

where

$$\begin{aligned} \mathcal{C}_1 &= \{W+M\,(s-1)+(3-s),\ \ s=1,2,\ldots,n_B\} \\ &= \{W+2,\ W+M+1,\ldots,W+(n_B-1)\,M+(3-n_B)\} \qquad (4.22) \\ \mathcal{C}_2 &= \{W+M\,(s-1)+(4-s),\ \ s=1,2,\ldots,n_B\} \\ &= \{W+3,\ W+M+2,\ldots,W+(n_B-1)\,M+(4-n_B)\} \qquad (4.23) \\ &\vdots \qquad\qquad \vdots \\ \mathcal{C}_{M-1} &= \{W+M\,(s-1)+(2+M-1-s),\ \ s=1,2,\ldots,n_B\} \\ &= \{W+M\,s+(1-s),\ \ s=1,2,\ldots,n_B\} \\ &= \{W+M,\ W+2M-1,\ldots,W+n_B\,M-(n_B-1)\}. \qquad (4.24) \end{aligned}$$

The resulting form of the extended transition probability matrix $\mathbf{P_E}$ now has a total of $W+(M-1)\,n_B+2$ states, and it can be partitioned as

$$\mathbf{P_E} = \begin{array}{c} \\ A \\ B \\ H \end{array} \begin{array}{c} \begin{array}{ccc} A & B & H \end{array} \\ \begin{pmatrix} \mathbf{P}_{AA} & \mathbf{P}_{AB} & \mathbf{P}_{AH} \\ \mathbf{P}_{BA} & \mathbf{P}_{BB} & \mathbf{P}_{BH} \\ \mathbf{P}_{HA} & \mathbf{P}_{HB} & \mathbf{P}_{HH} \end{pmatrix} \end{array}, \tag{4.25}$$

where

$\mathbf{P}_{AA}$ = a $(W+1) \times (W+1)$ matrix of transition probabilities from region $A$ to region $A$.

$\mathbf{P}_{AB}$ = a $(W+1) \times n_B(M-1)$ matrix of transition probabilities from region $A$ to region $B$.

$\mathbf{P}_{AH}$ = a $(W+1) \times 2$ matrix of transition probabilities from region $A$ to an absorbing state.

$\mathbf{P}_{BA}$ = a $n_B(M-1) \times (W+1)$ matrix of transition probabilities from region $B$ to region $A$.

$\mathbf{P}_{BB}$ = a $n_B(M-1) \times n_B(M-1)$ matrix of transition probabilities from region $B$ to region $B$.

$\mathbf{P}_{BH}$ = a $n_B(M-1) \times 2$ matrix of transition probabilities from region $B$ to an absorbing state.

$\mathbf{P}_{HA}$ = a $\mathbf{O}_{2\times(W+1)}$ matrix of zeros.

$\mathbf{P}_{HB}$ = a $\mathbf{O}_{2\times n_B(M-1)}$ matrix of zeros.

$\mathbf{P}_{HH}$ is an identity matrix that concerns the absorbing states.

To obtain the modified matrix $\mathbf{P_M}$, the probability of extremeness $\pi_{j,c_n}$ must be known. For an extended matrix $\mathbf{P_E}$ with $n_B$ buffer states based on the general $M$-in-a-row criteria, the probability of extremeness is

$$\begin{aligned}
\pi_{j,c_n} &= \Pr\left(\begin{array}{l}\text{Having } c_n - 1 \text{ additional consecutive points}\\ \text{in region } B \text{ to bring the process to a fixed}\\ \text{state } j \text{ in region } B, \text{ given that the process}\\ \text{is currently in region } B \text{ for the first time}\end{array}\right) \\
&= \Pr\left(S^+_{n+1,\,c_{n+1}} \in E_{(j,c_{n+1})} \mid S^+_{n,1} \in \mathcal{B}_1\right) \quad \text{(for fixed } j\text{)} \\
&= \mathbf{1}'_{n_B}\mathbf{P}^{c_n-1}_{BB}\boldsymbol{e}_j \quad (c_n = 2, 3, \ldots, M-1)
\end{aligned}$$

where $\mathbf{P}_{BB}$ is defined in Equation (4.25) as the matrix of probabilities for transitions from region $B$ to region $B$, $\boldsymbol{e}_j$ is a $n_B \times 1$ vector with 1 as the $j$th element and zeros elsewhere, and $\mathbf{1}_{n_B}$ is a vector of ones. The computation of $\pi_{j,c_n}$ is discussed further in a later section.

Expanded states that violate the $\pi_{j,c_n} \leq \pi_\alpha$ rule for a specified in-control mean are removed from the transient set and added to the absorbing set. The same set of alpha-expanded states will be removed from the transient set of any extended matrix with a specified out-of-control mean, which can then be used for comparison. The run length distribution is then based on this modified matrix. To simplify the calculations of the run length distribution and its moments (such as the ARL), the various types of absorbing states can be lumped together as $E^*_H$ (since these absorbing states form an $\mathbf{I}_{3\times 3}$ identity matrix in the lower right side of the modified matrix. The resulting

form of the modified matrix $\mathbf{P_M}$ can be partitioned as

$$\mathbf{P_M} = \begin{pmatrix} \mathbf{R_M} & (\mathbf{I} - \mathbf{R_M})\,\mathbf{1} \\ \mathbf{0}' & 1 \end{pmatrix} \tag{4.26}$$

where $\mathbf{R_M}$ is obtained from $\mathbf{P_M}$ by deleting the last row and last column. For any given implementation of the procedure however, the particular type of out-of-control condition can be specified.

The general control procedure described above is what we refer to as the *modified cumulative sum* (modified CUSUM) procedure, and its design and implementation follow the general algorithm provided earlier in this chapter. Both one-sided upper ($S^+_{n,\,c_n}$) and one-sided lower ($S^-_{n,\,c_n}$) modified CUSUMs can be designed, with and without the fast initial response feature, for different values of the parameters. Although the partitioning of the modified transition probability matrix (4.26) is similar to that provided in (3.24), there are considerable differences in their structures and forms. This is due to the fact that there are restrictions on the accessibility and communication involving some of the expanded states, which ultimately results in restrictions on the possible (or allowable) transitions among the states in the modified CUSUM scheme. The run length distribution, the ARL, and the variance of the run length can therefore be computed using similar methods described in chapter 3. Simple case examples are provided in the next section to illustrate the procedure.

Simple Illustrative Cases

The illustrative cases presented here are based on the *four-in-a-row* warning runs rules for a fixed $\pi_\alpha$, for which a flowchart is provided in Figure 2. Let $\mathcal{A}$ be the set consisting of all the states in region $A$ plus the warning level, i.e. $\mathcal{A} = \{E_0, E_1, E_2, \ldots, E_{W-1}, E_W\}$. Suppose we have $n_B$ (original) states in region $B$. Denote these states by $E_{W+1}, E_{W+2}, E_{W+3}, \ldots, E_{W+n_B}$. Also let the allowable counter values in region $B$ be $c_n = 1, 2, 3$. Then the expanded states in region $B$ are the (state, counter) combinations:

$$\mathcal{B}_1 = \{E_{(W+1,\,1)},\ E_{(W+2,\,1)},\ \ldots,\ E_{(W+n_B,\,1)}\} \tag{4.27a}$$

$$\mathcal{B}_2 = \{E_{(W+1,\,2)},\ E_{(W+2,\,2)},\ \ldots,\ E_{(W+n_B,\,2)}\} \tag{4.27b}$$

$$\mathcal{B}_3 = \{E_{(W+1,\,3)},\ E_{(W+2,\,3)},\ \ldots,\ E_{(W+n_B,\,3)}\} \tag{4.27c}$$

where $n_B = H - (W+1)$ is the number of buffer states (or original states in region $B$), $W$ is the warning level, $H$ is the decision interval, and $c_n$ is the counter value. The subsets $\mathcal{B}_1$, $\mathcal{B}_2$, and $\mathcal{B}_3$ are made up of expanded states in region $B$ containing one, two, and three consecutive points respectively, in region $B$. Remember that the states in $\mathcal{B}_4$, given by

$$\mathcal{B}_4 = \{E_{(W+1,\,4)},\ E_{(W+2,\,4)},\ \ldots,\ E_{(W+n_B,\,4)}\}, \tag{4.28}$$

are automatically classified as absorbing states due to the four-in-a-row rule. These are the states in region $B$ containing the maximum counter value of four consecutive CUSUM points in a row.

For example, $E_{(W+1,\,1)}$ means that the CUSUM has *just* moved from a state in region $A$ into state $E_{W+1}$ in region $B$. Similarly, $E_{(W+3,\,2)}$ means the CUSUM has stayed in region $B$ for the second consecutive sample (making a transition from a state in set $\mathcal{B}_1$ (say $E_{(W+1,\,1)}$ to $E_{(W+3,\,2)}$, all within region $B$). Each expanded state is associated with a column or a row in the extended transition probability matrix and is contained in subset $\mathcal{C}_{c_n}$, where

$$\mathcal{C}_{c_n} = \{3\,(s-1) + (W+1) + c_n\,,\ \ s = 1, 2, \ldots, n_B\},\ \ c_n = 1, 2, 3. \tag{4.29}$$

Thus, with respect to the extended transition matrix, the expanded states in the subsets $\mathcal{B}_1$, $\mathcal{B}_2$, and $\mathcal{B}_3$ in equation (4.27), will be contained, respectively, in rows or columns given by the subsets

$$\begin{aligned}
\mathcal{C}_1 &= 3\,(s-1) + (W+2),\ \ s = 1, 2, \ldots, n_B,\ \ c_n = 1 \\
&= \{W+2,\ \ W+5,\ \ W+8\,, \ldots, W+3\,n_B - 1\} \qquad (4.30a) \\
\mathcal{C}_2 &= 3\,(s-1) + (W+3),\ \ s = 1, 2, \ldots, n_B;\ \ c_n = 2 \\
&= \{W+3,\ \ W+6,\ \ W+9\,, \ldots, W+3\,n_B\} \qquad (4.30b) \\
\mathcal{C}_3 &= 3\,(s-1) + (W+4),\ \ s = 1, 2, \ldots, n_B;\ \ c_n = 3 \\
&= \{W+4,\ \ W+7,\ \ W+10\,, \ldots, W+3\,n_B + 1\}. \qquad (4.30c)
\end{aligned}$$

More specifically, $\mathcal{C}_1$ will contain the columns (or rows) in the extended transition matrix for the expanded states defined by $\mathcal{B}_1$ (states in region $B$ with counter value $c_n = 1$), whereas the columns (or rows) contained in $\mathcal{C}_2$ and $\mathcal{C}_3$ will be for the states with two ($\mathcal{B}_2$) and states with three ($\mathcal{B}_3$) consecutive points respectively, in region $B$.

EXAMPLE 4.1. Let H = 7, and W = 3. Then the original transition states for the CUSUM will be $\{0, 1, 2, \ldots, 6, 7\}$. The number of original states in region $B$ is $n_B = H - (W + 1) = 3$, corresponding to states $E_4$, $E_5$, and $E_6$. The expanded transition states are

$$\begin{aligned} \mathcal{B}_1 &= \{E_{(4,1)}, E_{(5,1)}, E_{(6,1)}\} \\ \mathcal{B}_2 &= \{E_{(4,2)}, E_{(5,2)}, E_{(6,2)}\} \\ \mathcal{B}_3 &= \{E_{(4,3)}, E_{(5,3)}, E_{(6,3)}\} \end{aligned}$$

and their rows or columns in the (extended) transition matrix are contained respectively in the corresponding subsets

$$\begin{aligned} \mathcal{C}_1 &= \{5,\ 8,\ 11\} \\ \mathcal{C}_2 &= \{6,\ 9,\ 12\} \\ \mathcal{C}_3 &= \{7,\ 10,\ 13\}. \end{aligned}$$

It should be obvious that state $E_0$ in region $A$ will correspond to column (or row) 1 in the transition matrix. Remember that although we have state zero ($E_0$) as a Markov state, zero is not a valid row (or column) number for a matrix. Note also the that absorbing states in $\mathcal{B}_4$ are given by

$$\mathcal{B}_4 = \{E_{(4,4)}, E_{(5,4)}, E_{(6,4)}\}.$$

Possible Transitions in the Modified CUSUM Scheme

Using the previous example with $H = 7$ and $W = 3$, suppose the CUSUM

enters state $E_5$ in region $B$ when the previous CUSUM is in region $A$. This will cause the counter value to increment to $c_n = 1$, resulting in the expanded state $E_{(5,\,1)}$ in subset $\mathcal{B}_1$. If the CUSUM stays in state $E_5$ after the next sample, the counter will increase to $c_n = 2$, resulting in the expanded state $E_{(5,\,2)}$. Thus the CUSUM cannot remain in $E_{(5,\,1)}$ after the next sample, and therefore transitions such as $E_{(5,\,1)} \to E_{(5,\,1)}, E_{(6,\,1)} \to E_{(6,\,1)}, \ldots$ are not possible. Observe that it is also not possible to move from $E_{(5,\,1)}$ to $E_{(5,\,3)}$ because the counter can increment by at most one per sample. That is, it must first increase to 2 before it can get to 3. Hence, there must be a run of two consecutive points first, before a run of three points can occur. Likewise, it is impossible to have transitions such as $E_{(5,\,1)} \to E_{(6,\,3)}$, $E_{(5,\,1)} \to E_{(7,\,3)}$, and so forth. On the other hand, the process can make a transition from state $E_{(5,\,1)}$ to any expanded state capable of taking two consecutive points, such as states $E_{(5,\,2)}$, $E_{(6,\,2)}$, $E_{(7,\,2)}$ and so on. Similarly it is possible to have transitions like $E_{(5,\,2)} \to E_{(5,\,3)}$, $E_{(5,\,2)} \to E_{(6,\,3)}$, $\ldots$, and $E_{(6,\,2)} \to E_{(6,\,3)}$, $E_{(6,\,2)} \to E_{(7,\,3)}$, and so on. Note that the process can also make a transition to the absorbing state $E_H^*$, or it can make a transition back to any other state in set $\mathcal{A}$ at any time. Each realization of this modified CUSUM scheme can thus be also regarded as a random walk over the states:

$$E_0,\ E_1,\ E_2,\ \ldots,\ E_W,\ E_{(W+1,1)},\ E_{(W+1,2)},\ E_{(W+1,3)},\ E_{(W+2,1)},\ \ldots,\ E_{(H-1,3)},\ E_H^*$$

where all the absorbing states have been combined into $E_H^*$. For a modified upper count data CUSUM defined by $S^+_{n+1,c_{n+1}} = \max[0,\ X_{n+1} - K + S^+_{n,c_n}]$, we set:

**(a)** $S^+_{n+1,c_{n+1}} = 0, \qquad$ if $X_{n+1} - K + S^+_{n,c_n} \leq 0,$

**(b)** $S^+_{n+1,c_{n+1}} = H, \qquad$ if $X_{n+1} - K + S^+_{n,c_n} \geq H, \quad$ and

**(c)** $S^+_{n+1,c_{n+1}} = X_{n+1} - K + S^+_{n,c_n}, \qquad$ otherwise.

Let $\{X_n, n = 1, 2, \ldots\}$ be independent random variables representing the observed number of counts in sampling interval $n$. Let $p_r = \Pr(X - K = r)$ and $F_r = \Pr(X - K \leq r)$ be respectively the probability mass function (pmf) and the cumulative distribution function (cdf) of $X$. The transition probabilities are again determined from the probability distribution of $X$ as discussed in Chapter 3, except for some minor changes resulting from the restrictions on the possible (or allowable) transitions.

DEFINITION 4.2. For a modified upper one-sided (count data) CUSUM, these transition probabilities are defined for the following transitions:

1. From state $E_{(i,\, c_n)}$, $(i = 0, 1, \ldots, H-1;\ c_n = 0, 1, 2, 3)$ to state $E_0$.

$$\begin{aligned} P_{(i,\, c_n),\, 0} &= \Pr(E_{(i,\, c_n)} \to E_0) \\ &= \Pr(S^+_{n,0} \leq 0 \mid S^+_{n-1,\, c_{n-1}} = i) \\ &= \Pr(X_n \leq K - i) \\ &= F_{-i} \end{aligned}$$

2. From states $E_i$ $(i = 0, 1, \ldots, W)$ to states $E_{(j,\, c_n)} \in \mathcal{B}_{c_n}$ $(c_n = 1, 2, 3)$.

$$\begin{aligned} P_{i,(j,1)} &= \Pr(E_i \to E_{(j,\,1)}) = \Pr(X = K + j - i) \quad (c_n = 1) \\ &= p_{j-i} \end{aligned}$$

$$P_{i,(j,c_n)} = \Pr(E_i \to E_{(j,\, c_n)}) = 0 \quad (c_n = 2,\, 3)$$

3. From states $E_{(i,\,c_n)} \in \mathcal{B}_{c_n}$ to states $E_{(j,\,c_n)} \in \mathcal{B}_{c_n}$ $(c_n = 1,\, 2,\, 3)$

$$\begin{aligned} P_{(i,c_n),(j,c_n+1)} &= \Pr(E_{(i,\,c_n)} \to E_{(j,\,c_n+1)}) = \Pr(X = K + j - i) \\ &= p_{j-i} \quad (\text{for } c_n = 1,\, 2) \\ P_{(i,c_n),(j,1)} &= \Pr(E_{(i,\,c_n)} \to E_{(j,\,1)}) = 0 \quad (c_n = 1,\, 2,\, 3) \\ P_{(i,c_n),(j,2)} &= \Pr(E_{(i,\,c_n)} \to E_{(j,\,2)}) = 0 \quad (c_n = 2,\, 3) \\ P_{(i,c_n),(j,3)} &= \Pr(E_{(i,\,c_n)} \to E_{(j,\,3)}) = 0 \quad (c_n = 1,\, 3) \end{aligned}$$

4. From states $E_{(i,\,c_n)}$ $(i \neq H,\ c_n = 0, 1, 2, 3)$ to state $E_j$ $(j = 1, 2, \ldots, W)$.

$$P_{(i,c_n),j} = \Pr(E_{(i,\,c_n)} \to E_j) = \Pr(X = K + j - i) = p_{j-i}$$

5. Transitions to the absorbing state $E_H^*$. Note in particular the different notation used for $\Pr(E_{(i,\,3)} \to E_H^*)$.

$$\begin{aligned} P_{(i,0),H} &= \Pr(E_i \to E_H^*) \\ &= \Pr(S_{n,0}^+ \geq H \mid S_{n-1,0}^+ = i) \quad (i = 0,\, 1,\, \ldots,\, W) \\ &= \Pr(X \geq K + H - i) \\ &= 1 - \Pr(X \leq K + H - i - 1) \\ &= 1 - F_{H-i-1} \\ P_{(i,c_n),H} &= \Pr(E_{(i,\,c_n)} \to E_H^*) = \Pr(X \geq K + H - i) \quad (c_n = 1,\, 2) \\ &= 1 - F_{H-i-1} \end{aligned}$$

$$\begin{aligned} P_{(i,3),H} &= \Pr(E_{(i,3)} \to E_H^*) = \Pr(X \geq K + W - i) \\ &= 1 - F_{W-(W+s)} \\ &= 1 - F_{-s}, \quad (s = 1, 2, \ldots, n_B) \end{aligned}$$

6. Transitions from the absorbing state $E_H^*$.

$$\begin{aligned} P_{H,(j,c_n)} &= \Pr(E_H^* \to E_{(j,c_n)}) = 0 \quad (j \neq H); \quad (c_n = 0, \ldots, 3) \\ P_{H,H} &= \Pr(E_H^* \to E_H^*) = 0. \end{aligned}$$

Construction of Extended Transition Probability Matrix

The transition probabilities defined in Definition 4.2 make up the entries of the one-step extended transition probability matrix ($\mathbf{P_E}$) which has the general form:

$$\mathbf{P_E} = \begin{array}{c} \\ 0 \\ 1 \\ \vdots \\ W \\ W{+}1{,}1 \\ W{+}1{,}2 \\ W{+}1{,}3 \\ \vdots \\ H \end{array} \begin{array}{c} \begin{array}{ccccccccc} 0 & 1 & \cdots & W & W{+}1{,}1 & W{+}1{,}2 & W{+}1{,}3 & \cdots & H \end{array} \\ \begin{pmatrix} F_0 & p_1 & \cdots & p_W & p_{W+1} & 0 & 0 & \cdots & 1 - F_{H-1} \\ F_{-1} & p_0 & \cdots & p_{W-1} & p_W & 0 & 0 & \cdots & 1 - F_{H-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ F_{-W} & p_{1-W} & \cdots & p_0 & p_1 & 0 & 0 & \cdots & 1 - F_{H-W-1} \\ F_{-W-1} & p_{-W} & \cdots & p_{-1} & 0 & p_0 & 0 & \cdots & 1 - F_{H-W-2} \\ F_{-W-1} & p_{-W} & \cdots & p_{-1} & 0 & 0 & p_0 & \cdots & 1 - F_{H-W-2} \\ F_{-W-1} & p_{-W} & \cdots & p_{-1} & 0 & 0 & 0 & \cdots & 1 - F_{-1} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \end{array} \tag{4.31}$$

The $(W+1) \times (W+1)$ square matrix in the upper left hand corner (for states 0, 1, 2, ..., $W$) ignores the $\pi_{n,c_n}$ rule, and it is the same as the corresponding part of the transition matrix $\mathbf{P}$ (given in Chapter 3) for the regular count data upward CUSUM. The main difference comes from the entries for the states in set $\mathcal{B}_{c_n}$, $c_n = 1, 2, 3$ defined in Equation (4.27). In particular, note the pattern of zeros and the $2 \times 2$ diagonal entries with $p_0$ on the main diagonal corresponding to states involved in the transitions $(\mathcal{B}_1 \rightarrow \mathcal{B}_2)$ and $(\mathcal{B}_2 \rightarrow \mathcal{B}_3)$ (based on a four-in-a-row criteria). For example, the probability that the process moves from state $E_{(W+1,1)}$ to state $E_{(W+1,2)}$ has the value $p_0$, and that from state $E_{(W+1,2)}$ to state $E_{(W+1,3)}$ is also $p_0$. Similarly, we will have probabilities $\Pr(E_{(W+1,1)} \rightarrow E_{(W+2,2)}) = p_1$, $\Pr(E_{(W+1,2)} \rightarrow E_{(W+2,3)}) = p_1$, $\Pr(E_{(W+2,1)} \rightarrow E_{(W+1,2)}) = p_{-1}$, and so on. Also, all the rows sum to one, and the absorbing states denoted by $E_H^*$ in the last line, still consists of zeros except for the last entry element. A similar transition matrix for the downward (lower) modified count data CUSUM will have the form:

$$
\begin{array}{c}
 \\
0 \\
1 \\
\vdots \\
W \\
W{+}1{,}1 \\
W{+}1{,}2 \\
W{+}1{,}3 \\
\vdots \\
H
\end{array}
\begin{array}{c}
\begin{array}{ccccccccc}
0 & 1 & \cdots & W & W{+}1{,}1 & W{+}1{,}2 & W{+}1{,}3 & \cdots & H
\end{array} \\
\begin{pmatrix}
1-F_{-1} & p_{-1} & \cdots & p_{-W} & p_{-W-1} & 0 & 0 & \cdots & F_{-H} \\
1-F_{0} & p_{0} & \cdots & p_{1-W} & p_{-W} & 0 & 0 & \cdots & F_{1-H} \\
\vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots \\
1-F_{W-1} & p_{W-1} & \cdots & p_{0} & p_{-1} & 0 & 0 & \cdots & F_{W-H} \\
1-F_{W} & p_{W} & \cdots & p_{1} & 0 & p_{0} & 0 & \cdots & F_{W-H+1} \\
1-F_{W} & p_{W} & \cdots & p_{1} & 0 & 0 & p_{0} & \cdots & F_{W-H+1} \\
1-F_{W} & p_{W} & \cdots & p_{1} & 0 & 0 & 0 & \cdots & F_{s-1} \\
\vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots \\
0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 1
\end{pmatrix}
\end{array}.
$$

Computing The Probability of Extremeness

The probability of extremeness $\pi_{j,c_n}$ has been defined earlier in Equation (4.26) for a general $M$-in-a-row criterion, as the probability of having $c_n - 1$ additional points in region $B$ that brings the process to reside in a fixed state (say $j$) also in region $B$, given that the process is currently in region $B$ for the first time. In other words, given that the process is in region $B$ for the first time ($c_n = 1$) in its current sojourn, we would like to know the sum of the probabilities of all paths that would lead the process to a specified state $j$ in region $B$ in $c_n - 1$ consecutive points. This probability corresponds to the sum of the probabilities in the $j$th column of the matrix $\mathbf{P}_{BB}^{c_n-1}$, for each $c_n$ ($c_n = 2, 3, \ldots, M-1$).

It can be seen that the matrix $\mathbf{P}_{BB}$ (of transitions from region $B$ to region $B$) has a lot of zeros in a systematic pattern. Let $\mathbf{P}_{bb}$ be the resulting matrix after stripping $\mathbf{P}_{BB}$ of its zeros. Then, for an extended transition matrix with $n_B$ buffer states following the four-in-a-row rule, $\mathbf{P}_{bb}$ has the form

$$\mathbf{P}_{bb} = \begin{array}{c} \\ W+1 \\ W+2 \\ W+3 \\ \vdots \\ W+n_B \end{array} \begin{array}{c} \begin{array}{ccccc} W+1 & W+2 & W+3 & \dots & W+n_B \end{array} \\ \begin{pmatrix} p_0 & p_1 & p_2 & \cdots & p_{n_B-1} \\ p_{-1} & p_0 & p_1 & \cdots & p_{n_B-2} \\ p_{-2} & p_{-1} & p_0 & \cdots & p_{n_B-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{1-n_B} & p_{2-n_B} & p_{3-n_B} & \cdots & p_0 \end{pmatrix} \end{array}. \tag{4.32}$$

Thus, for example:

1. If $n_B = 1$, then there is only one buffer state $j$ (where $j = W+1$). Therefore:

   $\pi_{W+1,2} = \Pr\left(E_{(W+1,1)} \to E_{(W+1,2)}\right) = p_0$ (using $c_n = 2$), and

   $\pi_{W+1,3} = \Pr\left(E_{(W+1,1)} \to E_{(W+1,2)} \to E_{(W+1,3)}\right) = p_0 p_0$ (with $c_n = 3$).

2. If $n_B = 2$, then $j = W+1, W+2$. Therefore, $\pi_{W+2,2}$ (with $j = W+2$ and $c_n = 2$) for instance, involves the transitions:

   $E_{(W+1,1)} \to E_{(W+2,2)}$ and $E_{(W+2,1)} \to E_{(W+2,2)}$,

   and therefore

   $$\pi_{W+2,2} = \sum_{i=W+1}^{H-1} p_{j-i} = p_1 + p_0.$$

Also, the computation of $\pi_{W+1,3}$ involves the transitions:

$$E_{(W+1,1)} \begin{cases} \nearrow E_{(W+1,2)} \to E_{(W+1,3)} & (p_0 p_0) \\ \searrow E_{(W+2,2)} \to E_{(W+1,3)} & (p_1 p_{-1}) \end{cases}$$

and

$$E_{(W+2,1)} \begin{cases} \nearrow E_{(W+1,2)} \to E_{(W+1,3)} & (p_{-1} p_0) \\ \searrow E_{(W+2,2)} \to E_{(W+1,3)} & (p_0 p_{-1}). \end{cases}$$

Hence

$$\begin{aligned} \pi_{W+1,3} &= \sum_{i=W+1}^{H-1} \sum_{k=W+1}^{H-1} p_{k-i}\, p_{j-k} \\ &= p_0 p_0 + p_1 p_{-1} + p_{-1} p_0 + p_0 p_{-1}, \end{aligned}$$

which is the sum of the first column (with $j = W + 1$) of the sub-matrix

$$\mathbf{P}_{bb}^2 = \begin{pmatrix} p_0 & p_1 \\ p_{-1} & p_0 \end{pmatrix}^2 = \begin{pmatrix} p_0 p_0 + p_1 p_{-1} & p_0 p_1 + p_1 p_0 \\ p_{-1} p_0 + p_0 p_{-1} & p_{-1} p_1 + p_0 p_0 \end{pmatrix}.$$

The probability of extremeness $\pi_{W+2,3}$ can similarly be computed as sum of the second column of $\mathbf{P}_{bb}^2$. Hence, in general, $\pi_{j,c_n}$ can be obtained as

$$\pi_{j,c_n} = \mathbf{1}'_{n_B} \mathbf{P}_{BB}^{c_n - 1} \boldsymbol{e}_j \quad (c_n = 2, 3, \ldots, M-1).$$

Run Length Distribution and Average Run Length

The run length of the CUSUM procedure is a random variable that represents the number of samples until the first out-of-control signal occurs. The run length distribution and the average run length (ARL) of the modified CUSUM scheme can thus be computed easily based on the partitioned modified transition matrix of Equation

(4.26), and using Equations (3.32) and (3.35) of the Markov chain approach of Brook and Evans (discussed in Chapter 3). Examples and computational results dealing with the modified CUSUM procedure are presented in the next chapter. The computations and plots were primarily done using the statistical software S-PLUS [76], and the programs are included in the appendix. The in-control ARLs (without any buffer states) were verified with the CUSARL program developed by Hawkins and Olwell [47]) and now available free from STATLIB.

### Choosing $K, H,\ W,$ and $\pi_\alpha$

It is known that an optimal traditional count data CUSUM can be designed to detect a specified shift in the parameter by appropriately choosing the reference value $K$, and the decision interval $H$. The determination of $K$ and $H$ in the modified CUSUM procedure is similar to what has been discussed in Chapter 3.

The parameter $\pi_\alpha$ is a pre-specified cut-off or threshold value for the probability of extremeness, and it serves as a rejection level. Like the rejection level often used in statistical tests of hypothesis, $\pi_\alpha$ also represents a risk level we want to take. Small $\pi_\alpha$ values, typically in the range 0.01 to 0.05, are chosen.

For some motivation and ideas behind the use of $\pi_\alpha$ and the probability of extremeness, consider for example, a regular Poisson upper CUSUM designed with parameters $K = 3$ and $H = 7$ to detect increases in the process mean from an in-control value of $\mu = 2.4$ to an unacceptatble value of $\mu = 3.7$. The probability that a Poisson

random variable with mean $\mu = 2.4$ is 10 or more is $\Pr(X \geq 10 \,|\mu = 2.4) = 0.0002$ (or 1 in 5000). If any single observed value is 10 or more, then the CUSUM crosses any decision interval $H$ of $X - K = 10 - 3 = 7$, regardless of where the CUSUM was prior to this large value (Hawkins and Olwell [47, pp. 109–110]). A single moderately large ("extreme") data value such as 6, 7, 8, or 9 is not enough to trigger the CUSUM from a zero start. Such extreme data values also result in small transition probabilities, and consequently in small probabilities of extremeness, and their frequent or consecutive occurrence can be an indication of problem in the process. It seems reasonable to us, therefore, that we can design a modified CUSUM by pre-specifying $\pi_\alpha = 0.05$ (say), and requiring that if the second (or the third) consecutive point falling inside region $B$ has a probability of exremeness $(\pi_{j,c_n})$ that is less than or equal to $\pi_\alpha$, then we have an indication of trouble or some instability in the process that need to be investigated.

Assuming $H$, $K$, and $\pi_\alpha$ have been determined, the only remaining choice is determining $W$, the warning level. The following discussion will be confined to modified count data CUSUMS that move in unit integer steps. CUSUMs that involve half- and quarter-integers can be handled similarly by an appropriate scaling. In choosing $W$, it is recommended that the difference $H - W$ be less than the reference value $K$. This means that if, for example, the CUSUM assumes only integer values, then one would typically choose $W$ such that $H - W - 1 \leq K$. Note in this case that if $H - W = 1$, then the modified procedure is identical to the traditional count data

CUSUM (moving in unit integer steps). This is explained further in the next chapter. If the CUSUM involves fractional integers of steps $1/b$, then we will need to have $H - W - (1/b) \leq K$.

A relatively small (or low) value of $W$ can have a potentially great impact (whether good or bad) on the ARL distribution of the CUSUM, while the effect of a relatively large $W$ is minimal. Obviously, we do not want to choose $W$ that is too small since that could result in numerous undesirable false alarms and would distort the interpretation and evaluation of the results from the modified CUSUM process. Based on empirical evidence from trials with different Poisson simulated data, I think choosing a warning level at $W = H/2$ (that is, 50 percent of the decision interval) is small enough for a given $K$ and $H$ values. The other extreme will be to implement just the regular (or traditional) count data CUSUM by choosing $W$ to be just *one step* lower than $H$ because there is no region $B$. The above considerations lead to the plausible choice of $W$ values in the range

$$\frac{H}{2} \leq W \leq H - 1. \tag{4.33}$$

Therefore, for an initial application of the proposed modified CUSUM method, one might find it useful to begin by choosing

$$W = \frac{1}{4}(3H - 2), \tag{4.34}$$

which is half-way between the two endpoints given in relation (4.33). Depending on the specific application, it might be appropriate to move a little above or below this

value. For processes where the count data CUSUM design has $K$ exceeding $H/2$ but less than $H$, it is reasonable to consider $W$ in the range

$$\max(K, H/2) \leq W \leq H - 1.$$

In general, let $\gamma > 0$ be some positive constant. Then consider values of $W$, which can be defined in terms of the difference in the two bounds given in relation (4.33) such that

$$\begin{aligned} W &= \gamma[(H-1) - H/2] \\ &= \frac{\gamma}{2}(H-2). \end{aligned} \tag{4.35}$$

If $W$ takes on the upper and lower bounds given in (4.33), then it is easily seen that the bounds for $\gamma$ will be

$$\frac{H}{H-2} \leq \gamma \leq \frac{2(H-1)}{H-2}. \tag{4.36}$$

Clearly, from relation (4.36), $\gamma$ is undefined for $H = 2$. Simply put, this means here that when $H = 2$ there is only one choice for the warning level, which is $W = 1$ (for a CUSUM moving only in unit integer steps). This default value is equivalent to the regular count data CUSUM, since it does not allow any buffer state in region $B$. Thus, in order to implement a count data CUSUM to use the features described in this procedure (with some buffer states in region $B$), we need $H$ to be 3 or more (assuming the CUSUM moves only in unit integer steps).

It is also easy to see that as $H$ gets larger ($H \to \infty$), we have the bounds

$$1 \leq \gamma \leq 2 \tag{4.37}$$

which, after plugging into relation (4.33), results in the range of $W$ values

$$\frac{H}{2} - 1 \leq W \leq H - 2. \tag{4.38}$$

In other words, a modified count data CUSUM designed for a fixed $K$, a large enough $H$, and one buffer state in region $B$ (i.e. $W = H-2$) will do "no harm" to the process. It could, however, improve the process's capability of detecting an off-aim process. Designing the CUSUM with the low extreme value $W = (H/2) - 1$ on the hand, can have a tremendous impact on the CUSUM process and it can result in some unreasonably low in-control ARLs. It may also be useful to work with a $W$ value that is half-way between the bounds specified in relation (4.38), which is

$$W = \frac{3}{4}(H - 2). \tag{4.39}$$

We can then adjust it little above or below this choice, depending on the specific needs of the process. Hence, for a CUSUM moving only in integer units, we would recommend for choice of $W$, the smallest integer greater than or equal to $(3/4)(H-2)$. There is not much problem for a CUSUM that involves half- and quarter-integers as that allows more flexibility in the choice of $W$.

# CHAPTER 5

## NUMERICAL EXAMPLES

In this chapter, examples of modified CUSUM charts and designs will be presented to illustrate the general procedure. The procedure can be used for both upper and lower one-sided count data CUSUMs with and without the fast initial response (FIR) feature. Three main computer programs (all written as S-Plus functions) are utilized. Program EXT.UP creates the extended transition matrix which is then used by function MOD.UP to compute the probabilities of extremeness, create the modified transition matrix, and calculate the ARLs. A third program MDPLOT.UP plots the modified CUSUM chart for a given sample of data. These functions are included in the appendices.

The Poisson distribution is used here partly out of convenience, and partly because it is the most common discrete distribution for modeling nonconformities. Since a CUSUM requires an explicit and precise statistical model for the observations (Hawkins and Olwell [47]), other discrete distributions could be used where appropriate. All computations are done using warning runs rules based on $M = 4$ consecutive points (four-in-a-row) rules, and with rejection level set at $\pi_\alpha = 0.05$.

### Examples: Based On Sample Data

The following examples illustrate the detailed modified CUSUM procedure for

simulated Poisson data. We will design an upper modified CUSUM for detecting an increase (a persistent upward shift) in the process mean from an acceptable value of $\mu_0 = 3.80$ to an unacceptable level of $\mu_1 = 4.21$. The modified CUSUM will be designed with parameters $K = 4$, $H = 6$, $\pi_\alpha = 0.05$, and three different warning levels $W = 3, 4$, and 5 based on the $M = 4$ warning runs rules. The reference value for this Poisson CUSUM is obtained from $K = (\mu_1 - \mu_0)/(\log \mu_1 - \log \mu_0) = (4.21 - 3.8)/(\log(4.21) - \log(3.8)) = 4.0015$, which is then rounded to $K = 4$. We will consider the following data set. The first twelve values (1, 5, 2, 2, 6, 6, 3, 4, 2, 2, 5, 8) were generated from a Poisson with $\mu = 3.8$, and the remaining twelve values (4, 4, 3, 4, 8, 5, 6, 6, 6, 5, 6, 6) were generated from a Poisson with mean $\mu = 4.21$.

With No Buffer State

EXAMPLE 5.1. The first example shows the modified upper CUSUM design with $\pi_\alpha = 0.05$, $K = 4$, $H = 6$, and $W = 5$. As this CUSUM moves in integer steps, it means that there is no buffer state in region $B$, since $n_B = H - W - 1 = 6 - 5 - 1 = 0$. Therefore, the design in this example is equivalent to a standard Poisson CUSUM with parameters $K = 4$, $H = 6$. Figure 3 (of page 118) shows a plot of the modified upper CUSUM chart with parameters $K = 4$, $H = 6$, and warning level $W = 5$ (which is identical to a regular one-sided Poisson count data CUSUM chart with parameters $K = 4$ and $H = 6$).

If we supply all these design parameters to the MOD.UP function, it returns the following:

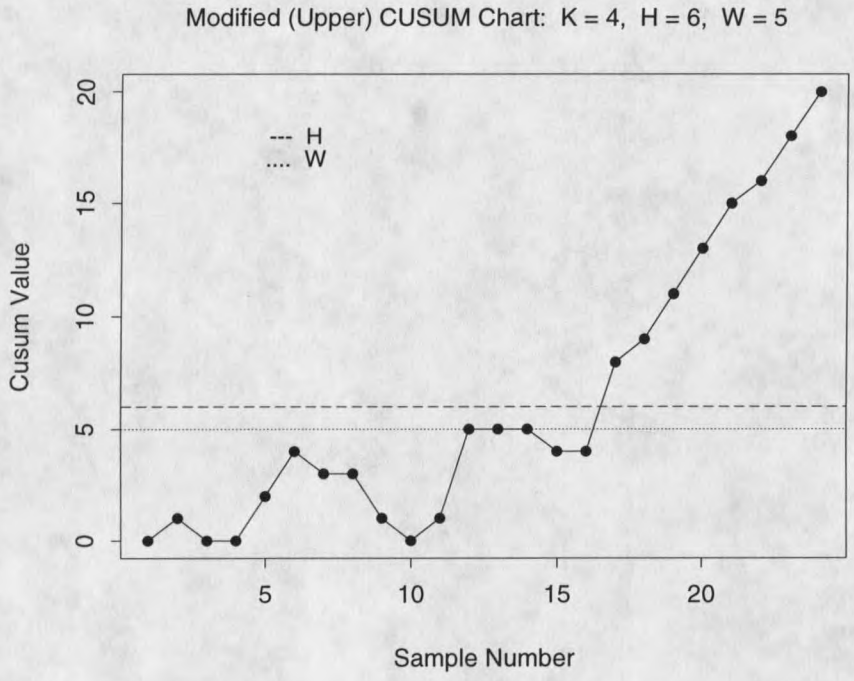


**Figure 3**: A Modified (Upper) CUSUM chart for simulated Poisson data with an in-control mean of $\mu = 3.8$ for Example 5.1. The design parameters are $K = 4$, $H = 6$, and $W = 5$ (illustrating the case where there is no buffer state in region $B$). Hence, this design is equivalent to a regular Poisson CUSUM chart with parameters $K = 4$, and $H = 6$. Out-of-control signal occurs at sample 17 with type H-ABS.

```
> MOD.up(mu0=3.8, mu1=4.21,K=4, H=6, W=5, alpha=.05, head-start=F)
$mat2.id:
NULL

$remove.st:
NULL

$P.EXT:
        0       1       2       3       4       5       6
0 0.66784 0.14771 0.09355 0.05079 0.02412 0.01019 0.00580
1 0.47348 0.19436 0.14771 0.09355 0.05079 0.02412 0.01598
2 0.26890 0.20459 0.19436 0.14771 0.09355 0.05079 0.04011
3 0.10738 0.16152 0.20459 0.19436 0.14771 0.09355 0.09089
4 0.02237 0.08501 0.16152 0.20459 0.19436 0.14771 0.18444
5 0.00000 0.02237 0.08501 0.16152 0.20459 0.19436 0.33216
6 0.00000 0.00000 0.00000 0.00000 0.00000 0.00000 1.00000
attr($P.EXT, "Ext. Up. Cusum"):
```

```
[1] "Transition Matrix P.EXT: 7 x 7"

$PP.m0:
        0       1       2       3       4       5       6
0 0.66784 0.14771 0.09355 0.05079 0.02412 0.01019 0.00580
1 0.47348 0.19436 0.14771 0.09355 0.05079 0.02412 0.01598
2 0.26890 0.20459 0.19436 0.14771 0.09355 0.05079 0.04011
3 0.10738 0.16152 0.20459 0.19436 0.14771 0.09355 0.09089
4 0.02237 0.08501 0.16152 0.20459 0.19436 0.14771 0.18444
5 0.00000 0.02237 0.08501 0.16152 0.20459 0.19436 0.33216
6 0.00000 0.00000 0.00000 0.00000 0.00000 0.00000 1.00000
attr($PP.m0, "Ext. Up. Cusum"):
[1] "Transition Matrix P.EXT: 7 x 7"

$PP.m1:
        0       1       2       3       4       5       6
0 0.58788 0.16362 0.11481 0.06905 0.03634 0.01700 0.01130
1 0.39355 0.19433 0.16362 0.11481 0.06905 0.03634 0.02829
2 0.20892 0.18464 0.19433 0.16362 0.11481 0.06905 0.06463
3 0.07735 0.13157 0.18464 0.19433 0.16362 0.11481 0.13368
4 0.01485 0.06250 0.13157 0.18464 0.19433 0.16362 0.24849
5 0.00000 0.01485 0.06250 0.13157 0.18464 0.19433 0.41212
6 0.00000 0.00000 0.00000 0.00000 0.00000 0.00000 1.00000
attr($PP.m1, "Ext. Up. Cusum"):
[1] "Transition Matrix P.EXT: 7 x 7"

$mu0:
[1] 3.8

$arls.m0:
     0
 21.32

$ARLs.occ:
       mm arls.m1
[1,] 4.21   12.09

Warning messages:
  No Region B, and hence, no Buffer state in: prog9D.up.fun(mu0 = 3.8,
mu1 = 4.21, K = 4, H = 6, W = 5, alpha = 0.05, headstart = F)
```

There are seven states in the Markov chain, and both the extended and modified transition matrices for the in-control mean ($\mu_0 = 3.8$) are exactly the same, indicating

| Sample | $X_i$ | Cusum | $E_i$ | $E_j$ | $C_i$ | $C_j$ | Probs | Counter | $\pi_{j,c_n}$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0.6678 | c = 0 | · |
| 2 | 5 | 1 | 0 | 1 | 1 | 2 | 0.1477 | c = 0 | · |
| 3 | 2 | 0 | 1 | 0 | 2 | 1 | 0.4735 | c = 0 | · |
| 4 | 2 | 0 | 0 | 0 | 1 | 1 | 0.6678 | c = 0 | · |
| 5 | 6 | 2 | 0 | 2 | 1 | 3 | 0.0936 | c = 0 | · |
| 6 | 6 | 4 | 2 | 4 | 3 | 5 | 0.0936 | c = 0 | · |
| 7 | 3 | 3 | 4 | 3 | 5 | 4 | 0.2046 | c = 0 | · |
| 8 | 4 | 3 | 3 | 3 | 4 | 4 | 0.1944 | c = 0 | · |
| 9 | 2 | 1 | 3 | 1 | 4 | 2 | 0.1615 | c = 0 | · |
| 10 | 2 | 0 | 1 | 0 | 2 | 1 | 0.4735 | c = 0 | · |
| 11 | 5 | 1 | 0 | 1 | 1 | 2 | 0.1477 | c = 0 | · |
| 12 | 8 | 5 | 1 | 5 | 2 | 6 | 0.0241 | c = 0 | · |
| 13 | 4 | 5 | 5 | 5 | 6 | 6 | 0.1944 | c = 0 | · |
| 14 | 4 | 5 | 5 | 5 | 6 | 6 | 0.1944 | c = 0 | · |
| 15 | 3 | 4 | 5 | 4 | 6 | 5 | 0.2046 | c = 0 | · |
| 16 | 4 | 4 | 4 | 4 | 5 | 5 | 0.1944 | c = 0 | · |
| 17 | 8 | 8 | 4 | 6 | 5 | 7 | 0.1844 | H-ABS * | · |

**Table 1**: Sample output for a modified upper CUSUM design with $K = 4$, $H = 6$, and $W = 5$, based on Poisson data with in-control mean $\mu = 3.8$. This output for Example 5.1 illustrates the case of no buffer state in region $B$.

that no state has been removed from the transient sets to the absorbing set. A notification is subsequently issued that there is no buffer state in this design. The transition matrix based on the out-of-control mean of $\mu_1 = 4.21$ is also provided. The in-control ARL is given as $\text{ARL}_0 = 21.32$ and the out-of-control ARL (for $\mu_1 = 4.21$) is $\text{ARL}_1 = 12.09$.

The entries of the sample process output Table 1 and the probability transition matrix in Table 2 were computed based on Poisson probabilities, using the general transition matrix for four-in-a-row design given in (4.31) and the transition probabilities defined in Definition 4.2. It can be seen, for instance, that the transitional probability of going from state $E_i = 0$ to $E_j = 0$ is found from row $C_j = 1$ and

| $i \backslash j$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| 0 | 0.6678 | 0.1477 | 0.0936 | 0.0508 | 0.0241 | 0.0102 | 0.0058 |
| 1 | 0.4735 | 0.1944 | 0.1477 | 0.0936 | 0.0508 | 0.0241 | 0.0160 |
| 2 | 0.2689 | 0.2046 | 0.1944 | 0.1477 | 0.0936 | 0.0508 | 0.0401 |
| 3 | 0.1074 | 0.1615 | 0.2046 | 0.1944 | 0.1477 | 0.0936 | 0.0909 |
| 4 | 0.0224 | 0.0850 | 0.1615 | 0.2046 | 0.1944 | 0.1477 | 0.1844 |
| 5 | 0.0000 | 0.0224 | 0.0850 | 0.1615 | 0.2046 | 0.1944 | 0.3322 |
| 6 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 1.0000 |

**Table 2**: Extended Transition Matrix for the modified upper CUSUM design in Example 5.1. The design has $K = 4$, $W = 5$, and $H = 6$, based on Poisson data with in-control mean $\mu = 3.8$. This matrix is identical to a transition matrix for a regular Poisson CUSUM with $K = 4$ and $H = 6$ (because there is no buffer state in region $B$).

column $C_i = 1$ of the transition matrix to be 0.6678. With an initial value of $S_0^+ = 0$, a reference value of $K = 4$, and Poisson mean of $\mu_0 = 3.8$, this transition probability is computed simply as

$$\begin{aligned}
\Pr(E_0 \to E_0) &= \Pr(S_{1,0}^+ = 0 | \, S_0^+ = 0) \\
&= \Pr(X_1 \leq 4 \,|\, X \sim Poisson(3.8)) \\
&= \sum_{i=0}^{4} \frac{(3.8)^i e^{-3.8}}{i\,!} \\
&\approx 0.6678.
\end{aligned}$$

The column labeled `Cusum` in the sample process output shown in Table 1 (page 120) gives the CUSUM values, `Probs` gives the observed transition probabilities, `Counter` gives the counter values at each sample, and $\pi_{j,c_n}$ gives the probability of extremeness which evidently is non-applicable in this particular situation because there is no region $B$. This CUSUM takes on only positive integer values, and so it will

take just one step to make a transition from the warning state $E_W$ (which is $E_5$) to the absorbing state $E_H^* = E_H$ (which is $E_6$). Also, since there is no buffer state, the counter value remains zero ($c_n = 0$) until the CUSUM crosses the decision interval $H$ at **sample 17**. This is the usual out-of-control situation, referred to as H-ABS in the output.

The transitions from state $E_i$ to state $E_j$, and their corresponding row (or column) numbers in the transition matrix (given by $C_i$ and $C_j$) are also shown in the sample process output. Remember that 0 is not a valid row or column number of a matrix. For instance, the observed transition probability at sample 12 of Table 1 (where the CUSUM moves from state 1 to state 5) is 0.0241. This value was obtained as

$$\begin{aligned}
\Pr(E_1 \to E_5) &= \Pr(S^+_{12,0} = 5 \mid S^+_{11,0} = 1) \\
&= \Pr(S^+_{12,0} = 5 \mid S^+_{11,0} = 1) \\
&= \Pr(X_{12} = 4 + 5 - 1 \mid X \sim Poisson(3.8)) \\
&= \frac{(3.8)^8 e^{-3.8}}{8!} \\
&\approx 0.0241.
\end{aligned}$$

With One Buffer State

EXAMPLE 5.2. If the previous modified upper CUSUM scheme (for $K = 4$, $H = 6$, and $\pi_\alpha = 0.05$) is designed with $W = 4$, it will allow one buffer state (which is state $E_5$) in region $B$ (so that $n_B = 1$). The function MOD.UP now produces:

```
> MOD.up(mu0=3.8, mu1=4.21,K=4, H=6, W=4, alpha=.05, head-start=F)
$mat2.id:
```

```
     id.state     pii.23 indx.c23
[1,] "5,2"    "0.19436"  "7"
[2,] "5,3"    "0.037776" "8"

$remove.st:
   pii.23 indx.c23
 0.037776        8

$P.EXT:
        0       1       2       3       4      5,1    5,2      5,3       6
0   0.66784 0.14771 0.09355 0.05079 0.02412 0.01019 0.00000 0.00000 0.00580
1   0.47348 0.19436 0.14771 0.09355 0.05079 0.02412 0.00000 0.00000 0.01598
2   0.26890 0.20459 0.19436 0.14771 0.09355 0.05079 0.00000 0.00000 0.04011
3   0.10738 0.16152 0.20459 0.19436 0.14771 0.09355 0.00000 0.00000 0.09089
4   0.02237 0.08501 0.16152 0.20459 0.19436 0.14771 0.00000 0.00000 0.18444
5,1 0.00000 0.02237 0.08501 0.16152 0.20459 0.00000 0.19436 0.00000 0.33216
5,2 0.00000 0.02237 0.08501 0.16152 0.20459 0.00000 0.00000 0.19436 0.33216
5,3 0.00000 0.02237 0.08501 0.16152 0.20459 0.00000 0.00000 0.00000 0.52652
6   0.00000 0.00000 0.00000 0.00000 0.00000 0.00000 0.00000 0.00000 1.00000
attr($P.EXT, "Ext. Up. Cusum"):
[1] "Transition Matrix P.EXT: 9 x 9"

$PP.m0:
          0       1       2       3       4      5,1     5,2      Abs
  0   0.66784 0.14771 0.09355 0.05079 0.02412 0.01019 0.00000 0.00580
  1   0.47348 0.19436 0.14771 0.09355 0.05079 0.02412 0.00000 0.01598
  2   0.26890 0.20459 0.19436 0.14771 0.09355 0.05079 0.00000 0.04011
  3   0.10738 0.16152 0.20459 0.19436 0.14771 0.09355 0.00000 0.09089
  4   0.02237 0.08501 0.16152 0.20459 0.19436 0.14771 0.00000 0.18444
  5,1 0.00000 0.02237 0.08501 0.16152 0.20459 0.00000 0.19436 0.33216
  5,2 0.00000 0.02237 0.08501 0.16152 0.20459 0.00000 0.00000 0.52652
 Abs  0.00000 0.00000 0.00000 0.00000 0.00000 0.00000 0.00000 1.00000
attr($PP.m0, "Mod. Upper Cusum"):
[1] "Matrix PP.m0:  8 x 8"

$PP.m1:
          0       1       2       3       4      5,1     5,2      Abs
  0   0.58788 0.16362 0.11481 0.06905 0.03634 0.01700 0.00000 0.01130
  1   0.39355 0.19433 0.16362 0.11481 0.06905 0.03634 0.00000 0.02829
  2   0.20892 0.18464 0.19433 0.16362 0.11481 0.06905 0.00000 0.06463
  3   0.07735 0.13157 0.18464 0.19433 0.16362 0.11481 0.00000 0.13368
  4   0.01485 0.06250 0.13157 0.18464 0.19433 0.16362 0.00000 0.24849
  5,1 0.00000 0.01485 0.06250 0.13157 0.18464 0.00000 0.19433 0.41212
  5,2 0.00000 0.01485 0.06250 0.13157 0.18464 0.00000 0.00000 0.60645
 Abs  0.00000 0.00000 0.00000 0.00000 0.00000 0.00000 0.00000 1.00000
attr($PP.m1, "Mod. Upper Cusum"):
```

```
[1] "Matrix PP.m1:  8 x 8"

$mu0:
[1] 3.8

$arls.m0:
     0
 21.03

$ARLs.occ:
       mm arls.m1
[1,] 4.21   11.97
```

The expanded states are now identified as (5,1), (5,2), (5,3) and the program computes the probabilities of extremeness for states $E_{(5,2)}$ and $E_{(5,3)}$ to be $\pi_{5,2} = 0.19436$ and $\pi_{5,2} = 0.037776$ respectively. The respective row (or column) numbers of these two states in the extended transition matrix are also given as 7 and 8. The program indicates that state $E_{(5,3)}$ corresponding to row (or column) number 8 has been removed from the transient set of states in the extended matrix to the absorbing set. Thus, the extended matrix based on the in-control mean of $\mu_0 = 3.8$ is now different from the modified matrix. More specifically, the extended matrix is a $9 \times 9$ square while the modified matrix is $8 \times 8$ in size. The in-control average run length is now $\text{ARL}_0 = 21.03$ and the out-of-control ARL is $\text{ARL}_1 = 11.97$. The modified matrix for the out-of-control mean at $\mu_1 = 4.21$ must also be of size $8 \times 8$, since it is based on the same number of states to facilitate comparisons between the two ARLs. Due to the use of one buffer state, the in-control ARL drops from 21.32 to 21.03 (about 1.36 %) while the out-of-control ARL drops from 12.09 to 11.97 (about 1.0%). Although there is little change in both $\text{ARL}_0$ and $\text{ARL}_1$, it will be seen later

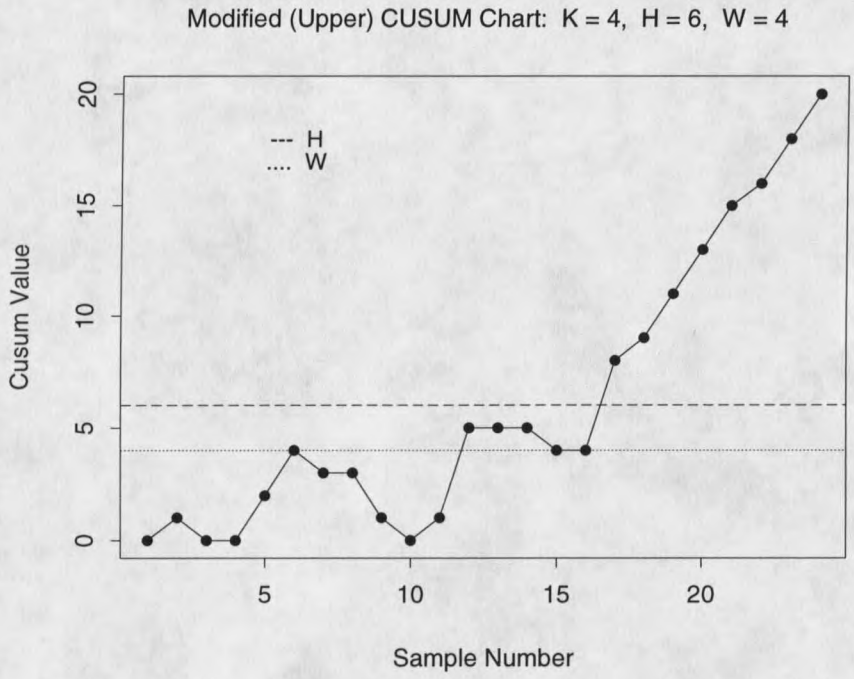


**Figure 4**: A Modified (Upper) CUSUM chart for simulated Poisson data with in-control mean $\mu = 3.8$. The design parameters are $K = 4$, $H = 6$, $W = 4$ and $\pi_\alpha = 0.05$. This is part of Example 5.2 and it illustrates the case of one buffer state in region $B$. Out-of-control signal occurs at sample 14, with type A-ABS.

that this modest change in the out-of-control ARL is still useful in helping to quickly detect out-of-control conditions.

The corresponding modified upper CUSUM chart for Example 5.2 is given in Figure 4. It is clear from Figure 4 that because the warning level is now state $E_W = E_4 = 4$, this allows some CUSUM points (here three in all) to fall inside the single buffer state. The sample process output displayed in Table 3 shows that this one buffer state in region $B$ results in three expanded states ($\mathcal{B}_1 = E_{(5,1)}$, $\mathcal{B}_2 = E_{(5,2)}$, and $\mathcal{B}_3 = E_{(5,3)}$) which are entered as `(5,1)`, `(5,2)`, and `(5,3)` in the transition

| Sample | $X_i$ | Cusum | $E_i$ | $E_j$ | $C_i$ | $C_j$ | Probs | Counter | $\pi_{j,c_n}$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0.6678 | c = 0 | · |
| 2 | 5 | 1 | 0 | 1 | 1 | 2 | 0.1477 | c = 0 | · |
| 3 | 2 | 0 | 1 | 0 | 2 | 1 | 0.4735 | c = 0 | · |
| 4 | 2 | 0 | 0 | 0 | 1 | 1 | 0.6678 | c = 0 | · |
| 5 | 6 | 2 | 0 | 2 | 1 | 3 | 0.0936 | c = 0 | · |
| 6 | 6 | 4 | 2 | 4 | 3 | 5 | 0.0936 | c = 0 | · |
| 7 | 3 | 3 | 4 | 3 | 5 | 4 | 0.2046 | c = 0 | · |
| 8 | 4 | 3 | 3 | 3 | 4 | 4 | 0.1944 | c = 0 | · |
| 9 | 2 | 1 | 3 | 1 | 4 | 2 | 0.1615 | c = 0 | · |
| 10 | 2 | 0 | 1 | 0 | 2 | 1 | 0.4735 | c = 0 | · |
| 11 | 5 | 1 | 0 | 1 | 1 | 2 | 0.1477 | c = 0 | · |
| 12 | 8 | 5 | 1 | 5,1 | 2 | 6 | 0.0241 | c = 1 | · |
| 13 | 4 | 5 | 5,1 | 5,2 | 6 | 7 | 0.1944 | c = 2 | $\pi_{5,2} = 0.1944$ |
| 14 | 4 | 5 | 5,2 | 5,3 | 7 | 8 | 0.1944 | c = 3* | $\pi_{5,3} = 0.0378$ |
| 15 | · | · | · | · | · | · | · | · | |
| 16 | · | · | · | · | · | · | · | · | |
| 17 | · | · | · | · | · | · | · | · | |

**Table 3**: Sample output for a modified upper CUSUM design with $K = 4$, $H = 6$, $W = 4$, and $\pi_\alpha = 0.05$, based on Poisson data with in-control mean $\mu = 3.8$. This is sample output for Example 5.2, illustrating the case with one buffer state in region $B$. The out-of-control signal is given at sample 14, where $\pi_{5,3} \leq 0.05$.

matrix (Table 4).

The rows (or columns) of the three expanded states $\{E_{(5,1)}, E_{(5,2)}, E_{(5,3)}\}$ in the extended transition matrix are given by $\mathcal{C}_1 = 6$, $\mathcal{C}_2 = 7$, and $\mathcal{C}_3 = 8$ respectively. The counter value shown in Table 3 accordingly registers 1, 2, and 3 respectively for these three consecutive points in region $B$.

The process stays in region $A$ for most part until it finally sojourns in region $B$ for the first time at sample 12 (in state $E_{(5,1)}$). Although this first entrance into region $B$ was caused by a very extreme observation ($X_{12} = 8$), it merely triggers a warning since our interest is in the next consecutive points thereafter, while ignoring

| $i\backslash j$ | 0 | 1 | 2 | 3 | 4 | 5,1 | 5,2 | 5,3 | 6 |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 0.668 | 0.148 | 0.094 | 0.051 | 0.024 | 0.010 | 0.000 | 0.000 | 0.006 |
| 1 | 0.474 | 0.194 | 0.148 | 0.094 | 0.051 | 0.024 | 0.000 | 0.000 | 0.016 |
| 2 | 0.269 | 0.205 | 0.194 | 0.148 | 0.094 | 0.051 | 0.000 | 0.000 | 0.040 |
| 3 | 0.107 | 0.162 | 0.205 | 0.194 | 0.148 | 0.094 | 0.000 | 0.000 | 0.091 |
| 4 | 0.022 | 0.085 | 0.162 | 0.205 | 0.194 | 0.148 | 0.000 | 0.000 | 0.184 |
| 5,1 | 0.000 | 0.022 | 0.085 | 0.162 | 0.205 | 0.000 | 0.194 | 0.000 | 0.332 |
| 5,2 | 0.000 | 0.022 | 0.085 | 0.162 | 0.205 | 0.000 | 0.000 | 0.194 | 0.332 |
| 5,3 | 0.000 | 0.022 | 0.085 | 0.162 | 0.205 | 0.000 | 0.000 | 0.000 | 0.526 |
| 6 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 |

**Table 4**: Extended Transition Matrix for the modified upper CUSUM design with $K = 4$, $H = 6$, $W = 4$, and $\pi_\alpha = 0.05$ for Poisson data with in-control mean $\mu = 3.8$ in Example 5.2. This design has one buffer state (state $E_5$) in region $B$.

| $i\backslash j$ | 0 | 1 | 2 | 3 | 4 | 5,1 | 5,2 | Abs |
|---|---|---|---|---|---|---|---|---|
| 0 | 0.66784 | 0.14771 | 0.09355 | 0.05079 | 0.02412 | 0.01019 | 0.00000 | 0.00580 |
| 1 | 0.47348 | 0.19436 | 0.14771 | 0.09355 | 0.05079 | 0.02412 | 0.00000 | 0.01598 |
| 2 | 0.26890 | 0.20459 | 0.19436 | 0.14771 | 0.09355 | 0.05079 | 0.00000 | 0.04011 |
| 3 | 0.10738 | 0.16152 | 0.20459 | 0.19436 | 0.14771 | 0.09355 | 0.00000 | 0.09089 |
| 4 | 0.02237 | 0.08501 | 0.16152 | 0.20459 | 0.19436 | 0.14771 | 0.00000 | 0.18444 |
| 5,1 | 0.00000 | 0.02237 | 0.08501 | 0.16152 | 0.20459 | 0.00000 | 0.19436 | 0.33216 |
| 5,2 | 0.00000 | 0.0223 | 0.08501 | 0.16152 | 0.20459 | 0.00000 | 0.00000 | 0.52652 |
| Abs | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 1.00000 |

**Table 5**: Modified Transition Matrix for the modified upper CUSUM design with $K = 4$, $H = 6$, $W = 4$, and $\pi_\alpha = 0.05$ for Poisson data with in-control mean $\mu = 3.8$ in Example 5.2. This design has one buffer state (state $E_5$) in region $B$.

isolated individual outliers. At sample 13, the process stays in state 5 in region $B$ (that is, state $E_{(5,2)}$) for the second consecutive time. Consequently, the $\pi_{j,c_n}$ rule is implemented at sample 13 but it fails, since $\pi_{5,2} = 0.1944 > 0.05$. The process continues to stay in region $B$ (in state $E_{(5,3)}$) for three consecutive points, and $\pi_{5,3}$ is calculated as 0.0378, which is again compared with $\pi_\alpha = 0.05$. Since $\pi_{5,3} \leq 0.05$, the $\pi_{j,c_n}$ rule has been violated, and so the process is declared out-of-control at **sample 14** with type A-ABS. Thus, even though the reduction in the out-of-control ARL was very modest, we are still able to detect the "out-of-control" condition three samples earlier than the regular Poisson CUSUM in this particular case.

Observe that since there is only one buffer state, the probability of extremeness $\pi_{5,2}$ can be computed easily as:

$$\begin{aligned}
\pi_{5,2} &= \Pr\left(E_{(5,1)} \to E_{(5,2)}\right) \\
&= \Pr\left(S^+_{13,2} = 5 \mid S^+_{12,1} = 5\right) \\
&= \Pr(X_{13} = 4 + 5 - 5 \mid X \sim Poisson(3.8)) \\
&= \frac{(3.8)^4 e^{-3.8}}{4!} \\
&\approx 0.1944.
\end{aligned}$$

The probability of extremeness for the third consecutive point in region $B$ can be

similarly computed as

$$\begin{aligned}
\pi_{5,3} &= \Pr\left(E_{(5,1)} \to E_{(5,2)} \to E_{(5,3)}\right) \\
&= \Pr\left(S^+_{14,3} = 5 \mid S^+_{13,2} = 5\right) \times \Pr\left(S^+_{13,2} = 5 \mid S^+_{12,1} = 5\right) \\
&= \Pr(X_{14} = 4 \mid X \sim Poi(3.8)) \times \Pr(X_{13} = 4 \mid X \sim Poi(3.8)) \\
&= \left(\frac{(3.8)^4 e^{-3.8}}{4!}\right)^2 \\
&\approx 0.0378.
\end{aligned}$$

With Two Buffer States

EXAMPLE 5.3. To conclude the current example with the sample data began in Example 5.1 (where $\mu_0 = 3.8$, $K = 4$, and $H = 6$), we will design the modified CUSUM with warning level $W = 3$. The average run lengths for this design are given as $\text{ARL}_0 = 20.43$ and $\text{ARL}_1 = 11.74$. Figure 5 shows the modified CUSUM chart for this example.

It can be seen that this design allows two ($n_B = 2$) buffer states (which are states $E_4$ and $E_5$) in region $B$, and therefore $(M-1)n_B = (3)2 = 6$ expanded states $(E_{(4,1)}, E_{(5,1)}, E_{(4,2)}, E_{(5,2)}, E_{(4,3)}, E_{(5,3)})$ in region $B$. None of probabilities of extremeness computed by the PROG.MOD function shows significance at the $\pi_\alpha = 0.05$ level, as shown in Table 6. These six expanded states have rows (or columns) in the extended transition matrix given by ($\mathcal{C}_1 = 5, 8$), ($\mathcal{C}_2 = 6, 8$), and ($\mathcal{C}_3 = 7, 10$) as shown in Table 7 and Table 8 of page 132. Because there is no violation of the $\pi_{j,c_n}$ rule, the modified transition matrix based on $\mu_0 = 3.8$ is exactly the same as the

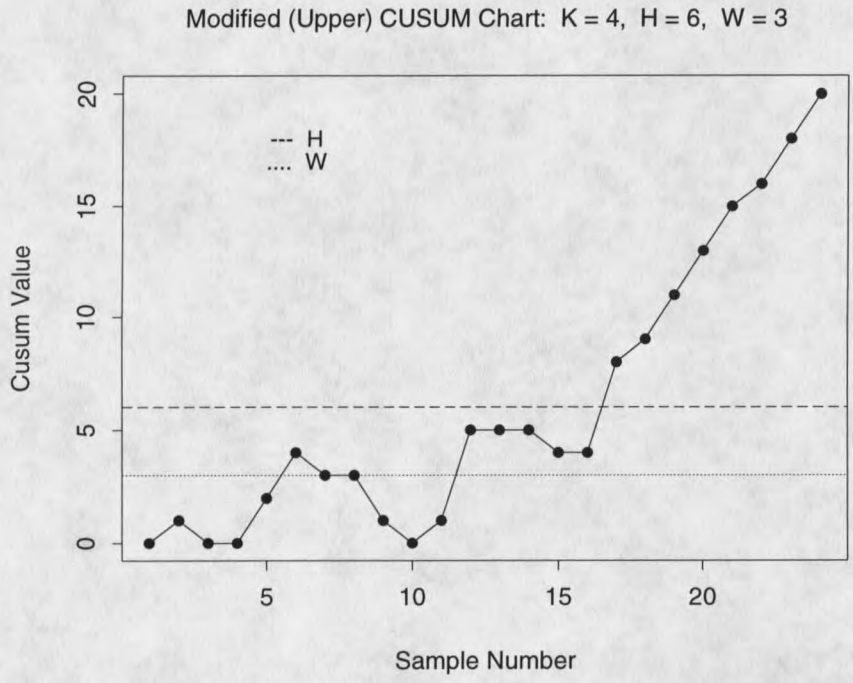


**Figure 5**: A Modified (Upper) CUSUM chart for simulated Poisson data with in-control mean $\mu = 3.8$. The design parameters are $K = 4$, $H = 6$, and $W = 3$ and it illustrates Example 5.3 with a four-in-a-row warning runs rule. The process signals out of control (with type C-ABS) at sample 15.

| State ID | $\pi_{j,\,c}$ | State Index |
|---|---|---|
| 4,2 | 0.398950 | 6 |
| 5,2 | 0.342070 | 9 |
| 4,3 | 0.147524 | 7 |
| 5,3 | 0.125414 | 10 |

**Table 6**: Output showing probabilities of extremeness for the design in Example 5.3, with $\mu_0 = 3.8$, $K = 4$, $H = 6$ and $W = 3$.

| Sample | $X_i$ | Cusum | $E_i$ | $E_j$ | $C_i$ | $C_j$ | Probs | Counter | $\pi_{j,c_n}$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0.6678 | c = 0 | · |
| 2 | 5 | 1 | 0 | 1 | 1 | 2 | 0.1477 | c = 0 | · |
| 3 | 2 | 0 | 1 | 0 | 2 | 1 | 0.4735 | c = 0 | · |
| 4 | 2 | 0 | 0 | 0 | 1 | 1 | 0.6678 | c = 0 | · |
| 5 | 6 | 2 | 0 | 2 | 1 | 3 | 0.0936 | c = 0 | · |
| 6 | 6 | 4 | 2 | 4 | 3 | 5 | 0.0936 | c = 1 | · |
| 7 | 3 | 3 | 4 | 3 | 5 | 4 | 0.2046 | c = 0 | · |
| 8 | 4 | 3 | 3 | 3 | 4 | 4 | 0.1944 | c = 0 | · |
| 9 | 2 | 1 | 3 | 1 | 4 | 2 | 0.1615 | c = 0 | · |
| 10 | 2 | 0 | 1 | 0 | 2 | 1 | 0.4735 | c = 0 | · |
| 11 | 5 | 1 | 0 | 1 | 1 | 2 | 0.1477 | c = 0 | · |
| 12 | 8 | 5 | 1 | 5,1 | 2 | 8 | 0.0241 | c = 1 | · |
| 13 | 4 | 5 | 5,1 | 5,2 | 8 | 9 | 0.1944 | c = 2 | $\pi_{5,2} = 0.3421$ |
| 14 | 4 | 5 | 5,2 | 5,3 | 9 | 10 | 0.1944 | c = 3 | $\pi_{5,3} = 0.1254$ |
| 15 | 3 | 4 | 5,3 | 4,1 | 10 | 11 | 0.7311 | C-ABS * | |
| 16 | · | · | · | · | · | · | · | · | |
| 17 | · | · | · | · | · | · | · | · | |

**Table 7**: Sample output for a modified upper CUSUM design with $K = 4$, $H = 6$, $W = 3$, and $\pi_\alpha = 0.05$, based on Poisson data with in-control mean $\mu = 3.8$. This output is part of Example 5.3, demonstrating the case with two buffer states. Out-of-control signal of type C-ABS occurs at sample 15.

extended transition matrix (Table 7), and hence any comparable transition matrix based on an out-of-control mean will also have the same size (which is $11 \times 11$ in this case).

A closer look at the sample process output in Table 7 reveals an interesting component of the modified CUSUM procedure. The process starts and remains in region $A$ until sample 6 when it makes a brief first sojourn into state 4 of region $B$ (that is, $E_{(4,1)}$). It quickly reverts back back to region $A$ and stays there until an extreme observation ($X_{12} = 8$) sends it back into state $E_{(5,1)}$ in region $B$. It stays in region $B$ for the second and third consecutive points but the probability of

| | 0 | 1 | 2 | 3 | 4,1 | 4,2 | 4,3 | 5,1 | 5,2 | 5,3 | 6 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0.668 | 0.148 | 0.094 | 0.051 | 0.024 | 0.000 | 0.000 | 0.010 | 0.000 | 0.000 | 0.006 |
| 1 | 0.474 | 0.194 | 0.148 | 0.094 | 0.051 | 0.000 | 0.000 | 0.024 | 0.000 | 0.000 | 0.016 |
| 2 | 0.269 | 0.205 | 0.194 | 0.148 | 0.094 | 0.000 | 0.000 | 0.051 | 0.000 | 0.000 | 0.040 |
| 3 | 0.107 | 0.162 | 0.205 | 0.194 | 0.148 | 0.000 | 0.000 | 0.094 | 0.000 | 0.000 | 0.091 |
| 4,1 | 0.022 | 0.085 | 0.162 | 0.205 | 0.000 | 0.194 | 0.000 | 0.000 | 0.148 | 0.000 | 0.184 |
| 4,2 | 0.022 | 0.085 | 0.162 | 0.205 | 0.000 | 0.000 | 0.194 | 0.000 | 0.000 | 0.148 | 0.184 |
| 4,3 | 0.022 | 0.085 | 0.162 | 0.205 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.526 |
| 5,1 | 0.000 | 0.022 | 0.085 | 0.162 | 0.000 | 0.205 | 0.000 | 0.000 | 0.194 | 0.000 | 0.332 |
| 5,2 | 0.000 | 0.022 | 0.085 | 0.162 | 0.000 | 0.000 | 0.205 | 0.000 | 0.000 | 0.194 | 0.332 |
| 5,3 | 0.000 | 0.022 | 0.085 | 0.162 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.731 |
| 6 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 |

**Table 8**: Transition table for the modified upper CUSUM design with $K = 4$, $H = 6$, $W = 3$ and $\pi_\alpha = 0.05$, for Poisson data with in-control mean $\mu_0 = 3.8$. This is part of Example 5.3 and it demonstrates the case of two buffer states.

extremeness ($\pi_{j,c_n}$) rule fails on both occasions. It stays in region $B$ for the fourth consecutive point but the persistency in the shift is immediately detected by the four-in-a-row warning runs rule criteria. Consequently, the process signals out-of-control with type (C-ABS) at **sample 15** when the maximum counter value ($M = 4$) is reached (for a four-in-a-row violation).

It is should be noted again that although there is very modest change in the ARLs, (the in-control ARL drops from 21.32 to 20.43 which is about about 4%, and the out-of-control ARL goes from 12.09 to 11.74 which is about 3%), we are still able to sacrifice this modest reduction in the in-control ARL to gain quicker detection of the out-of-control signal at **sample 15**, compared to sample 17 for the regular CUSUM.

Example: Changing $\pi_\alpha$ Values

EXAMPLE 5.4. Suppose the random variable $X$, representing counts of some sort, has a Poisson distribution with mean 3.5. A modified upper CUSUM scheme at $\pi_\alpha = 0.05$ will be used to examine the ARLs when the process shifts to unacceptably high count levels of 4.2, 5.6, 7.0, 8.4, 9.8, and 11.9 corresponding to 20%, 60%, 100%, 140%, 180%, and 240% of the specified in-control mean value. We will use $K = 7$, $H = 5$, and four warning levels $W = 4, 3, 2, 1$ to illustrate the changes in the ARLs for different number of buffer states $n_B$. The probabilities of extremeness $\pi_{j,c_n}$'s for this example are displayed in Table 9, together with the identification of those states chosen as alpha-expanded states. Obviously, the $\pi_{j,c_n}$'s are not applicable when $n_B = 0$ ($W = 4$), since there is no region $B$ in this case.

The ARLs for the designs in Example 5.4 are displayed in Table 10. The first ARL column (for $\mu = 3.5$) are the in-control ARL values for the different values of $W$. The table shows a general decreasing trend as the shift size increases, for all buffer states ($n_B$). Also, the ARLs decrease for decreasing values of $W$ (that is, the ARLs decrease as the number of buffer states increase). Notice from this example that the ARLs at $\mu = 3.5$ and $\mu = 4.2$ do not follow the general decreasing trend down the columns (with increasing $n_B$). Looking back at Table 9 reveals that although two alpha-expanded states (states (3,3) and (4,3)) were chosen (that is, removed from the transient set of the extended transition matrix to the absorbing set) using

| $W$ | $n_B$ | State ID | $\pi_{j,c}$ | Row ID | States Removed |
|---|---|---|---|---|---|
| 4 | 0 | NA | NA | NA | NA |
| 3 | 1 | (4,2) | 0.03855*** | 6 | $E_{(4,2)}$ |
| 3 | 1 | (4,3) | 0.00149*** | 6 | $E_{(4,3)}$ |
| 2 | 2 | (3,2) | 0.11565 | 5 | |
| 2 | 2 | (3,3) | 0.00873*** | 6 | $E_{(3,3)}$ |
| 2 | 2 | (4,2) | 0.05542** | 8 | |
| 2 | 2 | (4,3) | 0.00409*** | 9 | $E_{(4,3)}$ |
| 1 | 3 | (2,2) | 0.24782 | 6 | |
| 1 | 3 | (2,3) | 0.02796*** | 6 | $E_{(2,3)}$ |
| 1 | 3 | (3,2) | 0.13252 | 6 | |
| 1 | 3 | (3,3) | 0.01407*** | 6 | $E_{(3,3)}$ |
| 1 | 3 | (4,2) | 0.06198* | 6 | |
| 1 | 3 | (4,3) | 0.00625*** | 6 | $E_{(4,3)}$ |

$***$Significant at $\pi_\alpha = 0.05$; $**$Significant at $\pi_\alpha = 0.06$; $*$Significant at $\pi_\alpha = 0.07$; NA: not applicable.

**Table 9**: Comparison of Probabilities of Extremeness and Selected Alpha-Expanded States for Example 5.4, using $K = 7$, $H = 5$, $\pi_\alpha = 0.05$, and $W = 4, 3, 2, 1$.

| | | Out-of-Control Means ($\mu$) | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $n_B$ | $W$ | **3.5** | **4.2** | **5.6** | **7.0** | **8.4** | **9.8** | **11.9** |
| 0 | 4 | 2682.65 | 465.37 | 36.95 | 8.47 | 3.83 | 2.43 | 1.63 |
| 1 | 3 | 2473.25 | 422.36 | 34.18 | 8.10 | 3.74 | 2.39 | 1.62 |
| 2 | 2 | 2567.04 | 429.95 | 33.71 | 8.03 | 3.74 | 2.40 | 1.62 |
| 3 | 1 | 2138.29 | 334.50 | 27.30 | 7.19 | 3.56 | 2.36 | 1.62 |
| $*2*$ | 2 | 2213.98 | 370.06 | 30.69 | 7.62 | 3.61 | 2.35 | 1.61 |
| $*3*$ | 1 | 1820.85 | 291.83 | 25.37 | 6.86 | 3.43 | 2.29 | 1.59 |

$*2*$ for $n_B = 2$ $(W = 2)$ when $\pi_\alpha = 0.06$; $*3*$ for $n_B = 3$ $(W = 1)$ when $\pi_\alpha = 0.07$.

**Table 10**: ARLs at in-control mean ($\mu = 3.5$) and several shifts for Example 5.4, using a modified upper CUSUM with $K = 7$, $H = 5$, $\pi_\alpha = 0.05$, and $W = 4, 3, 2, 1$.

$n_B = 2$ buffer states, these states have very different effects on the CUSUM than the two states ((4,2) and (4,3)) selected when $n_B = 1$. Apparently, when $n_B = 2$, both state (3,3) with probability of extremeness $\pi_{3,3} = 0.00873$ and state (4,3) with $\pi_{3,3} = 0.00409$ have higher selection preferences (because they are states for three-consecutive points) than state (4,2) with $\pi_{4,2} = 0.05542$. Removing the two states with three-consecutive points therefore seems to have smaller effect on the CUSUM than removing state (4,2). If the rejection level was increased to $\pi_\alpha = 0.06$, that will cause state (4,2) also to be removed. This development will further reduce the number of transient states in the extended matrix for $n_B = 2$ buffer states, and consequently reduce the average run lengths, to make them follow the decreasing trend down the columns. The ARL values using $\pi_\alpha = 0.06$ and $\pi_\alpha = 0.07$ are also provided in the last row of Table 10.

Finally, recall from Chapter 4 that our general recommendation regarding the choice of $W$ (for a CUSUM moving only in integer units) is $H/2 < W < H - 1$, which yields $2.5 < W < 4$ for this particular example. Our specific recommendation for the choice of $W$ is then given by the smallest integer greater than or equal to $W = (3/4)(H - 2) = (3/4)(5 - 2) = 2.25$, which gives $W = 3$. Thus, an appropriate design for this example may be to use $K = 7$, $H = 5$, and $W = 3$ (for $n_B = 1$ buffer state) which results in about 9.2% reduction in the ARL at a shift of $\mu = 4.2$ (and a simultaneous reduction in the in-control ARL of 7.8%), when the modified CUSUM is designed with a rejection level of $\pi_\alpha = 0.05$.

This example has also shown that when the $\pi_\alpha$ level is increased from 0.05 to 0.06, the corresponding ARLs decrease (because the size of the modified transition matrix becomes smaller with the removal of alpha-expanded states from the transient sets to the absorbing set). Note further that the corresponding reduction in the ARL at the mean shift level of $\mu = 4.2$ is about 7.7% using $n_B = 2$ ($W = 2$), and 28% for using $n_B = 3$ ($W = 1$). Also, the larger shifts seem more difficult to detect as all the out-of-control ARLs converge around a value of about 1.62.

# CHAPTER 6

# COMPUTATIONAL RESULTS

The first section of this chapter will summarize and evaluate the modified CUSUM procedure by comparing it with the regular Poisson CUSUM with and without the fast initial response feature (FIR), with the aid of several plots and ARL tables. The second section will be concerned with design implementation and guidelines as to how one can use the several tables and plots to design and implement a modified CUSUM scheme. Design examples will also be provided.

## Evaluation and Summary of Results

The ARL performance of the modified CUSUM (based on the Poisson distribution) are now compared with the regular Poisson CUSUM, with and without the fast initial response feature (Lucas and Crosier [71]), by means of plots and tables, and through an example. Table 11 of page 138 shows the ARL's (without FIR) for various shift levels ($\mu$) when the in-control mean is $\mu_0 = 2.0$, 4.0, 6.0, 8.0, and 10.0, with $H = 7$ and $H = 10$. A more extensive set of ARL tables are provided in the APPENDIX B (ARLs without FIR) and APPENDIX C (ARLs with FIR).

The out-of-control mean levels ($\mu$) in the tables are given in multiples of the in-control means $\mu_0$, so that the actual out-of-control mean $\mu_1$ is found as $\mu_1 = \mu_0 \times \mu$. The in-control ARLs are given in column 5 of the ARL tables (or column

| $\pi_\alpha = 0.05$, $H = 7, 10$, $W = 3, 6$ (for $n_B = 3$ buffer states) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Parameters | | | | Mean $\mu$ as a multiple of $\mu_0$ (That is, $\mu = \mu_1/\mu_0$) | | | | | | | |
| $K$ | $H$ | $W$ | $\mu_0$ | 1.0 | 1.2 | 1.4 | 1.6 | 1.8 | 2.0 | 2.5 | 3.0 |
| 1 | 7 | 3 | 2 | 6.87 | 5.31 | 4.35 | 3.7 | 3.22 | 2.87 | 2.28 | 1.91 |
| 2 | 7 | 3 | 2 | 25.77 | 12.43 | 8.04 | 5.98 | 4.79 | 4.02 | 2.9 | 2.3 |
| 3 | 7 | 3 | 2 | 500.12 | 93.2 | 30.62 | 15 | 9.38 | 6.76 | 4.03 | 2.92 |
| 5 | 7 | 3 | 2 | * | 21939.31 | 3262.27 | 683.63 | 192.21 | 70.03 | 13.88 | 6.15 |
| 7 | 7 | 3 | 2 | * | * | * | * | * | * | * | * |
| 1 | 7 | 3 | 4 | 2.85 | 2.37 | 2.04 | 1.8 | 1.62 | 1.48 | 1.22 | 1.09 |
| 2 | 7 | 3 | 4 | 4.02 | 3.06 | 2.5 | 2.13 | 1.87 | 1.67 | 1.34 | 1.16 |
| 3 | 7 | 3 | 4 | 6.76 | 4.37 | 3.27 | 2.64 | 2.23 | 1.94 | 1.5 | 1.25 |
| 5 | 7 | 3 | 4 | 88.06 | 19.89 | 8.59 | 5.21 | 3.75 | 2.94 | 1.97 | 1.52 |
| 7 | 7 | 3 | 4 | 4606.48 | 445.76 | 77.12 | 22.73 | 10.19 | 6.03 | 2.92 | 1.98 |
| 1 | 7 | 3 | 6 | 1.91 | 1.62 | 1.41 | 1.26 | 1.16 | 1.09 | 1.02 | 1 |
| 2 | 7 | 3 | 6 | 2.29 | 1.87 | 1.59 | 1.4 | 1.25 | 1.16 | 1.04 | 1.01 |
| 3 | 7 | 3 | 6 | 2.91 | 2.23 | 1.83 | 1.57 | 1.38 | 1.25 | 1.07 | 1.02 |
| 5 | 7 | 3 | 6 | 6.5 | 3.75 | 2.67 | 2.1 | 1.75 | 1.52 | 1.19 | 1.06 |
| 7 | 7 | 3 | 6 | 42.5 | 10.61 | 5.09 | 3.3 | 2.46 | 1.99 | 1.4 | 1.15 |
| 1 | 7 | 3 | 8 | 1.47 | 1.26 | 1.13 | 1.06 | 1.03 | 1.01 | 1 | 1 |
| 2 | 7 | 3 | 8 | 1.67 | 1.39 | 1.22 | 1.11 | 1.05 | 1.02 | 1 | 1 |
| 3 | 7 | 3 | 8 | 1.93 | 1.56 | 1.33 | 1.18 | 1.09 | 1.04 | 1 | 1 |
| 5 | 7 | 3 | 8 | 2.93 | 2.1 | 1.67 | 1.4 | 1.23 | 1.13 | 1.02 | 1 |
| 7 | 7 | 3 | 8 | 6.21 | 3.3 | 2.28 | 1.78 | 1.48 | 1.29 | 1.07 | 1.01 |
| 1 | 7 | 3 | 10 | 1.22 | 1.09 | 1.03 | 1.01 | 1 | 1 | 1 | 1 |
| 2 | 7 | 3 | 10 | 1.34 | 1.16 | 1.06 | 1.02 | 1.01 | 1 | 1 | 1 |
| 3 | 7 | 3 | 10 | 1.49 | 1.25 | 1.11 | 1.04 | 1.02 | 1 | 1 | 1 |
| 5 | 7 | 3 | 10 | 1.96 | 1.52 | 1.27 | 1.13 | 1.06 | 1.02 | 1 | 1 |
| 7 | 7 | 3 | 10 | 2.94 | 1.99 | 1.54 | 1.29 | 1.15 | 1.07 | 1.01 | 1 |
| 1 | 10 | 6 | 2 | 9.87 | 7.45 | 6.02 | 5.06 | 4.38 | 3.87 | 3.03 | 2.52 |
| 2 | 10 | 6 | 2 | 51.39 | 19.62 | 11.78 | 8.48 | 6.67 | 5.52 | 3.9 | 3.05 |
| 3 | 10 | 6 | 2 | 5003.8 | 377.11 | 68.88 | 25.48 | 14.22 | 9.75 | 5.53 | 3.92 |
| 5 | 10 | 6 | 2 | * | * | * | * | * | * | * | * |
| 7 | 10 | 6 | 2 | * | * | * | * | * | * | * | * |
| 1 | 10 | 6 | 4 | 3.85 | 3.16 | 2.7 | 2.37 | 2.13 | 1.95 | 1.6 | 1.35 |
| 2 | 10 | 6 | 4 | 5.52 | 4.13 | 3.33 | 2.82 | 2.46 | 2.2 | 1.76 | 1.47 |
| 3 | 10 | 6 | 4 | 9.75 | 6.04 | 4.43 | 3.52 | 2.95 | 2.55 | 1.96 | 1.61 |
| 5 | 10 | 6 | 4 | 346.04 | 39.2 | 13.11 | 7.35 | 5.11 | 3.95 | 2.58 | 1.97 |
| 7 | 10 | 6 | 4 | * | 3805 | 299.66 | 50.65 | 16.95 | 8.86 | 3.92 | 2.6 |
| 1 | 10 | 6 | 6 | 2.52 | 2.13 | 1.87 | 1.66 | 1.49 | 1.35 | 1.12 | 1.03 |
| 2 | 10 | 6 | 6 | 3.04 | 2.46 | 2.09 | 1.84 | 1.64 | 1.47 | 1.19 | 1.05 |
| 3 | 10 | 6 | 6 | 3.91 | 2.95 | 2.4 | 2.05 | 1.8 | 1.61 | 1.27 | 1.09 |
| 5 | 10 | 6 | 6 | 9.44 | 5.11 | 3.55 | 2.76 | 2.29 | 1.97 | 1.49 | 1.21 |
| 7 | 10 | 6 | 6 | 120.05 | 17.47 | 7.19 | 4.45 | 3.26 | 2.6 | 1.79 | 1.39 |
| 1 | 10 | 6 | 8 | 1.94 | 1.66 | 1.44 | 1.27 | 1.15 | 1.08 | 1.01 | 1 |
| 2 | 10 | 6 | 8 | 2.19 | 1.83 | 1.58 | 1.38 | 1.23 | 1.13 | 1.02 | 1 |
| 3 | 10 | 6 | 8 | 2.54 | 2.05 | 1.73 | 1.5 | 1.32 | 1.19 | 1.04 | 1.01 |
| 5 | 10 | 6 | 8 | 3.93 | 2.76 | 2.17 | 1.81 | 1.56 | 1.38 | 1.11 | 1.02 |
| 7 | 10 | 6 | 8 | 9.05 | 4.45 | 3 | 2.31 | 1.9 | 1.63 | 1.22 | 1.06 |
| 1 | 10 | 6 | 10 | 1.6 | 1.35 | 1.18 | 1.08 | 1.03 | 1.01 | 1 | 1 |
| 2 | 10 | 6 | 10 | 1.76 | 1.47 | 1.26 | 1.13 | 1.05 | 1.02 | 1 | 1 |
| 3 | 10 | 6 | 10 | 1.95 | 1.61 | 1.36 | 1.19 | 1.09 | 1.04 | 1 | 1 |
| 5 | 10 | 6 | 10 | 2.57 | 1.97 | 1.62 | 1.38 | 1.21 | 1.11 | 1.01 | 1 |
| 7 | 10 | 6 | 10 | 3.95 | 2.6 | 1.99 | 1.63 | 1.39 | 1.22 | 1.04 | 1 |

*Value greater than $10^5$.

**Table 11**: Average run lengths for modified CUSUM (No FIR), for five in-control means $\mu_0 = 2.0$, 4.0, 6.0, 8.0, and 10.0; $H = 7.0, 10.0$; $W = 3, 6$; $K =$ 1,2, 3, 5, 7; and several shift levels.

6 if using ARL tables with FIR in APPENDIX C) where $\mu = 1.0$. From Table 11 (page 138) and the other tables provided in the appendices, it can be seen that for almost all shift levels, the modified CUSUM with number of buffer states $n_B = 1, 2, 3$ have noticeably shorter ARLs than the regular Poisson CUSUM. This general trend is clearly illustrated by the various ARL plots provided in APPENDIX A for different combinations of the design parameters. The modified CUSUM achieves these reductions in the out-of-control ARLs by sacrificing modest decreases in the in-control ARLs. This is consistent with our overall objective with the cattle feedlot problem, which is to achieve quick detection of morbid animals, and for that matter, we will be more willing to make a Type I error (rejecting the hypothesis that the process is in-control, when it is in fact true) than a Type II error (failing to reject the hypothesis that the process is out-of-control when in fact it is true).

The tables and figures provided in the appendices also show that, in general, the ARLs decrease with increasing number of buffer states (and therefore with decreasing values of the warning level parameter $W$). The modified CUSUM evidently has shorter ARLs (both in-control and out-of-control) than the regular Poisson CUSUM. Furthermore, the percentage reduction in the in-control ARL are very modest, in many cases. An example will now be considered to further illustrate and present the comparison of the modified CUSUM with the regular Poisson CUSUM (with and without the use of FIR feature).

Example: Using The FIR Feature

EXAMPLE 6.1. Consider a modified upper CUSUM designed with parameters $H = 7$, $K = 7$, and suppose we consider warning levels $W = 6, 5, 4$, and 3 corresponding to $n_B = 0, 1, 2, 3$ buffer states respectively, at $\pi_\alpha = 0.05$ level. For in-control mean $\mu_0 = 4$, the ARLs at several shift levels are summarized in Table 12 (without FIR) and Table 13 (with FIR) for comparison. The fast initial response is implemented with a head-start value of $S_0^+ = H/2 = 3.5$. Observe again that our criteria $(H/2) < W < H - 1$ implies that $W$ lies in the range $3.5 < W < 6$ for the CUSUM that moves only in integer values, so that our recommended warning level becomes the smallest integer greater than or equal to $(3/4)(H-2) = (3/4)(7-2) = 3.75$ (that is, $W = 4$). Using $W = 4$, $K = 7$, and $H = 7$ we find at $\pi_\alpha = 0.05$ level that the ARL at the shift of $\mu_1 = 4.8$ is 515.63, which is about 8.7% reduction. The in-control ARL correspondingly reduces by about 9.7% as in Table 12. Figure 6 shows a plot of the ARLs, and it shows (together with Table 12) that the ARLs for the modified CUSUM are shorter than those for the regular CUSUM (given by the first row of ARLs in Table 12), with the ARLs converging for larger shifts.

Table 12 further illustrates that generally, the percentage reductions in $\text{ARL}_0$ are modest. Overall, greater reductions in all ARLs are achieved with increasing number of buffer states (and therefore decreasing values of $W$) over different shift levels. Table 13 shows that the ARLs for the fast initial response CUSUM are shorter than both

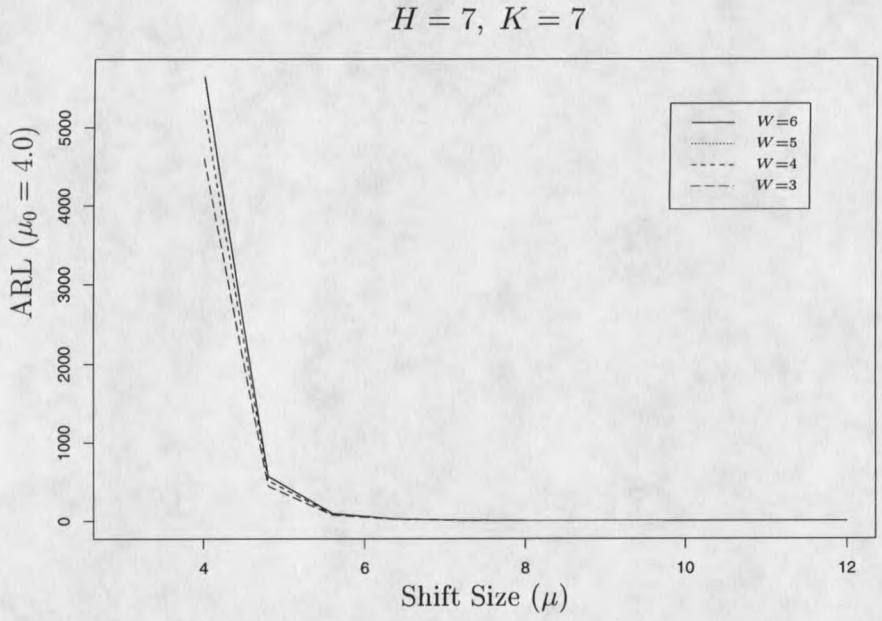


**Figure 6**: Decreasing trend of ARLs (with no FIR) for Modified (Upper) CUSUM scheme used in Example 6.1.

| | | Out-of-Control Means ($\mu$) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n_B$ | $W$ | **4.00** | **4.80** | **5.60** | **6.40** | **7.20** | **8.00** | **8.80** | **9.60** | **10.40** | **11.20** | **12.00** |
| 0 | 6 | 5647.6 | 571.35 | 95.46 | 26.33 | 11.19 | 6.4 | 4.37 | 3.32 | 2.70 | 2.28 | 1.99 |
| 1 | 5 | 5611.72 | 565.74 | 94.52 | 26.14 | 11.13 | 6.38 | 4.37 | 3.32 | 2.69 | 2.28 | 1.99 |
| | | (.98) | (1.0) | (.72) | (.54) | (.31) | | | | | | |
| 2 | 4 | 5214.6 | 515.63 | 87.14 | 24.71 | 10.75 | 6.24 | 4.31 | 3.29 | 2.68 | 2.28 | 1.99 |
| | | (9.7) | (8.7) | (6.2) | (4.0) | (2.5) | (1.4) | (.90) | (.74) | | | |
| 3 | 3 | 4606.48 | 445.76 | 77.12 | 22.73 | 10.19 | 6.03 | 4.21 | 3.25 | 2.66 | 2.27 | 1.98 |
| | | (21.9) | (19.2) | (13.7) | (9.0) | (5.8) | (3.7) | (2.1) | (1.5) | (.44) | (.50) | |

Values in Parentheses are the percentage reductions due to the use of buffer states.

**Table 12**: Percentage reductions in ARLs (with NO FIR) of a modified upper CUSUM scheme with in-control mean $\mu_0 = 4.0$, $H = 7$, $K = 7$, $\pi_\alpha = 0.05$, $W = 6, 5, 4, 3$, and several shift levels for Example 6.1. The first column of ARLs are the in-control ARLs values.

the regular CUSUM and the modified CUSUM. This due to the fact that the FIR CUSUM starts from $S_0^+ = 3.5$ which is mid-way (one-half) of the decision interval $H = 7$. The use of FIR however is recommended (only) for processes that restart after the occurrence of a shift, otherwise it can greatly distort the in-control ARL (Lucas and Crosier [71]; Montgomery [77]). Since the use of FIR starts the CUSUM from a rational number involving a half-integer, we have the criteria $bH < bW < bH - 1$, where $b = 2$. For this example, the range of values of $W$ then becomes $(H/2) < W < (bH - 1)/b = 3.5 < W < 6.75$, and the recommended value is obtained to be $(3/4b)(bH - 2) = (3/8)(14 - 2) = 3.75$. We will note here that although the FIR head-start value is 3.5, this impact soon zeros off, and the CUSUM begins to move in integer steps (since $K = 7$ is an integer).

From the summarized ARL tables (Table 12, Table 13, and Figure 6), it can be seen that in general, the modified CUSUM has better ARL performance than the regular Poisson CUSUM, while the FIR does consistently give shorter ARLs than the modified CUSUM. Similar trends prevail from the other figures and tables provided in the appendices. Using the modified procedure together with a FIR feature greatly enhances (improves the ARL performance of) the fast initial response CUSUM. The larger shift levels, as usual, seem more difficult to detect as all out-of-control ARLs tend to converge. Also, in general, the smaller the $\pi_\alpha$ value, the larger the ARLs are likely to become (as illustrated earlier in Example 5.4).

It can also be seen from the tables and figures that for fixed $H$ and $W$, the ARLs

| | | Out-of-Control Means ($\mu$) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n_B$ | $W$ | **4.00** | **4.80** | **5.60** | **6.40** | **7.20** | **8.00** | **8.80** | **9.60** | **10.40** | **11.20** | **12.00** |
| 0 | 6.5 | 5647.60 | 571.35 | 95.46 | 26.33 | 11.19 | 6.40 | 4.37 | 3.32 | 2.70 | 2.28 | 1.99 |
| | | *5624.42* | *560.45* | *89.62* | *22.82* | *8.84* | *4.74* | *3.13* | *2.35* | *1.91* | *1.63* | *1.45* |
| 1 | 6.0 | 5647.60 | 571.35 | 95.46 | 26.33 | 11.19 | 6.40 | 4.37 | 3.32 | 2.70 | 2.28 | 1.99 |
| | | *5624.28* | *560.33* | *89.54* | *22.77* | *8.84* | *4.73* | *3.12* | *2.34* | *1.90* | *1.63* | *1.45* |
| 2 | 5.5 | 5611.72 | 565.74 | 94.52 | 26.14 | 11.13 | 6.38 | 4.37 | 3.32 | 2.69 | 2.28 | 1.99 |
| | | *5588.57* | *554.85* | *88.68* | *22.63* | *8.81* | *4.72* | *3.12* | *2.34* | *1.90* | *1.63* | *1.45* |
| 3 | 5.0 | 5611.72 | 565.74 | 94.52 | 26.14 | 11.13 | 6.38 | 4.37 | 3.32 | 2.69 | 2.28 | 1.99 |
| | | *5587.08* | *553.76* | *88.01* | *22.25* | *8.61* | *4.62* | *3.07* | *2.32* | *1.89* | *1.63* | *1.45* |
| 4 | 4.5 | 5214.60 | 515.63 | 87.14 | 24.71 | 10.75 | 6.24 | 4.31 | 3.29 | 2.68 | 2.28 | 1.99 |
| | | *5191.83* | *504.91* | *81.37* | *21.23* | *8.44* | *4.60* | *3.07* | *2.32* | *1.89* | *1.63* | *1.45* |
| 5 | 4.0 | 5214.60 | 515.63 | 87.14 | 24.71 | 10.75 | 6.24 | 4.31 | 3.29 | 2.68 | 2.28 | 1.99 |
| | | *5189.66* | *503.35* | *80.43* | *20.70* | *8.16* | *4.45* | *3.00* | *2.29* | *1.88* | *1.62* | *1.44* |

Slanted values correspond to the use of FIR with $S_0^+ = H/2 = 3.5$.

**Table 13**: ARLs (With and Without) the FIR Feature for Poisson modified upper CUSUM. Italicized values correspond to the use of FIR with $S_0^+ = H/2 = 3.5$. The in-control mean is $\mu = 4.0$, with several shift values (for Example 6.1). Parameters are $H = 7$, $K = 7$, $\pi_\alpha = 0.05$, and various $W$ values.

(both in-control and out-of-control) generally increase with increasing $K$. A similar increasing trend of ARLs is seen by increasing $H$ for fixed $K$ and $W$.

## Design Implementation

This section discusses how one can utilize the various tables and figures in the appendices to implement a modified CUSUM design. The values of the parameter $K$ need not be restricted to an integer. In fact, the S-PLUS programs provided in APPENDIX D allows one to use half- and quarter-integer values as well for the reference value $K$ or the initial value $S_0^+$. It can also implement the fast initial response feature and can easily be converted into a lower CUSUM. Also, depending on

the needs of a particular application, one can choose a different $\pi_\alpha$ value (although the default value of 0.05 seems to work well for our purposes). It should also be mentioned that the programs have been written in such a way that a vector of out-of-control means (or shift levels) can be specified as input.

A modified CUSUM can be designed by appropriately choosing $K$, $H$, and $W$ values at a given in-control mean $\mu_0$ to achieve a desired ARL performance. The usual standard in CUSUM schemes is that the in-control ARL will generally be as high as possible and the out-of-control ARL will be as small as possible (eg. Lucas [68]). For count data CUSUMs moving in integer units, it is not often possible to achieve exactly a pre-specified in-control ARL because of the graininess due to the limited set of possible CUSUM values for CUSUMs based on discrete distributions (Hawkins and Olwell [47]).

For Poisson CUSUM, the parameter $K$ is easily determined for a known in-control mean and an out-of-control mean representing the shift that we want to detect quickly. That is, $K$ is usually chosen to be close to the ratio $K = (\mu_1 - \mu_0)/(\ln \mu_1 - \ln \mu_0)$. The value of $K$ will typically be rounded to the nearest integer if the CUSUM only moves in integer steps. The figures in APPENDIX A can also aid in choosing $K$ for a particular application, by looking at the $K$ values in the plots close to the desired in-control mean $\mu_0$ and $\text{ARL}_0$.

After $K$ has been selected, we can use the figures and tables in the appendices to choose the decision interval $H$ and warning level $W$ to give ARL properties close

to the desired levels. Alternatively, after $H$ has been selected, we can determine $W$ using $W = (3/4)(H - 2)$, and use table look-up to determine what the approximate $ARL_0$ and $ARL_1$ will be for this choice of parameters.

Design Example

As an illustration of how to use this procedure, suppose we want to design a modified CUSUM for a process with an in-control mean $\mu_0 = 4.0$, and it is desired to have a 'large' ARL of at least 300 when the process is operating near its in-control mean value. For the cattle feedlot problem, an ARL of over 300 will mean $(300/6 = 50)$ over 50 days of continuous sampling, which is still "quite" high, since we are dealing with animals and we cannot let them go on for that long without some intervention (on health grounds). We might also be interested in a shift of about 1.2 times the specified in-control mean (since it is essential to detect shifts as quickly as possible).

With $\mu_0 = 4.0$ and $\mu_1 = 4.8$, the value of $K$ works out to be 4.39, which can be suitably rounded to $K = 4$ or $K = 5$ in this case. If the CUSUM can assume half- or quarter-integers, we can also consider $K = 4.25, 4.5$, and 4.75 as well. The figures (and tables) for ARLs with $\mu_0 = 4.0$ in APPENDIX A suggest several possible combinations, but the one closest to the desired in-control ARL of 300 is $K = 5$ and $H = 10$. Note that we do not have tables for $K = 4$, but it is possible to determine the ARLs at $K = 4$ using interpolation between $K = 3$ and $K = 5$. The figures for the ARL plots with $\mu_0 = 4.0$ also suggest that for the combination $K = 5$ and

$H = 10$, a lower ARL at a shift of $\mu_1 = 4 \times 1.2 = 4.8$ can be achieved with $n_B = 3$ (corresponding to $W = 6$). Alternatively, with $H = 10$, our criteria for determining $W$ gives a range of $5 < W < 9$, with an estimate of $W = (3/4)(10 - 2) = 6$ (which corresponds to $n_B = 3$ buffer states). Looking up the tables with the combinations $K = 5$, $H = 10$, and $W = 6$ at $\pi_\alpha = 0.05$ then gives us the $\text{ARL}_0 = 346.04$ (for in-control) and $\text{ARL}_1(\mu_1 = 4.8) = 39.2$ (for out-of-control at a shift level of 4.8). This combination will therefore provide the quickest detection for a shift of size 1.2 times the in-control mean value of 4.0, and it represents a reduction in the out-of-control ARL of about 9.1% from the corresponding ARL (of 43.11) of the regular Poisson CUSUM at the same shift of 4.8.

# CHAPTER 7

# APPLICATION

In this chapter we will briefly outline how the modified CUSUM procedure can be implemented for the cattle feedlot problem discussed in Chapter 1. The underlying process distribution may be assumed to be either Poisson or negative binomial during an actual implementation. Because of the highly over-dispersed nature of the counts (involving many high counts), the state space of the count data CUSUM will be divided into several groups or sub-sets of some fixed mesh-size $\Delta$. The transition states will then be based on these *transition blocks* (from one group to another group). This will help reduce the number of states in the resulting Markov chain.

## Computing the Group Transition Matrix

Since the Lower CUSUM can be plotted upwards (by converting to the upper CUSUM) the following procedure will assume an upper one-sided CUSUM $S_n^+$, and the Markov chain approach discussed in chapters 3 and 4 will again be used. Let the state space of the traditional count data CUSUM be $\{0, 1, 2, 3, \ldots, \mathcal{H}\}$, where $\mathcal{H}$ (which is some large integer value) is the "raw" decision interval. Let $\mathcal{K}$ and $\mathcal{W}$ represent the raw reference value and the raw warning level. Define the corresponding "scaled" parameters by $H = \mathcal{H}/\Delta$, $K = \mathcal{K}/\Delta$, and $W = \mathcal{W}/\Delta$, where $\Delta$ is some fixed integer value.

A Markov chain with $\mathcal{H}$ states can be very large, and the corresponding transition matrix will also be large, with entries $\{P_{rs}\}$ denoting the transition probabilities from state $E_r$ to state $E_s$ $(r, s = 0, 1, 2, 3, \ldots, \mathcal{H})$. Since the transition matrix is a stochastic matrix, it satisfies the conditions

$$0 \leq P_{rs} \leq 1, \qquad \text{(for each } r, s) \tag{7.1a}$$

$$\sum_{s=1}^{\mathcal{H}} P_{rs} = 1 \qquad \text{(for each row } r). \tag{7.1b}$$

In order to reduce the number of states in the Markov chain, the approach taken here is to partition the discrete state space of the traditional CUSUM into groups or discrete sub-sets $\Delta_i$ $(i = 0, 1, 2, \ldots, H)$, using a mesh-width of fixed size $\Delta$. The states $E_i$ of the Markov chain are now formed by these distinct groups or sub-sets with

$$\begin{aligned}
E_0 &= \Delta_0 = 0 \\
E_1 &= \Delta_1 = \{1, \ldots, \Delta\} \\
&\vdots \qquad \vdots \qquad \vdots \\
E_i &= \Delta_i = \{(i-1)\Delta + 1, \ldots, i\Delta\} \\
&\vdots \qquad \vdots \qquad \vdots \\
E_H &= \Delta_H \geq (H-1)\Delta + 1
\end{aligned}$$

where each sub-set $\Delta_i$ $(i = 1, 2, \ldots, H-1)$ consists of $\Delta$ elements (or integer values). The exceptions are state $E_0$ corresponding to group $\Delta_0$ which represents a zero count,

and state $E_H$ corresponding to the absorbing state $\Delta_H$ which represents counts exceeding $(H-1)\Delta+1$. The transition probabilities for this grouped (regular) CUSUM can be computed with minor changes, and the transition matrix will have the form

$$
\begin{array}{c}
 \\
\Delta_0 \left\{ 0 \right. \\
\Delta_1 \left\{ \begin{array}{l} 1 \\ 2 \\ \vdots \\ \Delta \end{array} \right. \\
\Delta_2 \left\{ \begin{array}{l} \Delta+1 \\ \Delta+2 \\ \vdots \\ 2\Delta \end{array} \right. \\
\vdots \\
\Delta_H \left\{ \geq (H-1)\Delta+1 \right.
\end{array}
\begin{array}{c}
\overbrace{\phantom{0}}^{\Delta_0} \\
0 \\
\overbrace{1 \;\; 2 \;\; \cdots \;\; \Delta}^{\Delta_1} \\
\overbrace{\Delta{+}1 \; \Delta{+}2 \cdots 2\Delta}^{\Delta_2} \\
\overbrace{\geq (H-1)\Delta+1}^{\Delta_H}
\end{array}
\left( \begin{array}{ccccc}
F_0 & P_{\Delta_0\Delta_1} & P_{\Delta_0\Delta_2} & \cdots & 1-F_0 \\
F_1 & P_{\Delta_1\Delta_1} & P_{\Delta_1\Delta_2} & \cdots & 1-F_1 \\
F_2 & P_{\Delta_2\Delta_1} & P_{\Delta_2\Delta_2} & \cdots & 1-F_2 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 1
\end{array} \right)
\tag{7.2}
$$

which is very similar to the transition matrix of the traditional count data CUSUM given by Equation 3.23 (page 72). For an upper regular count data CUSUM defined by $S_n^+ = \max[0,\ X_n - \mathcal{K} + S_{n-1}^+]$, we set

- $S_n^+ = \quad 0, \qquad$ if $X_n - \mathcal{K} + S_{n-1}^+ \leq 0$,
- $S_n^+ = \quad \mathcal{H}, \qquad$ if $X_n - \mathcal{K} + S_{n-1}^+ \geq \mathcal{H}, \quad$ and
- $S_n^+ = \quad X_n - \mathcal{K} + S_{n-1}^+, \qquad$ otherwise.

The following minor changes in the computation of the transition probabilities can be noticed:

1. Transition from state $E_0$ to state $E_0$ is $P_{\Delta_0,\Delta_0}$, where

$$\begin{aligned} P_{\Delta_0,\Delta_0} &= \Pr\left(S_n^+ = 0 \mid S_{n-1}^+ = 0\right) \\ &= \Pr\left(X_n \leq \mathcal{K}\right) \\ &= F_0 \qquad \text{(as shown in Equation (7.2)).} \end{aligned}$$

2. Transition from state $E_0$ to state $E_j$ $(j = 1, 2, \ldots, H-1)$ is $P_{\Delta_0,\Delta_j}$ is given by $P_{\Delta_0,\Delta_j}$ (an example is $P_{\Delta_0,\Delta_1}$ in Equation (7.2)), where

$$\begin{aligned} P_{\Delta_0,\Delta_j} &= \Pr\left(S_n^+ \in \Delta_j \mid S_{n-1}^+ = 0\right) \\ &= \sum_{v \in \Delta_j} \Pr\left(S_n^+ = v \mid S_{n-1}^+ = 0\right) \\ &= \sum_{v \in \Delta_j} \Pr\left(X_n = \mathcal{K} + v\right) \\ &= \sum_{v \in \Delta_j} p_v \\ &= p_j. \end{aligned}$$

Note that here we are summing a bunch of probabilities from a fixed row (namely, $\Delta_0 = 0$) to a set of columns in the group $\delta_j$, and so $p_j$ satisfies the conditions in (7.1a)–(7.1b).

3. Transition from state $E_i$ to state $E_0$ $(i = 1, 2, \ldots, H-1)$ is given by $P_{\Delta_i,\Delta_0}$. Since we are summing a bunch of probabilities from a set of rows in group $\Delta_i$ to a fixed column (represented by $\Delta_0 = 0$), this sum $(P_{\Delta_i,\Delta_0})$ will not necessarily satisfy conditions (7.1a)–(7.1b), and the probabilities may not be valid. To ensure that the conditions stated above are satisfied, we use the "average" probability

which is obtained by dividing $P_{\Delta_i,\Delta_0}$ by the number of rows (or elements) in the sub-set $\Delta_i$. An example is the transition probability $P_{\Delta_2,\Delta_2}$ shown in Equation (7.2). Thus,

$$\begin{aligned} P_{\Delta_i,\Delta_0} &\approx \Pr\left(S_n^+ = 0 \mid S_{n-1}^+ \in \Delta_i\right) \times \frac{1}{\Delta} \\ &= \sum_{u\in\Delta_i} \Pr\left(S_n^+ = 0 \mid S_{n-1}^+ = u\right) \times \frac{1}{\Delta} \\ &= \sum_{u\in\Delta_i} \Pr\left(X_n \leq \mathcal{K} - u\right) \times \frac{1}{\Delta} \\ &= \sum_{u\in\Delta_i} p_u/\Delta \\ &= F_{-i}. \end{aligned}$$

4. Transition from state $E_i$ to state $E_j$ $(i, j = 1, 2, \ldots, H-1)$. Again, since a bunch of probabilities from all the rows in $\Delta_i$ are being summed across the columns in the sub-set $\Delta_j$, the "average" probability is used (as shown by the "block" transition probability $P_{\Delta_2,\Delta_2}$ in Equation (7.2). The is computed as

$$\begin{aligned} P_{\Delta_i,\Delta_j} &\approx \Pr\left(S_n^+ \in \Delta_i \mid S_{n-1}^+ \in \Delta_j\right) \times \frac{1}{\Delta} \\ &= \sum_{u\in\Delta_i}\sum_{v\in\Delta_j} \Pr\left(S_n^+ = v \mid S_{n-1}^+ = u\right) \times \frac{1}{\Delta} \\ &= \sum_{u\in\Delta_i}\sum_{v\in\Delta_j} \Pr\left(X_n = \mathcal{K} + v - u\right) \times \frac{1}{\Delta} \\ &= \sum_{u\in\Delta_i}\sum_{v\in\Delta_j} p_{v-u}/\Delta \\ &= p_{j-i} \end{aligned}$$

and it ensures that $p_{j-i}$ satisfies the conditions stated in (7.1a)–(7.1b).

The modified CUSUM procedure discussed in Chapters 4 and 5 can now be applied straight forwardly, and the form of the extended transition matrix will be the same as that discussed in Chapters 4 and 5. The actual numerical implementation results for the cattle feedlot study problem will not be provided here because there are major flaws which were discovered later, in the collection of the raw data, and these need to be addressed before any kind of charting procedure can be applied.

As mentioned earlier in Chapter 2, our main objective is not to "correctly" determine the underlying process distribution, but rather to quickly detect animal morbidity (preferably at the onset). This could be done by assuming both the Poisson and negative binomial models for the generic feedlot data.

# CHAPTER 8

# CONCLUSIONS, DISCUSSIONS AND FUTURE RESEARCH

## Conclusions

A modified cumulative sum procedure for count data has been presented in this dissertation. The method has been evaluated by comparing the regular count data CUSUM and the fast initial response CUSUM. The theoretical, computational and implementation details of the method have been described, and several examples have been used to illustrate the potential of this technique.

The plots and tables of average run lengths (ARLs) clearly indicate that the modified CUSUM scheme has better ARL performance than the regular traditional count data CUSUM (or the Poisson CUSUM). The ARLs decrease with increasing number of buffer states (which determine the warning level parameter). The fast initial response CUSUM (FIR), which starts at some level from the in-control mean value, consistently gives shorter ARLs at specified shifts compared to the modified procedure. It is noted here, however, that the ARL values for the fast initial response CUSUM are valid for the case when the process is out of control at the time the CUSUMs are reset (Montgomery [77, page 327]; Lucas and Crosier [71]). When the modified scheme is used in combination with the head-start feature, it greatly enhances the ARL performance of the fast initial response CUSUM.

The modified CUSUM scheme combines robustification with multiple out-of-control conditions and the probability of extremeness (which also serves as an additional signal probability) into one design feature, to quickly detect out-of-control conditions and other persistent trends.

Examples with simulated Poisson data indicate that even in cases where modest reductions occur in the out-of-control ARLs, these modest changes ultimately help greatly to quickly detect out-of-control conditions. If the probability of extremeness criteria fails to detect a shift and the shift persists for more than three samples, it is either detected immediately by the four-in-a-row warning runs rule criteria or else the process goes out-of-control beyond the usual decision boundary.

The modified CUSUM also seems to do remarkably well with the cattle feedlot data problem discussed in Chapter 1. Trial implementations conducted so far indicate that the modified CUSUM procedure indeed has great potential in the early detection of morbidity among group-fed animals.

Overall, I think some very useful ideas have been proposed, and the modified CUSUM method appears to be very promising and intuitively appealing in many ways. I believe this method can achieve its main purpose (perhaps after minor refinements), of being linked to an on-line electronic data acquisition and monitoring system, such as the Radio Frequency technology, to help in objectively monitoring and quickly detecting morbidity in group-fed animals.

## Discussions and Future Research

We have assumed throughout this study that the sampling (or aggregation) interval is wide enough in time, to greatly diminish the magnitude of the serial correlation and render the observations approximately independent. It has also been pointed out by Yashchin [132] that in most practical situations data manipulations and transformations do not eliminate the serial correlation; they only reduce its magnitude.

While even small correlations can have negative impacts on the ARL of control schemes, SPC procedures such as the CUSUM and EWMA may still be used directly with autocorrelated data (Faltin, Mastrangelo, Runger and Ryan [28]). This view is shared by many, including Crowder, Hawkins, Reynolds and Yashchin [24], who suggested that *"even when the data are inherently correlated, it may be still important to detect changes in the mean or correlation structure by using control chart techniques."* For the case where the serial correlation is strong and cannot be 'removed' from the process, they also cautioned that *"the traditional control charts should not be applied without modification to such process data."*

Yashchin [132] has however shown that if the autocorrelation in the observations is not too strong to begin with, it frequently pays to give up transformations and apply the CUSUM control statistic (which assumes independence) directly, even at the expense of some statistical power. He also suggested (p. 39) that *"in cases in which serial correlation is not expected to be strong, one does not lose too much in statistical power by working with the original observations."* In fact, simulation

results by Runger and Willemain [97] for instance, indicated that the effect of a lag one autocorrelation of 0.10 on CUSUM charts was negligible.

We hope that by this work on the modified CUSUM procedure, we have provided a standard from which to measure future extensions, such as departures from the assumption of independence. Further research on procedures for handling serially correlated data will be very useful in SPC, especially if those procedures can retain the original units of measurements to ensure better understanding of the underlying process and easy interpretation of the results.

A detailed simulation study to compare the modified CUSUM scheme with other existing monitoring schemes such as the regular cumulative sums (CUSUMs), exponentially weighted moving averages (EWMAs) and Shewhart individuals charts under some uniform conditions will be great, although such an exercise, if practically feasible, will be rather costly.

The limited number of possible ARLs (due to the discrete nature of the Poisson distribution), does not allow a straightforward evaluation of the modified procedure by simply comparing its ARL performance with other control schemes. If the in-control ARL could be constrained to be identical for the methods, then one could simply compare the out-of-control ARLs. In particular, it will be interesting to see how the probability of extremeness criterion (PEC) compares to the likelihood ratio criterion (LRC) with respect to ARLs.

Another criterion worth investigating will be the probability of extremeness ratio

criterion (PERC), where the probability of extremeness will be computed for the acceptable mean level (say, $PE_0$) and for the unacceptable mean level (say, $PE_1$) at which quick detection is desired. The PERC can then be based on the ratio of $PE_1$ to $PE_0$, which is a likelihood ratio-type criterion.

Exploring the possible extension of the modified CUSUM procedure to the more flexible (but probably also more challenging) case of continuous variables data CUSUMs will also be of interest. This might also include modeling or applying the modified scheme to waiting times (or time between clusters of feeding hits recorded for a steer).

Another interesting aspect will be an elaborate modeling (may be time series, dynamic or structural modeling) of the cattle feedlot process in order to gain better understanding of the underlying process. The modified CUSUM could then be applied to the resulting residuals (if they can be described as white noise).

An interesting component of this modeling exercise might be to consider other generalizations of the Poisson and negative binomial families. In particular, due to the unique nature of the generic feedlot data, it might be of interest to explore the use of zero-inflated negative binomial (or possibly, to consider further inflation of some additional low counts in the negative binomial) distribution.

Further investigation of the extent of robustness of the modified CUSUM scheme in comparison with other robust schemes (Lucas and Crosier [72]; Hawkins [45]) may be of interest.

The potential of combining the modified CUSUM with Shewhart individuals chart

to obtain a control scheme that can possibly possess good ARL properties for both small and large shifts (Lucas [67]) may also be investigated.

A final future thought concerning the modified CUSUM will be to determine the steady state ARLs, which assume that the shift occurs long after the process has been operating in a state of statistical control.

# REFERENCES CITED


[1] ADAMS, B. M., WOODALL, W. H. AND SUPERVILLE, C. R. Discussion on *Run-Length Distribution of Special-Cause Control Charts for Correlated Processes* by D. G. Wardell, H. Moskowitz, and R. D. Plante. *Technometrics*, **36**, (1994), 19–22.

[2] ALWAN, L. C., AND ROBERTS, H. V. Time-Series Modeling for Stationary Process Control. *Journal of Business & Economic Statistics*, **6**, 87–95.

[3] AMIN, R. W., REYNOLDS, MARION R., J., AND BAKIR, S. Nonparametric quality control charts based on the sign statistic. *Communications in Statistics, Part A – Theory and Methods*, **24**, (1995), 1597–1623.

[4] BARNARD, G. A. Control Charts and Stochastic Processes. *Journal of the Royal Statistical Society*, Series B, **21**, (1959), 239–257.

[5] BAGSHAW, M., AND JOHNSON, R. A. The effect of serial correlation on the performance of Cusum tests I. *Technometrics*, **17**, (1975), 73–80.

[6] BASARAB, J. A., MILIGAN, D., HAND, R., AND HUISMA, C. Automatic monitoring of watering behavior in feedlot steers: Potential use in early detection of respiratory disease and in predicting growth. In *Proceedings of the Canadian Society of Animal Science, 46th Annual Conference, July 7-11, Lethbidge, Alberta*, (1996), Canadian Society of Animal Science, p. 28.

[7] BESTERFIELD, DALE H. *Quality Control: A Practical Approach.* Prentice-Hall, Inc., Englewood Cliffs, N.J., 1979.

[8] BHAT, NARAYAN, U. *Elements of Applied Stochastic Processes.* John Wiley & Sons, Inc., New York, N.Y., 1984.

[9] BISSELL, A. F. Cusum techniques for quality control (with discussion). *Applied Statistics*, **18**, (1969), 1–30.

[10] BISSELL, A. F. The Performance of Control Charts and Cusums Under Linear Trend. *Applied Statistics*, **33**, (1984), 145–151.

[11] BISSELL, A. F. Estimation of Linear Trend from a Cusum Chart. *Applied Statistics*, , (1985),

[12] BISSELL, A. F. Weighted Cumulative Sums for Text Analysis Using Word Counts. *Journal of the Royal Statistical Society, Series A, General*, **158**, (1995), 525–545.

[13] BORROR, C. M., CHAMP, C. W., AND RIGDON, S. E. Poisson EWMA Control Charts. *Journal of Quality Technology*, **30**, (1998), 352–361.

[14] BOURKE, P. D. Performance of Cumulative Sum Schemes for Monitoring Low Count-level Processes. *Metrika*, **39**, (1992), 365–384.

[15] BOX, G. E. P., AND KRAMER, T Statistical Prpcess Control and Feedback Adjustment–A Discussion. *Technometrics*, **34**, (1992), 251–285.

[16] BROCKWELL, PETER J., AND DAVIS, RICHARD A. *Time Series: Theory and Methods.* Springer-Verlag Inc., New York, N.Y., 1987.

[17] BROOK, D., AND EVANS, D. A. An Approach to the Probability Distribution of Cusum Run Length. *Biometrika*, **59**, (1972), 539–549.

[18] BUCKLEY, M. J. Detecting a smooth signal: Optimality of cusum based procedures. *Biometrika*, **78**, (1991), 253–262.

[19] BURR, IRVING W. *Elementary Statistical Quality Control.* Marcel Dekker Inc., New York, N.Y., 1979.

[20] CHAMP, C. W., WOODALL, W. H. Exact Results for Shewhart Control Charts with Supplementary Runs Rules. *Technometrics*, **29**, (1987).

[21] CROSIER, R. B. Multivariate Generalizations of Cumulative Sum Quality Control Schemes. *Technometrics*, **30**, (1988).

[22] CROWDER, S. V. A Simple Method for Studying Run-Length Distributions of Exponentially Weighted Moving Average Charts. *Technometrics*, **29**, (1987), 401–407.

[23] CROWDER, S. V. Design of Exponentially Weighted Moving Average Scheme. *Journal of Quality Technology*, **21**, (1989), 155–162.

[24] CROWDER, S. V., HAWKINS, D. C., REYNOLDS, M. R. AND YASHCHIN, E. Process Control and Statistical Inference— A DISCUSSION ON STATISTICALLY-BASED MONITORING AND CONTROL. *Journal of Quality Technology*, **29**, (1997), 134–139.

[25] DEMING, W. EDWARDS *Quality, Productivity and Competitive Position.* Cambridge, MA: MIT Center for Advanced Engineering Study, (1982).

[26] DUNCAN, A. J. The Economic Design of $\overline{X}$-Charts Used to Maintain Current Control of a Process. *Journal of the American Statistical Association*, **51**, (1956).

[27] EWAN, W. D., AND KEMP, K. W. Sampling Inspection of Continuous Processes with No Autocorrelation Between Successive Results. *Biometrika*, **47**, 363–380.

[28] FALTIN, F. W., MASTRANGELO, C. M., RUNGER, G. C. AND RYAN, T. P. Considerations in the Monitoring of Autocorrelated and Independent Data—A DISCUSSION ON STATISTICALLY-BASED MONITORING AND CONTROL. *Journal of Quality Technology*, **29**, (1997), 131–133.

[29] FELLNER, WILLIAM H. Discussion on *Run-Length Distribution of Special-Cause Control Charts for Correlated Processes* by D. G. Wardell, H. Moskowitz, and R. D. Plante, *Technometrics*, **36**, (1994), 22–23.

[30] GAN, F. F. Combined Cumulative Sum and Shewhart Variance Charts. *Journal of Statistical Computation and Simulation*, **32**, (1989), 149–163.

[31] GAN, F. F. An Optimal Design of Cusum Quality Control Charts. *Journal of Quality Technology*, **23**, (1991), 279–286.

[32] GAN, F. F. Cusum Control Charts Under Linear Drift. *The Statistician*, **41**, (1992), 71–84.

[33] GAN, F. F. An Optimal Design of Cusum Control Charts for Binomial Counts. *Journal of Applied Statistics*, **20**, (1993), 445–460.

[34] GAN, F. F. Design of Optimal Exponential Cusum Control Charts. *Journal of Quality Technology*, **26**, (1994), 109–124.

[35] GAN, F. F., AND CHOI, K. P. Computing average run lengths for exponential Cusum schemes. *Journal of Quality Technology*, **26**, (1994), 134–143.

[36] GELMAN, ANDREW, C. J. B., STERN, H. S., AND RUBIN, D. B. *Bayesian Data Analysis.* Chapman & Hill, London, UK, 1998.

[37] GOEL, A. L., AND WU, S. M. Determination of A.r.l. and a contour nomogram for cusum charts to control normal mean. *Technometrics*, **13**, (1971), 221–230.

[38] GOLDSMITH, P. L., AND WHITFIELD, H. Average Run Lengths in Cumulative Chart Quality Control Schemes. *Technometrics*, **3**, (1961), 11–20.

[39] GRANT, EUGENE L., AND LEAVENWORTH, RICHARD S. *Statistical Quality Control.* McGraw-Hill Inc., 5th Edition, New York, NY, 1980.

[40] HARRIS, T. J. AND ROSS, W. H. Statistical Process Control Procedures for Correlated Observations. *The Canadian JOurnal of Chemical Engineering*, **69**, (1991), 48–57.

[41] HAWKINS, D. M. A Cusum for a Scale Parameter. *Journal of Quality Technology*, **13**, (1981), 228–231.

[42] HAWKINS, D. M. Self-starting Cusum Charts for Location and Scale. *The Statistician*, **36**, (1987), 299–316.

[43] HAWKINS, D. M. Multivariate Quality Control Based on Regression Adjusted Variables. *Technometrics*, **33**, (1991).

[44] HAWKINS, D. M. A Fast Accurate Approximation for Average Run Lengths of Cusum Control Charts. *Journal of Quality Technology*, **24**, (1992), 37–43.

[45] HAWKINS, D. M. Winsorizing Data to Make Cusum Charts Robust. In *ASA Proceedings of the Section on Physical and Engineering Sciences* (1993), American Statistical Association (Alexandria, VA), pp. 183–190.

[46] HAWKINS, D. M., AND OLWELL, D. H. Inverse Gaussian Cumulative Sum Control Charts for Location and Shape. *The Statistician*, **46**, (1997), 323–335.

[47] HAWKINS, D. M., AND OLWELL, D. H. *Cumulative Sum Charts and Charting for Quality Improvement.* Springer-Verlag, New York, NY, 1998.

[48] HEALY, J. D. Adaptive Cusum Procedures. In *ASQC Technical Conference Transactions* (1986), American Society for Quality Control (Milwaukee), pp. 291–296.

[49] HILL, D. Modified Control Limits. *Applied Statistics*, **5**, (1956).

[50] HUBELE, N. F., AND CHANG, S. I. Adaptive Exponentially Weighted Moving Average Schemes Using a Kalman Filter. *IIE Transactions* **22**, (1990), 361–369.

[51] HUGHES, C. S., REYNOLDS, MARION R., J., AND ARNOLD, J. C. Variable Sample Size Shewhart and Cusum Charts for Monitoring Process Variability. In *ASA Proceedings of the Section on Quality and Productivity* (1998), American Statistical Association (Alexandria, VA), pp. 13–18.

[52] JACKSON, J. E. All Count Distributions Are Not Alike. *Journal of Quality Technology*, **4**, (1972), 86–92.

[53] JOE, H. *Multivariate Models and Dependence Concepts.* Chapman & Hall, London, UK, 1997.

[54] JOHNSON, R. A., AND BAGSHAW, M. The Effect of Serial Correlation on The Performance of Cusum Tests. *Technometrics*, **16**, (1974), 103–112.

[55] JOHNSON, N. L. AND LEONE, F. C. Cumulative Sum Control Charts: Mathematical principles applied to their construction and use. Part I. *Industrial Quallity Control*, **18**, 15–21.

[56] JOHNSON, N. L. AND LEONE, F. C. Cumulative Sum Control Charts: Mathematical principles applied to their construction and use. Part II. *Industrial Quallity Control*, **19**, No. 1, 29–36.

[57] JOHNSON, N. L. AND LEONE, F. C. Cumulative Sum Control Charts: Mathematical principles applied to their construction and use. Part III. *Industrial Quallity Control*, **19**, No. 2, 22–28.

[58] JOHNSON, N. L., KOTZ, S. AND KEMP, A. W. *Univariate Discrete Distributions.* John Wiley & Sons, Inc., 2nd Edition, New York, N.Y., 1993.

[59] JOSEPH, J., AND BOWEN, V. A Cumulative Bayesian Technique for Use in Quality Control Schemes. In *ASA Proceedings of the Section on Quality and Productivity* (1991), American Statistical Association (Alexandria, VA), pp. 52–54.

[60] KEATS, J. BERT Process Control in Automated Manufacturing: Some Key Issues. *Statistical Process Control in Automated Manufacturing*, edited by J. B. Keats and N. F. Hubele, Marcel Dekker, Inc., Series in Quality and Reliability, New York, NY., 1989.

[61] KEMENY, JOHN G. AND SNELL, J. LAURIE. *Finite Markov Chains.* D. Van Nostrand Company, Inc., Princeton, New Jersey, 1960.

[62] KEMP, K. W. Formal Expressions Which can be Applied to Cusum Charts (With Discussion). *Journal of the Royal Statistical Society, Series B, Methodological*, **33**, (1971), 331–360.

[63] KRISHNAN, VENKATARAMA. *Nonlinear Filtering and smoothing: An Introduction to Martingales, Stochastic Integrals and Estimation.* John Wiley & Sons, Inc., New York, NY, 1984.

[64] LEON-GARCIA, A. *Probability and Random Processes for Electrical Engineering.* Addison-Wesley, Reading, Massachusetts, 1982.

[65] LORDEN, G. Procedures for Reacting to a Change in Distribution. *Annals of Mathematical Statistics*, **42**, (1971), 1897–1908.

[66] LOWRY, C. A., CHAMP, C. W., AND WOODALL, W. H. The Performance of Control Charts for Monitoring Process Variation. *Communications in Statistics, Part B – Simulation and Computation*, **24**, (1995), 409–437.

[67] LUCAS, J. M. Combined Shewhart-Cusum Quality Control Schemes. *Journal of Quality Technology*, **14**, (1982), 51–59.

[68] LUCAS, J. M. Counted Data Cusum's. *Technometrics*, **27**, (1985), 129–144.

[69] LUCAS, J. M. Cumulative Sum (Cusum) Control Schemes. *Communications in Statistics, Part A – Theory and Methods*, **14**, (1985), 2689–2704.

[70] LUCAS, J. M. Control Schemes for Low Count Levels. *Journal of Quality Technology*, **21**, (1989), 199–201.

[71] LUCAS, J. M., AND CROSIER, R. B. Fast Initial Response for Cusum Quality-Control Schemes: Give your Cusum a Head Start. *Technometrics*, **24**, (1982), 199–205.

[72] LUCAS, J. M., AND CROSIER, R. B. Robust Cusum: A Robustness Study for Cusum Quality Control Schemes. *Communications in Statistics, Part A – Theory and Methods*, **11**, (1982), 2669–2687.

[73] LUCAS, J. M., AND SACCUCCI, M. S. Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements. *Technometrics*, **32**, (1990), 1–12.

[74] LUCAS, J. M. Discussion on *Run-Length Distribution of Special-Cause Control Charts for correlated Processes* by D. G. Wardell, H. Moskowitz, and R. D. Plante, *Technometrics*, **36**, (1994), 3–17.

[75] MANDEL, B. J. The Regression Control Chart. *Journal of Quality Technology*, **1**, (1969), 1–9.

[76] MATHSOFT. *Splus Guide to Statistical and Mathematical Analysis.* MathSoft, Inc., Ver. 3.4 (SUNOS), Seattle, WA, 1995.

[77] MONTGOMERY, D. C. *Introduction To Statistical Quality Control.* John Wiley & Sons, Inc., New York, N.Y., 1997.

[78] MONTGOMERY, D. C., AND FRIEDMAN, J. J. Statistical Process Control in a Computer-Integrated Manufacturing Environment. *Statistical Process Control in Automated Manufacturing*, edited by J. B. Keats and N. F. Hubele, Marcel Dekker, Inc., Series in Quality and Reliability, New York, NY., 1989.

[79] MONTGOMERY, D. C., AND MASTRANGELO, C. M. Some Statistical Process Control Methods for Autocorrelated Data. *Journal of Quality Technology*, **23**, 179–193.

[80] MOOD, A. M., GRAYBILL, F. A., AND BOES, D. C. *Introduction to the Theory of Statistics.* McGraw-Hill, Inc., 3rd Edition, New York, NY, 1974.

[81] MOUSTAKIDES, G. V. Optimal Stopping Times for Dectecting Changes in Distributions. *Annals of Statistics*, **14**, (1986), 1379–1387.

[82] MUNFORD, A. G. A Control Chart Based on Cumulative Scores. *Applied Statistics*, **29**, 252–258.

[83] NCUBE, M. M., AND AMIN, R. W. Two parameter cuscore quality control procedures. *Communications in Statistics, Part A – Theory and Methods*, **19**, (1990), 2191–2205.

[84] NCUBE, M. M., AND WOODALL, W. H. A Combined Shewhart-Cumulative Score Quality Control Chart. *Applied Statistics*, **33**, (1984), 259–265.

[85] PALM, ANDREW C., RODRIGUEZ, R. N., SPIRING F. A., AND WHEELER, D. J. Some Perspectives and Challenges for Control Chart Methods— A DISCUSSION ON STATISTICALLY-BASED MONITORING AND CONTROL. *Journal of Quality Technology*, **29**, (1997), 122–127.

[86] PAGE, E. S. Continuous inspection schemes. *Biometrika*, **41**, (1954), 100–115.

[87] PAGE, E. S. Control Charts with Warning Lines. *Biometrika*, **42**, (1955), 243–257.

[88] PAGE, E. S. Cumulative sum charts. *Technometrics*, **3**, (1961), 1–9.

[89] PIJPERS, A., SCHOEVERS, E. J., VAN GOGH, H., VAN LEENGOED, L.A.M.G., VISSER, I.J.R., VAN MIERT, A.S.J.P.A.M. AND VERHEIJDEN, J.H.M. The influence of disease on feed and water consumption and on pharmacokinetics of orally administerd oxytetracycline in pigs. *Journal of Animal Science*, **69**, (1991), 2947–2954.

[90] PUTNAM, P. A., LEHMANN, R. AND DAVIS, R. E. Ration selection and feeding patterns of steers fed in drylot. *Journal of Animal Science*, **26**, (1967), 647–650.

[91] QUIMBY, W. F., SOWELL, B. F., BOWMAN, J. G. P., BRANINE, M. E., HUBBERT, M. E., AND SHERWOOD, H. W. Application of cumulative sums to predict morbidity of newly received calves in a commercial feedlot. To appear. *Canadian Journal of Animal Science*, (re-submitted in 2000).

[92] RADAELLI, GIOVANNI. Using the Cuscore Technique in the Surveillance of Rare Health Events. *Journal of Applied Statistics*, **19**, (1992), 75–81.

[93] RADAELLI, G. Poisson and negative binomial dynamics for counted data under Cusum-type charts. *Journal of Applied Statistics*, **21**, (1994), 347–356.

[94] RITOV, YA'ACOV. Decision Theoretic Optimality of the CUSUM Procedure. *Annals of Statistics*, **18**, (1990), 1464–1469.

[95] ROBERTS, S. W. Control Chart Tests Based on Geometric Moving Averages. *Technometrics*, **1**, (1959), 239–250.

[96] ROSS, SHELDON, M. *Introduction to Probability Models.* Academic Press, Inc., Orlando, Florida, 3rd Ed., 1985.

[97] RUNGER, G. C., WILLEMAIN, T. R. Model-Based and Model-Free Control of Autocorrelated Processes. *Journal of Quality Technology*, **27**, (1995), 283–292.

[98] RUNGER, G. C., WILLEMAIN, T. R., AND PRABHU, S. Average run lengths for Cusum control charts applied to residuals. *Communications in Statistics, Part A – Theory and Methods*, **24**, (1995), 273–282.

[99] SCHMID, W. Cusum control schemes for Gaussian processes. *Statistical Papers, 38*, (1997), 191–217.

[100] SCHWARTZKOPF-GENSWEIN, K. S., C., H., AND MCALLISTER, T. Validation of a radio frequnecy identification system for monitoring the feeding patterns of feedlot cattle. *Livestock Production Science*, **60**, (1999), 27–31.

[101] SHEWHART, W. A. *Economic Control of Quality of Manufactured Product.* Van Nostrand Reinhold Company, Inc., Princeton, N.J., 1931.

[102] SHIRYAYEV, A. N. *Probability.* Springer-Verlag, Inc., New York, N.Y., 1984.

[103] SMITH, R. A. Impact of disease on feedlot performance: a review. *Journal of Animal Science*, **76**, (1998), 272–274.

[104] SOWELL, B. F., BOWMAN, J. G. P., BRANINE, M. E., AND HUBBERT, M. E. Radio frequency technology to measure feeding behavior and health of feedlot steers. *Applied Animal Behavior Science*, **59**, (1998), 277–284.

[105] SOWELL, B. F., BRANINE, B. S., BOWMAN, J. G. P., HUBBERT, M. E., SHERWOOD, H. E., AND QUIMBY, W. Feeding and watering behavior of healthy and morbid steers in a commercial feedlot. *American Society of Animal Science*, **77**, (1999), 1105–1112.

[106] SRIVASTAVA, M. S. Cusum Procedure for Monitoring Variability. *Communications in Statistics, Part A – Theory and Methods*, **26**, (1997), 2905–2926.

[107] SRIVASTAVA, M. S., AND WU, Y. Evaluation of Optimum Weights and Average Run Lengths in Ewma Control. *Communications in Statistics, Part A – Theory and Methods*, **26**, (1997), 1253–1267.

[108] STEINER, S. H. Grouped Data Exponentially Weighted Moving Average Control Charts. *Applied Statistics*, **47**, (1998), 203–216.

[109] STEINER, S. H., GEYER, P. L., AND WESOLOWSKY, G. O. Grouped Data-Sequential Probability Ratio Tests and Cumulative Sum Control Charts. *Technometrics*, **38**, (1996), 230–237.

[110] TAYLOR, H. M. The Economic Design of Cumulative Sum Control Charts. *Technometrics*, **10**, (1968), 479–488.

[111] TAYLOR, HOWARD, AND KARLIN, SAMUEL. *An Introduction to stochastic Modeling* Academic Press, Inc., Orlando, Florida, 2nd Ed., 1994.

[112] VANBRACKLE, LEWIS N., AND REYNOLDS, MARION R., J. Ewma and Cusum control charts in the presence of correlation. *Communications in Statistics, Part B – Simulation and Computation*, **26**, (1997), 979–1008.

[113] VAN DOBBEN DE BRUYN, C. S. *Cumulative Sum Tests: Theory and Practice.* In *Griffins* Statistical Monographs and Courses, No. 24, New York: Hafner.

[114] VANDER WIEL, S. A. Monitoring Processes that Wander Using Integrated Moving Average Models. *Technometrics*, **38**, (1996), 139–151.

[115] WALDMANN, K.-H. Bounds for the distribution of the run length of one-sided and two-sided Cusum quality control schemes. *Technometrics*, **28**, (1986), 61–67.

[116] WARDELL, DON G., MOSKOWITZ, H., AND PLANTE, D. R. Run-Length Distribution of Special-Cause Control Charts for correlated Processes (with discussion). *Technometrics*, **36**, (1994), 3–22.

[117] WEI, WILLIAM W. *Time Series Analysis: Univariate and Multivariate Methods.* Addison-Wesley Company, Inc., Reading, MA, U.S.A., 1990.

[118] WEST, MIKE, AND HARRISON, JEFF. *Bayesian Forecasting and Dynamic Models.* Springer-Verlag, Inc., New York, NY, 2nd Ed., 1997.

[119] WESTERN ELECTRIC. *Statistical Quality Control Handbook* Western Electric Corp., Indianapolis, IN, 1956.

[120] WHITE, C. H., AND KEATS, J. B. ARLs and Higher-Order Run-Length Moments for the Poisson Cusum. *Journal of Quality Technology*, **28**, (1996), 363–369.

[121] WOODALL, W. H. The Distribution of the Run Length of One-Sided Cusum Procedures for Continuous Random Variables. *Technometrics*, **25**, (1983), 295–301.

[122] WOODALL, W. H. On the Markov Chain Approach to the Two-sided Cusum Procedure. *Technometrics*, **26**, (1984), 41–46.

[123] WOODALL, WILLIAM H. AND NCUBE, MATOTENG M. Multivariate CUSUM Quality-Control Procedures. *Technometrics*, **27**, (1985), 285–292.

[124] WOODALL, W. H. The Design of Cusum Quality Control Charts. *Journal of Quality Technology*, **18**, (1986), 99–102.

[125] WOODALL, W. H. Control Charts Based on Attribute Data: Bibliography and Review. *Journal of Quality Technology*, **29**, (1997), 172–183.

[126] WOODALL, WILLIAM H. Individual Contributions—A DISCUSSION ON STATISTICALLY-BASED MONITORING AND CONTROL. *Journal of Quality Technology*, **29**, (1997), 155–156.

[127] WOODBURY, MAX A. Inventing Modified Matrices. Cusum Procedure. *Memrandun Report*, **26**, Statistical Research Group, Princeton, N.J. (1950).

[128] XIAO, H. A Cumulative Score Control Scheme. *Applied Statistics*, **41**, (1992), 47–54.

[129] YASHCHIN, E. On the analysis and design of Cusum-shewhart control schemes. *IBM Journal of Research and Development*, 29, (1985), 377–391.

[130] YASHCHIN, E. Weighted Cusum schemes and their applications. In *Statistics Canada Symposium* (1990), Statistics Canada (Ottawa), pp. 303–317.

[131] YASHCHIN, E. Analysis of Cusum and other Markov-type control schemes by using empirical distributions. *Technometrics*, **34**, (1992), 54–63.

[132] YASHCHIN, E. Performance of CUSUM control schemes for serially correlated observations. *Technometrics*, **35**, (1993), 37–52.

[133] YASHCHIN, E. Statistical Control Schemes: Methods, Applications, and Generalizations. *International Statistical Review*, **61**, (1993), 41–66.

# APPENDICES

# APPENDIX A

Figures (Plots) of Average Run Lengths (ARLs) For Modified Upper CUSUM (With No FIR) Using In-Control Means $\mu_0 = 2.0, 4.0, 6.0, 8.0$ and $10.0$.

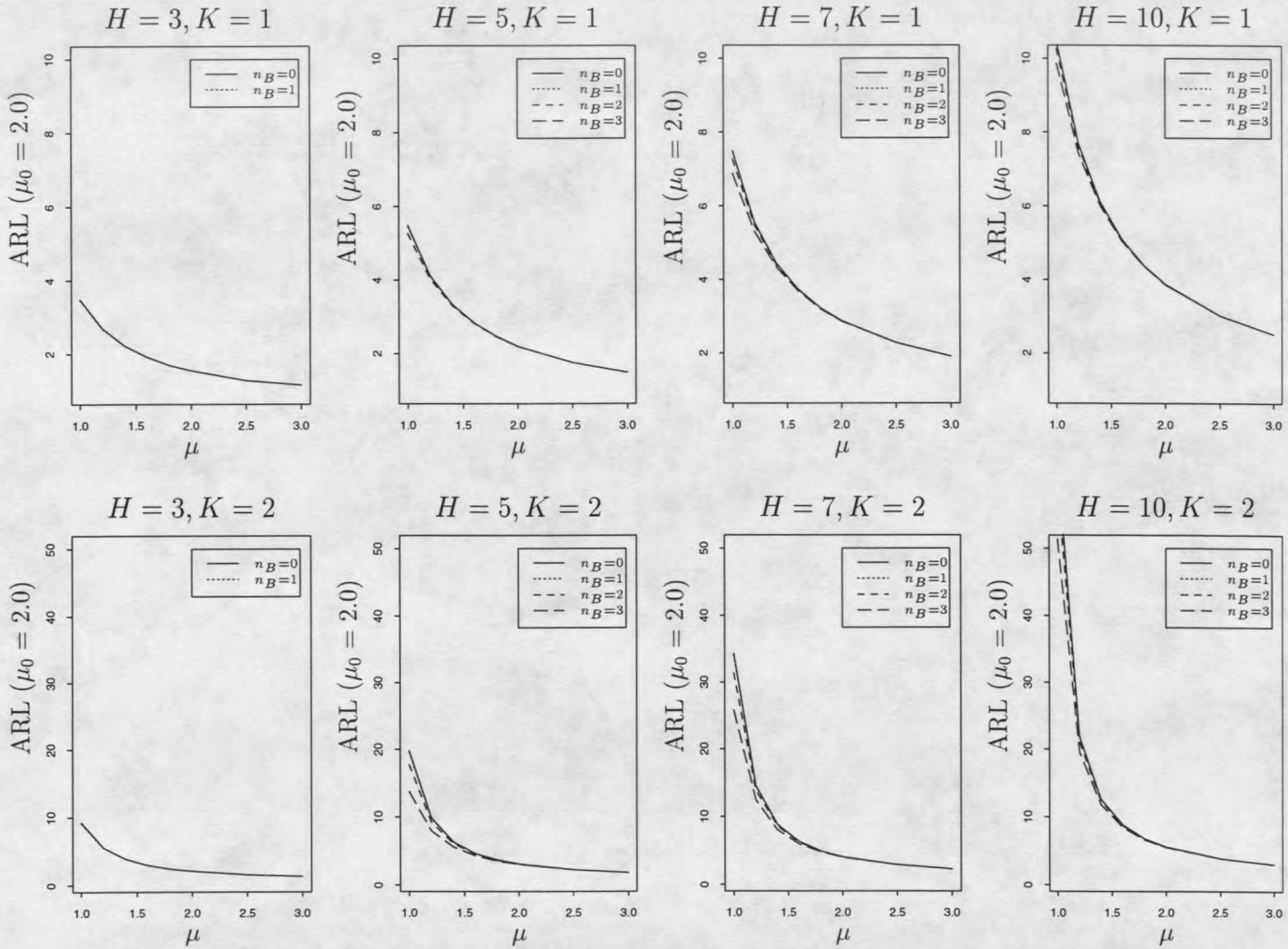


**Figure 7**: ARLs for Modified (Upper) CUSUM charts for various values of $H$, $K = 1, 2$, and $W$ (determined by $n_B$), when the process experiences a step shift of $\mu$ (given in multiples of the in-control mean $\mu_0 = 2.0$).

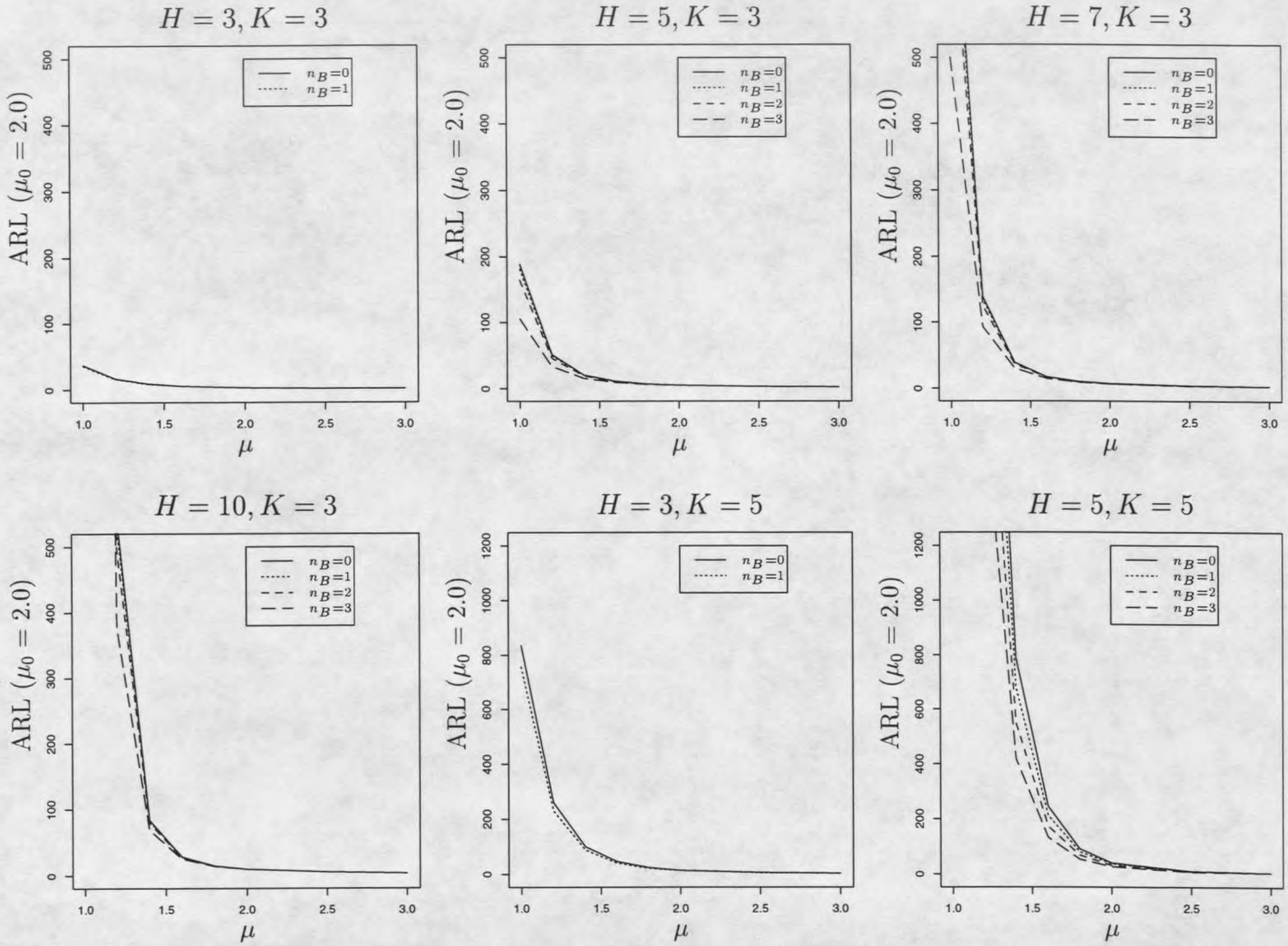


**Figure 8**: ARLs for Modified (Upper) CUSUM charts for various values of $H$, $K = 3, 5$, and $W$ (determined by $n_B$), when the process experiences a step shift of $\mu$ (given in multiples of the in-control mean $\mu_0 = 2.0$).

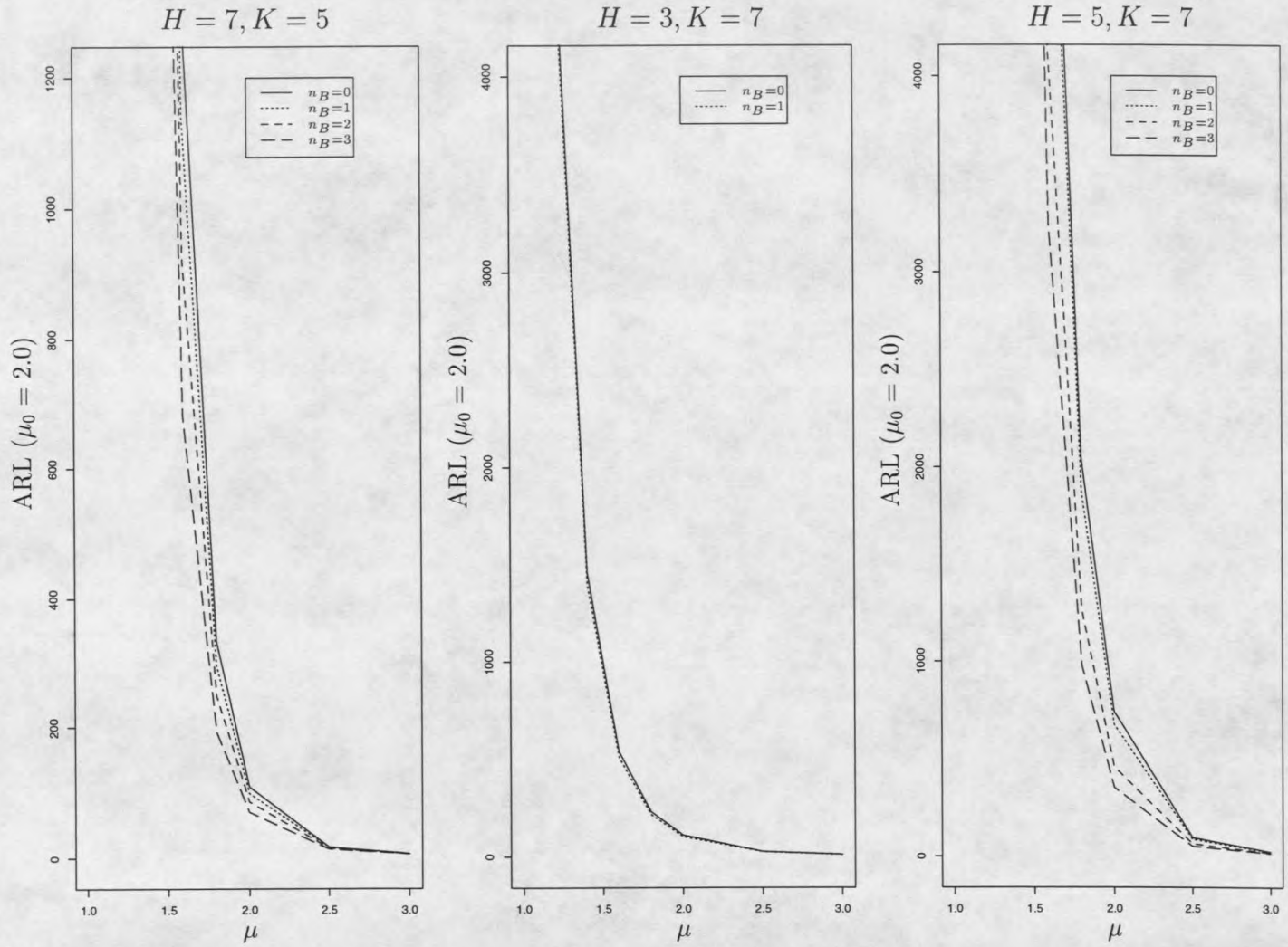


**Figure 9**: ARLs for Modified (Upper) CUSUM charts for various values of $H$, $K = 5, 7$, and $W$ (determined by $n_B$), when the process experiences a step shift of $\mu$ (given in multiples of the in-control mean $\mu_0 = 2.0$).

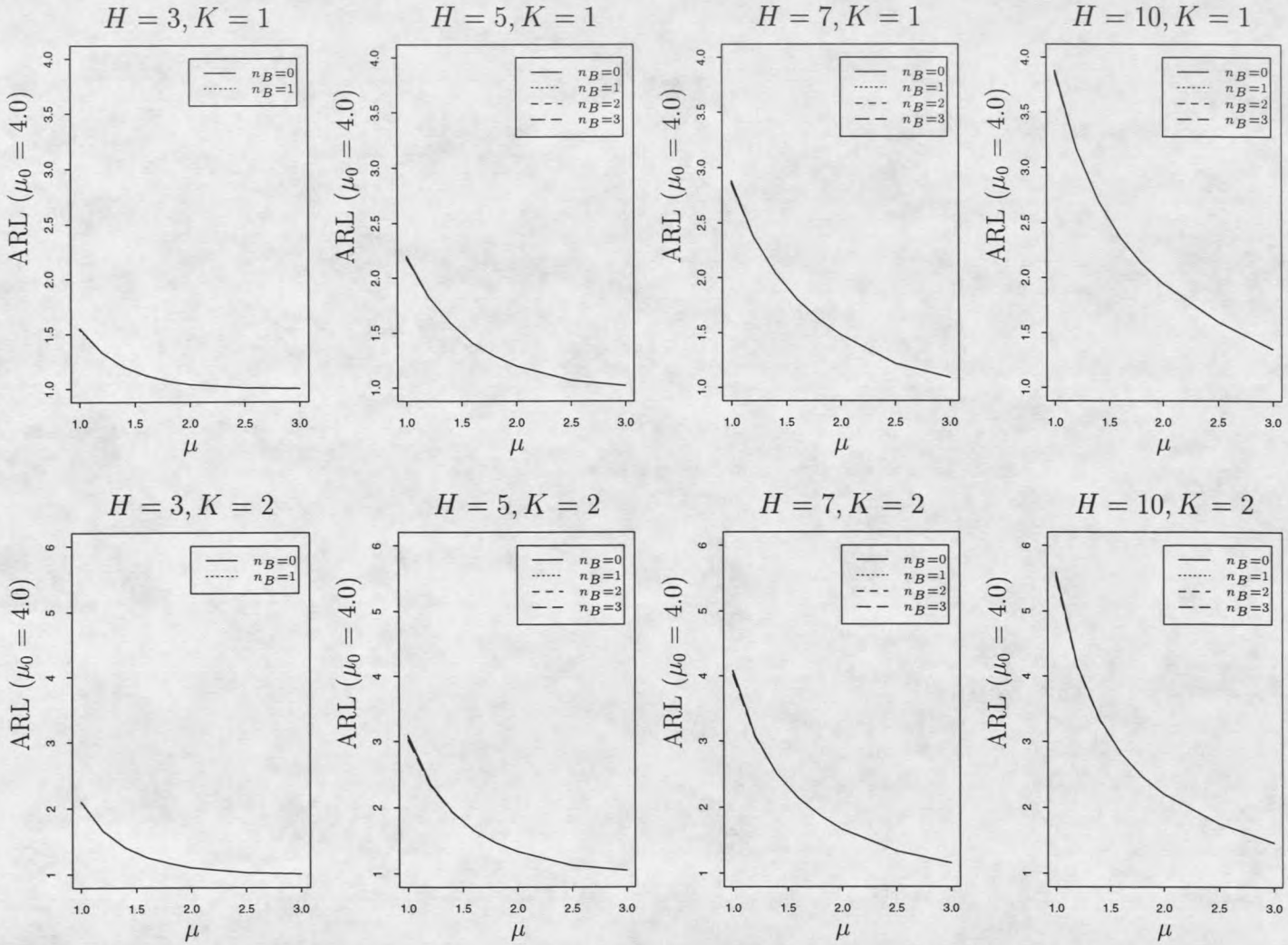


**Figure 10**: ARLs for Modified (Upper) CUSUM charts for various values of $H$, $K = 1, 2$, and $W$ (determined by $n_B$), when the process experiences a step shift of $\mu$ (given in multiples of the in-control mean $\mu_0 = 4.0$).

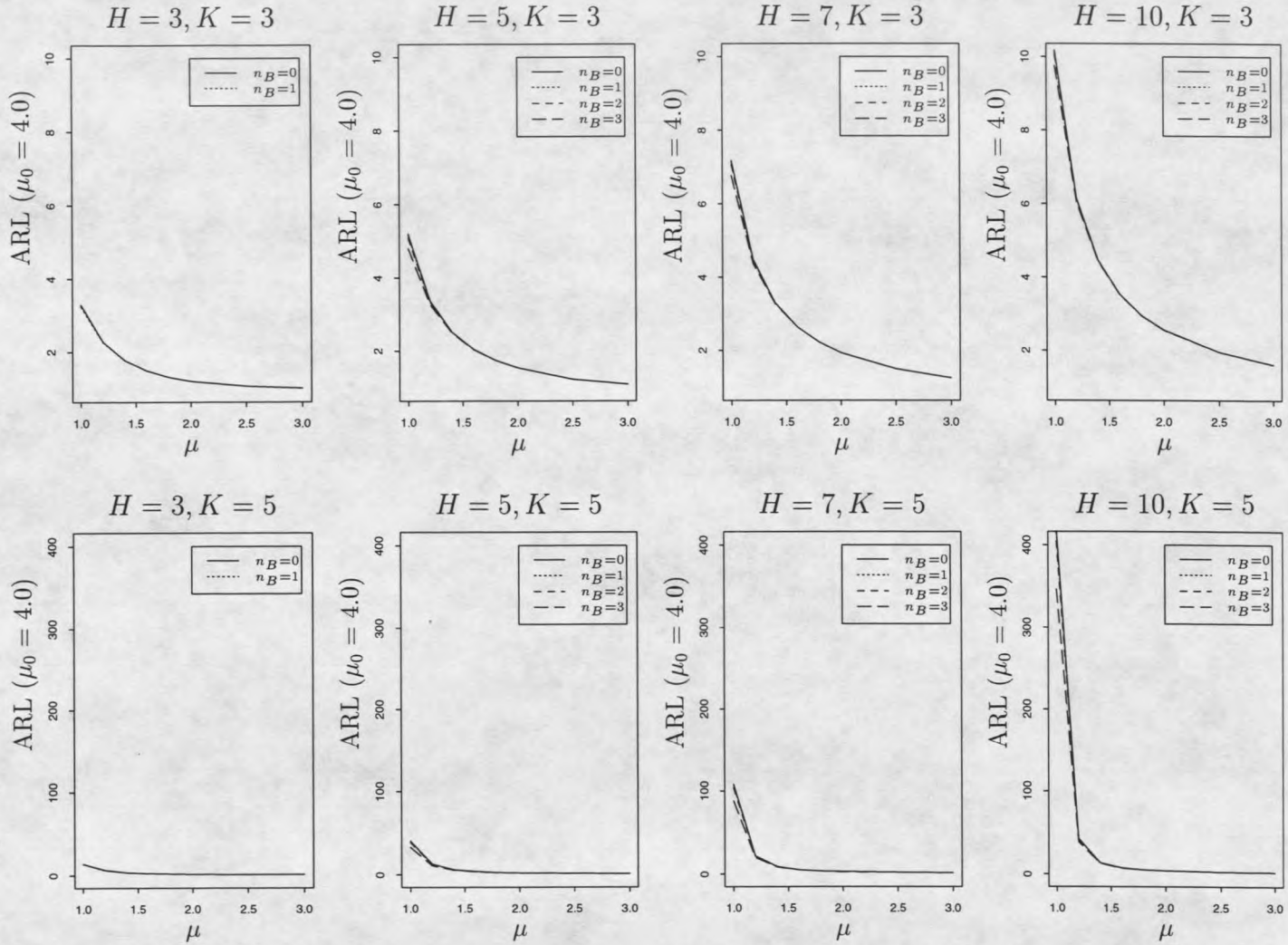


**Figure 11**: ARLs for Modified (Upper) CUSUM charts for various values of $H$, $K = 3, 5$, and $W$ (determined by $n_B$), when the process experiences a step shift of $\mu$ (given in multiples of the in-control mean $\mu_0 = 4.0$).

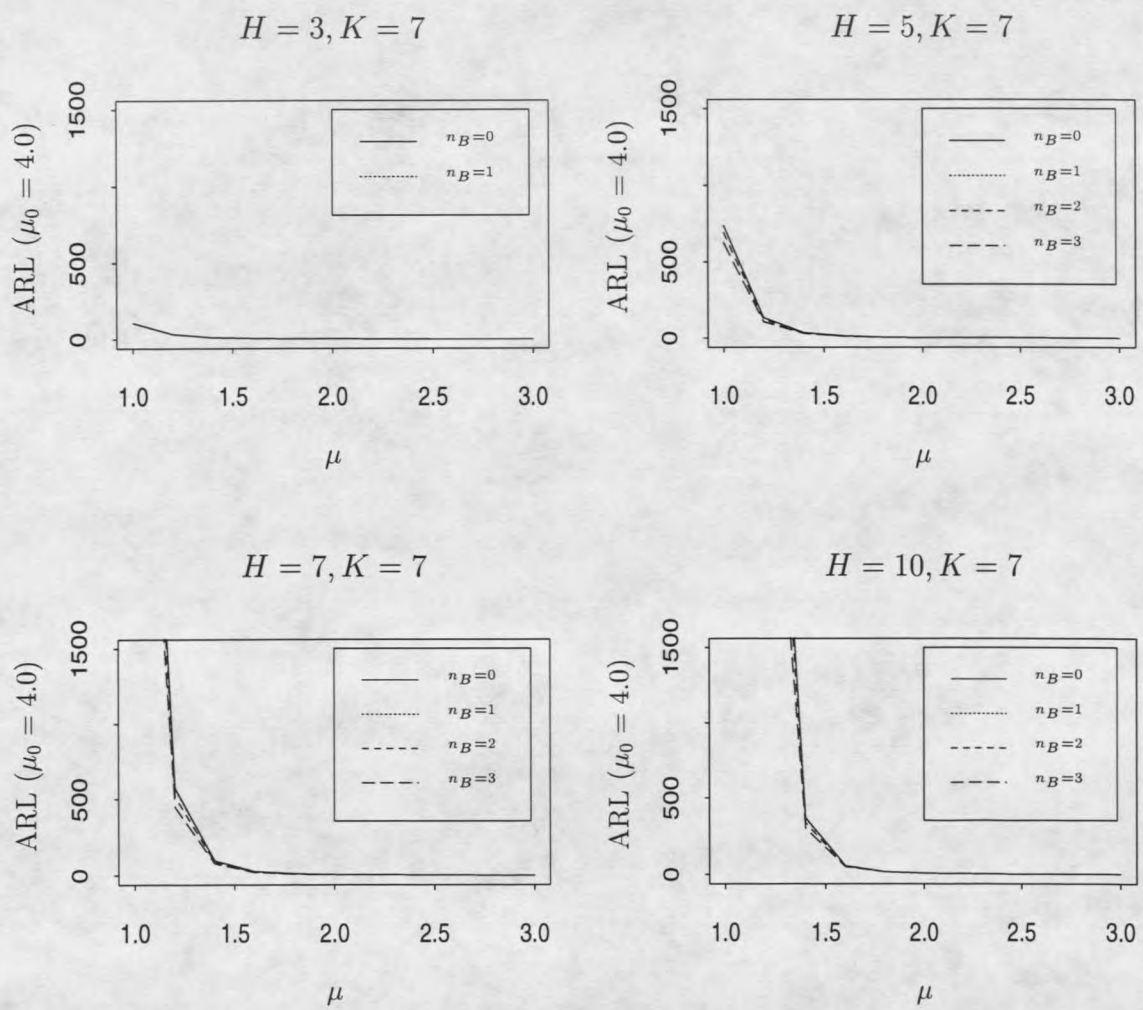


**Figure 12**: ARLs for Modified (Upper) CUSUM charts for various values of $H$, $K = 7$, and $W$ (determined by $n_B$), when the process experiences a step shift of $\mu$ (given in multiples of the in-control mean $\mu_0 = 4.0$).

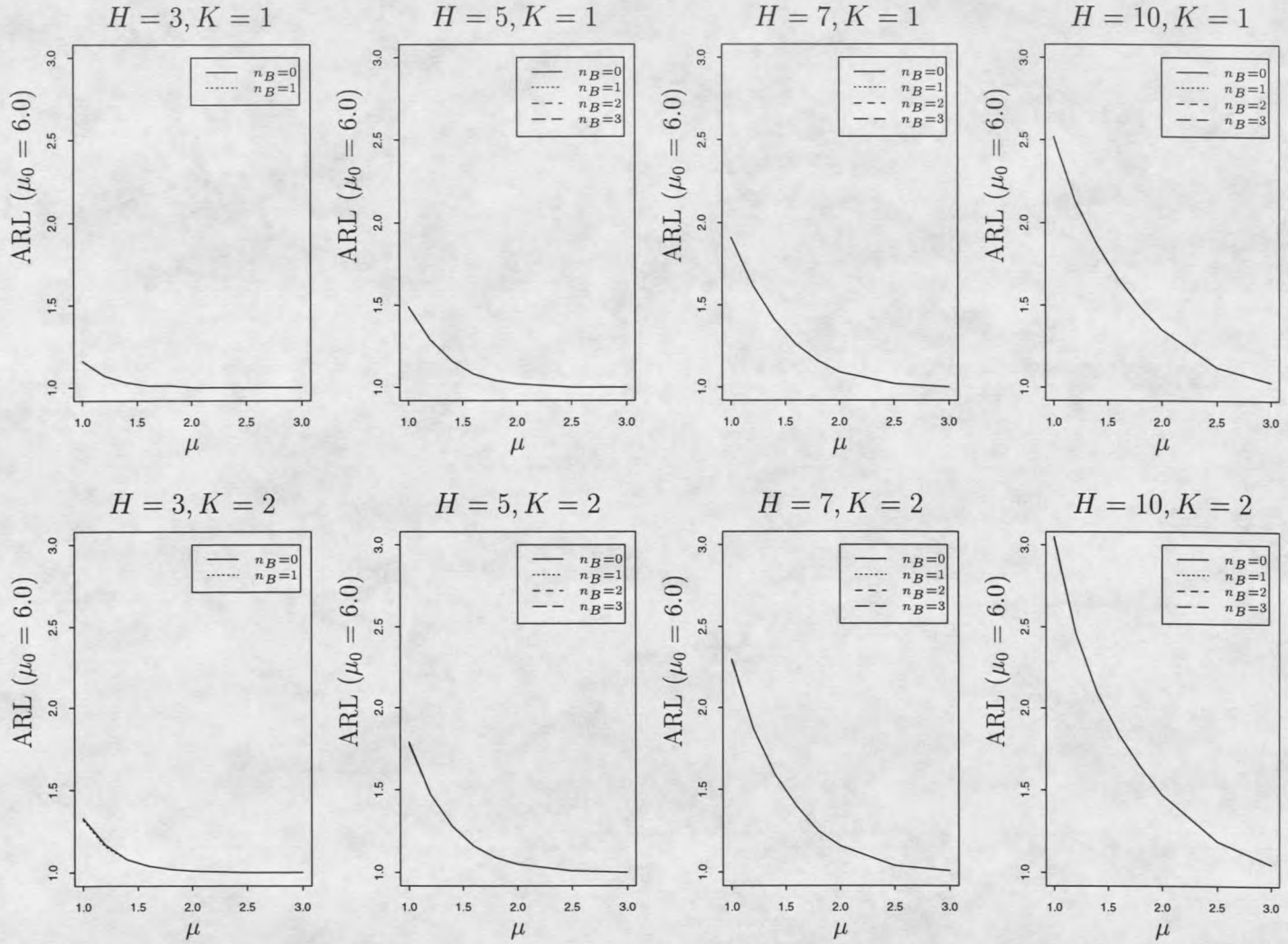


**Figure 13**: ARLs for Modified (Upper) CUSUM charts for various values of $H$, $K = 1, 2$, and $W$ (determined by $n_B$), when the process experiences a step shift of $\mu$ (given in multiples of the in-control mean $\mu_0 = 6.0$).

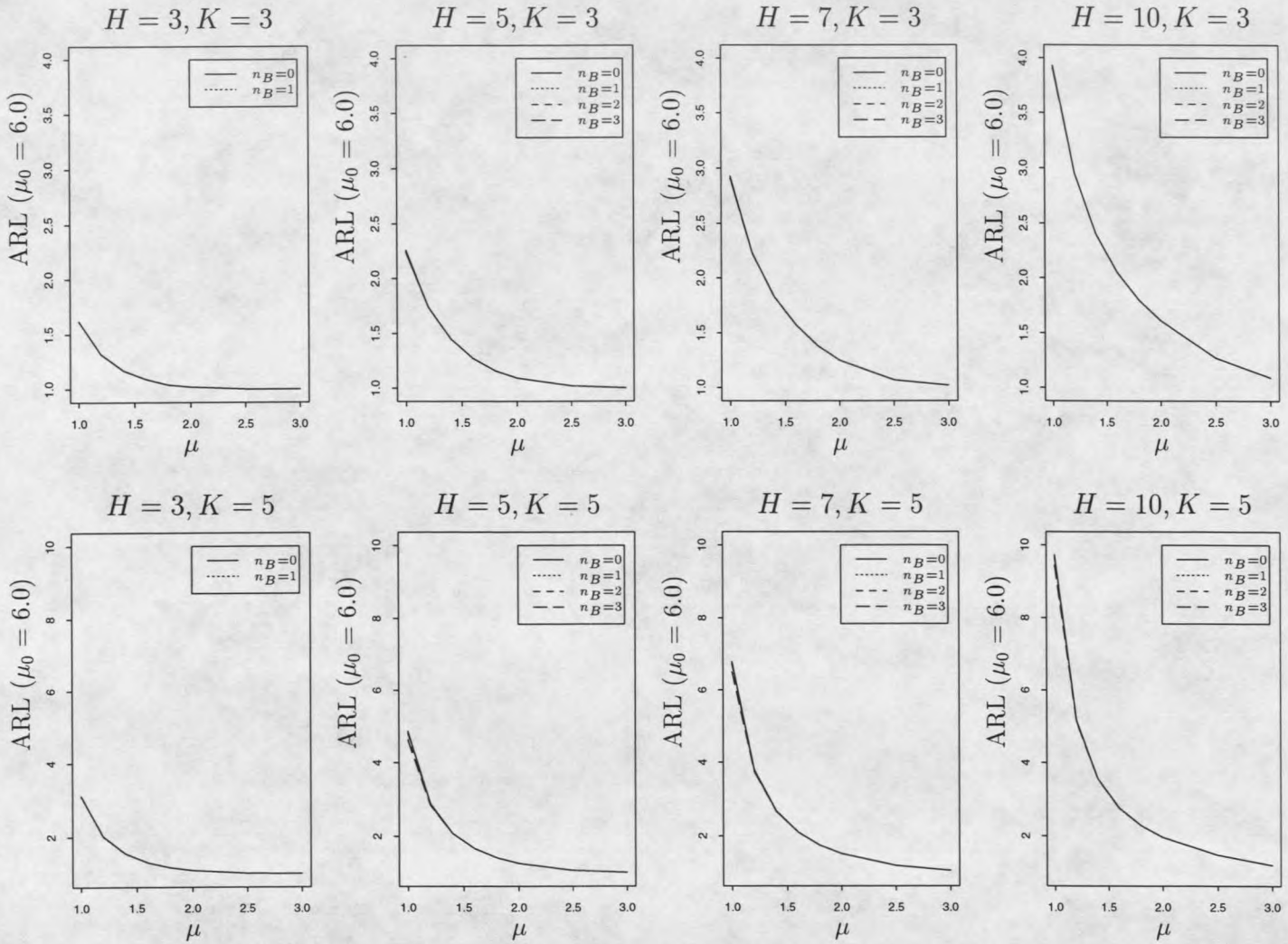


**Figure 14**: ARLs for Modified (Upper) CUSUM charts for various values of $H$, $K = 3, 5$, and $W$ (determined by $n_B$), when the process experiences a step shift of $\mu$ (given in multiples of the in-control mean $\mu_0 = 6.0$).