Surveillance of Negative Binomial and Bernoulli Processes

John L. Szarka, III

Dissertation submitted to the Faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Statistics

William H. Woodall, Chair
Scotland C. Leman
Marion R. Reynolds, Jr.
Eric P. Smith

March 25, 2011
Blacksburg, Virginia

Keywords: Adaptive threshold; Attributes data; Biosurveillance; CUSUM charts; High quality processes; Statistical process control; UGR chart; W2 method

Copyright 2011, John L. Szarka, III
Unless otherwise indicated
Surveillance of Negative Binomial and Bernoulli Processes

John L. Szarka, III

Abstract

The evaluation of discrete processes are performed for industrial and healthcare processes. Count data may be used to measure the number of defective items in industrial applications or the incidence of a certain disease at a health facility. Another classification of a discrete random variable is for binary data, where information on an item can be classified as conforming or nonconforming in a manufacturing context, or a patient’s status of having a disease in health–related applications.

The first phase of this research uses discrete count data modeled from the Poisson and negative binomial distributions in a healthcare setting. Syndromic counts are currently monitored by the BioSense program within the Centers for Disease Control and Prevention (CDC) to provide real–time biosurveillance. The Early Aberration Reporting System (EARS) uses recent baseline information comparatively with a current day’s syndromic count to determine if outbreaks may be present. An adaptive threshold method is proposed based on fitting baseline data to a parametric distribution, then calculating an upper–tailed p–value. These statistics are then converted to an approximately standard normal random variable. Monitoring is examined for independent and identically distributed data as well as data following several seasonal patterns. An exponentially weighted moving average (EWMA) chart is also used for these methods. The effectiveness of these methods in detecting simulated outbreaks in several sensitivity analyses is evaluated.

The second phase of research explored in this dissertation considers information that can be classified as a binary event. In industry, it is desirable to have the probability of a nonconforming item, $p$, be extremely small. Traditional Shewhart charts such as the $p$–chart, are not reliable for monitoring this type of process. A comprehensive literature review of control chart procedures for this type of process is given. The equivalence between two cumulative sum (CUSUM) charts, based on geometric and Bernoulli random variables is explored. An evaluation of the unit and group–runs (UGR) chart is performed, where it is shown that the in–control behavior of this chart is quite misleading and should not be recommended for practitioners.
Dedication

To my family and friends for their unwavering support.
Acknowledgments

I would like to thank my advisor Dr. William H. Woodall for his help and guidance throughout this process. I would also like to thank the other members of my committee, Dr. Scotland Leman, Dr. Marion Reynolds, Jr., and Dr. Eric Smith. None of this would be possible without the love and support of my family. Mom, Dad, Danny, and Steve, each of you mean so much to me in ways I cannot express. Thank you Pam for always standing by me.
Contents

Abstract ii

Contents v

List of Figures x

List of Tables xii

1 Introduction 1

2 An Adaptive Threshold Method for Monitoring Medical Event Counts 4
   2.1 Introduction 5
   2.2 Recurrence Intervals 7
   2.3 The W2c Method 8
      2.3.1 The W2c Statistic 9
   2.4 Adaptive Threshold Methods 9
      2.4.1 Poisson Distribution 10
      2.4.2 Negative Binomial Distribution 11
   2.5 Recurrence Interval Simulation Study 13
      2.5.1 Poisson Counts with W2c 13


<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5.2 Negative Binomial Counts with W2c</td>
<td>16</td>
</tr>
<tr>
<td>2.6 EWMA Method</td>
<td>21</td>
</tr>
<tr>
<td>2.7 Sensitivity Analysis</td>
<td>24</td>
</tr>
<tr>
<td>2.7.1 Poisson Counts</td>
<td>26</td>
</tr>
<tr>
<td>2.7.2 Negative Binomial Counts</td>
<td>27</td>
</tr>
<tr>
<td>2.8 Comments on Recurrence Intervals</td>
<td>28</td>
</tr>
<tr>
<td>2.9 Considering Differing Parameters on Weekdays and Weekends</td>
<td>30</td>
</tr>
<tr>
<td>2.10 Discussion</td>
<td>31</td>
</tr>
<tr>
<td>3 Effects of Seasonality on Some Biosurveillance Methods</td>
<td>34</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>35</td>
</tr>
<tr>
<td>3.2 Recurrence Intervals</td>
<td>36</td>
</tr>
<tr>
<td>3.3 Monitoring Schemes</td>
<td>37</td>
</tr>
<tr>
<td>3.3.1 The W2c Method</td>
<td>37</td>
</tr>
<tr>
<td>3.3.2 An Adaptive Threshold Method</td>
<td>38</td>
</tr>
<tr>
<td>3.3.3 EWMA Methods</td>
<td>38</td>
</tr>
<tr>
<td>3.4 Seasonal Effects</td>
<td>39</td>
</tr>
<tr>
<td>3.4.1 Poisson Model</td>
<td>40</td>
</tr>
<tr>
<td>3.4.2 Negative Binomial Model</td>
<td>41</td>
</tr>
<tr>
<td>3.5 Simulation Study</td>
<td>42</td>
</tr>
<tr>
<td>3.5.1 Overview of Setup</td>
<td>42</td>
</tr>
<tr>
<td>3.5.2 In-Control Results</td>
<td>44</td>
</tr>
<tr>
<td>3.5.3 Outbreak Results</td>
<td>46</td>
</tr>
<tr>
<td>3.6 Distribution of Signals</td>
<td>53</td>
</tr>
</tbody>
</table>
4 A Review and Perspective on Control Charting for High Quality Bernoulli Processes

4.1 Introduction

4.1.1 Background

4.1.2 Various Proposed Methods

4.1.3 Advice to Practitioners

4.1.4 Outline of Chapter

4.2 Criteria for Comparing Control Chart Performance

4.3 Transformations to Normality

4.4 Shewhart Control Charts Based on the Geometric Distribution

4.4.1 Basic Approach

4.4.2 Technical Issues with Two-Sided CCC Chart

4.4.3 Other Approaches

4.4.4 Fixed-Shift and Random-Shift Models

4.5 Control Charts Based on the Negative Binomial Distribution

4.5.1 Control Limits

4.5.2 Choice of $r$

4.5.3 Other Methods

4.6 Cumulative Sum Control Charts

4.6.1 Geometric CUSUM Chart

4.6.2 Binomial CUSUM Chart

4.6.3 Bernoulli CUSUM Chart
4.6.4 The Relationship Between the Geometric and Bernoulli CUSUM
Charts ................................................................. 81
4.7 Exponentially Weighted Moving Average Control Charts .......... 82
  4.7.1 One–Sided EWMA Control Charts ............................. 82
  4.7.2 Applications .................................................. 83
4.8 The Traditional Shewhart $p$–Chart .................................. 84
  4.8.1 Issues for Small Values of $p_0$ ................................ 84
  4.8.2 Modifications .................................................. 85
4.9 Summary of Performance Comparisons ............................... 88
  4.9.1 Comparisons with CUSUM Charts ............................. 88
  4.9.2 Comparisons Without CUSUM Charts ......................... 93
4.10 Discussion .......................................................... 94

5 On the Equivalence of the Bernoulli and Geometric CUSUM Charts 98
  5.1 Introduction ........................................................ 99
    5.1.1 Methods ...................................................... 99
    5.1.2 Performance Metrics ........................................ 99
    5.1.3 Chart Equivalence Issues ................................... 100
    5.1.4 Outline of Chapter .......................................... 100
  5.2 The Bernoulli CUSUM Chart ...................................... 101
  5.3 The Geometric CUSUM Chart ..................................... 102
  5.4 Equivalence Results ............................................... 102
  5.5 Zero–State Analysis ............................................. 104
  5.6 Steady–State Analysis ............................................ 105
5.7 Simulation Studies ............................................. 106
5.7.1 Example 5.1 ................................................... 107
5.7.2 Example 5.2 ................................................... 108
5.7.3 Example 5.3 ................................................... 109
5.8 Negative Binomial CUSUM ................................. 110
5.9 Conclusions ..................................................... 112

6 A Note on the Unit and Group–Runs Chart for Bernoulli Data 115
6.1 Introduction ..................................................... 116
6.2 Design of the UGR Chart ..................................... 117
6.3 UGR Chart Performance ..................................... 118
6.4 Simulation Study .............................................. 120
6.5 Markov Chain Representation ............................. 121
6.6 Steady–State Analysis ....................................... 123
6.7 Conclusions ..................................................... 124

7 Conclusions and Future Work ................................. 126

Bibliography ......................................................... 130
List of Figures

2.1 Threshold Curves Based on RI's: Poisson Counts .......................... 14
2.2 RI Thresholds for Different Baselines – Poisson Counts .................. 15
2.3 P-values – 7 days – Poisson Counts ............................................. 17
2.4 P-values – 14 days – Poisson Counts ............................................. 18
2.5 P-values – 28 days – Poisson Counts ............................................. 19
2.6 Thresholds Curves Based on RI's: Negative Binomial Counts .......... 20
2.7 RI Thresholds for Different Baselines – Negative Binomial Counts .... 22
2.8 P-values – 7 days – Negative Binomial Counts ................................. 23
2.9 P-values – 14 days – Negative Binomial Counts ................................. 24
2.10 P-values – 28 days – Negative Binomial Counts ............................... 25
2.11 Regular Versus Clustered Outbreaks ............................................. 31
3.1 Average Count for Poisson Seasonal Data During Two-Year Period .... 44
3.2 Average Count for Negative Binomial Seasonal Data During Two-Year
   Period .................................................................................. 45
3.3 W2c RI Threshold Functions for Simulated Data Sets and BioSense Standard 46
3.4 Shewhart RI Threshold Functions for Different Baselines – Poisson Counts 47
3.5 Shewhart RI Threshold Functions for Different Baselines – Negative Bi-
   nomial Counts .................................................................. 48
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.6</td>
<td>Adaptive Threshold EWMA Statistics for Seasonal Pattern N4 – 75 Days</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>of Monitoring with an Outbreak of $\delta = 0.5$ Lasting 7 Days</td>
<td></td>
</tr>
<tr>
<td>3.7</td>
<td>Distributions of False Alarm Rates, Seasonal Pattern P4, $n = 14$</td>
<td>55</td>
</tr>
<tr>
<td>3.8</td>
<td>Distributions of False Alarm Rates, Seasonal Pattern N3, $n = 7$</td>
<td>56</td>
</tr>
<tr>
<td>3.9</td>
<td>Distributions of Signal Rates, Seasonal Pattern P4, $n = 14$, $\delta = 0.5$</td>
<td>57</td>
</tr>
<tr>
<td>3.10</td>
<td>Distributions of Signal Rates, Seasonal Pattern N3, $n = 7$, $\delta = 0.5$</td>
<td>58</td>
</tr>
<tr>
<td>4.1</td>
<td>Nonmaximal In–Control ARL Plots for Two–Sided CCC Chart from Xie et al. (2000a)</td>
<td>70</td>
</tr>
<tr>
<td>4.2</td>
<td>ARL Curves Based on Two–Sided CCC Chart Illustrated by Chen (2009)</td>
<td>71</td>
</tr>
<tr>
<td>4.3</td>
<td>Random Shift Model – The “Shifting Sample” from Wu and Spedding (1999)</td>
<td>74</td>
</tr>
<tr>
<td>4.4</td>
<td>CUSUM Charts from Reynolds and Stoumbos (2000) Based on a Simulation Example</td>
<td>91</td>
</tr>
<tr>
<td>5.1</td>
<td>Illustration from Bourke (2001a) on the Fixed–Shift and Random–Shift Models</td>
<td>106</td>
</tr>
</tbody>
</table>
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Baseline Days used for Week $k$ when $n = 7$</td>
<td>8</td>
</tr>
<tr>
<td>2.2</td>
<td>Expected Number of Samples with all Counts Equal to Zero in Poisson Model</td>
<td>11</td>
</tr>
<tr>
<td>2.3</td>
<td>Negative Binomial Parameters</td>
<td>12</td>
</tr>
<tr>
<td>2.4</td>
<td>Negative Binomial – Proportion of Baseline Windows Where $\bar{x}_t \geq \hat{\sigma}_t^2$ Across Different Baseline Lengths</td>
<td>16</td>
</tr>
<tr>
<td>2.5</td>
<td>Proportion of Signals – Transient Shift in Poisson Case – Shewhart</td>
<td>27</td>
</tr>
<tr>
<td>2.6</td>
<td>Proportion of Signals – Transient Shift in Poisson Case – EWMA</td>
<td>28</td>
</tr>
<tr>
<td>2.7</td>
<td>Proportion of Signals – Transient Shift in Negative Binomial Case – Shewhart</td>
<td>29</td>
</tr>
<tr>
<td>2.8</td>
<td>Proportion of Signals – Transient Shift in Negative Binomial Case – EWMA</td>
<td>30</td>
</tr>
<tr>
<td>3.1</td>
<td>Seasonal Patterns Distributions</td>
<td>43</td>
</tr>
<tr>
<td>3.2</td>
<td>Negative Binomial Equivalence Conditions</td>
<td>46</td>
</tr>
<tr>
<td>3.3</td>
<td>Proportion of Signals – Transient Outbreak in Poisson Case – Shewhart</td>
<td>49</td>
</tr>
<tr>
<td>3.4</td>
<td>Proportion of Signals – Transient Outbreak in Poisson Case – EWMA</td>
<td>50</td>
</tr>
<tr>
<td>3.5</td>
<td>Proportion of Signals – Transient Outbreak in Negative Binomial Case – Shewhart</td>
<td>51</td>
</tr>
<tr>
<td>3.6</td>
<td>Proportion of Signals – Transient Outbreak in Negative Binomial Case – EWMA</td>
<td>52</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

Control charting procedures are a common method used in statistical process control (SPC) for evaluating the current state of a process. Attributes control charts may involve using count data, where in manufacturing processes, the number of flaws or errors present in a current sample are observed. In biosurveillance, the number of daily cases of a particular syndrome may be monitored at a health facility. In this class of charts, scenarios may also be considered where binary information is available for each item. In industrial applications, a value of one may represent a nonconforming (defective) item, whereas a zero indicates a conforming (non–defective) item. This extends to healthcare applications, where an event could be classified as a patient’s status of having a disease or congenital malformation.

In the next two chapters of this dissertation, the focus is for monitoring discrete count data in a healthcare setting. The Centers for Disease Control and Prevention (CDC) developed its BioSense program with the intent of providing real–time biosurveillance. The Early Aberration Reporting System (EARS) uses the W2count (W2c) method to determine if a current day’s syndromic count is large relative to a baseline period of counts, and hence, an indication of a possible outbreak. An adaptive threshold method is considered as an alternative method, where the baseline information was used to fit a parametric distribution. A one–sided p–value can be computed and converted to an approximately standard normal random variable by using the inverse cumulative distribution function (cdf) to use for monitoring. An exponentially weighted moving average
(EWMA) method may be a useful tool for each of these surveillance methods since they are variants of Shewhart methods, where only information from the most recent statistic is used in decision-making. These metrics were compared by their effectiveness in detecting transient outbreaks of syndromic counts. The underlying distribution of the daily syndromic counts were either Poisson or negative binomial.

The formatting of this dissertation is the work of several manuscripts that are the contents of the next five chapters. In Chapter 2, the daily syndromic counts follow a Poisson or negative binomial distribution and are assumed to be independent and identically distributed (iid). In many health-related applications, it is common for data to follow seasonal patterns. In Chapter 3, seasonal patterns are included in the data stream, which may be more representative of many diseases encountered in practice. A surveillance method will alert a possible outbreak if the current day’s statistic exceeds a fixed threshold. The thresholds were set such that these competing charts have equivalent performance when no outbreak is present based on their recurrence interval (RI) values. An empiric RI can be formed by taking a large baseline data set and calculating the inverse of the probability that a statistic exceeds a given threshold. The RI’s were formed using simulations and the appropriate thresholds were constructed.

In Chapters 4–6, it is of interest to monitor a random variable that can be classified as a binary event. This is often monitored as a sequence of Bernoulli random variables, where items are classified as nonconforming or conforming. In practice, it is often necessary that this process have a very small nonconforming rate, \( p \). This is known as a high quality process. Traditional SPC approaches for monitoring a proportion, such as the Shewhart \( p \)-chart or equivalent \( np \)-chart, are inadequate for monitoring this type of process. In Chapter 4, a comprehensive review and recommendations for practitioners for monitoring Bernoulli data is provided.

A cumulative sum (CUSUM) chart is a popular control charting technique where the chart can be set up to be optimal for detecting a specified shift in the parameter(s) of a given distribution from an in–control state to an out–of–control state. The Bernoulli CUSUM and geometric CUSUM charts have been shown in the literature to be effective methods for monitoring high quality Bernoulli data. These two charts can be setup to be mathematically equivalent. In Chapter 5, this equivalence is explored for a number of different monitoring scenarios.
The unit and group–runs (UGR) chart was introduced by Gadre and Rattihalli (2005) for monitoring high quality Bernoulli processes. Based on extensive analysis, the UGR chart has outperformed the Bernoulli CUSUM chart significantly in detecting an upward shift in $p$. In Chapter 6, this chart is carefully evaluated, where the in–control properties of the UGR chart are derived and shown to have misleading results.

Concluding remarks and proposals for future work are given in Chapter 7. This includes extensions of applications for the methods introduced in this dissertation that should be considered, as well as proposed methods for other related topics as well.
Chapter 2

An Adaptive Threshold Method for Monitoring Medical Event Counts

Abstract

In this chapter we examine some of the methodologies implemented by the Centers for Disease Control and Prevention’s (CDC) BioSense program. The program uses data from hospitals and public health departments to detect outbreaks using the Early Aberration Reporting System (EARS). The EARS method W2 allows one to monitor syndrome counts (W2count) from each source and the proportion of counts of a particular syndrome relative to the total number of visits (W2rate). We investigate the performance of the W2c method, which is designed using an empiric recurrence interval (RI), with simulated parametric data. Counts from the Poisson and negative binomial distributions are generated, and used to examine W2c properties. An adaptive threshold monitoring method is introduced based on fitting sample data to the above distributions, then converting the current value to a Z-score through a p-value. We compare the thresholds required to obtain given values of the recurrence interval for different sets of parameter values. We then simulate one-week outbreaks in our data and calculate the proportion of times these methods correctly signal an outbreak using Shewhart and exponentially weighted moving average (EWMA) charts. Our results indicate the adaptive threshold method gives more consistent statistical performance across different parameter sets and amounts of baseline historical data used for computing the statistics. For the sensitivity analysis, the EWMA chart is superior to its Shewhart counterpart in nearly all cases and the adaptive threshold method tends to outperform the W2c method.
2.1 Introduction

The Centers for Disease Control and Prevention (CDC) established the BioSense program with the intent of providing real-time biosurveillance for early event detection \((\text{CDC}(2009))\). The Early Aberration Reporting System (EARS) is used for finding statistical anomalies by the CDC. Currently, hundreds of hospitals and public health departments across the United States provide data to BioSense where the EARS methods are used for determining whether or not syndromic outbreaks have occurred \((\text{CDC}(2008))\).

There are two methodologies EARS uses for detecting these types of outbreaks. The \(W_2\) count \((W_2c)\) method focuses on the number of cases of a particular syndrome on a given day. The \(W_2\) rate \((W_2r)\) method is based on the proportion of visits corresponding to a particular syndrome which accounts for the total number of visits to a health facility on a given day. In this chapter we focus on the \(W_2c\) method. For further analysis of the \(W_2r\) method, see \(\text{Szarka et al.}(2011)\) and \(\text{Gan}(2010)\). The \(W_2\) statistics are based on 7-day moving windows. The short baseline is intended to accumulate recent information on a given syndrome. A 2-day lag is also incorporated in the calculation of the statistics, meaning the previous two days are not included in the baselines. If the current day’s value is large relative to the baseline data, this will result in a large \(W_2\) statistic. If a \(W_2\) value exceeds a specified threshold, an alarm is given.

The \(W_2\) statistics are calculated separately for weekdays and weekends. This is done because many healthcare facilities have fewer visits during weekends. However, in our study, we examine the simplified case where weekday and weekend counts follow the same distribution. Syndrome counts monitored by the \(W_2c\) method in our simulation studies are independent and identically distributed (iid) Poisson or negative binomial random variables.

The \(W_2c\) method is similar to the \(C_2\) method formerly used with BioSense. The previous methods for monitoring counts include methods \(C_1\), \(C_2\), and \(C_3\) and can be found in Table 5 of the CDC’s User Guide \((\text{CDC}(2006))\). These methods do not partition the data by weekday and weekend. The reader is referred to \(\text{Fricker et al.}(2008)\), \(\text{Hutwagner et al.}(2003, 2005a,b)\), \(\text{Zhu et al.}(2005)\), and \(\text{Watkins et al.}(2008, 2009b)\) for analyses of these methods. See \(\text{Tokars et al.}(2009)\) for a study on the \(W_2\) methods.
A modification of the adaptive threshold method proposed by Lambert and Liu (2006) for computer networking is also considered in our study. Using the baseline data, the parameters of a distribution can be fit using maximum likelihood (ML) or method of moments (MOM) estimators. The current day’s count has an upper–tail p–value calculated from the estimated distribution. A Z–score can be computed by taking the inverse standard normal cumulative distribution function (CDF) of one minus the p–value, giving an approximately standard normal statistic when there is no outbreak. The successive Z–scores are used for process monitoring.

The W2 methods rely on the use of an empiric recurrence interval (RI). Kleinman et al. (2005) explained that if monitoring of a process continues without interruption after any alarm, the RI is the fixed number of time periods for which the expected number of false alarms is one. Table 3 of the CDC’s Hospital User Guide (CDC (2009)) gives the W2 thresholds associated with a range of RI values from 10 to 2000 when \( n = 7 \), where \( n \) is the length of the baseline. Using our simulations, we compute our own empiric RI values for the W2c method and compare these to the results from BioSense. We also compare the RI results of the W2c and adaptive threshold methods across different parameter sets and baseline lengths.

The W2c method is evaluated using baselines of \( n = 7, 14, \) and 28 days. These baseline lengths were used by Tokars et al. (2009), but they used no more than 56 days of historical data. Therefore for weekends, only eight weeks of data were available, leading to only 16 days of data in their baseline. The current baseline of \( n = 7 \) used by BioSense is a short baseline that in many instances is insufficient for estimation. However, a baseline that is too long will mitigate the ability of the statistic to adjust to seasonal variation. This can lead to a decreased chance in signaling an outbreak.

A separate simulation study analyzes the ability of the W2c and adaptive threshold methods to detect outbreaks, a sensitivity analysis. This is done by generating samples from a reference distribution for several weeks, then systematically injecting a specified increase in the average number of syndrome counts. The outbreaks are assumed to last for 7 days. It is of interest to determine how frequently the two methods signal given different magnitudes in shifts and baseline window sizes. Traditional approaches for detecting outbreaks focus on the current day’s statistic exceeding a particular threshold. However, we can also use a statistic that accumulates information over time. In our
study we consider use of the exponentially weighted moving average (EWMA) statistic for detecting outbreaks as well.

We discuss the recurrence interval in more detail in Section 2.2. The W2c method is reviewed in Section 2.3. In Section 2.4, we examine the adaptive threshold methods for the Poisson and negative binomial distributions. The results of an initial simulation study are reviewed in Section 2.5. The EWMA approach is explained in greater detail in Section 2.6 with the sensitivity analysis in Section 2.7. A discussion follows in Section 2.8.

2.2 Recurrence Intervals

Using a large number of values of a surveillance method statistic, empiric recurrence intervals (RI’s) can be estimated. Nothing is explicitly mentioned by the CDC regarding which type(s) of syndromic data or how much data were used to calculate their RI thresholds for the W2 methods. It is mentioned that “the frequency distributions used to calculate the RI’s will be updated periodically” (CDC (2009)). For any frequency distribution, we can find the percentage of days with monitoring values that exceed a given threshold. Taking the reciprocal of this percentage gives the corresponding RI value. For instance, if five percent of our monitoring statistic values exceed 3.4 in a frequency distribution, this threshold corresponds to a RI of 20. Note that the thresholds for the W2c and W2r methods differ.

According to BioSense (CDC (2009)), parametric methods for developing recurrence intervals are not valid in this application because public health data are typically non-normal. Several discrete distributions may be of use, however, to better understand the W2c method. If different types of simulated data were to yield W2c RI thresholds similar to those found in the CDC handbook, the W2c method would be considered robust to differences in the characteristics of the different input data streams. Tokars et al. (2009) noted that BioSense uses past data to obtain empiric RI’s that may contain outbreaks. In our initial simulations we assume there are no outbreaks. The RI threshold functions for the W2c method will also be compared to those found for the adaptive threshold method through a simulation study. These results are discussed in Section 2.5.
2.3 The W2c Method

The W2c method is based on a centered and scaled statistic, using expected values and standard deviations estimated using past data. A minimum value of one is set for the standard deviation. We consider a baseline of $n$ days in our simulations, where $n = 7$, 14, and 28.

Using the 2-day lag when partitioning by weekday and weekend is more complicated than in the case of the C2 method used in the past. For example, for Tuesday, Sunday and Monday would be the lagged days using the C2 method. The W2c method uses the previous Friday as the most recent day in the baseline period. For a given week $k$, Table 2.1 shows all of the days used in the baselines when $n = 7$. There are four distinct baseline groups formed for each week.

The W2c method will signal whenever the corresponding statistic exceeds a given threshold, say $h_c$. These thresholds will be determined from the RI threshold functions obtained from our first simulation study, which is illustrated in Section 2.5.

<table>
<thead>
<tr>
<th>Day</th>
<th>Baseline Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mon($k$)</td>
<td>Thu-Fri($k - 2$), Mon-Fri($k - 1$)</td>
</tr>
<tr>
<td>Tue($k$)</td>
<td>Thu-Fri($k - 2$), Mon-Fri($k - 1$)</td>
</tr>
<tr>
<td>Wed($k$)</td>
<td>Thu-Fri($k - 2$), Mon-Fri($k - 1$)</td>
</tr>
<tr>
<td>Thu($k$)</td>
<td>Fri($k - 2$), Mon-Fri($k - 1$), Mon($k$)</td>
</tr>
<tr>
<td>Fri($k$)</td>
<td>Mon-Fri($k - 1$), Mon-Tue($k$)</td>
</tr>
<tr>
<td>Sat($k$)</td>
<td>Sun($k - 4$), Sat-Sun($k - 3$), Sat-Sun($k - 2$), Sat-Sun($k - 1$)</td>
</tr>
<tr>
<td>Sun($k$)</td>
<td>Sun($k - 4$), Sat-Sun($k - 3$), Sat-Sun($k - 2$), Sat-Sun($k - 1$)</td>
</tr>
</tbody>
</table>
2.3.1 The W2c Statistic

Let $x_t$ be the count of a specific syndrome for day $t$. The baseline data for day $t$ is dependent on its day of the week, as shown in Table 2.1. The W2c value for day $t$ is

$$W2c(t) = \frac{x_t - \bar{x}_t}{s_t},$$

where $\bar{x}_t$ and $s_t$ are the sample mean and standard deviation from the baseline period. These values are expressed as

$$\bar{x}_t = \frac{1}{n} \sum_{i=1}^{n} y_{it}$$

and

$$s_t = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (y_{it} - \bar{x}_t)^2},$$

where $y_{it}, i = 1, 2, \ldots, n$, correspond to the eligible baseline data counts for day $t$. If $s_t$ is less than one, it is reassigned a value of one. This will occur most often with rare syndromes, where we expect the counts to be low with many observed zeros. In some cases the standard deviation may equal zero, which would leave the W2c statistic undefined without this adjustment.

2.4 Adaptive Threshold Methods

An adaptive threshold approach used by Lambert and Liu (2006) for computer network monitoring leads to an alternative to the W2c method. Using the same baseline information as the W2c method, we can estimate the parameters of an assumed underlying parametric distribution. There are many methods of estimating these parameters, but we focus on the maximum likelihood (ML) and method of moments (MOM) estimators.

The simulations for this adaptive threshold method will demonstrate some of the same properties as the W2c method. With shorter baseline periods, we may have poor estimates of the true parameter values. Since we know the true distribution in our simulations, we can determine whether or not the estimates are accurate.

For day $t$, an upper–tail p–value, $P_t$, is computed based on the distribution with the set of estimated parameters, say $\hat{\theta}_t$. Then $P_t$ is approximately distributed uniformly over
[0,1] if the underlying distribution is stable. Using the inverse standard normal cdf, we obtain a standard normal Z-score, $Z_t$, with an approximate mean of zero and variance of one. The equations for these values for a random variable $S_t$ with observed value $s_t$ are

$$P_t = Pr(S_t \geq s_t | \hat{\theta}_t),$$

and

$$Z_t = \Phi^{-1}(1 - P_t).$$

For a Shewhart approach, a signal is given when $Z_t$ exceeds $h_{AT}$, where $h_{AT}$ is a specified threshold value. We also study the properties of an EWMA chart based on the Z-scores.

In the subsections below, the probability distributions we will use and their properties are introduced. The ML and MOM estimators are given for each distribution and we discuss the problems that may arise in estimation.

### 2.4.1 Poisson Distribution

One of the common discrete distributions used in modeling count data is the Poisson with probability mass function (pmf)

$$f(x_t | \lambda) = \frac{e^{-\lambda} \lambda^{x_t}}{x_t!},$$

where $x_t = 0, 1, 2, \ldots$ and $\lambda > 0$ is the expected value and variance of the distribution. Small values of $\lambda$ can be a problem for our calculation of W2c statistics since we expect to observe many zeros. Therefore, the lower bound of unity set on the standard deviation may need to be employed often in this situation.

With shorter baselines, the estimate of $\lambda$ will not always be accurate. Longer baselines will tend to give us better estimates of the parameter. The ML and MOM estimator for the Poisson parameter on day $t$ is $\hat{\lambda}_t = \bar{x}_t$. For small counts we have a problem if $\hat{\lambda}_t = 0$. This occurs if each count in the baseline equals zero. We would then expect all counts to be equal to zero with probability one. Any positive count for $x_t$ would then result in
Table 2.2: Expected Number of Samples with all Counts Equal to Zero in Poisson Model

<table>
<thead>
<tr>
<th>( \lambda )</th>
<th>( n = 7 )</th>
<th>( n = 14 )</th>
<th>( n = 28 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>18118.43</td>
<td>547.1292</td>
<td>0.498917</td>
</tr>
<tr>
<td>1.0</td>
<td>547.1292</td>
<td>0.498917</td>
<td>4.15 \times 10^{-7}</td>
</tr>
<tr>
<td>1.5</td>
<td>16.52187</td>
<td>4.55 \times 10^{-4}</td>
<td>3.45 \times 10^{-13}</td>
</tr>
<tr>
<td>2.0</td>
<td>0.498917</td>
<td>4.15 \times 10^{-7}</td>
<td>2.87 \times 10^{-19}</td>
</tr>
<tr>
<td>2.5</td>
<td>0.015066</td>
<td>3.78 \times 10^{-10}</td>
<td>2.39 \times 10^{-25}</td>
</tr>
<tr>
<td>3.0</td>
<td>4.55 \times 10^{-4}</td>
<td>3.45 \times 10^{-13}</td>
<td>1.98 \times 10^{-31}</td>
</tr>
<tr>
<td>3.5</td>
<td>1.37 \times 10^{-5}</td>
<td>3.15 \times 10^{-16}</td>
<td>1.65 \times 10^{-37}</td>
</tr>
<tr>
<td>4.0</td>
<td>4.15 \times 10^{-7}</td>
<td>2.87 \times 10^{-19}</td>
<td>1.37 \times 10^{-43}</td>
</tr>
<tr>
<td>4.5</td>
<td>1.25 \times 10^{-8}</td>
<td>2.62 \times 10^{-22}</td>
<td>1.14 \times 10^{-49}</td>
</tr>
<tr>
<td>5.0</td>
<td>3.78 \times 10^{-10}</td>
<td>2.39 \times 10^{-25}</td>
<td>9.48 \times 10^{-56}</td>
</tr>
</tbody>
</table>

an infinite value for \( Z_t \).

We must ensure that \( \lambda \) is sufficiently large so that \( \hat{\lambda}_t > 0 \), which occurs when each count in the baseline sample is zero. If the daily counts are iid, the probability of this event occurring can easily be calculated as

\[
P(x_1, \ldots, x_n = 0) = P(x = 0)^n.
\]

Multiplying this value by the number of groups in our study, we can find how often we expect this condition to occur. This result is summarized in Table 2.2, where 3.0 may be the smallest recommended value for \( \lambda \) to use with a 7-day baseline period. As \( n \) increases, a smaller value of \( \lambda \) is acceptable.

### 2.4.2 Negative Binomial Distribution

A more realistic discrete distribution we can use for public health count data is the negative binomial with pmf

\[
f(x_t|r, p) = \frac{\Gamma(x_t + r)}{\Gamma(r)x_t!}p^r(1 - p)^{x_t}.
\]
Table 2.3: Negative Binomial Parameters

<table>
<thead>
<tr>
<th>Case</th>
<th>$r$</th>
<th>$p$</th>
<th>$\mu$</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
<td>.5</td>
<td>20</td>
<td>6.32</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>.1</td>
<td>900</td>
<td>94.87</td>
</tr>
<tr>
<td>3</td>
<td>350</td>
<td>.8</td>
<td>87.5</td>
<td>10.46</td>
</tr>
<tr>
<td>4</td>
<td>60</td>
<td>.6</td>
<td>40</td>
<td>8.16</td>
</tr>
</tbody>
</table>

Note that $x_t = 0, 1, 2, \ldots$, $r > 0$ and $0 < p < 1$. The mean and variance of this distribution are $\mu = r(1-p)/p$ and $\sigma^2 = r(1-p)/p^2$, respectively. We note that the Poisson distribution is a limiting case of the negative binomial distribution. As $r \to \infty$ and $p \to 1$ such that $r(1-p)$ is a constant, then $x_t \to Poi(\lambda = r(1-p))$.

A two–parameter distribution allows us to have greater flexibility in fitting the data. This is an advantage over the Poisson, whose one–parameter fit may not be appropriate in many circumstances. The negative binomial best fits count data that is skewed–right and overdispersed relative to a Poisson distribution. We may expect this in public health settings. There are a number of sample data sets available on BioSense’s website CDC (2004), which provide useful summary statistics of syndromic data. However, these data sets contain outbreaks, trends, and seasonality, which are not considered in our simulation study for developing the empiric RI thresholds.

The adaptive threshold method is more difficult to implement in the negative binomial case since we have more restrictions on our parameter space, with $r > 0$ and $0 < p < 1$. When using a short baseline to construct these estimates, it is more likely to lead to problems. We use the MOM estimator for the negative binomial. The MOM estimators for the negative binomial distribution on day $t$ are

$$\hat{p}_t = \frac{\bar{x}_t}{\hat{\sigma}_t^2} \quad \hat{r}_t = \frac{\bar{x}_t^2}{\hat{\sigma}_t^2 - \bar{x}_t},$$

where $\hat{\sigma}_t^2 = \sum_{i=1}^n [(y_{it} - \bar{x}_t)^2]/n$. The domain of both parameters is violated when $\bar{x}_t \geq \hat{\sigma}_t^2$. Through simulation, we determine how likely this is to happen, and this is discussed in Section 2.5 for the four different combinations of $(r, p)$ considered in our study. These cases are shown in Table 2.3.
Using a technique implemented by Watkins et al. (2009a), if $\bar{x}_t \geq \hat{\sigma}_t^2$, we set $\hat{\sigma}_t^2 = 1.05\bar{x}_t$. This is an ad-hoc method that gives us valid estimates for our parameters. If this condition is used, then $\hat{p}_t = 1/1.05 = .9524$ and $\hat{r}_t = 20\bar{x}_t$. To mitigate problems with estimation, Lambert and Liu (2006) used a technique of outlier removal by replacing an extremely large outlier that is beyond the .9999 quantile of the estimated distribution with a randomly generated count beyond the .99 quantile of the fitted distribution. This was also done for extremely small counts. We focus only on large counts in our simulation study such that if a large count generates a value of $Z_t = \infty$, then $x_t$ is replaced with a random count beyond the 99th percentile of our fitted distribution. See Nadarajah and Kotz (2006) for a function for computing quantities from truncated distributions in the programming language R.

### 2.5 Recurrence Interval Simulation Study

In this section, we report the results of a simulation study for the Poisson and negative binomial distributions. A total of 150,000 weeks of data were simulated for each parameter combination considered. In all simulations, we examined the case where weekday and weekend counts follow the same distribution. The RI threshold functions are compared for the W2c, adaptive threshold, and BioSense methods. The BioSense threshold values were obtained from Table 3 of the CDC’s Hospital Data User Guide (CDC (2009)) when $n = 7$.

#### 2.5.1 Poisson Counts with W2c

For the Poisson simulation we used values of $\lambda = 5$, 10, 20, 50, and 100. Since $\lambda$ is sufficiently large, we need not worry about the case where $\hat{\lambda}_t = 0$. Figure 2.1 shows the thresholds needed to achieve specific RI values for the W2c method, while also displaying the BioSense standard. The BioSense threshold function differs greatly from those of the W2c method, lying to the far right in the plot. The BioSense data used in constructing these empiric RI’s are likely overdispersed compared to the Poisson distribution. This may explain why the thresholds differ so greatly, along with the fact that outbreak data are also used to determine the Biosense empiric thresholds.
Figure 2.1: Threshold Curves Based on RI’s: Poisson Counts

In Figure 2.2, we examine how the RI thresholds for the W2c and Z-score values change when different baseline sizes are used by constructing side-by-side plots for easy comparisons. For the $n = 7$ case, the thresholds for the adaptive threshold method tend to be lower than those for the W2c method for a given value of the RI. The adaptive threshold functions appear closer across values of $\lambda$ than those for the W2c method as well. As $n$ increases, the threshold values of the Z-scores stay reasonably consistent across $n$, while the W2c threshold values decrease as $n$ increases. Considering the set of parameters and a given window of historical data, the Z-score threshold functions are considerably less variable than those of the W2c method as $\lambda$ changes. Thus, the adaptive threshold method does, in fact, adapt to changes in the underlying average count.

A histogram of the p-values calculated using the adaptive threshold method are given in Figures 2.3–2.5. With a stable process, this distribution should closely follow a uniform distribution on [0,1]. The dashed line indicates the expected value of the density of p-values would be if this condition is satisfied. For smaller values of $\lambda$, this did not appear to be the case, particularly with p-values close to one. This phenomenon occurred because for any $x_t = 0$, $P_{t} = 1$ regardless of the estimate $\hat{\lambda}_t$. As $n$ increased, the
Figure 2.2: RI Thresholds for Different Baselines – Poisson Counts
Table 2.4: Negative Binomial – Proportion of Baseline Windows Where $\bar{x}_t \geq \hat{\sigma}_t^2$ Across Different Baseline Lengths

<table>
<thead>
<tr>
<th>Case</th>
<th>$(r, p)$</th>
<th>7-days</th>
<th>14-days</th>
<th>28-days</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(20, .5)</td>
<td>.26091</td>
<td>.09957</td>
<td>.02008</td>
</tr>
<tr>
<td>2</td>
<td>(100, 1)</td>
<td>.00556</td>
<td>.00002</td>
<td>.0</td>
</tr>
<tr>
<td>3</td>
<td>(350, .8)</td>
<td>.53143</td>
<td>.40675</td>
<td>.28758</td>
</tr>
<tr>
<td>4</td>
<td>(60, .6)</td>
<td>.35235</td>
<td>.18464</td>
<td>.06648</td>
</tr>
</tbody>
</table>

distribution of p-values appeared to become more uniform, with the exception of cases where $P_t = 1$.

2.5.2 Negative Binomial Counts with W2c

The parameters listed in Table 2.3 are used for our negative binomial study. With all cases in the negative binomial distribution, $\mu < \sigma^2$. However, this is not necessarily going to be reflected in the parameter estimates when sampling from this distribution. As $p$ approaches one, we have a case where the mean and variance are nearly equal. Therefore, it would be more likely to obtain a sample where $\bar{x}_t \geq \hat{\sigma}_t^2$, indicating the sample is “underdispersed”.

Over the course of our simulation study, there are four different baselines for each week, giving many parameter estimates ($\hat{r}_t, \hat{p}_t$). Table 2.4 shows how often $\bar{x}_t \geq \hat{\sigma}_t^2$, affecting the estimation of our parameters. It is most common in Case 3 (350, .8), where it occurs a majority of the time when $n = 7$. For smaller values of $p$ and larger values of $n$, we obtain better results.

Since we are estimating a two-parameter distribution with a limited baseline, it is not surprising that we are not obtaining good estimates in some cases. A larger window is needed in most cases since it becomes more likely the sample is overdispersed as $n$ increases. For larger values of $n$, these estimates are greatly improved. The distribution of $\hat{r}_t$ tends to still be bimodal, but with lower frequencies of unusually large estimates. With fewer estimates of .9524, the distribution of $\hat{p}_t$ is also improved. Tokars et al. (2009) found that use of $n = 14$ and $n = 28$ led to better W2 performance based on their study.
Figure 2.3: P–values – 7 days – Poisson Counts
Figure 2.4: P–values – 14 days – Poisson Counts
Figure 2.5: P–values – 28 days – Poisson Counts
The W2c threshold functions are compared to BioSense’s standard in Figure 2.6. Note that the threshold functions are closer to BioSense’s than in the Poisson case. With $n = 7$, the use of the negative binomial distribution may result in more extremely large W2c and adaptive threshold statistic values than with the Poisson distribution.

We examine the RI threshold functions for the W2c and Z-score methods as $n$ varies in Figure 2.7. When $n = 7$, the W2c threshold functions are closer to one another than those for the adaptive threshold method based on the Z-scores. While Case 2 results in the best estimates of its parameters, its large mean and variance separate it from the rest of the negative binomial threshold functions in the adaptive threshold case. The W2c threshold functions are grouped tighter together than those for the Z-scores, except when $n = 28$. If there is a sufficient amount of historical data, we can effectively estimate the parameters and have Z-score threshold values that are close to one another over different sets of parameters. Similar to the Poisson case, there is a significant drop in the threshold values of W2c when going from $n = 7$ to $n = 14$. Figure 2.7 shows that the performance of the adaptive threshold method is not as robust to changes in
parameter values relative to the W2c method with the negative binomial distributions as with the Poisson distribution.

The p–values are now examined for the negative binomial case in Figures 2.8 – 2.10. There is more variability in the negative binomial counts relative to the Poisson, and as a result there are more frequent p–values near the boundaries of zero and one. The issue with p–values near the lower and upper bounds occurred most frequently for Case 2, which may be a result of this distribution having the greatest amount of overdispersion. As $n$ increased, this became less of a problem because of the improved estimation of the reference distribution with more observations in the baseline.

## 2.6 EWMA Method

The EWMA chart has been widely used in traditional quality control applications since it was first proposed by Roberts (1959). While the Shewhart decision rule relies on using one observation at a time, the EWMA statistic incorporates information using past observations with observations closer to the current time point given larger weights than those further back in time. For standardized variables, say $v_t$, $t = 1, 2, \ldots$, the EWMA statistics $E_t$ are

$$E_t = \alpha v_t + (1 - \alpha)E_{t-1}, \ t = 1, 2, \ldots,$$

where $\alpha$ is the weight given to the current observation and $E_0 = 0$. When $\alpha = 1$, the EWMA method reduces to a Shewhart chart. Montgomery (2009) recommended using weights between .05 and .25 for EWMA charts. Smaller values of $\alpha$ are recommended for detecting smaller shifts, and larger values are recommended for larger shifts.

In most industrial applications, a two–sided EWMA chart is used, signaling for abnormally low or large values of the EWMA statistic. However, we are only concerned with outbreaks in our application, so a one–sided chart is used. The one–sided EWMA statistics are expressed as

$$E_t = \max[0, \alpha v_t + (1 - \alpha)E_{t-1}], \ t = 1, 2, \ldots \quad (2.1)$$
Figure 2.7: RI Thresholds for Different Baselines – Negative Binomial Counts
A signal is given if $E_t \geq h_{ET}$, where $h_{ET}$ is a specified threshold. The reflecting barrier at zero is used so that the statistic does not become very small. If this is not done and an outbreak occurred when the statistic is very small, it would be more difficult to signal. For more on a one–sided EWMA method, see Crowder and Hamilton (1992).

We study the effect of using an EWMA versus the traditional Shewhart method for the W2 and adaptive threshold methods for our sensitivity analysis in Section 2.7. Lambert and Liu (2006) recommended using a one–sided EWMA chart, but did not use the reflecting barrier at zero shown in Equation (2.1) that we will consider in our sensitivity analysis. Failure to use a reflecting barrier in a one–sided EWMA chart can lead to serious inertia problems, a topic discussed by Woodall and Mahmoud (2005).
2.7 Sensitivity Analysis

We examine different types of syndromic outbreaks in this section. We are only interested in an increase in syndrome counts and rates, so one-sided methods are used. Baseline data were simulated from an in-control distribution for several weeks, then an outbreak lasting...

seven days was injected. This process was repeated 10,000 times for each scenario. For each of these transient shifts, we determined the proportion of times the methods signal during the outbreak. We must have all of the methods set to the same in–control false alarm rate to make fair comparisons. The value of the signaling threshold was derived from the results in Section 2.5, where for a given $n$ and parameter set, a threshold was found so that the RI equals 500. Therefore, we expect that when no outbreak is present, we would expect one false alarm every 500 days.

The second methods we compare are the EWMA control chart methods based on the approach explained in Section 2.6. From the simulations used in Section 2.5, we calculated the EWMA statistics from the W2c and Z–score values. A weight of $\alpha = .2$ 

Figure 2.10: P–values – 28 days – Negative Binomial Counts
was used. We constructed similar threshold functions as shown in Figures 2.2 and 2.7. We then found the thresholds corresponding to a RI of 500. The outbreak detection rates of the Shewhart and EWMA charts for the W2c and adaptive threshold methods are compared in the subsections below.

We consider a percentage shift of size 100δ in a parameter for this study, which will increase the average amount of syndromic counts per day over the course of the outbreak. The values used are δ = 0, .1, .2, .5, 1, and 2, where a zero shift indicates analysis when there is no outbreak. This no outbreak simulation was done to ensure that the proportion of signals should be approximately the same across parameter values and baseline lengths for the in–control case.

2.7.1 Poisson Counts

The increase in the Poisson parameter λ is assumed to be 100(1 + δ)% of the in–control parameter value. For Tables 2.5 and 2.6, when δ = 0, the proportion of signals is slightly higher for the Shewhart method. The signaling thresholds chosen for both the Shewhart and EWMA methods were such that if we are in–control, we expect one false alarm in 500 days. However, the EWMA statistics are autocorrelated, so we are more likely to have successive signals than with the Shewhart method. Therefore, the δ = 0 case has the same number of days signal for the Shewhart and EWMA methods, but since the EWMA method likely signals an outbreak in clusters, it signals an outbreak in fewer weeks than the Shewhart method. This occurs for all cases in our sensitivity analysis.

Table 2.5 for the Shewhart case shows the adaptive threshold method is superior to the W2c method. Similar results are shown in Table 2.6 for the EWMA method. We see that in most cases, as n increases, so does the probability of detecting an outbreak. In general, the EWMA method is superior to the Shewhart method, with much larger outbreak signaling rates in some cases. The exceptions where the Shewhart method outperforms the EWMA method is for the W2c method where δ = .1, and λ = 5, 10, and 20 when n = 7, as well as when λ = 5 and n = 14.
Table 2.5: Proportion of Signals – Transient Shift in Poisson Case – Shewhart

<table>
<thead>
<tr>
<th>δ</th>
<th>n</th>
<th>( W_{2c} )</th>
<th>Adaptive Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>In–control ( \lambda )</td>
<td>5</td>
</tr>
<tr>
<td>0.0</td>
<td>7</td>
<td>0.157</td>
<td>0.159</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>0.0141</td>
<td>0.0143</td>
</tr>
<tr>
<td>0.1</td>
<td>7</td>
<td>0.0239</td>
<td>0.0262</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>0.0258</td>
<td>0.0307</td>
</tr>
<tr>
<td>0.2</td>
<td>7</td>
<td>0.0414</td>
<td>0.0441</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>0.0491</td>
<td>0.0775</td>
</tr>
<tr>
<td>0.5</td>
<td>7</td>
<td>0.1085</td>
<td>0.1577</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>0.1905</td>
<td>0.3615</td>
</tr>
<tr>
<td>1.0</td>
<td>7</td>
<td>0.3324</td>
<td>0.5185</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>0.6260</td>
<td>0.9180</td>
</tr>
<tr>
<td>2.0</td>
<td>7</td>
<td>0.8374</td>
<td>0.9775</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>0.9949</td>
<td>1.000</td>
</tr>
</tbody>
</table>

2.7.2 Negative Binomial Counts

The increase in the negative binomial parameter \( r \) is assumed to be 100(1 + \( \delta \))% of the in–control value. Unlike the Poisson case, there are situations when the \( W_{2c} \) method outperforms the adaptive threshold method. Referring to Table 2.7, there are cases where the \( W_{2c} \) Shewhart method outperforms its counterpart. This happens most frequently for Case 2, uniformly for \( \delta \) equal to .2 and .5 in particular. When \( n = 28 \) for Case 1, \( W_{2c} \) is best for most values of \( \delta \), although the differences are small. Case 3 is the only one of the four cases that the adaptive threshold method is uniformly best.

The EWMA results shown in Table 2.8 lead to a different set of conclusions from the Shewhart case. Only for one situation using the EWMA statistic, Case 2 when \( \delta = .5 \) and \( n = 7 \), does the \( W_{2c} \) method signal an outbreak more frequently on average than
The adaptive threshold method. This is a much different result than the performance for the Shewhart case. The results of the EWMA method versus the Shewhart method are similar to that of the Poisson counts. The EWMA approach is nearly uniformly better than the Shewhart approach, with the only exception being for Case 1, when \( \delta = .1 \) and \( n = 7 \) for \( W_2c \).

### 2.8 Comments on Recurrence Intervals

Fraker et al. (2008) noted that we should be wary of the use of the RI metric. If a process over time signals multiple times in a row, in practice this would typically be considered to be one alarm. However, in this BioSense application, all monitoring statistic values
Table 2.7: Proportion of Signals – Transient Shift in Negative Binomial Case – Shewhart

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>$n$</th>
<th>W2c Case</th>
<th>Adaptive Threshold Case</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1 2 3 4</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td>0.0</td>
<td>7</td>
<td>.0132 .0116 .0121 .0135</td>
<td>.0113 .0119 .0123 .0130</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>.0132 .0120 .0141 .0120</td>
<td>.0123 .0121 .0124 .0124</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>.0129 .0131 .0151 .0139</td>
<td>.0121 .0134 .0152 .0131</td>
</tr>
<tr>
<td>0.1</td>
<td>7</td>
<td>.0242 .0624 .0568 .0322</td>
<td>.0260 .0628 .0963 .0415</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>.0292 .0936 .0846 .0402</td>
<td>.0317 .0906 .1054 .0449</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>.0285 .1286 .1072 .0519</td>
<td>.0281 .1243 .1173 .0524</td>
</tr>
<tr>
<td>0.2</td>
<td>7</td>
<td>.0392 .1812 .1582 .0602</td>
<td>.0506 .1756 .3401 .1061</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>.0554 .3328 .2881 .1017</td>
<td>.0591 .3118 .4034 .1213</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>.0632 .4545 .3980 .1306</td>
<td>.0608 .4383 .4409 .1328</td>
</tr>
<tr>
<td>0.5</td>
<td>7</td>
<td>.1329 .8563 .7753 .3040</td>
<td>.2130 .8114 .9866 .5651</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>.2162 .9889 .9740 .5314</td>
<td>.2363 .9814 .9970 .6093</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>.2903 1.000 1.000 .7015</td>
<td>.2738 .9994 .9990 .6970</td>
</tr>
<tr>
<td>1.0</td>
<td>7</td>
<td>.4504 1.000 1.000 .8562</td>
<td>.6569 .9997 1.000 .9782</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>.7160 1.000 1.000 .9917</td>
<td>.7201 1.000 1.000 .9938</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>.8560 1.000 1.000 .9997</td>
<td>.8172 1.000 1.000 .9993</td>
</tr>
<tr>
<td>2.0</td>
<td>7</td>
<td>.9529 1.000 1.000 .9999</td>
<td>.9836 1.000 1.000 1.000</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>.9990 1.000 1.000 1.000</td>
<td>.9978 1.000 1.000 1.000</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>1.000 1.000 1.000 1.000</td>
<td>1.000 1.000 1.000 1.000</td>
</tr>
</tbody>
</table>

are used without regard to order, and the empiric RI is calculated based strictly on the percentage of days with W2 values beyond a given threshold. As an example, suppose we have 100 observations over time. If we signal at observations 27, 45, 79, 89, and 96, the five signals would result in a RI value of 20. Suppose a similar scenario is considered, but observations 93–97 signal instead. Under the traditional RI approach, the RI value is also equal to 20. However, it may be more appropriate to consider this cluster of signals as one event rather than as five separate events because they occur consecutively. Large W2 values may be found in clusters in real BioSense data, yielding large thresholds for the RI’s. Refer to Figure 2.11 for an illustration of this phenomenon, where the solid horizontal lines represent signaling thresholds. Another metric to use may be the average time between signals (ATBS), also used by Fraker et al. (2008). Using the ATBS approach, threshold functions may be calculated as well.
Table 2.8: Proportion of Signals – Transient Shift in Negative Binomial Case – EWMA

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>$n$</th>
<th>W2c Case</th>
<th></th>
<th></th>
<th></th>
<th>Adaptive Threshold Case</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1  2  3  4</td>
<td></td>
<td>1  2  3  4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>7</td>
<td>.0079  .0088  .0086  .0096</td>
<td></td>
<td></td>
<td>.0097  .0083  .0099  .0107</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>.0100  .0101  .0097  .0085</td>
<td></td>
<td></td>
<td>.0112  .0099  .0095  .0107</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>.0111  .0092  .0097  .0095</td>
<td></td>
<td></td>
<td>.0113  .0093  .0094  .0098</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>7</td>
<td>.0198  .1028  .0786  .0347</td>
<td></td>
<td></td>
<td>.0318  .1113  .1640  .0581</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>.0296  .1811  .1579  .0553</td>
<td></td>
<td></td>
<td>.0336  .1816  .2092  .0658</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>.0362  .2755  .2183  .0741</td>
<td></td>
<td></td>
<td>.0372  .2783  .2412  .0761</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>7</td>
<td>.0420  .4232  .3156  .0924</td>
<td></td>
<td></td>
<td>.0779  .4426  .6526  .2062</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>.0753  .7450  .6383  .2070</td>
<td></td>
<td></td>
<td>.0902  .7457  .7693  .2567</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>.1081  .8763  .7975  .2791</td>
<td></td>
<td></td>
<td>.1132  .8787  .8342  .2980</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>7</td>
<td>.2415  .9977  .9884  .6327</td>
<td></td>
<td></td>
<td>.4545  .9968  1.000  .9013</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>.4850  1.000  1.000  .9193</td>
<td></td>
<td></td>
<td>.5609  1.000  1.000  .9573</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>.6535  1.000  1.000  .9783</td>
<td></td>
<td></td>
<td>.6726  1.000  1.000  .9827</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>7</td>
<td>.7957  1.000  1.000  .9964</td>
<td></td>
<td></td>
<td>.9537  1.000  1.000  .9999</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>.9785  1.000  1.000  1.000</td>
<td></td>
<td></td>
<td>.9879  1.000  1.000  1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>.9982  1.000  1.000  1.000</td>
<td></td>
<td></td>
<td>.9985  1.000  1.000  1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>7</td>
<td>.9992  1.000  1.000  1.000</td>
<td></td>
<td></td>
<td>1.000  1.000  1.000  1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>1.000  1.000  1.000  1.000</td>
<td></td>
<td></td>
<td>1.000  1.000  1.000  1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>1.000  1.000  1.000  1.000</td>
<td></td>
<td></td>
<td>1.000  1.000  1.000  1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.9 Considering Differing Parameters on Weekdays and Weekends

For the studies in this chapter, we considered the weekday and weekend counts to follow the same parametric distribution. We let a subscript of 1 denote a weekday, and 2 denote a weekend for this subsection. Threshold functions for the W2c and the adaptive threshold methods may differ because multiple distributions are now considered in the same data stream. It may be of interest to determine if the RI functions can be expressed as a function of the previously calculated functions. Consider the Poisson case where $\lambda_1 = 100$ and $\lambda_2 = 50$. A simulation may be run with these parameters to determine the threshold functions. However, this could be expressed using a weighted average.
of the previously calculated functions when the separate simulations were run. The weights would be proportional, 5/7 for weekdays and 2/7 for weekends. Our RI threshold functions may be calculated as

\[ RI(\lambda_1 = 100, \lambda_2 = 50) = \frac{5}{7}RI(\lambda_1 = 100) + \frac{2}{7}RI(\lambda_2 = 50). \]

The properties of this type of data stream may be of use in future work. A more complicated procedure may be to use different signaling thresholds for weekdays and weekends when monitoring.

\section*{2.10 Discussion}

We have reviewed the EARS W2c method the CDC uses in its BioSense program. These methods provide simple tools for biosurveillance based on a moving window approach. A modification to an adaptive threshold method was introduced and its Z–score thresholds were compared to the W2c thresholds for given RI values over different sets of parameter values and baseline lengths.

For the Poisson case, the RI threshold functions were tighter across values of \( \lambda \) for the adaptive threshold method than the W2c method, particularly when \( n = 7 \). The RI
thresholds for the adaptive threshold method tend to stay in the same neighborhood of values despite the window size $n$ increasing from 7 to 28 days, while the W2c method saw a sharp decrease in their threshold values as $n$ increases. For the negative binomial, Case 2’s ($r = 100, p = .1$) Z-score thresholds gave significantly different values than the other three parameter sets when $n = 7$. The W2c threshold functions were closer together in this case. However, as in the Poisson case, as $n$ increased, the Z-score threshold functions stayed closer to one another than those for W2c, with the Case 2 threshold function changing significantly. The W2c threshold values also dropped a sufficient amount again as $n$ increased. It is reasonable to expect that the W2c RI thresholds need to be changed when different baseline window lengths are used in practice. Tokars et al. (2009) used this varying threshold approach for different baseline lengths as well as for different average count categories.

The initial negative binomial simulation study displayed some estimation issues. With the constraints on the parameter space, an ad-hoc method used by Watkins et al. (2009a) to give us valid MOM estimators was implemented. However, even for some cases when $n = 28$, these estimators were often far from desirable.

A separate simulation study was performed to test how well the W2c and adaptive threshold methods signal an outbreak that lasts 7 days for different magnitudes of outbreaks. Shewhart and EWMA charts were used for the sensitivity analysis. The EWMA was superior to the corresponding Shewhart method in most comparisons. In these EWMA comparisons, the adaptive threshold method outperformed the W2c method nearly uniformly. In general, the adaptive threshold approach offered the greatest improvement in the case of the rate data.

Future studies on this topic are needed. Many different parameter combinations can be used in the data models. It would also be beneficial if the parameter choices were close to those obtained from simulated data provided by the CDC (CDC (2004)) for the negative binomial case, since public health count data are likely to be overdispersed relative to the Poisson distribution. There will have to be a tradeoff between having smaller variances (large values of $p$) and the frequency of $\bar{x}_t \geq \hat{\sigma}_t^2$ (small values of $p$). The modification used by Watkins et al. (2009a) may need to be improved upon. Also, an alternative to the MOM estimators may be considered.
We may also want to closely examine the behavior of the monitoring statistics over time. A time–series plot may give us indications of whether or not the W2c and Z–score statistics are autocorrelated. Another future study may include the implementation of separate distributions of counts for weekdays and weekends. Healthcare facilities typically have fewer visits on weekends and this can be reflected in a simulation study. It can be determined how the RI functions are affected when these data are pooled together to form one empirical distribution of monitoring statistics. The study of cyclical data is important, although we expect that methods based on the same size window of past data will be affected similarly by seasonal cyclical patterns.

The results from our studies indicate that RI threshold functions change relative to the data stream used to calculate the W2c and adaptive threshold statistics. Our results show that using $n = 14$ or $n = 28$ instead of $n = 7$ results in much more consistent non–outbreak performance across varying distributions of data streams. This supports in another way the baseline length recommendation of Tokars et al. (2009). Our sensitivity analysis shows that the EWMA approach with a reflecting lower boundary may be very useful for detecting syndromic outbreaks. Finally, it would be very informative to test the adaptive threshold methods on BioSense data as done in Tokars et al. (2009) for the W2 methods.
Chapter 3

Effects of Seasonality on Some Biosurveillance Methods

Abstract

The Centers for Disease Control and Prevention (CDC) established its BioSense program with the intent of providing real-time biosurveillance using syndromic data. Surveillance has been implemented using the Early Aberration Reporting System (EARS). The EARS method W2count (W2c) is used for monitoring daily counts of a syndrome that may be reported at a hospital or other health facility. If the daily count is large relative to a baseline period of seven days of count data, then a signal is raised for a possible outbreak. A threshold for a signal is based on an empiric recurrence interval (RI) metric. We incorporate an adaptive threshold method by using the baseline data to estimate a parametric distribution, then calculate a standard normal Z-score through a p-value. We consider several cases of varying amounts of seasonality combined with Poisson and negative binomial distributions to simulate syndromic counts. A sensitivity analysis is performed, where systematic outbreaks are injected into the data and the detection rate is calculated. An exponentially weighted moving average (EWMA) approach is also used to compete with traditional Shewhart methods.
3.1 Introduction

The Centers for Disease Control and Prevention (CDC) established BioSense, a national public health surveillance program to improve the capability of real–time biosurveillance with access of real–time health data (CDC (2008)). The Early Aberration Reporting System (EARS) is a free program available from the CDC that has been implemented as a standard monitoring method (http://www.bt.cdc.gov/surveillance/ears). Many major metropolitan areas use the BioSense system, including hospitals and public health departments to help determine if syndromic outbreaks have occurred (CDC (2008)).

The W2 count (W2c) EARS method employs the use of daily count data of a particular syndrome. The W2c statistic uses a seven–day moving window of background data, where the short baseline is intended to provide recent information on a given syndrome. The moving baseline approach includes a two day lag. If the current day’s count is large relative to the baseline data, a large W2c statistic is calculated. If a W2c value exceeds a specified threshold, an alarm is raised, indicating a possible outbreak.

The W2c method uses separate baseline data stratified by weekday and weekend. This is because some health facilities may expect sharp declines in patient visits on weekends. For simplicity, we considered the case where the data streams do not have weekday effects. Previous work on this method considered daily counts modeled by the Poisson and negative binomial distributions (Szarka et al. (2011)). However, the counts were assumed to be independent and identically distributed (iid) random variables. For many applications in health surveillance, syndromic data follow seasonal patterns. In this chapter, we considered yearly seasonal patterns for Poisson and negative binomial data.

The W2c monitoring method is similar to the C2 method used by the CDC (CDC (2006)). However, in the case of the C2 method, the partition of baseline data by weekday and weekend was not used. Two similar methods, C1 and C3, have also been used by the CDC for monitoring count data (CDC (2006), Hutwagner et al. (2005a,b)).

The adaptive threshold method first used by Lambert and Liu (2006) for computer networking applications is considered in our study. Using baseline data, the parameters of a distribution may be fit using maximum likelihood or method of moments estimators.
Once this distribution is estimated, an upper–tail p–value is calculated based on the current day’s count. The inverse standard normal cumulative distribution function of one minus the p–value will be an approximately standard normal statistic when no outbreak is present. These Z–scores are used for monitoring.

An empiric recurrence interval (RI) is used when monitoring using the W2c method. Kleinman et al. (2005) noted that if monitoring of a process continues without interruption after any alarm, the RI is the fixed number of time periods for which the expected number of false alarms is one. Using large amounts of in–control, seasonal data, RI thresholds can be generated based on the various seasonal patterns considered for the W2c and adaptive threshold methods.

The W2c method uses a background of \( n = 7 \) days for monitoring. We also considered longer baselines of \( n = 14 \) and 28 days. A sensitivity analysis was performed, where a transient outbreak lasted for seven days. We investigated the effectiveness of these methods to detect the shifts using the traditional Shewhart method with the W2c and Z–score statistics, as well as an exponentially weighted moving average (EWMA) method based on these statistics.

In Section 3.2, the recurrence interval is discussed. In Section 3.3, we introduce the W2c and adaptive threshold methods, as well as the EWMA modifications. The seasonal pattern models based on the Poisson and negative binomial distributions are given in Section 3.4. A simulation study to determine the effectiveness of detecting an outbreak lasting seven days is explored in Section 3.5. A discussion follows in Section 3.6.

3.2 Recurrence Intervals

If a large number of statistics are given for a surveillance method, an empiric recurrence interval (RI) can be calculated. For a large number of syndromic counts, the appropriate W2c and adaptive threshold statistics were first computed under the assumption of no outbreaks. Then a RI value can be calculated as the reciprocal of the proportion of days when the statistics exceed a particular threshold. Hence, the 95th percentile of W2c values computed from the data set is the threshold for a RI of 20. The empiric RI
thresholds used by the CDC for the W2c method are available (CDC (2009)). The data used to form the empiric RI’s by the CDC are updated periodically.

Tokars et al. (2009) noted that BioSense used outbreak data in forming their empiric RI’s, which may inflate the values of RI thresholds. Szarka et al. (2011) showed that RI thresholds for iid Poisson and negative binomial data using the W2c method were quite different from the BioSense thresholds for larger RI values. If an underlying model for a data stream of in–control counts can be identified, then a large simulation can be run to develop the RI thresholds for the given data set. The thresholds given by BioSense represent an attempt to use a single threshold across numerous data streams. However, the appropriate RI thresholds may vary considerably for different types of syndromes. Fraker et al. (2008) noted that the RI metric has some undesirable features because signals may happen in clusters and be considered separate events, when it may be more appropriate to consider it to be a single event. The authors recommended using the average time between signals as an alternative metric. We chose the RI metric because it is currently used by the CDC.

3.3 Monitoring Schemes

3.3.1 The W2c Method

The W2c method uses a baseline of \( n = 7 \) days in BioSense applications with a two day lag and stratification by weekday and weekend. For example, a W2c statistic for a Friday uses baseline data starting three days before on Tuesday, then Monday, and then the five weekdays from the previous week.

Let \( x_t \) be a syndromic count on day \( t \). The W2c statistic on day \( t \) is calculated as

\[
W2c(t) = \frac{x_t - \bar{x}_t}{s_t},
\]

where \( \bar{x}_t \) and \( s_t \) represent the sample mean and standard deviation, respectively, for the baseline data set. A lower bound of one is set for \( s_t \) so that the W2c statistic is never undefined. The adjustment of \( s_t \) will prevent the presence of one or two more patients
with a specific syndrome on day \( t \) to cause a signal. A signal is raised at time \( t \) if \( W2c(t) \geq h_W \), where \( h_W \) is a specified threshold.

### 3.3.2 An Adaptive Threshold Method

An adaptive threshold technique used by Lambert and Liu (2006) for computer network monitoring can be used as an alternative monitoring procedure. Using the same baseline data as the W2c method, the parameters of an underlying parametric distribution are fit using maximum likelihood or method of moments estimators. The estimated parameter set at time \( t \) is denoted as \( \hat{\theta}_t \).

Let \( X_t \) be a random variable that follows a given parametric distribution. An upper-tailed \( p \)-value can be calculated, \( P_t \), based on the current day’s count, \( x_t \), and the estimated parameter set, \( \hat{\theta}_t \) as

\[
P_t = Pr(X_t \geq x_t | \hat{\theta}_t).
\]

If there is a stable, no-outbreak process, the \( p \)-values may be assumed to follow a uniform \((0,1)\) distribution. The inverse cumulative distribution function (cdf) is applied so approximately standard normal \( Z \)-scores are obtained:

\[
Z_t = \Phi^{-1}(1 - P_t).
\]

A signal is given at time \( t \) if \( Z_t \geq h_{AT} \).

### 3.3.3 EWMA Methods

Roberts (1959) introduced the EWMA chart, a widely-used method in traditional quality control applications. The decision rules introduced in the previous section rely on the use of the most recent statistic and are considered Shewhart methods. The EWMA statistic incorporates information using past data where statistics nearest the current time point are given larger weights. For standardized monitoring statistics \( V_t, t = 1, 2, \ldots \), the
EWMA statistics $E_t$ for detecting an increase in the process parameter are computed as

$$E_t = \max[0, \tau v_t + (1 - \tau)E_{t-1}],$$

where $\tau$ is the weight given to the most recent statistic, and $E_0 = 0$. When $\tau = 1$, the EWMA statistic reduces to a Shewhart statistic. Montgomery (2009, p. 423) recommended using $0.05 \leq \tau \leq 0.25$. A weight of $\tau = 0.2$ was chosen for our analysis. A signal is given for the EWMA method at time $t$ if $E_t \geq h_E$.

In most industrial applications, a two–sided EWMA chart is used. This chart is used because both an increase and decrease in a process parameter are typically of interest. In health surveillance, we are only interested in detecting outbreaks, so a one–sided method is used. The reflecting barrier at zero is included in Equation (3.1) so that the EWMA statistic does not become very small. If this condition is not set and an outbreak occurred when the statistic is far from the charting limit, it would be more difficult to signal. See Crowder and Hamilton (1992) for further information on this implementation. Failure to use a reflecting barrier in a one–sided EWMA chart can lead to serious inertia problems, a topic discussed by Woodall and Mahmoud (2005). We note that Lambert and Liu (2006) did not use a reflecting barrier for their EWMA approach.

In industrial statistical process control applications, the EWMA statistic is reset to zero after a signal. However, the reset occurs under the assumption that a process is stopped and corrective action is taken. The EWMA statistic is usually not reset after a signal in health surveillance where immediate process intervention and recovery are not possible. In our analysis, the EWMA statistic was not reset after a signal.

### 3.4 Seasonal Effects

In many public health applications, it is common for syndromic counts to follow a seasonal pattern. The Poisson and negative binomial distributions are commonly used for monitoring discrete data, so we propose adding various seasonal trends for these types of data streams.
3.4.1 Poisson Model

The probability mass function (pmf) for a Poisson random variable at time $t$ is given as

$$f(x_t|\lambda_t) = \frac{e^{-\lambda_t} \lambda_t^{x_t}}{x_t!},$$

where $x_t = 0, 1, 2, \ldots$, and $\lambda_t > 0$. The mean and variance at time $t$ equal $\lambda_t$. For iid non-outbreak data, $\lambda_t = \lambda^{(0)}$, where $\lambda^{(0)}$ is a constant. The maximum likelihood estimator using the baseline data for the adaptive threshold method is $\hat{\lambda}_t = \bar{x}_t$. For this model to be useful in some of our methods, it must be assumed that the Poisson parameter is sufficiently large so that the baseline data does not consist of all zeros, yielding $\hat{\lambda}_t = 0$, since in this case a single syndromic count on day $t$ would cause a signal.

Consider monitoring where $\lambda^{(0)} = 50$. The overall yearly average for a syndromic count may equal 50, the average counts could be quite different depending on the time of year. A seasonal effect is considered using the following model of the average count at time $t$:

$$\lambda_t = \lambda^{(0)} + \sum_{i=1}^{a} [\alpha_i \cos(hit) + \beta_i \sin(hit)]. \tag{3.2}$$

We define $a$ as the number of sine and cosine terms in the model, and $h = 2\pi/\rho$. We let $\rho = (52 \times 7)/w$, where $w$ is the size of the time period (in days) for which counts are available. In our case, $w = 1$. Using this representation for the data allows the same seasonal cycles to appear yearly, e.g., the average count on day 1 of year 1 is the same average as day 1 of year 4. A trend could be considered as well by adding a regression component or day–of–the–week effects in Equation (3.2).

Many forms could have been chosen for the seasonal component of the Poisson parameter. The link function used between the seasonal terms and mean, $\lambda_t$ in Equation (3.2), is the identity function. The canonical link for the Poisson distribution is the log link, where Equation (3.2) would model $\log(\lambda_t)$ rather than $\lambda_t$. We chose the identity function so that the overall yearly mean is equal to $\lambda^{(0)}$. The quantities $\alpha_i$ and $\beta_i$ were chosen such that $\lambda_t > 0$ for all values of $t$. Syndromic counts tend to follow irregular seasonal patterns. Since we do not consider any biosurveillance methods that involve an attempt
to model the seasonal effects, however, we believe our simpler monitoring methods will be adequate for comparisons.

3.4.2 Negative Binomial Model

In health surveillance, it is common to have data that are overdispersed with respect to the Poisson distribution. The negative binomial distribution can be expressed as a mixture between the Poisson and gamma distributions. If

\[ X_t | \gamma \sim \text{Poi}(\gamma \lambda_t) \]

and

\[ \gamma \sim \text{Gam}(\psi, \psi), \]

the counts \( X_t \) follow a negative binomial distribution. The mean and variance of \( X_t \) are now derived.

\[
E(X_t) = E[E(X_t | \gamma)] = E[\gamma \lambda_t] = \lambda_t E[\gamma] = \lambda_t \frac{\psi}{\psi} = \lambda_t
\]

\[
V(X_t) = V[E(X_t | \gamma)] + E[V(X_t | \gamma)]
\]

\[
= V[\gamma \lambda_t] + E[\gamma \lambda_t]
\]

\[
= \lambda_t^2 V[\gamma] + \lambda_t E[\gamma]
\]

\[
= \lambda_t^2 \frac{\psi}{\psi^2} + \lambda_t \frac{\psi}{\psi}
\]

\[
= \lambda_t \left( \frac{\lambda_t}{\psi} + 1 \right)
\]

The pmf of the negative binomial distribution at time \( t \) is then expressed as

\[
 f(x_t | \lambda_t, \psi) = \frac{\Gamma(x_t + \psi)}{\Gamma(\psi) x_t!} \left( \frac{\psi}{\lambda_t + \psi} \right)^\psi \left( \frac{\lambda_t}{\lambda_t + \psi} \right)^{x_t}
\]
for \( x_t = 0, 1, 2, \ldots, \lambda_t > 0 \) and \( \psi > 0 \). As \( \psi \to \infty \), the negative binomial distribution converges to a Poisson distribution. For seasonal data, the same form for \( \lambda_t \) is used as shown in Equation (3.2).

Szarka et al. (2011) used the parameters \((r, p)\) to represent a negative binomial distribution with pmf

\[
f(x_t | r, p) = \frac{\Gamma(x_t + r)}{\Gamma(r)x_t!} p^r (1 - p)^{x_t},
\]

where \( r > 0 \) and \( 0 < p < 1 \). These two distributions are equivalent if \( r = \psi \) and \( p = \psi / (\lambda_t + \psi) \). The method of moments estimators for \( \psi \) and \( \lambda_t \) at time \( t \) are \( \hat{\psi}_t = \bar{x}_t / \hat{\sigma}_t^2 \) and \( \hat{\lambda}_t = \bar{x}_t \), where \( \hat{\sigma}_t^2 = ((n - 1)s_t^2) / n \). Overdispersion for the negative binomial distribution does not occur in a baseline when \( \bar{x}_t \geq \hat{\sigma}_t^2 \). If the baseline data is not overdispersed, a technique from Watkins et al. (2009a) was used where we set \( \hat{\sigma}_t^2 = 1.05\bar{x}_t \). This is an ad-hoc method that gives us valid estimates for our parameters.

### 3.5 Simulation Study

#### 3.5.1 Overview of Setup

Consider for the Poisson distribution a case where \( \lambda^{(0)} = 50 \), and the negative binomial distribution where \( \lambda^{(0)} = 40 \) and \( \psi = 60 \), which were used by Szarka et al. (2011) for iid data. The parameters chosen for the seasonal terms for these cases are shown in Table 3.1. Graphs of the average syndromic count over a two–year period are given for each distribution in Figures 3.1 and 3.2, where the average yearly counts for the Poisson and negative binomial seasonal patterns are 50 and 40, respectively. The empiric RI values were formed by simulating a large number (3,000 years) of counts from these seasonal patterns, then computing the appropriate W2c and adaptive threshold values for the Shewhart and EWMA methods.

While the CDC uses a single data set for constructing the RI thresholds, separate functions were constructed for each seasonal pattern considered across all values of the baseline window size \( n \). The threshold functions shown by Szarka et al. (2011) for the
### Table 3.1: Seasonal Patterns Distributions

<table>
<thead>
<tr>
<th>Seasonal Patterns</th>
<th>$\alpha_i$</th>
<th>$\beta_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>1</td>
<td>2, 4</td>
</tr>
<tr>
<td>P2</td>
<td>2</td>
<td>8, 3, 6, 2</td>
</tr>
<tr>
<td>P3</td>
<td>3</td>
<td>3, 11, 4, 7, 9, 6</td>
</tr>
<tr>
<td>P4</td>
<td>4</td>
<td>9, 5, 8, 5, 4, 6, 4, 2</td>
</tr>
<tr>
<td>N1</td>
<td>1</td>
<td>6, 8</td>
</tr>
<tr>
<td>N2</td>
<td>2</td>
<td>9, 4, 2</td>
</tr>
<tr>
<td>N3</td>
<td>3</td>
<td>5, 4, 10, 5, 5</td>
</tr>
<tr>
<td>N4</td>
<td>4</td>
<td>7, 6, 4, 3, 6, 3, 7, 8</td>
</tr>
</tbody>
</table>

Iid data case were used for comparisons.

The threshold where the RI value is equal to 500 was set as a signaling threshold when monitoring each data stream for each value of $n$. When monitoring Poisson data, a shift from $\lambda^{(0)}$ to $\lambda^{(1)}$ was considered, where $\lambda^{(1)} = (1 + \delta)\lambda^{(0)}$. The values of $\delta$ used in the sensitivity analysis were 0, .1, .2, .5, and 1. When $\delta = 0$, the in–control properties of the method were observed. We note that for a shift $\delta$ in the underlying parameter will result in larger increases when seasonality is high and smaller increases when seasonality is low. When monitoring the negative binomial data, the same shift from $\lambda^{(0)}$ to $\lambda^{(1)}$ was considered, as well as a shift from $\psi^{(0)}$ to $\psi^{(1)}$, where $\psi^{(1)} = (1 + \delta)\psi^{(0)}$, so we can consider the same type of shift as in Szarka et al. (2011), where a shift was considered from $r_0$ to $r_1$. The equivalence of the shifts considered for the negative binomial distribution are shown in Table 3.2.

In our simulations, two years of in–control data from the seasonal distributions were generated. At the beginning of the second year, the surveillance statistics were computed. Then $k$ additional in–control days were simulated from a third year of this data stream, where $k = 0, 1, \ldots, 363$. This was done so that the outbreak detection was evaluated across the entire year. Afterwards, the parameter(s) shifted to the outbreak state for seven days. An iteration for each value of $k$ was simulated one thousand times. The proportion of simulations with signals during the outbreak for each case was then calculated.
Figure 3.1: Average Count for Poisson Seasonal Data During Two-Year Period

3.5.2 In–Control Results

A comparison of the Shewhart threshold functions for the iid case, the four seasonal patterns, and BioSense standard given by CDC (2008) when $n = 7$ are displayed in Figure 3.3. It is important that the RI threshold functions be relatively close together if one uses a single set of thresholds for all data streams, as in BioSense. The BioSense RI threshold function (on the far right) appears to be similar to the other functions for smaller RI values, but is quite different as the RI becomes very large. The RI threshold functions for the Shewhart versions of the two types of surveillance methods for Poisson and negative binomial distributions across $n$ are shown in Figures 3.4 and 3.5. The iid data case’s threshold function (on the far left) is closer to that with the seasonal patterns
with fewer terms in the model. For both distributions, the two seasonal patterns with the most dramatic seasonality (P3, P4, N3, N4) have thresholds that are larger than those for the iid data and other seasonal patterns. There appears to be a distinct difference in the RI threshold functions across seasonal patterns, especially for the Poisson adaptive threshold approach.

The adaptive threshold method’s RI functions for seasonal patterns P3 and P4 get larger as \( n \) increases in Figure 3.4. The large amount of seasonality produces more variability with a longer baseline period. The Poisson distribution has lighter tails than the negative binomial distribution, which results in larger Z-scores for large counts relative to the baseline period. The RI threshold functions for the adaptive threshold methods
Table 3.2: Negative Binomial Equivalence Conditions

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>$(r, p)$</th>
<th>$(\lambda(0), \alpha)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>60, .6</td>
<td>40, 60</td>
</tr>
<tr>
<td>0.1</td>
<td>66, .6</td>
<td>44, 66</td>
</tr>
<tr>
<td>0.2</td>
<td>72, .6</td>
<td>48, 72</td>
</tr>
<tr>
<td>0.5</td>
<td>90, .6</td>
<td>60, 90</td>
</tr>
<tr>
<td>1.0</td>
<td>120, .6</td>
<td>80, 120</td>
</tr>
<tr>
<td>2.0</td>
<td>180, .6</td>
<td>120, 180</td>
</tr>
</tbody>
</table>

Figure 3.3: W2c RI Threshold Functions for Simulated Data Sets and BioSense Standard (On Far Right)

shown in Figure 3.5 do not show the wide variation present for the Poisson data.

3.5.3 Outbreak Results

In Tables 3.3 – 3.6, the results of our outbreak simulation study for the W2c versus adaptive threshold methods are given by charting method (Shewhart or EWMA) and distribution. For each value of $\delta$ and seasonal pattern, the highest detection rate among the four method and baseline window size $n$ combinations is shown in boldface. The signaling threshold for each method across $n$ was chosen as the quantity that obtained a RI value of 500 in Section 3.5.2. If the in–control RI is equal to 500, then approximately
Figure 3.4: Shewhart RI Threshold Functions for Different Baselines – Poisson Counts
Figure 3.5: Shewhart RI Threshold Functions for Different Baselines – Negative Binomial Counts
Table 3.3: Proportion of Signals – Transient Outbreak in Poisson Case – Shewhart

<table>
<thead>
<tr>
<th>δ</th>
<th>n</th>
<th>W2c</th>
<th>Adaptive Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Seasonal Pattern</td>
<td>Seasonal Pattern</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P1</td>
<td>P2</td>
</tr>
<tr>
<td>0.0</td>
<td>7</td>
<td>0.0134</td>
<td>0.0132</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.0132</td>
<td>0.0139</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>0.0143</td>
<td>0.0136</td>
</tr>
<tr>
<td>0.1</td>
<td>7</td>
<td>0.0465</td>
<td>0.0445</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.0658</td>
<td>0.0601</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>0.0782</td>
<td>0.0627</td>
</tr>
<tr>
<td>0.2</td>
<td>7</td>
<td>0.1182</td>
<td>0.1128</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.2043</td>
<td>0.1802</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>0.2678</td>
<td>0.1951</td>
</tr>
<tr>
<td>0.5</td>
<td>7</td>
<td>0.6194</td>
<td>0.5943</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.8923</td>
<td>0.8334</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>0.9648</td>
<td>0.8626</td>
</tr>
<tr>
<td>1.0</td>
<td>7</td>
<td>0.9955</td>
<td>0.9921</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td><strong>1.0000</strong></td>
<td>0.9999</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td><strong>1.0000</strong></td>
<td>0.9999</td>
</tr>
</tbody>
</table>

0.2 percent of the monitoring statistics exceed the threshold when there is no outbreak. Assuming that these data are not autocorrelated, the probability of obtaining at least one signal in a seven day period can be calculated using the binomial distribution: \(1 - (1 - .02)^7 = .0139\). For the Shewhart methods when \(\delta = 0\), this weekly false alarm rate is close to the given value for all cases. However, for the EWMA method, the statistics are autocorrelated. Hence, if a false alarm is raised during the specified seven day period, it is likely that several of those days result in a signal. Thus, the EWMA method signals during a seven day window with less frequency when there is no outbreak. If we had considered a transient outbreak lasting for one day, the in–control Shewhart and EWMA results would be more similar.
Detection Rates in Poisson Case

The results of the sensitivity analysis show variability in the best overall method and differences in performance across the studies. In general, for seasonal patterns P1 and P2, the EWMA adaptive threshold method provided the highest detection rates. However, for a given method, the value of $n$ that yielded the largest power was either $n = 7$ or $n = 14$. The $n = 28$ case never yielded the largest power for this particular method. This differs from the results shown by Szarka et al. (2011), where a longer baseline window was always desired with iid data. For seasonal patterns P3 and P4, the W2c Shewhart method when $n = 14$ and $n = 28$ performed well in many cases, but the Shewhart adaptive threshold method when $n = 7$ typically had the highest overall detection rate.

The EWMA performance is more difficult to interpret. The W2c method improved in sensitivity as $n$ increased for seasonal pattern P1, and had its best performance when $n = 14$ for seasonal pattern P2. For seasonal patterns P3 and P4, the $n = 28$ baseline
Table 3.5: Proportion of Signals – Transient Outbreak in Negative Binomial Case – Shewhart

<table>
<thead>
<tr>
<th>δ</th>
<th>n</th>
<th>W2c Seasonal Pattern</th>
<th>Adaptive Threshold Seasonal Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>N1</td>
<td>N2</td>
</tr>
<tr>
<td>0.0</td>
<td>7</td>
<td>0.0141</td>
<td>0.0134</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.0141</td>
<td>0.0137</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>0.0142</td>
<td>0.0133</td>
</tr>
<tr>
<td>0.1</td>
<td>7</td>
<td>0.0342</td>
<td>0.0326</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.0418</td>
<td>0.0383</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>0.0429</td>
<td>0.0395</td>
</tr>
<tr>
<td>0.2</td>
<td>7</td>
<td>0.0710</td>
<td>0.0662</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.1002</td>
<td>0.0895</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>0.1084</td>
<td>0.0950</td>
</tr>
<tr>
<td>0.5</td>
<td>7</td>
<td>0.3199</td>
<td>0.3013</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.5009</td>
<td>0.4510</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>0.5477</td>
<td>0.4860</td>
</tr>
<tr>
<td>1.0</td>
<td>7</td>
<td>0.8469</td>
<td>0.8218</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.9720</td>
<td>0.9391</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>0.9870</td>
<td>0.9382</td>
</tr>
</tbody>
</table>

rarely led to the highest detection rate. Relative to its Shewhart counterpart, the W2c EWMA method only yielded improvement for seasonal patterns P1 and P2, whereas it was generally worse for P3 and P4.

The adaptive threshold EWMA performance was best when \( n = 14 \) for seasonal pattern P1 and \( n = 7 \) for seasonal pattern P2. For the last two seasonal patterns, the drop-off in outbreak detection as \( n \) increased was substantial, especially for large shifts. This is evident for seasonal pattern P3, when the detection rate was 0.7644 when \( n = 7 \), but only 0.1959 when \( n = 28 \). This phenomenon also occurred for the Shewhart case, but was not quite as extreme.

Overall, it is difficult to recommend any one monitoring method or baseline window size to use. However, because of the deterioration in detection when \( n = 28 \) for the adaptive threshold method, this baseline should not be chosen for this monitoring scheme. In traditional quality control applications, the EWMA chart is usually better than a Shewhart chart for detecting small changes in the parameter. The advantage of the
Table 3.6: Proportion of Signals – Transient Outbreak in Negative Binomial Case – EWMA

<table>
<thead>
<tr>
<th>δ</th>
<th>n</th>
<th>W2c Seasonal Pattern</th>
<th>Adaptive Threshold Seasonal Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>7</td>
<td>0.0093 0.0085 0.0080 0.0074</td>
<td>0.0102 0.0092 0.0088 0.0083</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.0091 0.0090 0.0080 0.0074</td>
<td>0.0093 0.0091 0.0080 0.0075</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>0.0092 0.0087 0.0079 0.0080</td>
<td>0.0092 0.0087 0.0077 0.0077</td>
</tr>
<tr>
<td>0.1</td>
<td>7</td>
<td>0.0339 0.0303 0.0233 0.0185</td>
<td><strong>0.0533</strong> <strong>0.0458</strong> 0.0301 0.0227</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.0455 0.0385 0.0241 0.0182</td>
<td>0.0517 0.0421 0.0244 0.0179</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>0.0472 0.0353 0.0237 0.0186</td>
<td>0.0483 0.0345 0.0226 0.0177</td>
</tr>
<tr>
<td>0.2</td>
<td>7</td>
<td>0.1012 0.0875 0.0592 0.0432</td>
<td><strong>0.1908</strong> <strong>0.1599</strong> 0.0826 0.0548</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.1566 0.1265 0.0577 0.0408</td>
<td>0.1861 0.1444 0.0579 0.0382</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>0.1555 0.1105 0.0532 0.0368</td>
<td>0.1594 0.1080 0.0504 0.0348</td>
</tr>
<tr>
<td>0.5</td>
<td>7</td>
<td>0.5986 0.5365 0.3539 0.2588</td>
<td>0.8500 <strong>0.7825</strong> <strong>0.5101</strong> 0.3509</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.7927 0.6775 0.3170 0.2398</td>
<td><strong>0.8519</strong> 0.7251 0.3239 0.2297</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>0.7301 0.6032 0.2139 0.1616</td>
<td>0.7413 0.5963 0.2131 0.1585</td>
</tr>
<tr>
<td>1.0</td>
<td>7</td>
<td>0.9878 0.9699 0.8922 0.7809</td>
<td>0.9996 <strong>0.9975</strong> <strong>0.9680</strong> <strong>0.8702</strong></td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.9998 0.9939 0.9008 0.7732</td>
<td><strong>0.9999</strong> 0.9967 0.9077 0.7567</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>0.9999 0.9791 0.8233 0.6918</td>
<td>0.9999 0.9759 0.8151 0.6659</td>
</tr>
</tbody>
</table>

EWMA method is not applicable for many cases in this study, particularly for data with more pronounced and complicated seasonal patterns.

Detection Rates in Negative Binomial Case

The adaptive threshold methods appear to be the better class of methods for monitoring negative binomial data. For seasonal patterns N1 and N2, the EWMA method yields the best overall outbreak detection performance. The Shewhart monitoring method may be preferred for seasonal patterns N3 and N4. The \( n = 7 \) baseline provided the best detection rates for these cases.

The W2c Shewhart method gives consistent performance, although it is rarely optimal. For seasonal patterns N1 and N2, outbreak detection rates improve as \( n \) increases. The \( n = 14 \) window appears best for monitoring seasonal patterns N3 and N4 using the W2c Shewhart method.
Similar to the Poisson case, the effect of $n$ on outbreak detection using the EWMA method can be dramatic. The decrease in power as $n$ increased is evident for seasonal patterns N3 and N4 for the W2c method and seasonal patterns N2–N4 for the adaptive threshold method. For large shifts, $\delta$, large decreases in detection rates exist for the adaptive threshold method as $n$ increases. The $n = 28$ baseline should not be used for the EWMA adaptive threshold method with negative binomial data, as we found with Poisson data. The EWMA method for the adaptive threshold approach where $n = 7$ appears to have the best overall performance.

Initially, we found the poor results of the EWMA method for data streams with heavy seasonality and a long baseline to be somewhat surprising. However, with a fluctuating mean across the year, it is possible that the baseline counts are large when no outbreak is present, and the current count, $x_t$, may not be large by comparison when an outbreak is present. If an outbreak occurs at this point, it may be difficult to signal in this case.

In Figure 3.6, we illustrated the monitoring statistics for seasonal pattern N4 where $n = 28$ for the adaptive threshold EWMA method for the first seventy-five days of the year for a simulated example. From Figure 3.2, the average count increases above 70, then begins to decline during this time of the year for seasonal pattern N4. In Figure 3.6, the EWMA statistics reflected this pattern. The values began to decline steadily towards zero after two weeks because the average count is decreasing for this time period of the year. An outbreak is inserted for the last seven days of this data set where $\delta = .5$ and is represented by the dashed line in the plot. Despite a sizable increase in the average count over this period, the EWMA statistic did not come close to reaching the signaling threshold of approximately 2.25. We note that the opposite trend may occur at other points during this year. This phenomenon may occur as a result of having low counts present in the baseline, where the current count will be very large relative to this baseline regardless of whether or not there is an outbreak in the data stream.

3.6 Distribution of Signals

The sensitivity analysis discussed in the previous section provides a summary of the probability that an outbreak is detected during a seven day window. However, it is also
important to consider if an outbreak is more likely to be detected depending on the time of year in which it occurs. Frisen (1992) considered this type of analysis for monitoring data from a normal distribution.

In Figures 3.7 and 3.8, the proportion of weeks containing false alarms by the day of the year where the week begins is shown for the Poisson and negative binomial data streams. We chose seasonal patterns P4 \((n = 14)\) and N3 \((n = 7)\) because of the heavy amount of seasonality in each case. For all four monitoring schemes, false alarms are more likely to appear late in the year, when these seasonal patterns have spikes in their average counts. The false alarm rate can be as large as 25 percent for both adaptive threshold approaches. The false alarm rate is very close to zero for most of the year using these monitoring methods.

We also considered the distribution of signal rates when \(\delta = 0.5\) for each case. These results are shown in Figures 3.9 and 3.10. Frisen (1992) referred to the metric displayed as the probability of successful detection. The seasonal patterns of the signal rates are more pronounced in this case. The range of probabilities is large, where it may be close
Figure 3.7: Distributions of False Alarm Rates, Seasonal Pattern P4, \( n = 14 \)

to zero or one for some methods, depending on the value of \( k \). This is an undesirable feature. To reduce the effect of seasonality it would be necessary to estimate the seasonal pattern and incorporate this into the surveillance method. Wieland et al. (2007) used an expectation–variance model in order to obtain a constant false alarm rate for public health data for seasonal patterns but used a much longer baseline of six years for their applications.
Figure 3.8: Distributions of False Alarm Rates, Seasonal Pattern N3, \( n = 7 \)

3.7 Discussion

In this chapter, we have evaluated the performance of two competing methods, the CDC’s W2c method, and an adaptive threshold method, for detecting outbreaks. The RI threshold functions across different seasonal patterns for Poisson data varied considerably using the adaptive threshold method. This differs from previous work on RI threshold functions for Poisson iid data (Szarka et al. (2011)), where the threshold functions were closer together across different values of the Poisson parameter. The CDC uses a single RI threshold function across all data streams monitored. Our results in the Poisson case show that the appropriate threshold to obtain a particular RI value could be quite
different depending on the amount of seasonality in the model. Smaller, but noticeable differences appeared in other RI threshold functions examined.

Szarka et al. (2011) showed that longer baseline windows and the implementation of an EWMA feature help in outbreak detection. However, the results shown in our sensitivity analyses with seasonal data differed somewhat. If the underlying data stream is Poisson, the adaptive threshold method performed well for the Shewhart and EWMA methods, but with more pronounced seasonal patterns, these methods deteriorated in outbreak detection as the baseline window size increased. The W2c Shewhart method when \( n = 28 \) was a competing method for the most complicated seasonal patterns with Poisson data. The monitoring of negative binomial data should be based on adaptive
threshold methods with $n = 7$ for outbreak detection. The adaptive threshold EWMA method appeared to have the best performance or was extremely competitive across all four seasonal patterns evaluated when $n = 7$. However, the EWMA feature is not as helpful for data with seasonal patterns as it is with iid data, and the use of longer baselines often gave worse outbreak detection performance.

While the overall detection rate is an important metric when comparing methods, the power of these methods across the entire year is insightful as well. The false alarm rates and the probability of signals varied considerably depending on the time of year. This is troubling in biosurveillance because this indicates that existing methods, such as W2c are more vulnerable to missing outbreaks for certain periods. It would be interesting
to study the alarm patterns with real BioSense data with injected outbreaks to see how heavily the signaling rates depend on the time of year.

Future work may consider having seasonal terms built-in to the adaptive threshold surveillance method as in Lambert and Liu (2006). But in practice, this modeling may be difficult to implement effectively since the seasonal pattern often does not repeat in a regular pattern from year to year. The data stream is more likely to be overdispersed in healthcare data, hence the negative binomial distribution may be more reasonable to model than the Poisson distribution in practice. Fricker et al. (2008) considered a cumulative sum (CUSUM) chart based on residuals from a forecasting model that used linear and day-of-the-week terms to model recent counts. The performance of this method needs to be further studied under the assumption of seasonal effects. The W2rate (W2r) method also used by the CDC would be expected to have more robust performance with seasonal data because this method accounts for the total number of daily visits to a facility. An extensive analysis of this method for data with seasonal patterns would be beneficial as well.
Chapter 4

A Review and Perspective on Control Charting for High Quality Bernoulli Processes

Abstract

High quality Bernoulli processes have been monitored using a wide variety of techniques in statistical process control (SPC). The data consist of information on successive items classified as conforming (non-defective) or nonconforming (defective), where the probability of obtaining a nonconforming item is very small. This area of SPC is also applied to health–related monitoring, where the incidence rate of a rare medical problem such as a congenital malformation is of interest. In these applications, standard Shewhart control charts based on the binomial distribution are no longer useful. In our expository chapter we review the methods implemented for these scenarios and present ideas for future work in this area. We offer advice to practitioners and present a comprehensive literature review for researchers.
4.1 Introduction

4.1.1 Background

A review of control charts designed for attribute data was given by Woodall (1997), who gave a comprehensive bibliography of methods categorized by the type of problem that may be of interest to a researcher or practitioner. More recent reviews of attribute control charts were given by Duran and Albin (2009) and Topalidou and Psarakis (2009).

In this chapter, we consider a sequence of independent Bernoulli random variables, $X_i, i = 1, 2, \ldots$, representing whether items sampled are classified as conforming or nonconforming. For the in-control state, a constant probability of a nonconforming item, $p$, is often assumed in this type of monitoring scenario. This data structure is referred to as Bernoulli data for the remainder of this chapter. Analysis of Bernoulli data has been performed using 100% inspection, where all items are considered, as well as interval sampling, where items are inspected at scheduled periods. We primarily analyze 100% inspection applications. In most cases, interest is on detecting a sustained increase in $p$ from an in-control nonconforming rate, $p_0$, to an out-of-control nonconforming rate, $p_1 = \gamma p_0$, where $1 < \gamma < 1/p_0$ because it is an indication of a deterioration in the process. However, a downward shift in $p$ ($0 < \gamma < 1$) may be of interest to detect as well.

In this chapter, we focus on “high quality processes” for Bernoulli data, where the in-control nonconforming rate, $p_0$, is very small. As systems improve and the demand for high quality continues, methods for monitoring these types of processes become more important. In industrial applications, some systems tend to have small nonconforming rates. In health-related monitoring, interest may be on the incidence rate of a rare disease or medical problem, which could be very small as well. As $p_0$ becomes smaller and smaller, however, it becomes increasingly useful for the practitioner to identify an appropriate surrogate continuous variable, if possible, to monitor instead.

There is variation in the statistical process control (SPC) literature corresponding to what values of $p_0$ represent a high quality process. Goh and Xie (1994) set a benchmark that a high quality process has an in-control nonconforming rate of no more than 1000 parts per million (ppm) ($p_0 = .001$). However, it is common in many papers to consider ‘small’ values of $p_0$ to be as large as $.05$. There is no definitive value of $p_0$ that can be
used as a reasonable cutoff to constitute a high quality process across all applications; it will naturally vary accordingly with the processes being considered.

Applications of methods that monitor the nonconforming rate, such as the Shewhart $p$-chart (or equivalent $np$-chart) may be reasonable for larger values of $p_0$, but do not work well in a high quality setting. However, some control chart procedures reviewed in this chapter will be useful for very small values of $p_0$.

4.1.2 Various Proposed Methods

An ideal approach for monitoring high quality processes is to provide a simple and effective method for a practitioner to apply. The $p$-chart is easy to understand, but, as discussed later in this chapter, has weaknesses that result in disadvantages for detecting a shift from $p_0$ to $p_1$, when $p_0$ is very small. Many adaptations of this chart have been proposed because of its simplicity and ease of interpretation of the control chart statistic. Generally, however, the $p$-chart is not effective in the monitoring of high quality processes with Bernoulli data. The aggregation of data over time for a subgroup of $n > 1$ items results in a loss of information and unnecessary delays in detecting changes in the underlying proportion. The $p$-chart methods are covered for completeness of our review.

If Bernoulli data can be inspected continuously and sequentially, then another class of methods may be applied. Calvin (1983) originated the idea to chart the number of conforming items that were inspected between successive nonconforming items. This approach is beneficial with Bernoulli data because the observations are not pooled into subgroups of size $n$. This chart is referred to as the cumulative count of conforming (CCC) chart, and is statistically–based on the geometric distribution. A negative binomial extension to this chart is known as the CCC–$r$ chart, where it is of interest to plot counts of the number of items until $r$ nonconforming items are found. Other common names of charts based on the CCC and CCC–$r$ concept are the conforming run length (CRL) chart and the sum of conforming run lengths chart (SCRL), respectively. The $RL_1$ approach of Bourke (1991) is equivalent to the CCC chart. For the remainder of this chapter the CCC and CCC–$r$ notation is used.

The CCC chart is easy to design and interpret. Its weakness is in its ability to quickly
detect small shifts in the underlying proportion. We do not recommend CCC–$r$ charts ($r > 1$) when $p_0$ is small, however, for reasons discussed later in this chapter.

Cumulative sum (CUSUM) and exponentially weighted moving average (EWMA) charts have been proposed for high quality processes as well. The Shewhart–type methods such as the $p$–chart and CCC–type charts only use the most recent information for decision making. The CUSUM and EWMA charts accumulate more information over time. This has made them superior methods for detecting small changes in process parameters. These charts are used primarily for Phase II analysis, when it is of interest to detect parameter shifts in a process quickly. The CUSUM chart also has optimality properties as discussed by Moustakides (1986) and Hawkins and Olwell (1998).

A transformation of geometric counts to random variables that are approximately normally–distributed has been used in the high quality area as well. This allows the use of basic control charting methods to be used, making it a simple approach to implement for practitioners. For some transformations, however, the normal approximation is not accurate. These inaccuracies may give misleading results for expected statistical performance under some circumstances.

### 4.1.3 Advice to Practitioners

A transformation to normality may yield good statistical properties and allow a practitioner to use traditional SPC methods for their data. This would be more straightforward to implement than other charts presented later in this chapter, making it the easiest approach for practitioners. However, transforming data to a normal random variable may complicate interpretation on a control chart (Chang and Gan 2001), and in some cases lead to inaccuracies in obtaining desired in–control performance. If having known in–control performance is important, then the CCC charts should be used. If having known in–control performance and the best out–of–control performance is desired, then CUSUM charts should be used.

If our in–control parameter $p_0$ is unknown, it is common to estimate it using Phase I historical data. Yang et al. (2002) noted that for high quality processes, a very large historical data set is needed in order to estimate $p_0$ accurately because nonconforming
items rarely appear. An insufficient background sample size may result in a false alarm rate, $\alpha$, that is much larger than the specified value for Shewhart–type control charts.

### 4.1.4 Outline of Chapter

Many acronyms are used throughout this chapter to document the names of charts and performance criteria considered. A table of these acronyms are given as a reference.

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANDO</td>
<td>Average Number of Defectives Observed</td>
</tr>
<tr>
<td>ANOS</td>
<td>Average Number of Observations to Signal</td>
</tr>
<tr>
<td>AOQ</td>
<td>Average Outgoing Quality</td>
</tr>
<tr>
<td>ARL</td>
<td>Average Run Length</td>
</tr>
<tr>
<td>ATS</td>
<td>Average Time to Signal</td>
</tr>
<tr>
<td>CCC</td>
<td>Cumulative Count of Conforming</td>
</tr>
<tr>
<td>CRL</td>
<td>Conforming Run Length</td>
</tr>
<tr>
<td>CSP</td>
<td>Continuous Sampling Plan</td>
</tr>
<tr>
<td>CUSUM</td>
<td>Cumulative Sum</td>
</tr>
<tr>
<td>EWMA</td>
<td>Exponentially Weighted Moving Average</td>
</tr>
<tr>
<td>FSI</td>
<td>Fixed Sampling Interval</td>
</tr>
<tr>
<td>MBCUSUM</td>
<td>Markov Binary CUSUM</td>
</tr>
<tr>
<td>RL</td>
<td>Run Length</td>
</tr>
<tr>
<td>SCRL</td>
<td>Sum of Conforming Run Lengths</td>
</tr>
<tr>
<td>SDNOS</td>
<td>Standard Deviation of Number of Observations to Signal</td>
</tr>
<tr>
<td>SPRT</td>
<td>Sequential Probability Ratio Test</td>
</tr>
<tr>
<td>UGR</td>
<td>Unit and Group Runs</td>
</tr>
<tr>
<td>VSI</td>
<td>Variable Sampling Interval</td>
</tr>
</tbody>
</table>

The format of this chapter is as follows. In Section 4.2, we consider the criteria that should be used for evaluating the statistical performance for these types of control charts. Transformations to normality that are made on data from high quality processes to justify the use of standard control charting methods are then considered in Section 4.3. A discussion of control charts based on the geometric and negative binomial distributions follows in Sections 4.4 and 4.5. The uses of CUSUM and EWMA methods are then explored in Sections 4.6 and 4.7. The traditional Shewhart $p$–chart and modifications for
high quality processes are examined in Section 4.8. In Section 4.9, we review some of the comparisons made between methods. Some concluding remarks, advice to practitioners, and ideas for future research topics are given in Section 4.10.

4.2 Criteria for Comparing Control Chart Performance

As the methods for high quality processes are introduced and developed, it is important to make comparisons to determine which methods have the best performance over a range of values for $p_1$, given that the methods had roughly equivalent performance at $p_0$. Standard SPC applications use the average run length (ARL) as a criterion for determining the optimal method over a range of shifts. The ARL is defined as the expected number of plotted points on a control chart before the first signal is raised. This is an appropriate measure when using equal sample sizes and sampling intervals for all methods being compared.

A much more suitable metric for comparing across methods designed for high quality processes is the average number of observations to signal (ANOS). When using the CCC charts, for example, one plots a control chart statistic at irregularly spaced points in time. Several $p$-charts may also be compared using different values of $n$. The ANOS metric considers the total number of items inspected, whereas the ARL metric only takes into account the number of times a charting statistic is plotted. When the process is in control, the control chart limits should be set such that the in–control ANOS, $ANOS_0$, is equal or close to equal for all competing methods. Because of the discrete nature of this problem, equality may not be obtained exactly. Determining the best methods of performance is based on having the smallest out–of–control ANOS ($ANOS_1$) values across a range of shifts in $p$.

Other names for the ANOS metric include the average number inspected (ANI) and the average number of items to signal (ANIS). The ANOS notation is used throughout this chapter. The reader should give considerable attention to the metrics used in the various papers referenced for evaluating the effectiveness of the proposed methods.
4.3 Transformations to Normality

Various functions have been proposed for transforming observed geometric random variables from high quality processes to approximate normality. Generally, these methods work better for smaller values of $p_0$. Once approximate normality is achieved, standard control charting methods, such as the individuals control chart, can be used. Note that when $p$ is small, the geometric distribution is well-approximated by the exponential distribution. See Glushkovsky (1994) for additional information on the relationship between these two distributions.

Bourke (1991) proposed transforming geometric counts to approximately uniform-distributed values on the interval $(0, 1)$, when the process is in-control, using the cumulative distribution function (cdf). This leads to an easily constructed chart referred to as the $\alpha$-chart.

Nelson (1994) considered the time between events of nonconforming items to follow an exponential distribution. Therefore, if $W$ is an exponential random variable, then $W^{1/3.6}$ follows a Weibull distribution. This transformation was chosen by Nelson (1994).
because the parameters of this Weibull distribution yield skewness and kurtosis values that are close to that for a normal distribution. Use of this transformation is the only approach covered by Montgomery (2009, p. 324) to monitor high quality processes.

When \( p_0 < .01 \), Quesenberry (1995) noted that the control charts based on the binomial distribution were inadequate for monitoring a process. A transformation of a geometric count, \( Y \), to an approximately standard normal random variable is obtained by the use of the inverse normal cdf, \( Q = \Phi^{-1}(F(Y)) \), where \( F(\cdot) \) is the cumulative distribution function for a random variable. Control charts derived from this transformation are known as \( Q \) charts. As \( p_0 \) gets smaller, the normal approximation improves. After the transformation is applied, traditional analyses using the normal distribution can be used. The author considered the cases where \( p_0 \) is known and unknown. An application of the geometric \( Q \) chart in health surveillance was given by Quesenberry (2000). Yun and Youlin (1996) applied the methods presented by Quesenberry (1995) to the negative binomial distribution.

McCool and Joyner-Motley (1998) proposed a transformation of a geometric random variable using the natural log to obtain an approximately normal random variable. This was compared to the transformation used by Nelson (1994). McCool and Joyner-Motley (1998) compared the properties of the power and log transformations, \( W^{1/3.6} \), and \( \ln(W) \), respectively. However, both charts were shown to have poor performance when using a Shewhart chart based on the normal distribution. An exponentially weighted moving average (EWMA) chart was used for each of these transformed random variables to improve the detection of changes in \( p \). The authors concluded that the power transformation with an EWMA chart was the better method because the normal approximation was more accurate and it led to superior ARL performance for most process shifts. Other EWMA methods for transformed geometric random variables are discussed in Section 7.

Ryan (2000) provided a review of several arcsin transformations, where each charting statistic had similar properties. In the most comprehensive paper on the use of transformations with a geometric random variable, Xie et al. (2000b) considered a log transformation and a double-square root transformation of geometric counts. Two examples were explored to illustrate the use of these transformations, along with the \( Q \) chart by Quesenberry (1995), but no sensitivity analyses were used to directly compare the performance of these transformations. The authors recommended the double-square
root transformation \((Y^{1/4})\) because it is well–approximated by the normal distribution and was easy to compute. This transformation, and the similar \(W^{1/3.6}\) transformation, are simple tools that a practitioner may apply for this type of data. These two similar transformations appear to be the most widely accepted.

### 4.4 Shewhart Control Charts Based on the Geometric Distribution

#### 4.4.1 Basic Approach

The concept of the CCC chart was first developed by Calvin (1983). Suppose a data stream of Bernoulli data is monitored where items are inspected continuously. Let \(Y\) be the count of the number of items until a nonconforming item is observed. Then \(Y\) is a geometric random variable with probability mass function (pmf)

\[
f(y) = p(1-p)^{y-1},
\]

where \(y = 1, 2, \ldots\). This pmf could also be parameterized as the number of conforming items observed before the nonconforming item. Control limits were initially set by Calvin (1983) and Goh (1987b) using approximations based on the geometric distribution. The control limits most commonly used for the CCC chart are derived from the cumulative distribution function (cdf) of \(Y\), i.e.,

\[
F(y) = 1 - (1-p)^y,
\]

for \(y = 1, 2, \ldots\). Bourke (1991) and later Xie et al. (1995) set the control limits for an in–control value of \(p_0\) based on the probability limits, where \(F(\text{LCL}) \approx \alpha/2\) and \(F(\text{UCL}) \approx 1 - \alpha/2\) to obtain an overall false alarm rate close to \(\alpha\), with roughly equal tail probabilities. This methodology for finding the control limits of a control chart is known as the use of probability limits. These control limits are better to use than 3–sigma limits for the CCC chart because the geometric distribution is quite skewed. This
approach results in the following set of control limits:

\[
\begin{align*}
LCL &= \ln(1 - \alpha/2) / \ln(1 - p_0); \\
UCL &= \ln(\alpha/2) / \ln(1 - p_0).
\end{align*}
\] (4.3)

The specified \( \alpha/2 \) tail probabilities, particularly in the lower tail, may be difficult to obtain exactly because of the discreteness of the geometric distribution. A simple example of this is shown in Table 4.2, where \( p_0 = .001 \) and \( \alpha = .0027 \). The upper tail probability is close to .00135, but the lower tail probability is nearly 50% larger than specified and cannot be improved.

**Table 4.2: Probability Limits Example for CCC Chart**

<table>
<thead>
<tr>
<th>Limit</th>
<th>Value</th>
<th>Tail Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCL</td>
<td>1.350</td>
<td>.0019990</td>
</tr>
<tr>
<td>UCL</td>
<td>6604.346</td>
<td>.0013491</td>
</tr>
</tbody>
</table>

Smaller values of \( Y \) indicate a larger value of \( p \), while large values of \( Y \) indicate a smaller value of \( p \). Therefore, when \( p_1 > p_0 \), a signal is expected to occur below the LCL for the CCC chart. As stated earlier, the ANOS should be used as the performance metric, not the ARL.

The use of the CCC chart can be justified based on its simplicity. Control limits can be obtained easily based on the geometric distribution. There are some technical issues with respect to performance discussed in the next section, but these may not be of high importance to practitioners.

**4.4.2 Technical Issues with Two–Sided CCC Chart**

**Non–Maximal Metrics and Bias**

The two–sided CCC chart was designed to detect any shift in \( p \), but the skewness of the geometric distribution may cause problems with the ANOS performance. Earlier papers
on this topic focused on ARL performance, but the zero–state ANOS values may be calculated from the ARL values as well. The CCC control chart limits are set up such that there is non–maximal performance at $p_0$. Therefore, ANOS$_1 >$ ANOS$_0$ for some small upward shifts, $p_1 > p_0$. This issue was presented for ARL values by Xie and Goh (1997).

Xie et al. (2000a) discussed the issues with ARL$_1 >$ ARL$_0$ for small increases in $p$. An illustrative example is shown in Figure 4.1 where $p_0 = 50$ ppm, where the ARL values are not the maximum at $p_0$. The three curves are based on three separate false alarm rates considered. A solution to provide the largest ARL value at $p = p_0$ was provided by Xie et al. (2000a) for maximizing the ARL where the control limits are determined by taking the derivative of the ARL function with respect to $p$ and solving to find the value of $p$. A maximal ARL at $p_0$ was obtained by multiplying the original limits from Equation (4.3) by an adjustment factor. It was shown, however, that when the adjustment was made, the ARL$_0$ value could be far from $1/\alpha$, resulting in ‘biased’ limits, despite the ARL being maximized at $p_0$. A graph to better visualize this change was given by Chen (2009) and is shown in Figure 4.2(a).

![Figure 4.1](image.png)

**Figure 4.1:** Nonmaximal in–control ARL plots for two–sided CCC chart from Xie et al. (2000a). Used with permission by World Scientific.
(a) Results with Biased Limits from Xie et al. (2000a)

(b) Results with Proposed Limits from Chen (2009)

Figure 4.2: ARL Curves Based on Two-Sided CCC Chart Illustrated by Chen (2009)

Possible Resolutions

This technical problem was discussed further by Zhang et al. (2004). The control limits were set such that the sum of the tail probabilities using a new set of probability limits was as close as possible to $\alpha$. The two tail probabilities were not necessarily equal in this case as well. The authors stated the procedure results in the maximal ARL at $p_0$, but no comparisons were made.
Chen (2009) presented an iterative method for obtaining near-maximal and near-unbiased ARL₀ values. This has the disadvantage that no closed form solution exists for the control limits. Two metrics that measured the effectiveness of this approach were the bias of the ARL₀ values and the deviation percentage from p₀ to the value of p where the ARL was maximized. These measures were calculated for the limits from Equation (4.3), the adjusted limits constructed by Xie et al. (2000a), and the new limits calculated in their paper, based on a range of values of p₀ and fixed α. The results showed that the new method was best based on both criteria. A graph of the ARL function versus p for all three methods from Chen (2009) are given in Figure 4.2(b). The ‘Initiative’ curve is based on the original control limits used for the CCC chart. This method may provide better results than previous methods considered to resolve this issue.

The problems with non-maximal and biased ARL results for the two-sided CCC chart are not present for the one-sided case, where only a shift p₁ > p₀ is of interest to detect. For this case, only the LCL is needed for monitoring a process. The LCL can be obtained from Equation (4.3) by replacing α/2 with α.

4.4.3 Other Approaches

Lucas (1989) suggested a runs rule approach, where a signal is given if a nonconforming item appears in 2 of s items. This rule results in a method equivalent to the use of the CCC Shewhart chart, but can be represented in a CUSUM framework. A runs rule method considers the most recent b (b > 1) sample statistics for determining if a shift has occurred. Markov chains are typically required to represent charts that incorporate runs rules. The Markov chain approach was given by Lucas (1989), as well as a recommendation for choosing the value of s. However, some of the other papers reviewed in this chapter, e.g. Kuralmani et al. (2002), do not use a correct method for evaluating a runs rule method. For a discussion of the use of Markov chains with runs rules, see Champ and Woodall (1987).

A number of researchers have proposed modifications of the CCC chart in order to improve performance. A two-stage control chart used by Chan et al. (1997) will signal if a CCC count, Y, is below a specified limit, or if two consecutive CCC counts are below a warning limit. Chan et al. (2003) considered using a two-stage control chart as well,
however only the properties of this chart were presented and no sensitivity comparisons
were performed. This should be considered a runs rule method as well.

Kuralmani et al. (2002) used a conditional CCC chart that signaled if \( Y \) was outside of
the control limits, or if 2 of \( s \) counts were outside of less–restrictive warning limits. This
is also an application of a runs rule. The problems of the non–maximal ARL\(_0\) values at
\( p_0 \) were noted for this two–sided chart. Therefore, a similar type of adjustment that was
made by Xie et al. (2000a) was applied to obtain new control limits. Tables containing
the new control limits based on \( p_0 \) and \( \alpha \) were given in this paper.

The conditional procedure used by Kuralmani et al. (2002) was discussed by Noorossana
et al. (2007), who showed that the properties of the conditional procedure were deter-
mined based on an incorrect assumption of independence. The ARLs obtained by Kural-
mani et al. (2002) tended to underestimate the true ARL quantities. Noorossana et al.
(2007) presented tables of new control limits based on their approach.

Ranjan et al. (2003) analyzed the two–sided CCC chart when inspection errors were
present. Depending on the magnitude of the inspection error rates, the ARLs may have
a very large bias when the control limits were adjusted to be maximal at \( p_0 \).

A two–sided CCC chart using supplementary runs rules, evaluated using a Markov
chain approach, was considered by Cheng and Chen (2008). A sensitivity analysis com-
pared this chart to the traditional CCC chart for \( p_0 = 50 \) ppm with a range of \( p_1 \) from 10
to 500 ppm. ARL comparisons between the two methods showed that the supplementary
runs rules method was superior when \( p_1 > p_0 \), but the traditional CCC chart was slightly
better when \( p_1 < p_0 \). The authors also noted that the issue of non–maximal in–control
ARLs was present for both charts as well.

Although the use of runs rules can improve the performance of the CCC approach,
we recommend the use of CUSUM charts. Generally the runs rule approaches are more
difficult to design and are less efficient than CUSUM methods.

When the independence assumption is removed for Bernoulli data, a dependence
structure is present, and was considered by Lai et al. (1998), where a correlated model
for the binomial, geometric, and negative binomial distributions were used. Lai et al.
(2000) used a two–state Markov model to represent the dependent process. The false
alarm rate may be much different from \( \alpha \), depending on the size of the autocorrelation
4.4.4 Fixed–Shift and Random–Shift Models

A steady–state analysis will typically consist of simulating in–control runs, then having a shift occur during monitoring. For control chart methods based on the geometric distribution, a number of geometric random variables may be simulated before a shift is present. However, in this model, there is an implicit assumption that a shift occurs immediately after a nonconforming item is found. This is known as the fixed–shift model.

Wu and Spedding (1999) considered the random–shift model for the CCC chart in detail. The random–shift model incorporates the case where a shift from $p_0$ to $p_1$ occurs at any time during monitoring, rather than immediately after a nonconforming item is found. A simple illustration of this is shown in Figure 4.3. Note that the CRL terminology is used in this figure. The plotted CCC value in this figure is $k + d$, but the count is based on a process involving two different nonconforming rates, $p_0$ and $p_1$. With a large number of conforming items between nonconforming ones with 100% inspection, it is more reasonable to assume that a shift may occur between conforming items.

![Figure 4.3: Random Shift Model – The “Shifting Sample” from Wu and Spedding (1999)](image)

If this type of shift is not considered, the steady–state ANOS (SSANOS) values may be quite different. We find the random–shift model to be much more realistic than the fixed–shift model when steady–state performance comparisons are made.

If items are aggregated somewhat arbitrarily into subgroups of size $n$, such as for
the $p$–chart with 100% sampling, the random–shift model should be applied as well. If interval sampling is used, this may not be necessary, since use of rational subgroups may be assumed in some cases, where any process change would occur between samples.

### 4.5 Control Charts Based on the Negative Binomial Distribution

If $Y_1, Y_2, \ldots$ are independent geometric random variables with pmf in Equation (4.1), then $Z = \sum_{i=1}^{r} Y_i$ is a negative binomial random variable with pmf

$$f(z) = \binom{z-1}{r-1} p^r (1-p)^{z-r},$$

for $z = r, r+1, \ldots$. This distribution can be used when monitoring the number of items until $r$ nonconforming items are found. A special case is the geometric distribution, where $r = 1$. In general, however, these charts should not be used for monitoring high quality processes since aggregating the geometric CCC values results in a loss of information and poor statistical performance compared, for example, to CUSUM approaches. For larger values of $p_0$, the negative binomial charts may be useful.

The first use of the negative binomial distribution for monitoring Bernoulli data was by Bourke (1991). The sums of independent non–overlapping geometric random variables to form negative binomial random variables was used by Kaminsky et al. (1992). A Shewhart–type chart, the $g$–chart was developed, with $k$–sigma control limits based on the negative binomial distribution.

#### 4.5.1 Control Limits

Xie and Goh (1997) investigated the problems of using $k$–sigma limits with the $g$–chart. In many cases, the computed LCL was negative, so no LCL could be used. In this context, process deterioration may not be detected. The authors recommended using probability limits similar to what was proposed by Xie et al. (1995) for the CCC chart. The technical issues with the $g$–chart were also discussed by Xie et al. (2000b).
A control chart for a negative binomial random variable was given by Xie et al. (1999). The authors proposed a general form for the two–sided CCC chart, where \( r > 1 \), called the CCC–\( r \) chart. The control limits were based on the probability limits from the cdf of \( Z \) and were presented for values of \( r \) from 1 to 6 and \( p_0 \) ranging from .0001 to .10. As \( p_0 \) decreases and \( r \) increases, the control limits become quite wide. This is because very few nonconforming items are expected to appear in this case. Since a decision cannot be made until \( r \) nonconforming items are detected, there may be extremely long waiting times until a statistic can be obtained. Hence, for extremely small nonconforming rates \( (p_0 < 100 \text{ ppm}) \), the authors recommended using \( r = 1 \).

Wu et al. (2001b) also considered the probability limits for constructing a CCC–\( r \) chart. The design of all parameters for this chart are considered and evaluations are made using the fixed–shift and random–shift models.

### 4.5.2 Choice of \( r \)

Schwertman (2005) provided extensive tables of the LCL, UCL, and ARL values for the CCC–\( r \) chart. A range for \( p_0 \) from .005 to .10 was considered, with values of \( r \) as large as 20. The author recommended using \( r = 2 \) or 3 for “increased sensitivity”. In those tables, ARL\(_1 \) values were given for four different shifts. From their tables, one may conclude that a large value of \( r \) should be used because of the lower ARL\(_1 \) values. However, we reiterate our point that comparisons should be made based on the ANOS, not ARL, because the number of items between plotted points can vary significantly. These tables should not be used for comparisons.

The ANOS performance of the CCC–\( r \) chart may be similar for several values of \( r \). Di Bucchianico et al. (2005) used the standard deviation of the number of observations to signal (SDNOS) as an additional performance criterion. This was applied for a case study in order to determine the best value of \( r \). The authors also explained the problem associated with using the ARL to measure the performance of the CCC–\( r \) chart. A case study was examined for this problem, but the only recommendation given is to use a smaller value of \( r \) if two choices yielded similar performance.

Lai and Govindaraju (2008) examined a design for the CCC–\( r \) chart, where a signal
was not raised until there were \( v \) breaches outside of the control limits. The default case used for the standard control charts is \( v = 1 \). The optimal choices of \( v \) and \( r \) were analyzed when the ANOS\(_0\) value was specified. The ANOS\(_1\) values and two measures of the variability of the number of observations to signal were considered for chart performance. Several examples were studied by Lai and Govindaraju (2008), but no general conclusions could be reached to optimize performance with respect to \( r \) and \( v \).

Ohta et al. (2001) stated that the CCC–\( r \) chart was more sensitive to small upward shifts in \( p \), but long waiting times are an issue. An economic model was constructed to find the best value of \( r \) to use as a function of a set of cost and time variables. In this paper, only an increase in \( p \) was of interest to be detected. For an economic model approach for the CCC chart, see Xie et al. (2001). A method to determine the optimal value of \( r \) for the CCC–\( r \) chart was also considered by Albers (2010).

### 4.5.3 Other Methods

In general, the negative binomial count \( Z_1 \) is constructed from counts \( Y_1 \) to \( Y_r \), \( Z_2 \) is constructed from counts \( Y_{r+1} \) to \( Y_{2r} \), and so forth. Therefore all values of \( Z_i \) \((i = 1, 2, \ldots)\) are independent from one another. Bourke (1991) used a negative binomial random variables by summing the two most recent geometric random variables, with the values plotted on the \( RL_2 \) chart. The \( RL_2 \) statistic at time \( t \) is expressed as

\[
RL_2(t) = Y_t + Y_{t-1},
\]

\( t = 2, 3, \ldots \). In general, the statistical performance of this chart is not as desirable as that of CUSUM–based methods.

Ohta and Kusukawa (2004) developed a confirmation sample CCC–\( r \) chart when interval sampling is used. If a CCC–\( r \) value \( Z \) is outside of the control limits, then an additional sample is drawn to obtain a second CCC–\( r \) value. If both quantities are outside of the control limits, then a signal is raised.

In these past two sections, methods based on the number of conforming items between two nonconforming items were considered. An overview of this area was given by Xie et al. (2002), who covered many of these approaches, provided values for the control
limits of specific charts, and discussed issues for monitoring with the CCC and CCC–r charts. The technical issues with the two–sided CCC chart discussed in Section 4.2 also apply to the CCC–r chart.

4.6 Cumulative Sum Control Charts

It has been well–documented in the SPC literature that cumulative sum (CUSUM) control charts have better ARL performance than Shewhart control charts for detecting small to medium–sized shifts in the parameter of interest. For a comprehensive overview of CUSUM charts for numerous distributions, see Hawkins and Olwell (1998).

A CUSUM chart is designed to be optimal for detecting a specified change from \( p_0 \) to \( p_1 \). However, \( p_1 \) should be chosen such that the chart has good statistical performance for a wide range of shifts. In many cases for high quality processes, only an increase in \( p \) is of interest to detect, so a one–sided CUSUM chart is used.

Bourke (1991) first proposed a CUSUM chart based on geometric random variables. Xie et al. (1998) considered the equivalent CUSUM V–mask representation based on geometric observations. The V–mask is drawn on a control chart with two arms such that a signal is raised if any of the previous cumulative sum statistics are outside of the arms (Montgomery 2009, p. 415). It has been argued that the V–mask chart is difficult to construct, is not very flexible, and can be difficult to interpret since there is a separate V–mask for each observation (Woodall and Adams 1993; Montgomery 2009, p. 416). We do not recommend the use of the V–mask procedure.

4.6.1 Geometric CUSUM Chart

The upper–sided CUSUM statistics, \( G_t, t = 1, 2, \ldots \) for a geometric random variable are

\[
G_t = \max(0, G_{t-1} - Y_t + k_G),
\]

(4.6)

where \( G_0 = 0, Y_t \) is the CCC count at time \( t \), and \( k_G \) can often be conveniently specified to be an integer. The formula for \( k_G \) is based on the log likelihood ratio for optimally
detecting a shift from \( p_0 \) to \( p_1 \). This value may be expressed as

\[
k_G = \frac{\ln \left( \frac{p_1(1-p_0)}{p_0(1-p_1)} \right)}{\ln \left( \frac{1-p_0}{1-p_1} \right)}.
\]

(4.7)

A signal for this chart is raised if \( G_t \geq h_G \), where \( h_G \) is known as the decision interval or decision limit.

For a high quality process, Bourke (1991) used a geometric CUSUM chart, which was referred to as the run–length CUSUM (RL–CUSUM) chart. Another name for the CUSUM chart based on geometric random variables is the CRL–CUSUM chart. A large–scale analysis of the ANOS values for this chart were presented for \( p_0 \) ranging from .002 to .02, with different charting parameters. The range for \( \gamma \) considered by Bourke (1991) was from .4 to 5. A one–sided CUSUM chart was used to detect \( p_1 > p_0 \), so for the cases of process improvement \( (p_1 < p_0) \), the performance of this chart was rather poor, as would be expected.

Sun and Zhang (2000) considered both the lower and upper–sided geometric CUSUM charts for detecting a change in \( p \). A table of parameters to use in order to obtain a specified ARL\(_0\) value was presented. A similar approach, where the ANOS\(_0\) values were specified instead, was presented by Chang and Gan (2001). The optimal parameters used by Chang and Gan (2001) were based only on detecting an increase in \( p \) for \( p_0 \) varying from 10 to 500 ppm.

### 4.6.2 Binomial CUSUM Chart

If information on successive individual items is not available, and items must be collected in subgroups of size \( n \), the binomial CUSUM chart may be appropriate to use. This chart was studied extensively by Gan (1993). Optimal parameters for this chart were given for \( p_0 \) spanning from .05 to .20. However, \( n \) was often quite large, where the minimum value considered was \( n = 100 \), so there would be a reasonable chance that at least one nonconforming item would appear in each sample. If \( p_0 \) is very small though, \( n \) would need to be extremely large for this condition to be met.
A binomial CUSUM chart designed to detect large shifts in $p$ was considered by Wu et al. (2008). The CUSUM statistic was altered in order to improve on the performance for these shifts. This chart compared favorably to the traditional binomial CUSUM chart for large values of $\gamma$ for two cases. However, the traditional binomial CUSUM chart would be expected to perform best overall because of its advantage in detecting smaller shifts.

### 4.6.3 Bernoulli CUSUM Chart

Reynolds and Stoumbos (1999) studied the binomial CUSUM chart and a special case of this chart where $n = 1$, which was termed the Bernoulli CUSUM chart. A primary advantage of this method, providing calculations are automated, is that a CUSUM statistic is calculated after each item is inspected. The upper-sided Bernoulli CUSUM statistics, $B_t$, $t = 1, 2, \ldots$, are

$$B_t = \max(0, B_{t-1} + X_t - k_B),$$

where $B_0 = 0$ and $k_B = 1/k_G$. A signal is raised if $B_t \geq h_B$. It was shown by Reynolds and Stoumbos (1999) that the Bernoulli CUSUM chart had much better ANOS performance for detecting an increase in $p$ than binomial CUSUM charts with $n = 51$ and 100. Megahed et al. (2011) considered an ARL property of the two-sided Bernoulli CUSUM chart.

Bourke (2001b) considered finding the optimal $n$ for a binomial CUSUM chart and concluded that the Bernoulli CUSUM chart had the best statistical performance. In general, the statistical performance of the Bernoulli CUSUM chart is the best of all competing methods. The performance is expected to be good based on the CUSUM optimality results of Moustakides (1986).

An extension of the Bernoulli CUSUM chart may consider the random variables to be correlated, removing the standard assumption of independence for our Bernoulli data. Mousavi and Reynolds (2009) addressed this problem using a Markov binary CUSUM (MBCUSUM) chart. It was assumed that a positive correlation exists with a first-order dependence, represented by a two-state Markov model. As the autocorrelation coefficient, $\rho$, gets larger in magnitude, the in-control ANOS values based on the Bernoulli
CUSUM charts become quite small. Hence, the standard Bernoulli CUSUM methods become unreliable, even for smaller values of \( \rho \), because the actual ANOS\(_0\) values are much lower than for the special case with independent observations, where \( \rho = 0 \).

### 4.6.4 The Relationship Between the Geometric and Bernoulli CUSUM Charts

Reynolds and Stoumbos (1999) noted that the geometric CUSUM chart is equivalent to the Bernoulli CUSUM chart, but only if the Bernoulli CUSUM chart has a head-start feature. If monitoring starts at time \( t = 1 \), then the geometric CUSUM chart is based on the implicit assumption that a nonconforming item was found at time \( t = 0 \). When comparing the two charts for detecting an increase in \( p \), the zero-state ANOS performance of the geometric CUSUM chart is always better than the Bernoulli CUSUM chart without the head-start. An example of this concept was given by Bourke (2001a). Chang and Gan (2001) provided a proof showing the equivalence of these charts when

\[
B_0 = \frac{(k_G - 1)}{k_G} \quad \text{and} \quad h_B = \frac{(h_G + k_G - 1)}{k_G}.
\]

It may be beneficial to use the Bernoulli CUSUM chart rather than the geometric CUSUM chart in many applications. The head-start feature should be applied to the Bernoulli CUSUM chart, which will give fairer zero-state comparisons to many competing methods. For detecting a decrease in \( p \), a geometric CUSUM chart requires that a nonconforming item be found before plotting a point on a control chart unless curtailed sampling is used, for which a signal is given if a required number of conforming items are observed based on the current CUSUM value. As \( p \) gets very small, it will take much longer to obtain a nonconforming item. Therefore, for large decreasing shifts, where \( \gamma << 1 \), there will be a delay in obtaining a signal (Chang and Gan 2001) without curtailed sampling. However, since the Bernoulli CUSUM chart is updated after every item is inspected, the charting statistic can result in a signal before a nonconforming item is obtained. The random-shift model further complicates steady-state evaluation of the geometric CUSUM chart. For more information on the relationship between the geometric and Bernoulli CUSUM charts, see Szarka and Woodall (2011a).
4.7 Exponentially Weighted Moving Average Control Charts

Like the CUSUM chart, an EWMA chart dominates the Shewhart chart in detecting small to moderate-sized shifts in the parameter of interest in terms of statistical performance. The choice between use of a CUSUM chart and an EWMA chart depends largely on the personal preference of the user. We prefer the CUSUM approach for high quality applications due to its optimality properties.

The general form for two–sided EWMA statistics, $E_t$, $t = 1, 2, \ldots$, is

$$E_t = \lambda V_t + (1 - \lambda)E_{t-1},$$

for random variables $V_t$, $t = 1, 2, \ldots$, and $E_0 = \mu_V$, the in–control mean of $V$. The parameter $\lambda$ is an assigned weight to be given to the most recent data point. When $\lambda = 1$, the EWMA chart is equivalent to the Shewhart chart. Typically, a weight of $0.05 \leq \lambda \leq 0.25$ is recommended (Montgomery 2009, p. 423). The lower and upper control limits for an EWMA chart at time $t$ are expressed as

$$LCL(t) = \mu_V - L\sigma_V \sqrt{\frac{\lambda}{2 - \lambda} [1 - (1 - \lambda)^{2t}]};$$

$$UCL(t) = \mu_V + L\sigma_V \sqrt{\frac{\lambda}{2 - \lambda} [1 - (1 - \lambda)^{2t}]},$$

for $t = 1, 2, \ldots$, where $L$ is a specified multiplier and $\sigma_V$ is the standard deviation of $V$. Asymptotic limits for the EWMA chart are sometimes considered, where the term $[1 - (1 - \lambda)^{2t}]$ is removed from Equation (4.10).

4.7.1 One–Sided EWMA Control Charts

If the primary focus is to detect a change in a parameter in only one direction, then one control limit may be used. A reflecting barrier should be used such that the control chart statistic has a set maximum distance from the control limit. Otherwise, when a shift in the parameter occurs for the steady–state case, $E_t$ could be quite far from the control
limit. For further discussion of the resulting inertial effect, see Woodall and Mahmoud (2005). A one–sided upper EWMA control chart statistic at time $t$ may be expressed as

$$E_t = \max \left[ \mu V, \lambda V_t + (1 - \lambda)E_{t-1} \right],$$

(4.11)

for $t = 1, 2, \ldots$. Crowder and Hamilton (1992) first proposed an EWMA method with a reflecting barrier, but in another context.

### 4.7.2 Applications

The EWMA chart for high quality processes with Bernoulli data was first presented by McCool and Joyner-Motley (1998) based on a transformation to an approximately normal random variable. Spliid (2010) considered an EWMA chart for a transformation of a geometric random variable using an adjusted log transformation to the Gumbel distribution. Comparisons were made to the double–square root transformation ($Y^{1/4}$) to the normal distribution. The former method had better ANOS properties for large increases in $p$, but the latter method was best for small increases in $p$.

Sun and Zhang (2000) considered developing an EWMA chart with the asymptotic control limits for monitoring geometric random variables. A large range of values for $L$ were determined given a fixed value of $\lambda$ in order to obtain a desired in–control ARL value for their chart.

An EWMA chart for monitoring a negative binomial random variable was presented by Kotani et al. (2005). The asymptotic control limits were used in this paper. The ARL values for this chart were derived using a Markov chain approach, although the authors refer to this metric erroneously as the ANOS. For small values of $\lambda$ (.05 to .08), $L$ can be chosen such that a particular $ARL_0$ value is obtained. The results were applied for $r = 2$ and 5.

EWMA charts were derived by Yeh et al. (2008) for the geometric, Bernoulli, and binomial distributions. One–sided charts were used for detecting an increase in $p$, but the reflecting barrier presented in Equation (4.11) was not used for these charts. The authors stated that this may give the charts an advantage in signaling faster when using a zero–state analysis. However, it was discovered that the EWMA charts had worse
steady-state ANOS performance versus the zero-state case.

4.8 The Traditional Shewhart $p$–Chart

The traditional SPC approach for monitoring a proportion is the use of the Shewhart $p$–chart based on the binomial distribution. At each sampling interval, suppose a fixed sample of size $n$ is obtained. The plotted Phase II control chart statistics are the sample proportions of defective items in the samples, $\hat{p}_i$, $i = 1, 2, \ldots$. The control limits for monitoring the nonconforming rate with the $p$–chart are usually given as

$$\text{LCL} = p_0 - k \sqrt{\frac{p_0(1 - p_0)}{n}};$$
$$\text{UCL} = p_0 + k \sqrt{\frac{p_0(1 - p_0)}{n}}. \quad (4.12)$$

The in–control value $p_0$ is assumed to be known. Typically, $k = 3$ is chosen to coincide with the “3–sigma” limits used for traditional Shewhart charts. This standard approach does not work when $p_0$ is small, so some modifications have been proposed. In general, however these methods should not be used by practitioners for monitoring high–quality Bernoulli data.

4.8.1 Issues for Small Values of $p_0$

Unless $n$ is very large, the computed LCL in Equation (4.12) is negative when $p_0$ is small, meaning that there is no LCL, and one is unable to use the chart to detect process improvements. To ensure a non–negative LCL when $k = 3$, $n > 9(1 - p_0)/p_0$ must be satisfied. If $p_0 = .01$, $n$ must be greater than 891. In most cases, the sample size required is unreasonably large for high quality processes.

Another undesirable feature of some $p$–charts is that the UCL may be so small that the chart signals any time a single nonconforming item is found in a sample (Montgomery 2009, p. 298). This can result in a higher false alarm rate than desired. One solution to this approach is to choose a value of $n$ large enough such that the probability at least
one nonconforming item is present in a sample is reasonably large (Montgomery 2009, pp. 298–299). The false alarm rate may also be quite different from \( \alpha \) because of the discreteness of the binomial distribution. Lucas et al. (2006) recommended adding a value of \( 1/n \) to the calculated UCL in order to obtain a smaller false alarm rate.

Acosta-Mejia (1999) showed that the traditional \( p \)-chart with control limits from Equation (4.12) when \( k = 3 \) has non–maximal properties. Two examples of this condition occurring were for \((n, p_0) = (1000, .01) \) and \((5000, .01) \). A graph was shown to illustrate the ARL values for a range of \( p \) for these charts; the maximum ARL was not attained at \( p = .01 \) for these cases. The ARL may be used for comparisons to other adaptations of this chart if \( n \) is equal for all cases. Otherwise, the ANOS metric should be used.

The \( p \)-chart is also unable to quickly detect small shifts from \( p_0 \) to \( p_1 \). For further reading on problems with the \( p \)-chart with high quality processes, see Goh (1987a,b), Goh and Xie (1994, 2003), Ryan and Schwertman (1997), Xie et al. (1999), Ryan (2000), and Chan et al. (2003).

### 4.8.2 Modifications

#### Probability Limits

Because of the discrete nature of the process, combinations of \((n, p_0) \) may not lead to in–control lower and upper tail probabilities summing to a value near the specified value of \( \alpha \). The lower and upper tail probabilities are expressed as \( P(\hat{p} < \text{LCL}) \) and \( P(\hat{p} > \text{UCL}) \), respectively. Several examples of this phenomenon were given by Xie and Goh (1993). For small values of \( p_0 \), the traditional \( p \)-chart with \( k = 3 \) may have a much different false alarm rate than the specified value of .0027, which is somewhat arbitrary, but commonly used for many Shewhart charts. Ryan and Schwertman (1997) considered a range of values for \( p_0 \) from .005 to .10, where \( n \), the LCL, and the UCL were given such that the total tail probability was close to .0027 using probability limits. A program for obtaining these limits was given by Schwertman and Ryan (1997).

A sensitivity analysis by Ryan and Schwertman (1997) considered the ARL values for \( \gamma = .75 \) and 1.25. The ARL values for the chart using the fixed probability limits were compared to the ARLs using the traditional 3–sigma limits from Equation (4.12). The
results using the probability limits were more consistent than the traditional limits. A $p$–chart with adjusted control limits was also introduced using a Poisson approximation to the binomial distribution for $p_0 \leq .03$. For larger values of $p$, the computer program from Schwertman and Ryan (1997) was recommended. See Chen (1998) for adjusted control limits based on the Cornish–Fisher expansion.

An adjusted confidence interval approach with a better coverage probability was considered by Wang (2009), who used the Agresti–Coull confidence interval for adjusting the $p$–chart for small values of $p$.

For detecting process improvements using the $p$–chart, Xie and Goh (1993) considered finding a minimum $n$ for a given $p_0$ such that the probability of zero nonconforming items appearing in a sample is equal to a specified value with a fixed certainty level. The subgroup size $n$ can be extremely large for small values of $p_0$ and the specified probability. If there are no nonconforming items in a sample, a signal is given that $p$ has decreased. The values of $p_0$ considered ranged from .005 to .30 in their paper. Smaller values of $p_0$ were considered previously by Xie and Goh (1992).

Supplementary Runs Rules

Goh (1987a) proposed a ‘pattern recognition’ for the $p$–chart, where if $a$ of the last $b$ samples ($a \leq b$) contained $d$ nonconforming items, a signal was raised. This is an application of a runs rule. Obtaining a false alarm rate close to a specified value of $\alpha$ was considered for a few cases, although use of a Markov chain would be needed to determine meaningful properties for the method.

The issues with a non–maximal in–control ARL was discussed by Acosta-Mejia (1999). A method for obtaining an unbiased run length chart was introduced by using a runs rule method to replace the LCL. If $m$ consecutive samples of size $n$ resulted in sample proportions below a modified center line, a signal was raised. The Markov chain for this rule was constructed to obtain the ARL values.

The use of probability limits for detecting a shift in $p$ for a $p$–chart was considered by Chang and Gan (2007). A runs rule was used in this paper with its properties determined from a Markov chain approach. A signal indicating process improvement occurred if $w$
consecutive samples of size $n$ were observed with $d = 0$ nonconformities. To detect process deterioration, a signal was issued if $d > 1$ for any sample, or if $a$ of $b$ consecutive samples had $d = 1$ nonconforming item. The authors developed separate charts according to three groupings of increases in $p$ that were of interest to detect based on the severity of a shift. Tables with values of $p_0$ between 100 and 500 ppm were constructed based on fixed $\text{ARL}_0$ and $\text{ANOS}_0$ quantities. From these tables, the reader may choose a sampling scheme to use for a particular type of shift to detect. A large number of parameters $(n, p_0, \gamma, a, b, w)$ must be considered when creating this design. Practical issues may arise since the recommended value for $n$ may vary greatly depending on some minor changes in settings. A program for determining supplementary runs rules for the $p$–chart was also provided by Nelson (1997).

Tang and Cheong (2006) introduced a chain inspection scheme by groups for detecting a shift in $p$ where the data is considered to be correlated. A signal for detecting an increase in $p$ are the same as that used by Chang and Gan (2007) with a runs rule approach. The ANOS values were compared to the CCC chart.

Lucas et al. (2006) recommended policies for detecting $p_1 < p_0$ based on $P(\hat{p} = 0)$ for the $p$–chart given $(n, p_0)$. For high quality processes, $P(\hat{p} = 0)$ for a moderate value of $n$ will be large. A decrease in $p$ is signaled using one of three methods, depending on the values of $(n, p_0)$ chosen. These methods are: if $w$ consecutive samples contain $d = 0$ nonconforming items, if 2 in $s$ samples have $d = 1$ nonconformity, and if a plotted value falls below the traditional LCL calculated in Equation (4.12).

Other Methods

Wu et al. (2006) considered a $p$–chart to detect an increase in $p$ using curtailed sampling. It was shown in their paper that this method was more powerful than the traditional $p$–chart, based on smaller out–of–control SSANOS values across a variety of increases in $p$, where $p_0$ ranged from .005 to .02.

Weiss (2009) used a Markov $np$–chart where a serial dependence was considered. For a wide range of values $(n, p_0)$, tables are given with control limits, where the effect of $\rho$ on $\text{ARL}_0$ was illustrated. Adjustments to the control limits were suggested for these cases so that a specified $\text{ARL}_0$ value may be obtained. However, the minimum value of
Recap of the $p$–Chart

Again, we have discussed for completeness various modifications to the $p$–chart for use in monitoring high quality processes. The statistical performance of these methods, however, will not be competitive with the other methods discussed in this chapter. We do not advise the use of $p$–charts with Bernoulli data from high quality processes.

4.9 Summary of Performance Comparisons

With many proposed methods for monitoring high quality processes, the list of papers that contain comparisons of methods is quite large. The papers reviewed in this section are not exhaustive of the comparisons that can be found in the literature. Many papers focus only on detecting a sustained increase in $p$. Overall, the performance comparisons favor the use of the Bernoulli or the equivalent geometric CUSUM approaches, as would be expected from optimality results. The $p$–chart fares very poorly. The zero–state and steady–state performance based on the fixed–shift model is potentially misleading due to implicit head–starts associated with many of the proposed methods, as well as the often unwarranted assumption of a parameter shift only occurring after a nonconforming item is found.

4.9.1 Comparisons with CUSUM Charts

Bourke (1991) compared the $RL_2$, geometric CUSUM, and $p$–charts. Only an increase in $p$ was considered, and evaluations were based on the zero–state ANOS performance. The in–control nonconforming rate used was $p_0 = .01$, with values of $\gamma$ as large as 15. The geometric CUSUM chart appeared to be the best overall method for this case. A $p$–chart with a fixed UCL was also shown to be superior to the $RL_2$ chart for smaller shifts in $p$ as well.
Wu et al. (2000) provided an overview of the $p$–chart, geometric CUSUM, and CCC–$r$ charts for the two–sided case. The $p$–chart was quickly disregarded as a competitive method. The random shift model was considered for the remaining charts. For $p_0$ ranging from .0001 to .01, and $\gamma$ from 0.1 to 10, the geometric CUSUM chart was the best for detecting a decrease and large increases in $p$. But for small, upward shifts in $p$, the CCC–$r$ chart ($r = 3$) was preferred. Very narrow control limits were used with these charts, which may have a negative effect on chart performance.

Bourke (2001a) considered sampling items at fixed intervals. Since each item was no longer being inspected, there were nonconforming items produced that were undetected. A metric used in this case was the average number of defectives observed (ANDO), which is an ARL–like metric. The ANOS criterion was used as well. Zero and steady–state analysis with a random–shift model was considered for the geometric CUSUM chart and $p$–chart, where the geometric CUSUM chart was shown to be the better method for detecting an increase in $p$ for $p_0 = .01$ and .002.

Wu et al. (2001a) considered the $p$–chart with a fixed UCL, the CCC chart, and a proposed synthetic chart that combines the use of the former two charts for detecting an increase in $p$. Using the zero–state ANOS criterion, the synthetic chart typically had the best performance for various simulation studies. Bourke (2008) compared the $p$–chart, synthetic, $RL_2$ and geometric CUSUM charts using the zero–state and steady–state ANOS criteria for the random–shift model. In the steady–state case, the synthetic chart performed only marginally better than the $p$–chart. The author noted that when this method was proposed by Wu et al. (2001a), an implicit head–start was given to this method, which gave it a large overall zero–state advantage. Other issues with the synthetic chart were addressed by Davis and Woodall (2002). The performance of the $RL_2$ chart was greatly affected by the SSANOS$_0$ value chosen. The geometric CUSUM chart was the best overall method.

Wu et al. (2006) used curtailed sampling for the $p$–chart. The analysis contained two studies and an example using the $p$–chart, $p$–chart with curtailed sampling, $RL_2$, and geometric CUSUM charts. Steady–state analysis with the random–shift model was considered. The geometric CUSUM chart was considered the best overall method using the SSANOS metric. The authors recommended curtailed sampling based on its simplicity. However, an iterative procedure must be performed for choosing $n$ for this method.
A unit and group runs (UGR) chart was presented by Gadre and Rattihalli (2005). It is similar to a synthetic chart because depending on the state of the process, a CCC chart in combination with a $p$–chart switching rule with a fixed UCL is used. Many comparisons were made to the geometric CUSUM and Bernoulli CUSUM charts, where the ANOS performance for the UGR chart for a zero–state analysis was overwhelmingly dominant. A Markov chain approach was used, and comparisons were given to the $p$–chart and synthetic chart under steady–state conditions, where the UGR chart was the best method. However, the in–control performance of this chart is misleading due to the sampling scheme being considered due to the presence of outliers in the run–length distribution. See Szarka and Woodall (2011a) for more information on the evaluation of the UGR chart.

A control chart based on applying a sequential probability ratio test (SPRT) at each sampling point was introduced by Reynolds and Stoumbos (1998) for interval sampling. They compared their chart to the $p$–chart and binomial CUSUM chart. The SPRT chart is flexible due to the use of sequential testing, and it gives mostly superior results over the two competing methods using the steady–state average time to signal (SSATS) metric when considering an increase from $p_0 = .01$ for a wide range of shifts. Reynolds and Stoumbos (1999) evaluated the properties of the $p$–chart, the binomial CUSUM, and Bernoulli CUSUM charts. One–sided charts were used in the analysis, focusing on an increase in $p$, where $p_0 = .01$. Using the zero–state ANOS to evaluate performance, the Bernoulli CUSUM chart was shown to be the uniformly best method for detecting all shifts in $p$. The optimal performance for a binomial CUSUM chart when $n = 1$ (the Bernoulli CUSUM chart) was also shown by Bourke (2001b) using zero–state and steady–state analyses with the random–shift model.

Monitoring with the binomial and Bernoulli CUSUM charts where 100% inspection was not present was also considered by Reynolds and Stoumbos (2000). These CUSUM charts were also compared to a $p$–chart for detecting an increase in $p$. Because the time intervals between samples differ, the SSATS was used as a metric so that fair comparisons as $n$ varied could be made. The random–shift model was used. When $n$ observations were collected at each sampling interval, the Bernoulli CUSUM chart was updated after each item was inspected, while the other two charts required grouping of data into their specified subgroup size before charting a statistic.
One example considered by Reynolds and Stoumbos (2000) had 2500 observations of Bernoulli data. For the first half of the data set $p_0 = .01$. Then a shift to $p_1 = .04$ was sustained for the rest of the data set. Every fifth item was sampled and each of the methods plot control chart statistics when applicable. These results are shown in Figure 4.4, where the Bernoulli CUSUM chart is most responsive in detecting this shift because of its sequential updating.

Figure 4.4: CUSUM charts from Reynolds and Stoumbos (2000) based on a simulation example. Shift from $p_0 = .01$ to $p_1 = .04$ occurs at item 1251.

The CCC, geometric CUSUM, binomial CUSUM, and Bernoulli CUSUM charts were evaluated by Chang and Gan (2001). For the binomial CUSUM chart, two different quantities for $n$ were considered for each analysis. Based on the zero–state ANOS for detecting an increase from $p_0 = 100$ ppm, the geometric CUSUM chart has the best performance except for signaling large shifts, where the CCC chart was considered the best method. For detecting a decrease in the nonconforming rate, the binomial CUSUM
chart based on the largest value of $n$ was considered best, although the performance of the Bernoulli CUSUM chart was comparable. In this case, the CCC and geometric CUSUM charts have poor performance for smaller values of $\gamma$, particularly when $\gamma < 0.5$. Again, this delay in detection of a downward shift in $p$ for the geometric CUSUM chart is why the Bernoulli CUSUM chart should be preferred.

Many comparisons were performed by Chang and Gan (2007) for a $p$–chart with runs rules. The ARL and ANOS criteria were used. The $p$–chart, synthetic, CCC, conditional CCC, Bernoulli CUSUM, and geometric CUSUM charts were analyzed. Both zero–state and steady–state analyses were used, but with the fixed–shift model. The runs rule method worked best against all except the CUSUM methods, but the authors preferred the runs rules methods due to their simplicity. However, many design parameters must be chosen for their runs rule approach. The geometric CUSUM chart had the best overall SSANOS performance because the random–shift model was not considered.

Quesenberry (1995) focused on testing the shift detection from $p_0$ to $p_1$ using the $Q$ chart based on runs rules, EWMA charts, and CUSUM charts. This was measured by calculating the probability that a chart detects the specified change by the time that thirty nonconforming items were found. However, the results varied greatly depending on whether the chart was designed for detecting process improvement or process deterioration. There appeared to be no clear method that should be used for both types of shifts from the results given in his paper.

Sun and Zhang (2000) analyzed the EWMA and CUSUM charts based on the geometric distribution, as well as the two–stage control chart considered by Chan et al. (1997) for detecting an increase in the nonconforming rate. A simulation study based on the zero–state ANOS performance showed that the two–stage control chart was an inferior method.

The zero–state and steady–state properties of EWMA charts have also been examined by Yeh et al. (2008). Zero–state analysis of the geometric EWMA chart and two EWMA charts based on transformations to normality dominated in ANOS performance over the geometric CUSUM chart. For the steady–state case with a fixed–shift model, an EWMA chart based on the geometric distribution was inferior to the geometric CUSUM chart, likely due to the fact that a reflecting barrier was not used. The geometric EWMA chart
was also inferior in many cases to the other two EWMA charts as well.

The EWMA charts based on transformations studied by Spliid (2010) were compared to the Bernoulli CUSUM chart. Using zero–state ARL analysis for detecting an increase in $p$, where $p_0 = 10$ ppm, the Bernoulli CUSUM chart performed well for small shifts, but the fourth–root transformation EWMA chart worked well for larger shifts.

Joner et al. (2008) examined a scan statistic and compared this to the Bernoulli CUSUM chart. Scan statistics are used primarily in health–related monitoring. The Bernoulli scan method is based on a moving window of past data. If the number of nonconforming items in the most recent window of items exceeds a specified value, a signal is given. Considering a large range of shifts, the Bernoulli CUSUM statistic had better performance for detecting an increase in $p$ based on the random–shift SSANOS metric.

Sego et al. (2008) considered three different ‘sets’ methods versus the Bernoulli CUSUM chart. The sets methods are based on runs rules–type procedures based on the successive counts of conforming items between nonconforming items. An implicit head–start feature exists for these sets methods, which leads to very good zero–state performance. However, this advantage was lost when a steady–state analysis was considered. The methods studied were the traditional sets method and two modifications. Of the three methods, the first used a runs rule method, the second a series of two ‘flags’ in order to signal, and the last method assigned scores based on the geometric counts. The Bernoulli CUSUM chart was inferior for zero–state analysis (even when a head–start feature was added), but was uniformly best when the random–shift steady–state analysis was considered. The reader is referred to Sego et al. (2008) for more information on sets methods and their modifications.

4.9.2 Comparisons Without CUSUM Charts

Performance comparisons should include CUSUM methods due to the CUSUM optimality properties. We briefly discuss some performance comparisons, however, which did not include CUSUM charts in this section.
Benneyan (2001a) provided the theoretical framework for the $g$-chart that was introduced by Kaminsky et al. (1992). The focus of this paper was for health-related applications and the results were compared strictly to the $p$-chart in a subsequent paper (Benneyan 2001b). In this latter paper, the zero-state ARL and ANOS performance of these charts were shown. The results showed that the $p$-chart was ineffective for detecting a decrease in $p$, while the $g$-chart was unable to detect increases in $p$ effectively because the computed LCL was negative. Both of these charts are inefficient for monitoring high quality processes.

Kotani et al. (2005) compared the EWMA CCC–$r$ chart to the confirmation sample CCC–$r$ chart developed by Ohta and Kusukawa (2004). The EWMA chart was found to be superior based on zero-state ARL performance for several simulation studies where $r = 2$ and 5.

Liu et al. (2006) considered the use of interval sampling only for the two-sided CCC chart. Typically, fixed sampling intervals (FSIs) are used, but a variable sampling interval (VSI) was proposed. If the previous data point is close to the LCL, the sampling interval is shortened; if it is near the UCL, the interval is lengthened. Many cases were evaluated based on zero-state ATS and ANOS comparisons. The authors concluded that the VSI method resulted in better performance. Chen et al. (2009) also studied this problem, using the zero-state ATS metric. The authors stated that using VSIs gives about a 15% improvement in detection of a shift in $p$ over use of a FSI.

The zero-state and steady-state analyses of the $RL_2$ and $p$-charts were performed by Bourke (2006). The author also considered the random-shift model and interval sampling for a process. Three charts of each type were considered. The $RL_2$ charts had better steady-state ANDO performance in general, except for small shifts in the parameter $p$.

4.10 Discussion

It is surprising how much work has been done on monitoring of Bernoulli data. We have tried to provide a comprehensive review of this topic, which we hope will be useful to both practitioners and researchers. In particular, our major points are the following:
1. Practitioners face a wide variety of choices in methods for monitoring Bernoulli data. We recommend the choice be made based on desired statistical performance, taking into account simplicity and ease in finding the values of the design parameters of the chart. The use of a transformation to approximate normality combined with a standard individuals control chart, the geometric–based CCC chart, and the Bernoulli CUSUM chart are in the order of increasing efficiency and decreasing simplicity.

2. Chart performance comparisons should be based on the ANOS metric, not the ARL. The ARL typically corresponds to the average number of points plotted on a control chart until a signal is given. Thus, the ARL is not appropriate when the number of Bernoulli observations observed between plotted points varies from chart–to–chart and even between plotted points on the same chart.

3. The statistical performance of the various chart modifications based on runs rules should be evaluated based on the use of Markov chains or simulation. Generally, however, the runs rules approaches are harder to design and less efficient than CUSUM methods.

4. Due to well–established theory, it is very difficult to compete with the overall steady–state statistical performance of the Bernoulli CUSUM chart. Results claiming to show the statistical performance of various proposed methods to be significantly better than that of the CUSUM chart are often based on zero–state analyses with the proposed method having a significant headstart feature.

5. The geometric and Bernoulli CUSUM for detecting increases in the proportion can be designed to be mathematically equivalent. For detecting decreases in the proportion, however, the Bernoulli CUSUM is preferable to a geometric CUSUM chart. The geometric CUSUM can only signal after a nonconforming item is observed unless curtailed sampling is used, while the Bernoulli CUSUM chart faces no such restriction.

6. Performance comparisons of methods should be based on the random–shift steady–state model. Use of the zero–state model gives an unfair advantage to methods with headstart features. The fixed–shift model is much less realistic with the assumption that a parameter shift must occur immediately after a nonconforming
item is observed. Use of the steady–state fixed–shift model also gives an unfair advantage to some methods.

7. Aggregating Bernoulli data over time to obtain binomial data or aggregating geometric counts to obtain negative binomial data generally leads to less efficient methods in the high quality area. Thus, we do not recommend the use of the $p$–chart or the CCC–$r$ chart with Bernoulli data.

We see then need for further research in this area. In particular,

1. More work needs to be done to assess the effect of estimating $p_0$ on chart performance. Jensen et al. (2006) provided a general discussion of the effect of estimation error on control chart performance.

2. Little work has been done on the analysis of Phase I Bernoulli data and more work is needed. Pettitt (1980), Worsley (1983), Wallenstein et al. (1994), and Balakrishnan et al. (2002) have analyzed the Bernoulli parameter, but this has not been discussed in the statistical quality control literature. How does one best use historical data to assess the stability of a Bernoulli process over time?

3. The performance of many of the standard methods has not been adequately compared based on the steady–state ANOS metric with the random–shift model. We strongly recommend that the Bernoulli CUSUM chart be included in all comparisons.

4. Besides the work of Ryan et al. (2011), no research has been done on the extensions of the Bernoulli methods to multinomial data with more than two outcomes.

5. More work is needed on the robustness of the various methods with respect to violations of the assumption of independence of the Bernoulli observations and the assumption of a constant in–control value of $p$.

6. The risk–adjusted approach of Steiner et al. (2000) allows the in–control Bernoulli probability to vary from item–to–item. In their application, 30–day mortality rates for surgery patients depended on a patient risk covariate through a logistic regression model. Although risk–adjustment has proven indispensable in health–related monitoring, it has yet to be used in industrial monitoring.
7. Finally, almost all of the work on Bernoulli methods has been based on a sustained step shift in the underlying probability as the out–of–control state. Trends in the probability or transient shifts are also likely to occur in practical applications.
Chapter 5

On the Equivalence of the Bernoulli and Geometric CUSUM Charts

Abstract

The Bernoulli cumulative sum (CUSUM) chart has been shown to have good properties in detecting an increase in the nonconforming rate, $p$, based on the average number of observations to signal (ANOS). The geometric CUSUM chart has also been proposed, in particular for high quality processes. The geometric CUSUM chart has very good properties in zero–state analysis because of a built–in headstart feature. In steady–state performance analyses, often a shift in the process has been assumed to occur immediately after a nonconforming item has been found. In this case, the performance of the geometric CUSUM chart is said to be better than that of the Bernoulli CUSUM chart. We explore the case where a process shift may occur at any time. We show that the steady–state properties of the Bernoulli CUSUM and geometric CUSUM charts under these more realistic circumstances give equivalent results. In fact, the two types of CUSUM charts can be designed to be mathematically equivalent.
5.1 Introduction

5.1.1 Methods

Many control chart methods have been proposed to monitor $p$, the probability of obtaining a nonconforming (defective) item when the observations are assumed to be independent Bernoulli random variables. In this chapter, 100% inspection is considered, where the time order of these observations is preserved. When the nonconforming rate is small, this is known as a high quality process, since few nonconforming items are manufactured. Traditional control charts such as the $p$–chart, or the equivalent $np$–chart, have been shown to be inadequate for detecting a shift in $p$ for this process. For a discussion of these issues, see Xie et al. (1999), Goh and Xie (2003), and Szarka and Woodall (2011b).

Many competing methods have been proposed for quickly detecting an increase in $p$ for high quality processes. It has been shown that a well–designed cumulative sum (CUSUM) chart is better than a competing Shewhart chart for detecting a shift in $p$. For a review of these methods and results, see Szarka and Woodall (2011b).

We consider a shift from an in–control nonconforming rate, $p_0$, to an out–of–control nonconforming rate, $p_1$, where $p_1 > p_0$. Bourke (1991) used the geometric CUSUM chart for monitoring this process. Reynolds and Stoumbos (1999) developed a special case of the binomial CUSUM chart where the subgroup size $n$ is equal to one. This is known as the Bernoulli CUSUM chart. In this chapter we clarify the relationship between these two types of charts and their relative statistical performance.

5.1.2 Performance Metrics

In traditional statistical process control (SPC) applications, the average run length (ARL) is often used as a criterion for determining the best method for detecting a shift to an out–of–control state based on equal in–control performance of competing methods. However, the ARL is defined as the expected number of plotted points until a signal is raised. The ARL for a geometric CUSUM chart only provides a measure of the number of nonconforming items and a point on a control chart is plotted at nonconstant intervals. A more useful metric is the average number of observations to signal (ANOS). The ANOS
gives us more information than the ARL because the total number of items inspected is considered.

### 5.1.3 Chart Equivalence Issues

Reynolds and Stoumbos (1999), Bourke (2001a), and Chang and Gan (2001) have shown that the Bernoulli CUSUM and geometric CUSUM charts can be set up to be equivalent. We note that any geometric count can be expressed as a sequence of Bernoulli counts and vice-versa. When monitoring begins, the geometric CUSUM chart has a built-in headstart feature, however, that gives it an advantage over the Bernoulli CUSUM chart with an initial value of zero. If an appropriate headstart and control limit are used with the Bernoulli CUSUM chart, these charts have equivalent performance in detecting an increase in $p$. Even though the Bernoulli and geometric CUSUM charts can be set up to be equivalent, this is most often not done in the literature. This leads to conflicting conclusions on the relative performance of the two charts.

The effect of a headstart no longer holds when a steady-state analysis is performed. A large number of in-control Bernoulli counts may be given for the Bernoulli CUSUM chart, then a shift for $p$ occurs, and monitoring continues until the chart statistic crosses a decision interval. For the geometric CUSUM chart, a similar model is usually assumed, using geometric counts. However, the implicit assumption for the geometric chart case is that a shift must occur immediately after a nonconforming item occurs. This is known as the fixed-shift model, and also leads to dominant performance over the Bernoulli CUSUM chart for which other assumptions are usually made.

A more realistic approach is to consider a shift occurring after any item is produced for both charts. This is known as the random-shift model. We show equivalence of the two methods under these conditions.

### 5.1.4 Outline of Chapter

In this chapter, we review the Bernoulli CUSUM and geometric CUSUM charts as well as their equivalence properties in Sections 5.2–5.4. Zero-state and steady-state analyses
show differences in performance in Section 5.5. The steady-state analysis of the geometric CUSUM chart is performed for the fixed-shift and random-shift models in Sections 5.6 and 5.7. We investigate the assumptions of each model and give details on the differences in ANOS performance for each case using simulation studies. The geometric distribution is a special case of the negative binomial distribution. Therefore, a CUSUM chart can be implemented based on a sequence of negative binomial counts. A negative binomial CUSUM chart for monitoring a high quality process is introduced in Section 5.8. Concluding remarks are given in Section 5.9.

5.2 The Bernoulli CUSUM Chart

Let $X_k$ be a Bernoulli random variable at time $k$, $k = 1, 2, \ldots$. We assume independence of these random variables. Then the upper-sided Bernoulli CUSUM statistics can be expressed as

$$B_k = \max[0, B_{k-1} + X_k - k_B],$$

(5.1)

where $B_0 = 0$. The formula for $k_B$ is based on the log likelihood ratio for optimally detecting a shift from $p_0$ to $p_1$. This value was expressed by Chang and Gan (2001) as

$$k_B = \frac{\ln \left( \frac{1-p_0}{1-p_1} \right)}{\ln \left( \frac{p_1(1-p_0)}{p_0(1-p_1)} \right)}.$$

To investigate performance using Markov chains, $k_B$ can be conveniently expressed as $1/m$, where $m$ is an integer. A signal is raised for this chart if $B_k \geq h_B$, where $h_B$ is known as the decision interval. A primary advantage of the Bernoulli CUSUM chart is that the CUSUM statistic is updated sequentially after every item is inspected, giving a constant stream of information on the current status of the process.
5.3 The Geometric CUSUM Chart

The parameterization of the geometric distribution has two different forms. Let $Y$ be expressed as the total number of inspections until a nonconforming item is found, with probability mass function (pmf) of

$$f(y) = p(1 - p)^{y-1},$$

where $y = 1, 2, \ldots$. Another parametrization used in the literature is where a geometric random variable is the number of inspections between nonconforming items, $Y - 1$. In this case, the nonconforming item observed is not included in the geometric count. We let $Y_1, Y_2, \ldots$, represent the number of items observed between nonconforming items in the Bernoulli sequence $X_1, X_2, \ldots$. Thus, a geometric count is not obtained until a nonconforming item is found and points are plotted on a control chart at nonconstant time intervals. The form of the upper-sided geometric CUSUM statistic given by Chang and Gan (2001) after the $j^{th}$ nonconforming item ($j = 1, 2, \ldots$) is

$$G_j = \max[0, G_{j-1} - Y_j + k_G],$$

where $G_0 = 0$, $k_G = 1/k_B = m$, and a signal is raised if $G_j \geq h_G$.

The notation of the geometric CUSUM chart by Bourke (1991) was based on the alternative form of the geometric distribution, resulting in different parameters for the CUSUM chart. We denote this new set of parameters by $(k_G^*, h_G^*)$, where $k_G^* = k_G - 1$. The form of this geometric CUSUM statistic is the same as that shown in Equation (5.2), but with the adjusted set of parameters and decision interval.

5.4 Equivalence Results

For detecting an increase in $p$, the Bernoulli CUSUM and geometric CUSUM charts can be set up such that they are equivalent. Reynolds and Stoumbos (1999) stated that if the Bernoulli CUSUM chart is given a headstart of $B_0 = 1$, then it is equivalent to a geometric CUSUM chart where $G_0 = 0$ if $h_G = mh_B + m - 1$. However, the correct
formula here should have been \( h_G = m(h_B - 1) \).

We offer an artificial example to illustrate that their use of \( B_0 = 1 \) leads to a lack of equivalence between the two CUSUM charts. Suppose, for example, \( m = 4 \) and \( h_B = 11/4 \). Thus, \( h_G = 7 \) using \( h_G = m(h_B - 1) \). If a sequence of geometric counts are 2, 1, and 2, the CUSUM statistics from Equations (5.1) and (5.2) are tabulated in columns [1] and [3] of Table 5.1. In both cases, each charting statistic signals exactly at the respective control limit at item 5, illustrating equivalent performance for the two methods. A similar sequence of geometric counts are 5, 2, 1, and 2. The CUSUM statistics for this data set are also shown in columns [1] and [3] of Table 5.2. Because of the larger initial count, the Bernoulli CUSUM statistic resets to zero. When a nonconforming item is found, the statistic increases by a value of \((m - 1)/m\), rather than 1. This explains why the Bernoulli CUSUM statistic is a distance of \(1/m\) from its control limit at item 10, but the geometric CUSUM statistic results in a signal at its control limit exactly. Hence, in this case, the two CUSUM charts are not equivalent.

**Table 5.1:** Bernoulli and Geometric CUSUM Statistics with Headstarts Using First Sequence of Data

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>– 1 3/4</td>
<td>– 0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0 3/4 2/4</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1 6/4 5/4</td>
<td>2 2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1 9/4 8/4</td>
<td>1 5</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0 8/4 7/4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1 11/4 10/4</td>
<td>2 7</td>
<td></td>
</tr>
</tbody>
</table>

Chang and Gan (2001) developed the correct equivalence between the two CUSUM charts. If \( B_0 = (m - 1)/m \) and \( G_0 = 0 \), these charts are equivalent if \( h_G = mh_B - m + 1 \). The proof of this equivalence was shown in the Appendix of Chang and Gan (2001). If designed this way, the Bernoulli and geometric CUSUM charts have equivalent zero–state and steady–state performance. In the previous example, consider \( B_0 = 3/4 \) and \( h_B = 10/4 \). In columns [2] and [3] of Tables 5.1 and 5.2, a signal is raised in both cases at the exact control limits, an example of the equivalence between these two methods.
If we are interested in detecting a decrease in $p (p_1 < p_0)$, the equivalence no longer holds. Because of the sequential updating of the Bernoulli CUSUM statistic, a signal can be raised after a sufficiently long sequence of conforming items is found. However, since a geometric count can only be obtained once a nonconforming item is found, the most recent item produced must be nonconforming in order for the geometric CUSUM chart to signal. If $p_1 << p_0$, the waiting time until a nonconforming item appears could be very long, as shown by Chang and Gan (2001). In order to have equivalence between the two types of charts, sampling for the geometric CUSUM chart would have to be curtailed at an appropriate value.

### 5.5 Zero–State Analysis

In a zero–state sensitivity analysis, we consider a shift in the parameter of interest as soon as monitoring begins or when the chart is at its initial value. As previously discussed, a headstart feature may be used such that the initial CUSUM statistic is equal to some value between zero and the control limit. If no headstart is used with the Bernoulli
CUSUM chart, this gives a large advantage to the geometric CUSUM chart over the Bernoulli CUSUM chart in ANOS performance for detecting an upward shift in $p$ for a Bernoulli process. The results of such a comparison were shown by Chang and Gan (2001, 2007).

5.6 Steady–State Analysis

A headstart feature no longer provides an advantage in detection when a steady–state analysis is used. For the Bernoulli CUSUM chart, we would consider having a large number, say $s$, in–control items before a shift in $p$ occurs. Then monitoring continues until a signal is given. In many CUSUM applications, steady–state results give more realistic results compared to the zero–state analysis because the CUSUM statistic may be closer to the signaling limit when the shift occurs than in the zero–state case. If a signal is given during the in–control runs in simulating steady–state performance, it is customary to either reset the CUSUM statistic to zero, or discard the sample and run a new iteration in the simulation where this does not occur.

The steady–state properties of the geometric CUSUM chart can be determined in two ways. The first is to run a large number, say $v$, in–control geometric runs before a shift occurs. This approach was taken by Chang and Gan (2001, 2007) and Yeh et al. (2008). However, the underlying assumption of this analysis is that the shift occurs immediately after the $v^{th}$ nonconforming item. This is known as the fixed–shift model.

With 100% inspection, it is more reasonable to assume that a shift may occur at any time. With very low nonconforming rates in high quality processes, there may be long periods of time between nonconforming items being produced. It is unnecessarily restrictive to assume that a shift occurs only after a nonconforming item has been found. It is also a restriction that is not used for the Bernoulli CUSUM chart. This model was introduced by Wu and Spedding (1999) and is known as the random–shift model. A visual representation of these models were shown by Bourke (2001a) in Figure 5.1, where the shaded boxes represent nonconforming items.

The steady–state ANOS (SSANOS) results of the fixed and random–shift models for the geometric CUSUM chart may be quite different. This has led to misleading
conclusions in the literature that the geometric CUSUM chart is the better method, when in fact, it is equivalent to an appropriately designed Bernoulli CUSUM chart. For instance, Chang and Gan (2007) stated, “Comparisons of the Shewhart chart with runs rules with geometric CUSUM, Bernoulli CUSUM and synthetic charts showed that the geometric CUSUM chart is the most sensitive in detecting increases in $p$. ” Additionally, Yeh et al. (2008) validated this claim, stating, “Under the steady–state scenarios, Chang and Gan’s geometric counts based CUSUM charts provided the best overall performance across various out–of–control levels.”

![Graph](image)

Figure 5.1: Illustration from Bourke (2001a) on the Fixed–Shift and Random–Shift Models

### 5.7 Simulation Studies

In this section we study the zero–state and steady–state performance comparisons with the Bernoulli CUSUM and geometric CUSUM charts. In all of our simulations, one million runs were used to obtain each ANOS and SSANOS value. In order to be consistent in steady–state analysis with the methods presented, we used the same procedure as the respective authors did for their example when burn–in runs resulted in a signal. In Examples 5.1 and 5.3, the run was discarded. In Example 5.2, the CUSUM statistic was reset to zero.
5.7.1 Example 5.1

The zero–state analysis for the Bernoulli CUSUM chart was considered by Reynolds and Stoumbos (1999). A headstart feature was not used for this chart. The steady–state analysis for the Bernoulli CUSUM chart was performed by Reynolds and Stoumbos (2000). The in–control nonconforming rate was $p_0 = .01$, and the charts’ parameters were designed to optimize detection of a sustained shift to $p_1 = .025$. We ran three separate simulation studies based on the geometric CUSUM chart, the zero–state analysis and the steady–state analysis for the fixed and random–shift models. For steady–state burn–in runs, 50 geometric counts and 5,000 Bernoulli counts were used for the fixed and random–shift models, respectively.

The results of these simulations are shown in columns [1–3] of Table 5.3. The results given by Reynolds and Stoumbos (1999, 2000) are shown in columns [4] and [5], respectively. For the zero–state analysis, it is clear that with the headstart feature, the geometric CUSUM chart dominates the Bernoulli CUSUM chart with no head start. For the steady–state analysis, the fixed–shift model leads to the best performance uniformly, with the lowest SSANOS values across all values of $p_1$. The random–shift model in column [3] shows equivalent performance to the steady–state analysis of the Bernoulli CUSUM chart in column [5] apart from simulation error. We also note that the SSANOS values for the random–shift model are larger than the ANOS values for the zero–state case when $p_1 \geq .025$.

Because the random–shift model is more realistic, the analysis based on this model versus the fixed–shift model should be preferred. The performance of the geometric CUSUM chart under this model fares much worse than under the other two models. This is because of what Bourke (2001a) termed the “crossover run length.” In Figure 5.1(b), the first geometric count after the shift occurs is based on items subject to in–control and out–of–control probabilities, based on the time of the shift. Hence, there can be a delay in detection under the random–shift model because the crossover run length may include items with the in–control probability of being nonconforming and thus be increased in value.
Table 5.3: Comparison of CUSUM Chart Performance to ANOS and SSANOS Results Shown by Reynolds and Stoumbos (1999, 2000)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Geometric CUSUM Chart</td>
<td>Bernoulli CUSUM Chart</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$k_G = 61, h_G = 260$</td>
<td>$k_B = 1/61, h_B = 320/61$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.010</td>
<td>29201.6 29201.6 29201.6</td>
<td>29248.6 29249</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.015</td>
<td>2783.4 2582.2 2752.1</td>
<td>2847.2 2753</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.020</td>
<td>901.3 803.8 894.9</td>
<td>951.7 897</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.025</td>
<td>486.8 424.1 488.8</td>
<td>526.6 489</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.030</td>
<td>325.9 281.5 329.8</td>
<td>359.5 330</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.040</td>
<td>194.1 166.4 198.9</td>
<td>219.2 199</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.050</td>
<td>137.8 117.8 142.1</td>
<td>157.8 143</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.060</td>
<td>106.6 91.2 110.6</td>
<td>123.3 111</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.070</td>
<td>87.0 74.4 90.6</td>
<td>101.2 91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.080</td>
<td>73.3 62.7 76.8</td>
<td>85.8 77</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.090</td>
<td>63.3 54.2 66.5</td>
<td>74.4 67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.100</td>
<td>55.6 47.8 58.6</td>
<td>65.7 59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.150</td>
<td>34.5 29.8 36.8</td>
<td>41.2 37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.200</td>
<td>25.2 21.8 26.9</td>
<td>30.2 27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.300</td>
<td>16.7 14.3 17.7</td>
<td>20.0 18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.500</td>
<td>10.0 8.5 10.5</td>
<td>12.0 11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.750</td>
<td>6.7 5.6 6.9</td>
<td>8.0 7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.000</td>
<td>5.0 4.2 5.2</td>
<td>6.0 6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5.7.2 Example 5.2

Yeh et al. (2008) considered steady–state performance without specifying the type of model used, where $p_0 = .0001$ and the geometric CUSUM chart was optimized to detect a shift to $p_1 = .0002$. Fifteen burn–in geometric counts were used by Yeh et al. (2008) for steady–state analysis. We also simulated the steady–state ANOS values for the fixed–shift model. In the random–shift model case, we used 150,000 Bernoulli counts for burn–in. Table 5.4 shows that the results are quite different for the fixed and random–shift models. The results from columns [2] and [4] are very similar, which is an indication that the fixed–shift model was used by Yeh et al. (2008). Again, the random–shift model in
Table 5.4: Geometric CUSUM Chart ANOS and SSANOS Results Compared to Results Shown by Yeh et al. (2008) (* An Approximate Value)

<table>
<thead>
<tr>
<th>$p$</th>
<th>Geometric CUSUM Chart $k_G = 6931, h_G = 16483$</th>
<th>Simulation Study Results</th>
<th>Yeh et al. (2008) Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Steady–State</td>
<td>Steady–State</td>
</tr>
<tr>
<td>0.00010</td>
<td>299242.1</td>
<td>299242.1</td>
<td>299242.1</td>
</tr>
<tr>
<td>0.00015</td>
<td>79478.3</td>
<td>67470.2</td>
<td>79676.3</td>
</tr>
<tr>
<td>0.00020</td>
<td>37928.1</td>
<td>31319.4</td>
<td>38803.0</td>
</tr>
<tr>
<td>0.00025</td>
<td>23686.7</td>
<td>19279.6</td>
<td>24575.7</td>
</tr>
<tr>
<td>0.00030</td>
<td>16947.4</td>
<td>13662.8</td>
<td>17815.4</td>
</tr>
<tr>
<td>0.00035</td>
<td>13120.9</td>
<td>10534.8</td>
<td>13931.5</td>
</tr>
<tr>
<td>0.00040</td>
<td>10647.0</td>
<td>8562.0</td>
<td>11414.9</td>
</tr>
<tr>
<td>0.00045</td>
<td>8963.1</td>
<td>7181.6</td>
<td>9660.9</td>
</tr>
<tr>
<td>0.00050</td>
<td>7727.5</td>
<td>6205.6</td>
<td>8375.2</td>
</tr>
<tr>
<td>0.00100</td>
<td>3198.3</td>
<td>2599.5</td>
<td>3591.8</td>
</tr>
<tr>
<td>0.00500</td>
<td>598.9</td>
<td>479.2</td>
<td>663.1</td>
</tr>
<tr>
<td>0.10000</td>
<td>30.0</td>
<td>23.7</td>
<td>32.7</td>
</tr>
</tbody>
</table>

column [3] results in much larger SSANOS values than for the other two models.

5.7.3 Example 5.3

Bourke (2008) considered the random–shift model for a number of simulations where the interest was in detecting an increase in $p$. However, the SSANOS results were only shown at the specified value of $p_1$ where the chart was optimized, rather than over a range of values for $p_1$. One case considered was $p_0 = .0020$ and $p_1 = .0060$. The author stated the SSANOS value at $p_1 = .0060$ was 1282.2. We ran steady–state simulations with the shift occurring after 50 geometric and 25,000 Bernoulli counts for the fixed and random–shift models, respectively. In Table 5.5, the results are similar to those of Bourke (2008) based on our simulation study for the random–shift model, with a SSANOS value of 1282.9 at $p_1 = .0060$. The fixed–shift model simulations resulted in a much smaller quantity
Table 5.5: Geometric CUSUM Chart ANOS and SSANOS Results Compared to Results Shown by Bourke (2008)

<table>
<thead>
<tr>
<th>$p$</th>
<th>Simulation Study Results</th>
<th>Geometric CUSUM Chart $k^<em>_G = 301$, $h^</em>_G = 956$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fixed–Shift</td>
<td>Random–Shift</td>
</tr>
<tr>
<td>0.0020</td>
<td>50231.5</td>
<td>50231.5</td>
</tr>
<tr>
<td>0.0025</td>
<td>17608.3</td>
<td>16330.7</td>
</tr>
<tr>
<td>0.0030</td>
<td>8353.2</td>
<td>7546.8</td>
</tr>
<tr>
<td>0.0035</td>
<td>4839.9</td>
<td>4287.5</td>
</tr>
<tr>
<td>0.0040</td>
<td>3223.6</td>
<td>2803.3</td>
</tr>
<tr>
<td>0.0050</td>
<td>1823.2</td>
<td>1555.5</td>
</tr>
<tr>
<td>0.0060</td>
<td>1239.7</td>
<td>1048.2</td>
</tr>
<tr>
<td>0.0080</td>
<td>743.4</td>
<td>622.8</td>
</tr>
<tr>
<td>0.0100</td>
<td>527.4</td>
<td>441.2</td>
</tr>
<tr>
<td>0.0200</td>
<td>213.5</td>
<td>178.9</td>
</tr>
<tr>
<td>0.1000</td>
<td>40.0</td>
<td>32.6</td>
</tr>
<tr>
<td>0.2000</td>
<td>20.0</td>
<td>16.2</td>
</tr>
<tr>
<td>0.5000</td>
<td>8.0</td>
<td>6.4</td>
</tr>
<tr>
<td>0.7500</td>
<td>5.3</td>
<td>4.3</td>
</tr>
<tr>
<td>1.0000</td>
<td>4.0</td>
<td>3.2</td>
</tr>
</tbody>
</table>

of 1048.2 for the SSANOS value. Again, the choice of the model can greatly affect our understanding of the performance.

5.8 Negative Binomial CUSUM

A more general form for the geometric CUSUM chart is the negative binomial CUSUM chart. Let $Z$ be the number of items inspected until $r$ nonconforming items are found. Then $Z$ is a negative binomial random variable with pmf

$$f(z) = \binom{z - 1}{r - 1} p^r (1 - p)^{z-r},$$
where \( z = r, r+1, r+2, \ldots \). When \( r = 1 \), this reduces to the geometric distribution. The form of the negative binomial CUSUM statistic after \( i \) negative binomial counts is

\[
N_i = \max[0, N_{i-1} - W_i];
\]

\[
= \max[0, N_{i-1} - Z_i + k_N];
\]

where \( N_0 = 0 \) and a signal is raised if \( N_i \geq h_N \). The form of \( W_i \) is given using theory from the sequential probability ratio test (SPRT) introduced by Page (1954). The quantity \( W_i \) for a fixed value of \( r \) is expressed as

\[
W_i = \ln \left( \frac{f(z_i|p_1)}{f(z_i|p_0)} \right).
\]

The value of \( W_i \) can be partitioned into two parts as shown in Equation (5.3) so that we may solve for \( k_N \).

\[
W_i = \ln \left( \frac{f(z_i|p_1)}{f(z_i|p_0)} \right)
= \ln \left( \frac{(z_i-1)p_1^r(1-p_1)^{z_i-r}}{(z_i-1)p_0^r(1-p_0)^{z_i-r}} \right)
= r \ln(p_1) + (z_i - r) \ln(1-p_1) - r \ln(p_0) - (z_i - r) \ln(1-p_0)
= z_i \ln \left( \frac{1-p_1}{1-p_0} \right) - r \ln \left( \frac{p_0(1-p_1)}{p_1(1-p_0)} \right)
= \ln \left( \frac{1-p_1}{1-p_0} \right) \left[ z_i - \frac{r \ln \left( \frac{p_0(1-p_1)}{p_1(1-p_0)} \right)}{\ln \left( \frac{1-p_1}{1-p_0} \right)} \right]
= \ln \left( \frac{1-p_1}{1-p_0} \right) [z_i - rG]
\]

The \( \ln((1-p_1)/(1-p_0)) \) term is a constant that is usually disregarded for this exercise. We may conclude that \( k_N = rG \).

For a high quality process, aggregating geometric counts to fit a negative binomial distribution may result in extremely large counts. If we revisit Example 5.1, we can make comparisons between the equivalent Bernoulli and geometric CUSUM charts and negative binomial CUSUM charts where \( r = 2, 5, \) and \( 10 \). The control limits will be
narrower for the negative binomial case in order to obtain the same in–control ANOS values. The number of burn–in runs used for steady–state analysis must also be altered. For the fixed–shift model, 50/r in–control negative binomial runs were used and 5,000 Bernoulli random variables were generated for the random–shift model.

In Table 5.6, the results from Example 5.1 are presented for the zero–state, fixed–shift, and random–shift models respectively for the geometric CUSUM (r = 1), and the three negative binomial CUSUM charts. In the zero–state case for small shifts in p, larger values of r yielded better ANOS results. However, for the r = 10 case, there is a great aggregation of data and resulted in worse performance as p1 increased. The r = 5 case obtained the best overall results across p1. For the fixed–shift model, the geometric CUSUM chart was nearly uniformly best, with the exception of the largest shifts. In the random–shift model, the geometric CUSUM chart had the best SSANOS performance, and the sensitivity of the negative binomial CUSUM chart greatly deteriorated as r increased. The results shown in Table 5.6 indicate that the aggregation of geometric counts should not be implemented for high quality processes.

5.9 Conclusions

In this chapter we have examined the Bernoulli CUSUM and geometric CUSUM charts for detecting an increase in the nonconforming rate of a process, p. Previous studies have clarified that these two charts can be made equivalent if the Bernoulli CUSUM chart is given the appropriate headstart. However, the Bernoulli CUSUM chart has often been examined in sensitivity analyses without this feature. Using the headstart value given by Reynolds and Stoumbos (1999), a resetting of the Bernoulli CUSUM statistic negates the equivalence of the two charts. Chang and Gan (2001) correctly established conditions for equivalence between the two charts.

The advantage of a headstart feature is lost when a steady–state analysis is performed. Hence, in our simulations of the Bernoulli CUSUM and geometric CUSUM chart with a random–shift model, we were able to obtain similar results based on the SSANOS metric when the appropriate control limits were used. Many authors who have studied the geometric CUSUM chart have misleadingly, in our opinion, applied the fixed–shift
Table 5.6: ANOS and SSANOS Values – Negative Binomial CUSUM Chart Results for Example 5.1 from Reynolds and Stoumbos (1999, 2000)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>k = 61</td>
<td>k = 122 k = 305 k = 610</td>
<td>k = 61 k = 122 k = 305 k = 610</td>
</tr>
<tr>
<td></td>
<td>h = 260</td>
<td>h = 234 h = 185 h = 105</td>
<td>h = 260 h = 234 h = 185 h = 105</td>
</tr>
<tr>
<td></td>
<td>r = 1</td>
<td>r = 2 r = 5 r = 10</td>
<td>r = 1 r = 2 r = 5 r = 10</td>
</tr>
<tr>
<td>0.010</td>
<td>29201.6</td>
<td>29051.8 29410.3</td>
<td>29295.8 2752.1</td>
</tr>
<tr>
<td>0.015</td>
<td>2783.4</td>
<td>2769.1 2826.6</td>
<td>2663.7 2582.2</td>
</tr>
<tr>
<td>0.020</td>
<td>901.3</td>
<td>891.6 890.3</td>
<td>858.4 803.8</td>
</tr>
<tr>
<td>0.025</td>
<td>486.8</td>
<td>479.0 475.5</td>
<td>486.9 424.1</td>
</tr>
<tr>
<td>0.030</td>
<td>325.9</td>
<td>320.2 318.1</td>
<td>354.8 281.5</td>
</tr>
<tr>
<td>0.040</td>
<td>194.1</td>
<td>191.0 186.9</td>
<td>250.9 166.4</td>
</tr>
<tr>
<td>0.050</td>
<td>137.8</td>
<td>136.2 127.9</td>
<td>200.0 117.8</td>
</tr>
<tr>
<td>0.060</td>
<td>106.6</td>
<td>106.7 95.6</td>
<td>166.7 91.2</td>
</tr>
<tr>
<td>0.070</td>
<td>87.0</td>
<td>88.3 76.6</td>
<td>142.8 74.4</td>
</tr>
<tr>
<td>0.080</td>
<td>73.3</td>
<td>75.9 64.5</td>
<td>124.9 62.7</td>
</tr>
<tr>
<td>0.090</td>
<td>63.3</td>
<td>66.9 56.3</td>
<td>111.1 54.2</td>
</tr>
<tr>
<td>0.100</td>
<td>55.6</td>
<td>59.9 50.3</td>
<td>100.0 47.8</td>
</tr>
<tr>
<td>0.150</td>
<td>34.5</td>
<td>39.4 33.3</td>
<td>66.7 29.8</td>
</tr>
<tr>
<td>0.200</td>
<td>25.2</td>
<td>28.8 25.0</td>
<td>50.0 21.8</td>
</tr>
<tr>
<td>0.300</td>
<td>16.7</td>
<td>17.7 16.7</td>
<td>33.3 14.3</td>
</tr>
<tr>
<td>0.500</td>
<td>10.0</td>
<td>8.7 10.0</td>
<td>20.0 8.5</td>
</tr>
<tr>
<td>0.750</td>
<td>6.7</td>
<td>5.3 6.7</td>
<td>13.3 5.6</td>
</tr>
<tr>
<td>1.000</td>
<td>5.0</td>
<td>4.0 5.0</td>
<td>10.0 4.2</td>
</tr>
</tbody>
</table>
model, where a shift in $p$ occurs in the process immediately after a nonconforming item is found. Chang and Gan (2007) and Yeh et al. (2008) have stated that this case has led to the geometric CUSUM chart having the best overall performance for detecting an increase in $p$. However, we believe that in the case of 100% inspection, it is much more reasonable to assume that a shift may occur at any time. A Bernoulli CUSUM chart, however, could be designed to be equivalent to a geometric CUSUM chart under the fixed–shift model as well.

A negative binomial CUSUM chart was introduced as a possible competing method for detecting an increase in $p$. A negative binomial count is obtained as a sequence of $r$ geometric random variables. With values very small values of $p$ in high quality processes, there can be a delay before a negative binomial statistic is available. These long waiting times yielded poor SSANOS performance for large $r$ when using the random–shift model. We do not recommend this charting technique for high quality Bernoulli data.

Since the two–sided Bernoulli CUSUM chart is also able to more quickly detect a decrease in $p$, we believe it is more appealing to use than the geometric CUSUM chart in applications involving Bernoulli data.
Chapter 6

A Note on the Unit and Group–Runs Chart for Bernoulli Data

Abstract

In this chapter we examine the unit and group–runs (UGR) chart introduced by Gadre and Rattihalli (2005) for monitoring an increase in the nonconforming rate of a process, $p$, for Bernoulli data. Previous work on monitoring this type of data stream has shown that the Bernoulli cumulative sum (CUSUM) chart is the overall best method for detecting a change in the parameter based on the average number of observations to signal (ANOS) metric. For a shift from an in–control nonconforming rate, $p_0$, to an out–of–control nonconforming rate, $p_1$, it was shown by Gadre and Rattihalli (2005) that the UGR chart overwhelmingly dominates the Bernoulli CUSUM chart in ANOS performance. However, we believe that this result is misleading because the in–control ANOS values for the UGR chart are inflated by the presence of extreme outliers. Once the out–of–control case is considered, these outliers no longer occur, and the out–of–control ANOS can become quite small. We provide a review of the UGR chart and demonstrate its unusual in–control performance using a simulation study.
6.1 Introduction

In statistical process control (SPC) it may be of interest to monitor the nonconforming rate of a process, $p$. Often items are analyzed using 100% inspection, and it is assumed that observations are Bernoulli random variables, where the in–control nonconforming rate is constant and no autocorrelation exists. We refer to this type of data stream as Bernoulli data for the remainder of this chapter. Process deterioration is often of importance to detect, where a sustained change from an in–control nonconforming rate, $p_0$, to a specified out–of–control nonconforming rate of $p_1$ ($p_1 > p_0$) should be detected quickly. Many control chart procedures have been proposed for this problem. For a review of methods for this type of process, see Szarka and Woodall (2011b).

In this chapter we focus on the unit and group–runs (UGR) chart introduced by Gadre and Rattihalli (2005). This method is based on a combination of two types of sampling schemes for determining if an upward shift in $p$ has occurred for Bernoulli data. The first type of sampling used is sequential unit sampling until a nonconforming item is obtained. The second type of sampling uses a pooling of items into groups of size $n$, where the number of nonconforming items in each group is recorded.

The Bernoulli cumulative sum (CUSUM) chart introduced by Reynolds and Stoumbos (1999) has been shown in the literature to be the best method for detecting an increase in $p$ for Bernoulli data against many competing methods (Reynolds and Stoumbos (2000), Joner et al. (2008), Sego et al. (2008)). A CUSUM chart has optimality properties, as shown by Moustakides (1986) that give this class of charts excellent statistical performance. If control charts have roughly equivalent performance at $p_0$, the CUSUM chart can be designed to have the best performance at a specified value of $p_1$. The proper metric used for evaluating the performance of a control chart using Bernoulli data is the average number of observations to signal (ANOS). The ANOS considers all items that have been inspected from the beginning of monitoring until the chart signals.

Gadre and Rattihalli (2005) showed that the UGR chart dominates the Bernoulli CUSUM chart when detecting an increase in $p$ when a zero–state analysis is considered. After investigating this method, we have determined that the in–control ANOS (ANOS$_0$) values of the UGR chart are inflated by the presence of very extreme outliers. This leads the user to a misguided notion of the true in–control behavior of the UGR chart. We
believe the well-known Bernoulli CUSUM chart is a better option for monitoring this type of process.

In Section 6.2, the design of the UGR chart is explained. The performance of the UGR chart versus the Bernoulli CUSUM chart for an upward shift in $p$ is given in Section 6.3. In Section 6.4, a simulation study is conducted to show the variability in the UGR chart’s in-control performance. A more appropriate Markov chain representation of the UGR chart is presented in Section 6.5. A steady-state analysis of the UGR chart is conducted in Section 6.6. Concluding remarks and recommendations are detailed in Section 6.7.

6.2 Design of the UGR Chart

The implementation of the UGR chart begins monitoring by inspecting items individually as units. Let $X$ be defined as the number of items inspected until a nonconforming item is found. Then $X$ is a geometric random variable with parameter $p$. A small value of $X$ implies that $p$ may be large. Therefore, if $X < L_1$, where $L_1$ is a charting limit, a signal is raised.

If $X \geq L_1$, then a signal is not raised and the monitoring method is switched to group sampling. Items are now aggregated into groups of size $n$, where $n > 1$. In each group, the number of nonconforming items found, $d$, is observed. Monitoring continues until a group contains more than $c$ nonconforming items. Let $Y$ be the number of groups examined until a group has more than $c$ nonconforming items. If $Y < L_2$, where $L_2$ is another charting limit, sampling is switched back to unit monitoring. If $Y \geq L_2$, $Y$ is reset to zero and group sampling is resumed.

There are several design constants that must be specified for the UGR chart. Gadre and Rattihalli (2005) specified that the ANOS$_0$ value should be at least as large as a specified constant $\tau$. The authors provided a methodology when using given values of $p_0$, $p_1$, and $\tau$, for choosing the values of $L_1$, $L_2$, $n$, and $c$ in order to minimize the out-of-control ANOS at $p_1$ (ANOS$_1$).

Exact expressions of the ANOS values of the UGR chart were provided by Gadre and Rattihalli (2005). Let $P$ equal the probability that a group of size $n$ has greater than $c$
nonconforming items. This value can be computed using the binomial distribution:

\[ P = Pr(d > c) = \sum_{i=c+1}^{n} \binom{n}{i} p^i q^{n-i}, \]  

(6.1)

where \( q = 1 - p \). We also define \( Q = 1 - P \). The ANOS values were expressed as

\[ \text{ANOS} = \left( \frac{1}{1 - q^{L_1-1}} \right) \left( \frac{nq^{L_1-1}}{P(1 - Q^{L_2-1}) + \frac{1}{p}} \right). \]  

(6.2)

### 6.3 UGR Chart Performance

Many design constant sets for the UGR method were given by Gadre and Rattihalli (2005) to use for comparisons. We focus on Example 4 in their paper, where the UGR and Bernoulli CUSUM charts were analyzed for two cases. In Table 6.1, we provide the constants for the two cases considered that were compared to the results given by Reynolds and Stoumbos (1999).

**Table 6.1:** Charting Constants for the UGR Chart

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Case 6.1</th>
<th>Case 6.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_0 )</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>( p_1 )</td>
<td>0.04</td>
<td>0.025</td>
</tr>
<tr>
<td>( \tau )</td>
<td>29100</td>
<td>29250</td>
</tr>
<tr>
<td>( L_1 )</td>
<td>842</td>
<td>1674</td>
</tr>
<tr>
<td>( L_2 )</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>( n )</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>( c )</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

It is much easier for the UGR chart to raise a signal if the initial unit count of \( X \) is less than \( L_1 \). Once monitoring begins in the group–runs phase, it can become difficult, depending on the design chosen, to return to unit monitoring, where a signal may only occur. Since \( X \) is a geometric random variable, we can find the probability that a signal is given when the first nonconforming item is found by using the cumulative distribution.
function (cdf) of this distribution. The cdf of $X$ is given as

$$F(x) = 1 - (1 - p)^x,$$

where $x = 1, 2, \ldots$. In Case 6.1 for the in–control nonconforming rate, $P(X < 842) = F(841) = .9997865979$. The probability that group sampling is used for Case 6.1 is $1 - F(841) = 2.134 \times 10^{-4}$. The value of $L_1$ is much larger in Case 6.2 with the same in–control nonconforming rate of $p_0 = .01$. We calculate $F(1673) = .9999999501$ and group sampling is used with probability $4.985 \times 10^{-8}$. The group–runs approach described by the authors is used very rarely for the in–control case and it is very likely that a signal will occur once the first nonconforming item is inspected.

In the rare instance that this monitoring scheme leads to the group–runs approach, it is extremely difficult given the chart’s design to return to unit monitoring when the process is in–control. This is more restrictive for Case 6.2, where it is required that a group of size $n = 3$ must contain more than $c = 2$ nonconforming items and $Y < L_2 = 6$. In Case 6.2, all three items must be nonconforming in a group to meet this requirement; this will occur with probability $.000001$ when $p = .01$ using Equation (6.1). With a small value of $L_2 = 6$, it is likely that $Y \geq L_2$. Therefore, $Y$ is reset to zero, and group–runs monitoring continues. It is likely that an extremely long period of time will pass before $Y < L_2$, so the in–control ANOS will be inflated because of this strict condition.

The zero–state ANOS results for the UGR and Bernoulli CUSUM charts are displayed in Table 6.2, where the UGR chart’s results were given based on a simulation study conducted by Gadre and Rattihalli (2005) and the exact results from Equation (6.2). We note glaring differences in the ANOS values in columns [2] and [3], where the exact ANOS value was much larger than the simulated case for the UGR chart for $p = .01$ and $.015$. For all values of $p > .01$ the UGR chart was dominant in ANOS performance over the Bernoulli CUSUM chart. This is most easily apparent for small, upward shifts in $p$. These results implied that the UGR chart provided an enormous improvement over the Bernoulli CUSUM chart. Since the Bernoulli CUSUM chart has optimality properties, it is important to delve into this performance comparison more deeply.
Table 6.2: ANOS Values for the Bernoulli CUSUM and the UGR Chart

<table>
<thead>
<tr>
<th></th>
<th>Case 6.1</th>
<th></th>
<th></th>
<th>Case 6.2</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bernoulli</td>
<td>UGR</td>
<td>UGR</td>
<td>Bernoulli</td>
<td>UGR</td>
<td>UGR</td>
</tr>
<tr>
<td></td>
<td>CUSUM</td>
<td>Gadre</td>
<td>Exact</td>
<td>CUSUM</td>
<td>Gadre</td>
<td>Exact</td>
</tr>
<tr>
<td>0.010</td>
<td>29050.8</td>
<td>29172.3</td>
<td>42789.5</td>
<td>29248.6</td>
<td>30819.7</td>
<td>30011.1</td>
</tr>
<tr>
<td>0.015</td>
<td>3875.3</td>
<td>68.1</td>
<td>185.9</td>
<td>2847.2</td>
<td>72.0</td>
<td>67.2</td>
</tr>
<tr>
<td>0.020</td>
<td>1201.2</td>
<td>50.0</td>
<td>50.5</td>
<td>951.7</td>
<td>50.0</td>
<td>50.0</td>
</tr>
<tr>
<td>0.025</td>
<td>587.4</td>
<td>40.0</td>
<td>40.0</td>
<td>526.6</td>
<td>40.0</td>
<td>40.0</td>
</tr>
<tr>
<td>0.030</td>
<td>366.6</td>
<td>33.3</td>
<td>33.3</td>
<td>359.5</td>
<td>33.3</td>
<td>33.3</td>
</tr>
<tr>
<td>0.040</td>
<td>202.6</td>
<td>25.0</td>
<td>25.0</td>
<td>219.2</td>
<td>25.0</td>
<td>25.0</td>
</tr>
<tr>
<td>0.050</td>
<td>139.0</td>
<td>20.0</td>
<td>20.0</td>
<td>157.8</td>
<td>20.0</td>
<td>20.0</td>
</tr>
<tr>
<td>0.060</td>
<td>105.8</td>
<td>16.7</td>
<td>16.7</td>
<td>123.3</td>
<td>16.7</td>
<td>16.7</td>
</tr>
<tr>
<td>0.070</td>
<td>85.4</td>
<td>14.3</td>
<td>14.3</td>
<td>101.2</td>
<td>14.3</td>
<td>14.3</td>
</tr>
<tr>
<td>0.080</td>
<td>71.6</td>
<td>12.5</td>
<td>12.5</td>
<td>85.8</td>
<td>12.5</td>
<td>12.5</td>
</tr>
<tr>
<td>0.090</td>
<td>61.6</td>
<td>11.1</td>
<td>11.1</td>
<td>74.4</td>
<td>11.1</td>
<td>11.1</td>
</tr>
<tr>
<td>0.100</td>
<td>54.2</td>
<td>10.0</td>
<td>10.0</td>
<td>65.7</td>
<td>10.0</td>
<td>10.0</td>
</tr>
<tr>
<td>0.150</td>
<td>34.0</td>
<td>6.7</td>
<td>6.7</td>
<td>41.2</td>
<td>6.7</td>
<td>6.7</td>
</tr>
<tr>
<td>0.200</td>
<td>25.1</td>
<td>5.0</td>
<td>5.0</td>
<td>30.2</td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td>0.300</td>
<td>16.7</td>
<td>3.3</td>
<td>3.3</td>
<td>20.0</td>
<td>3.3</td>
<td>3.3</td>
</tr>
<tr>
<td>0.500</td>
<td>10.0</td>
<td>2.0</td>
<td>2.0</td>
<td>12.0</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td>0.750</td>
<td>6.7</td>
<td>1.3</td>
<td>1.3</td>
<td>8.0</td>
<td>1.3</td>
<td>1.3</td>
</tr>
<tr>
<td>1.000</td>
<td>5.0</td>
<td>1.0</td>
<td>1.0</td>
<td>6.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

6.4 Simulation Study

In order to investigate the ANOS performance, a large simulation study for the in–control case was performed for the UGR chart using the Case 6.1 design from Table 6.1 with a total of one million runs simulated. A much larger study would be needed for Case 6.2 because of the extremely low probability that an iteration of the simulation would enter the group–runs phase.

In the Case 6.1 simulation study, we estimated the ANOS₀ to be 46,404.6, much closer to the exact value of 42,789.5 shown in column [3] of Table 6.2. We believe our estimated ANOS₀ value was slightly larger than the exact value due to more iterations of our simulation reaching the group–runs inspection phase than expected.
There is a large difference in the number of observations to signal between iterations that did and did not require group sampling. The largest count that can be observed without group sampling is $L_1 - 1 = 841$. However, our next largest count greater than 841 was 800,561. The largest number of observations to signal observed based on our simulation was 1,140,377,617. In total, there were 138 counts greater than 100 million, and 3 greater than 1 billion. It is clear that these large outliers inflated the ANOS₀ value. The outliers will be more extreme for Case 6.2, as we project that the number of observations to signal when the group phase is entered will be larger than 600 billion in order to obtain the desired ANOS₀ value. These ANOS₀ values are misleading to a user who may try to implement this approach. Despite the ANOS₀ value equaling 46,404.6, the median number of observations to signal for our in–control simulation study was only 69. With this type of in–control ANOS performance, the ANOS metric is not useful.

When there is an upward shift in $p$, it becomes even more difficult for the UGR chart to enter group sampling. Therefore, nearly always in these cases, a signal will be given based on the initial geometric random variable $X$. For all values of $p > .015$ in Table 6.2, the ANOS values for the UGR chart are equal to $1/p$, implying a signal in the out–of–control case was always raised once the first nonconforming item is found. If an increase in the probability of a defective item occurred after the group sampling had begun, however, then detection of this shift would be much slower.

### 6.5 Markov Chain Representation

The UGR chart can be expressed using a Markov chain containing transition probabilities. Gadre and Rattihalli (2005) provided two representations of the UGR chart using a Markov chain, but it is incomplete because it does not provide a summary of all possible states in the process. Their Markov chains can be used to compute the zero–state ANOS values. Our representation is helpful when calculating ANOS values with headstart features and steady–state calculations because each state is clearly defined. The ANOS is represented as a column vector, $A$ that can be computed as

$$A = (I - R)^{-1}1,$$  

(6.3)
where \( I \) is an identity matrix and \( \mathbf{1} \) is a column vector of ones. We note that \( R \) is a subset of the transition matrix \( M \), where the last row and column are eliminated from this matrix. This last row and column correspond to the absorbing state when a signal has been alarmed for this chart. The elements of the ANOS vector correspond to the ANOS value for the given initial state of the process. The first element of the ANOS vector is commonly used when no headstart feature is utilized. The transition matrix, \( M \), for Case 6.1 can be expressed using a total of 849 states. Here are the definitions for each state \( j \).

- States 0–840: Current unit count \( X = j \).
- State 841: Current unit count \( X \geq 841 \).
- State 842: Group–runs inspection, subgroup count \( Y = 1 \), item one.
- State 843: Group–runs inspection, subgroup count \( Y = 1 \), item two, where the first item in the subgroup was nonconforming.
- State 844: Group–runs inspection, subgroup count \( Y = 1 \), item two, where the first item in the subgroup was conforming.
- State 845: Group–runs inspection, subgroup count \( Y > 1 \), item one.
- State 846: Group–runs inspection, subgroup count \( Y > 1 \), item two, where the first item in the subgroup was nonconforming.
- State 847: Group–runs inspection, subgroup count \( Y > 1 \), item two, where the first item in the subgroup was conforming.
- State 848: Absorbing state.

The dimension of \( R \) is \( 848 \times 848 \), eliminating the last row and column corresponding to the absorbing state. The transition matrix \( M \) is given below. For a Bernoulli process with 100% inspection, the probability of transitioning from state \( j \) to state \( j' \) will equal zero, \( p \), or \( 1 - p \) for all states, with the exception of the absorbing state, which always remains in its current state.
6.6 Steady–State Analysis

The implicit assumption of the UGR chart is that a shift in $p$ is present once monitoring begins. Hence, all previous analysis was based on the zero–state model. A steady–state analysis considers a process to be running without a signal for a period of time before a shift is present. Since the UGR chart may signal as soon as the first nonconforming item is found, the length of the burn–in period, $s$, may greatly affect the performance of this chart. For burn–in lengths less than $L_1$, the steady–state analysis is equivalent to a headstart feature since all items inspected must be conforming. The SSANOS values are calculated as the $(s+1)^{th}$ element of $A$ from Equation (6.3). An analysis of the SSANOS values for Case 6.1 is shown in Table 6.3. When $s = 0$, this was equivalent to the zero–state ANOS and were calculated the same as column [3] of Table 6.2 from Equation (6.2). As $s$ increased, the SSANOS values increased dramatically for smaller values of $p$ since monitoring was more likely to enter the group–runs phase. This is a highly undesirable feature for this control chart procedure since in practice many processes are monitored in a steady–state context.
### Table 6.3: Exact SSANOS Values for the UGR Chart – Case 6.1

<table>
<thead>
<tr>
<th>$p$</th>
<th>0</th>
<th>5</th>
<th>10</th>
<th>50</th>
<th>100</th>
<th>300</th>
<th>500</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.010</td>
<td>42789.5</td>
<td>44989.6</td>
<td>47303.0</td>
<td>70660.5</td>
<td>116727.8</td>
<td>870588.8</td>
<td>6497270.4</td>
</tr>
<tr>
<td>0.015</td>
<td>185.9</td>
<td>195.3</td>
<td>205.4</td>
<td>320.6</td>
<td>607.3</td>
<td>11175.5</td>
<td>228322.4</td>
</tr>
<tr>
<td>0.020</td>
<td>50.5</td>
<td>50.6</td>
<td>50.6</td>
<td>51.4</td>
<td>53.9</td>
<td>274.0</td>
<td>12785.1</td>
</tr>
<tr>
<td>0.025</td>
<td>40.0</td>
<td>40.0</td>
<td>40.0</td>
<td>40.0</td>
<td>40.0</td>
<td>45.8</td>
<td>951.7</td>
</tr>
<tr>
<td>0.030</td>
<td>33.3</td>
<td>33.3</td>
<td>33.3</td>
<td>33.3</td>
<td>33.3</td>
<td>33.5</td>
<td>109.5</td>
</tr>
<tr>
<td>0.040</td>
<td>25.0</td>
<td>25.0</td>
<td>25.0</td>
<td>25.0</td>
<td>25.0</td>
<td>25.0</td>
<td>25.7</td>
</tr>
<tr>
<td>0.050</td>
<td>20.0</td>
<td>20.0</td>
<td>20.0</td>
<td>20.0</td>
<td>20.0</td>
<td>20.0</td>
<td>20.0</td>
</tr>
<tr>
<td>0.060</td>
<td>16.7</td>
<td>16.7</td>
<td>16.7</td>
<td>16.7</td>
<td>16.7</td>
<td>16.7</td>
<td>16.7</td>
</tr>
<tr>
<td>0.070</td>
<td>14.3</td>
<td>14.3</td>
<td>14.3</td>
<td>14.3</td>
<td>14.3</td>
<td>14.3</td>
<td>14.3</td>
</tr>
<tr>
<td>0.080</td>
<td>12.5</td>
<td>12.5</td>
<td>12.5</td>
<td>12.5</td>
<td>12.5</td>
<td>12.5</td>
<td>12.5</td>
</tr>
<tr>
<td>0.090</td>
<td>11.1</td>
<td>11.1</td>
<td>11.1</td>
<td>11.1</td>
<td>11.1</td>
<td>11.1</td>
<td>11.1</td>
</tr>
<tr>
<td>0.100</td>
<td>10.0</td>
<td>10.0</td>
<td>10.0</td>
<td>10.0</td>
<td>10.0</td>
<td>10.0</td>
<td>10.0</td>
</tr>
<tr>
<td>0.150</td>
<td>6.7</td>
<td>6.7</td>
<td>6.7</td>
<td>6.7</td>
<td>6.7</td>
<td>6.7</td>
<td>6.7</td>
</tr>
<tr>
<td>0.200</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td>0.300</td>
<td>3.3</td>
<td>3.3</td>
<td>3.3</td>
<td>3.3</td>
<td>3.3</td>
<td>3.3</td>
<td>3.3</td>
</tr>
<tr>
<td>0.500</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td>0.750</td>
<td>1.3</td>
<td>1.3</td>
<td>1.3</td>
<td>1.3</td>
<td>1.3</td>
<td>1.3</td>
<td>1.3</td>
</tr>
<tr>
<td>1.000</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

### 6.7 Conclusions

The UGR chart has been shown by Gadre and Rattihalli (2005) to have overwhelm-
ingly dominant zero–state ANOS performance over competing methods when inspecting
Bernoulli data. However, the results from the use of this chart were misleading due to
its very odd in–control performance. When observing the in–control properties of this
chart, there were extreme differences between the number of observations to signal based
on only the unit sampling method, and those that involved the group–runs sampling
method. For many cases, only one nonconforming item was necessary in order to signal.
However, for other cases, a large number of nonconforming items may be present before
signaling.

Once inspection reached the group–runs phase when the process was in–control, it
was extremely difficult to meet the necessary criteria so that a switch could be made to return to unit inspection. This led to prolonged monitoring times before a signal could be given. Based on our simulation study, this can lead to the number of items observed to signal to be larger than one billion items despite a specified average near 30,000. These outliers artificially inflated the $\text{ANOS}_0$ values. Group sampling rarely occurred because there was a very small probability that the criterion to use this sampling method was met based on the parameters chosen for the UGR chart.

For the out–of–control case, the issues with large outliers no longer existed. Since $p_1 > p_0$, it was even less likely that a switch from unit to group–runs inspection would occur because we expected our initial defective item to appear more quickly. Hence, nearly all numbers of observations to signal were based on the initial geometric random variable. Once the first nonconforming item appeared, a signal was given.

The unit and group–runs method should not be used due to its complexity and unreliable in–control performance. We believe that the Bernoulli CUSUM chart should be used due to its proven strength against other methods and familiarity of its implementation based on zero–state and steady–state analysis.
Chapter 7

Conclusions and Future Work

Many surveillance methods have been evaluated for monitoring various types of data streams for healthcare and industrial processes. In the first phase of research in Chapters 2 and 3, we focused on how to effectively monitor daily syndromic counts at a health facility to detect outbreaks. The CDC’s EARS program was established in order to quickly detect outbreaks for these types of data streams. The use of the W2c method has provided a center–scaled statistic based on a moving baseline of seven days to monitor the current state of a process. The effectiveness of detecting simulated outbreaks may be improved upon. A modification of an adaptive threshold method proposed by Lambert and Liu (2006) was shown to be a suitable method for syndromic surveillance. Using the same baseline data as the W2c method, parameters of a discrete distribution were estimated, and an upper–tailed p–value was computed, then converted to a standard normal Z–score. While information was used from the baseline information in the calculation for these monitoring statistics for the current day, they were not true cumulant statistics. An EWMA method was implemented for each monitoring method.

For the surveillance of iid Poisson and negative binomial data in Chapter 2, the adaptive threshold method, in general, was the most effective method for monitoring daily counts, particularly in the Poisson case. However, issues arose when using the adaptive threshold method for data that is greatly overdispersed, where the variance is much larger than the mean. In these cases, it was difficult to estimate the parameters of a negative binomial distribution with little baseline data, such as when \( n = 7 \) days in
the baseline. A longer baseline window \((n = 28)\) yielded better signaling rates in this case. The EWMA method vastly improved the sensitivity of these monitoring schemes, where the adaptive threshold method had the best outbreak detection capabilities for both distributions.

Numerous seasonal patterns were generated for Poisson and negative binomial data to simulate syndromic counts in Chapter 3 because this type of data stream is common in many public health applications. The effectiveness of the W2c and adaptive threshold methods for several sensitivity analyses were quite different from that of the iid case. With more seasonality present in the model, there was more variability in the data sets. For the Poisson cases with longer baseline periods, the adaptive threshold method RI threshold functions varied considerably for cases with the greatest amount of seasonality.

The sensitivity analyses for data with seasonal patterns provided interesting results. The adaptive threshold method resulted in the highest overall detection rates for most cases considered. However, the EWMA method worked best for data streams with mild seasonal patterns, while the Shewhart method often gave the best results for more complicated seasonal patterns. The smallest baseline window size \((n = 7)\) frequently yielded the highest signaling rates, a sharp contrast to the iid case. As \(n\) increased, the signaling rates were much smaller in some cases. A study of these surveillance methods showed that the signaling rates were very different depending on the time of year the outbreak occurred. This was an undesirable feature for an existing monitoring technique such as the W2c method, which indicated that it is susceptible to missing outbreaks during certain time periods.

While syndromic data may follow complicated models with seasonal patterns, simple models may be used to model a small baseline of data. Fricker et al. (2008) used an adaptive regression model to predict daily counts using baseline data. These models included linear and day–of–the–week effects. A negative binomial regression model may be more appropriate in biosurveillance because the assumption of constant variance used in standard regression techniques may be an issue when monitoring seasonal patterned data. A fitted model may be given for day \(t\) and the appropriate p–value can be calculated from the mean function and estimated overdispersion parameter of the negative binomial distribution. This may provide a more stable process across the year for signaling outbreaks since a time component is considered in the underlying model.
Research in biosurveillance would improve if real health data could be used for analysis, but this may be unlikely in many instances. Joint work with healthcare professionals would help test monitoring methods so that they could be implemented in practice. The EARS method also consists of the W2rate (W2r) method, where the proportion of daily syndromic counts relative to the total counts at a health facility are recorded. The W2r method was studied extensively by Gan (2010), but the effects of seasonality on this method have not been evaluated in the literature. Monitoring with both the W2c and W2r methods together versus competing methods may also be considered. Real–time monitoring in biosurveillance also requires analyzing numerous syndromes simultaneously. Multivariate extensions in this area are needed. Parametric models should be evaluated using overdispersed distributions such as the negative binomial, which are more likely to occur in public health data.

The literature on high quality processes for binary attributes data is extensive. The traditional Shewhart $p$–chart does not work well for small values of $p$ as the subgroup size $n$ is typically not large enough to yield proper control limits and has numerous performance issues for detecting a sustained change in $p$. A well–designed CUSUM chart is the most effective way for monitoring this type of data stream. While the Bernoulli and geometric CUSUM charts may be set up such that they are equivalent, this is often not considered in the literature. The geometric distribution has a naturally built–in headstart feature, based on the assumption that a nonconforming item was detected before monitoring began. With a proper headstart for the Bernoulli CUSUM chart, these charts are mathematically equivalent.

Issues with the equivalence of the Bernoulli and geometric CUSUM charts was generated from the type of steady–state analysis used to evaluate ANOS performance. A fixed–shift model for the geometric CUSUM chart has used a burn–in of geometric runs so that the statistics were fully updated and charted once a shift in $p$ occurred. However, with this model, the implicit assumption was that a shift had occurred immediately after a nonconforming item was inspected, and the SSANOS results were better for the geometric CUSUM chart in this case. If $p$ is very small, this model is unnecessarily restrictive and is extremely unlikely to happen. The random–shift model has accounted for the case where a shift may transpire at any time, and has led to an equivalence between the Bernoulli and geometric CUSUM charts.
The Bernoulli CUSUM chart is advantageous because of its ability to detect a decrease in $p$ as well in a timely fashion. The updating of a geometric random variable requires a nonconforming item to be found so that a control chart statistic can be calculated. But for cases where $p_1 << p_0$, this may result in a large delay. Hence, the Bernoulli CUSUM chart should be the preferred monitoring technique in this application.

Competing methods typically cannot match the performance of the Bernoulli CUSUM chart because of the optimality properties associated with CUSUM charts. The UGR chart developed by Gadre and Rattihalli (2005) had dominating zero–state ANOS performance over the Bernoulli CUSUM chart. The ANOS metric was misleading for this application because of the influence of several extreme outliers on the in–control ANOS value. The outliers inflated the in–control ANOS value and did not properly illustrate the true in–control distribution of the number of observations to signal in general. In the out–of–control state, these outliers were no longer present, thus a very small ANOS value was computed. A simple steady–state analysis showed the deterioration of this monitoring method.

If $p_0$ is extremely small, it is important to monitor a continuous random variable if possible since little information is extracted from a large sequence of zeros. The value of $p_0$ may also not be known in practice. More Phase I analysis for estimating the nonconforming rate and assumption of iid random variables is needed. Information on an observed nonconforming item being monitored may contain additional information with the number of defectives on an item that is classified as nonconforming. This is commonly referred to as zero–inflated data, where the additional count information can be used to supplement the Bernoulli random variable. It may be beneficial to monitor a zero–inflated process using multiple CUSUM control charts, where the proportion of nonconforming items and the number of defects on nonconforming items can be evaluated. See He et al. (2011) for an application with a zero–inflated Poisson random variable. An extension to the negative binomial distribution may be appropriate, where an additional control chart would be needed to monitor the two parameters of this distribution. This method would also be beneficial in healthcare applications where zero frequencies of syndromic counts are prevalent in a health facility, but an outbreak may be best modeled by an overdispersed distribution.
Bibliography


