ANALYTICAL AND NUMERICAL METHODS FOR DETECTION OF A CHANGE IN DISTRIBUTIONS

SAOWANIT SUKPARUNGSEE

A thesis submitted for the degree of

Doctor of Philosophy

in Mathematical Sciences

Department of Mathematical Sciences, Faculty of Science

University of Technology, Sydney

December 2008

Principal Supervisor: Professor Alexander Novikov

UNIVERSITY OF TECHNOLOGY, SYDNEY

CERTIFICATE OF AUTHORSHIP/ORIGINALITY

Date: **Dec 2008**

Author: **Saowanit Sukparungsee**

Title: **Analytical and Numerical Methods for Detection of a Change in Distributions**

Department: **Mathematical Sciences**

Degree: **Ph.D.** Convocation: **Mar** Year: **2005**

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Production Note:
Signature removed prior to publication.

Signature of Candidate

Dedication

To Dad & Mom

and Family

Acknowledgments

I would like first to express my deepest gratitude to my advisor, Professor Dr. Alexander Novikov, for introducing me to the wonderful world of martingales, and for his constant support and guidance during my Ph.D. study and thesis preparation. It has been a great pleasure to work under his supervision. I would also like to thank, Dr. Nino Kordazakhia from Macquarie University for all her support.

I am very grateful to Professor Dr. Tony Moon, not only for his invaluable guidance and support throughout my Ph.D. study at UTS but also for their great care and help during my staying in Australia.

I also wish to express my grateful appreciation to all faculty members in the Department of Mathematical Science at UTS, especially Dr. Tim Langtry, Dr. Yakov Zinder and Dr. Nicolae Nicorovici. I can never forget their kindness in assisting me with my Ph.D. study. The warmth of all the staffs and post-graduate students, in particular Gabrile Mititelu here at UTS will live in my heart.

Special thanks go to Dr. Elvin James Moors, Lectures at the Department of Mathematics, King Mongkut's University of Technology, North Bangkok (KMUTNB), Thailand for his careful proof-reading of my thesis and for the helpful suggestions he provided, resulting in substantial improvements to the language of my thesis.

I would like to express my gratitude to the Government of Thailand and the University of Technology, Sydney for their financial support during my study at UTS.

I am also grateful for the generous support from Department of Applied Statistics, KMUTNB, Thailand, where I have worked and developed my academic mental. Thank for giving me the opportunity to study in Australia.

Last, but not least, I would like to thank my mom and family, for their unconditional love, support and patience. To my special friend, Mus, thank you for being there. Finally, I want to dedicate this thesis to my late father, Chanin Sukparungsee, who always called me Doctor Kik in his lifetime.

Table of Contents

List of Figures

List of Tables

List of Notation and Abbreviations

Notation and abbreviations are used in this thesis are:

ARL	Average Run Length
AD	Average Delay
EWMA	Exponentially Weighted Moving Average
CUSUM	Cumulative Sum
MCA	Markov Chain Approach
IE	Integral Equations
MC	Monte Carlo Simulations
H	control limit/boundary
θ	moment of change-point
τ	stopping time/alarm time
α_0	in-control parameter
α	out-of-control parameter
λ	weighted smoothing parameter in EWMA statistic, $\lambda \in (0, 1)$
χ	overshoot
φ	characteristic function
ψ	cumulant generating function (logarithm of the moment generating function)

List of Publications and Conference Presentations

Parts of the work presented in this thesis have been previously published as:

1. Sukparungsee, Saowanit, & Novikov, Alexander. 2006. On EWMA Procedure for Detection of a Change in Observations via Martingale Approach. In *Proceedings of the KMITL International Conference on Science and Applied Science, 8-10 March*, Bangkok, Thailand (KSAS 2006).

2. Sukparungsee, Saowanit, & Novikov, Alexander. 2006. On EWMA Procedure for Detection of a Change in Observations via Martingale Approach. *KMITL Science Journal; An International Journal of Science and Applied Science*, 6(2a): 373-380.

3. Sukparungsee, Saowanit, & Novikov, Alexander. 2007. Analytical Approximations for Average Run Lengths in EWMA Charts in Case of Light-tailed Distributions. In *Proceedings of International Conference of Mathematical Sciences, 28-29 November*, Bangi-Putrajaya, Malaysia (ICMS 2007).

4. Sukparungsee, Saowanit, & Novikov, Alexander. 2008. Analytical Approximations for Detection of a Change-Point in Case of Light-tailed Distributions. *Journal of Quality Measurement and Analysis*, 4(2): 49-56.

Abstract

This thesis aims to derive analytical approximations and numerical algorithms for analysis and design of control charts used for detection of changes in distributions. In particular, we present a new analytical approach for evaluating of characteristics of the "Exponentially Weighted Moving Average (EWMA)" procedure in the case of Gaussian and non-Gaussian distributions with light-tails.

The main characteristics of a control chart are the mean of false alarm time or, Average Run Length (ARL), and the mean of delay for true alarm time or, Average Delay time (AD). ARL should be sufficiently large when the process is in-control and AD should be small when the process is out-of-control. Traditional methods for numerical evaluation ARL and AD are the Monte Carlo simulations (MC), Markov Chain Approach (MCA) and Integral Equations method (IE). These methods have the following essential drawbacks: the crude MC is very time consuming and difficult to use for finding optimal designs; MCA requires matrix inversions and, in general, is slowly convergent; IE requires intensive programming even for the case of Gaussian observations.

In this thesis we develop an approach based on a combination of the martingale technique and Monte Carlo simulations. With the use of a popular symbolic/numerical software Mathematica®, this new approach allows to obtain accurate procedures for finding the optimal weights, alarm boundaries and approximations ARL and AD for the case of Gaussian observations. Further, we show that our approach can be used also for non-Gaussian distributions with light-tails and, in particular, for Poisson and Bernoulli distributions.

Chapter 1

Introduction

Statistical Process Control (SPC) charts play a vital role in quality improvement and are widely used for monitoring, measuring, controlling and improving quality in many fields of application. These charts are used to track the performance of output processes, (e.g. manufacture etc.) in order to detect a possible change and then bring the specific process back to a target value as quickly as possible.

In 1931 W.A. Shewhart suggested simple SPC charts to use for quality control, (see e.g. Montgomery (2005)). SPC charts are now widely used, not only in industry, but also in many other areas with real applications, such as:

- Health care (Frisén, 1992; Hawkins and Olwell, 1998; Hillson et al., 1998)
- Epidemiology (Sitter et al., 1990)
- Clinical Chemistry (Westgard et al., 1977)
- Bioterrorism (see e.g. Hutwagner et al., 2003)
- Finance and Economics (Anderson, 2002; Ergashev, 2003)
- Computer intrusion detection (see e.g. Ye et al., 2002, 2003)
- Environmental Sciences (see for detail Basseville and Nikiforov, 1993)

A major aim of SPC charts is to detect a change in a process as soon as possible, but at the same time it would be desirable to have a low rate of false alarms. It is assumed that the parameter of an in-control process should be sustained at some specified target value but the target value might change at an unknown time point θ after which the process is out-of-control. A controller observes the process up to an alarm time τ (see Figure 1.1) before deciding that the process is out-of-control.

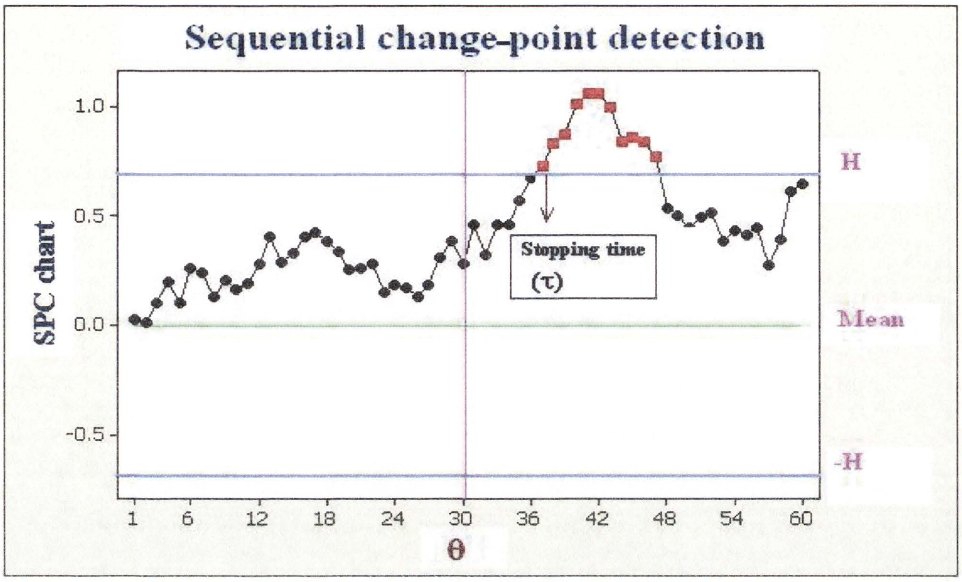


Figure 1.1: Sample of control chart

1.1 Formal Setting of Problem

Assume that sequential observations $\xi_1, \xi_2, ...,$ are independent random variables with the distribution function $F(x, \alpha)$, where α is a parameter. Further suppose that before the change-point θ the process is "in-control" and the distribution function of ξ_t is $F(x, \alpha_0)$, whereas after the change-point ("out-of-control") the distribution function of ξ_t is $F(x, \alpha)$, where $\alpha \neq \alpha_0$:

$$P(\xi_t < x) = \begin{cases} F(x, \alpha_0), & t = 1, 2, ..., \theta - 1 \\ F(x, \alpha), & t = \theta, \theta + 1, ..., \end{cases} \quad \text{for all} \quad x \in \mathbb{R}.$$

An alarm time of the "Shewhart" charts is based only on the latest observation, so earlier observations are completely disregarded. For this reason these charts are not efficient in the monitoring of small changes.

During the past few decades, two effective alternatives to the Shewhart chart, namely, the "Cumulative Sum (CUSUM)" and "Exponential Weighted Moving Average (EWMA)" charts have been developed. The CUSUM chart was introduced by Page (1954) (see for detail Lucas, 1976; Siegmund, 1985; Pollak, 1985; Srivastava and Wu, 1993). The EWMA chart was initially presented by Roberts (1959) (see also Roberts, 1966; Crowder, 1987; Hunter, 1986; Lucas and Saccucci, 1990; Srivastava and Wu, 1997). Both the CUSUM and EWMA charts are known to be more sensitive to the detection of small changes compared with the Shewhart chart (because they use all current observations). The CUSUM chart is known to be an efficient (or, at least, asymptotically efficient) tool, but only under some settings, (Lorden, 1971; Moustakides, 1986; Shiryaev, 1996). However, EWMA charts are inherently simpler and are also believed to be more robust than CUSUM charts.

The detection of the change-point θ is a trade-off between two limitations imposed on true alarm and false alarm times. The detection of large changes in the parameter α might only require a small number of observations, whereas the detection of small changes could require a much larger number of observations. One of the most popular criteria for choosing an alarm time (to claim that a process is out-of-control) is the "Average Run Length (ARL)"- the expectation of the stopping time given that the process is in-control. The ARL should be large enough when the process is in-control state.

To formalise this setup we use the probability space $(\Omega, \mathcal{F}, \mathbb{P}_\theta)$ and the concept of the natural filtration $\mathcal{F}_t = \sigma\{\xi_i, i \leq t\}$ (which is σ-algebra generated by the observed process $\xi_i, \{i = 1, 2, ...\}$). The distribution $\mathbb{P}_\theta(A)$, $A \in \mathcal{F}$ corresponds to the assumption that the change-point happened at time $t = \theta$, $\{\theta = 1, 2, ...,\}$. We set $\theta = \infty$ for the case when ξ_t has the distribution $F(x, \alpha_0)$ for all $t = 1, 2, ...$ (no change-point), α_0 is a parameter assigned to an "in-control" state. We shall use the symbol $E_\theta(.)$ for the expectation under the distribution $\mathbb{P}_\theta(.)$. Any stopping time τ is a random time such that the event $\{\tau > t\} \in \mathcal{F}_t$ for any $t \geq 0$. Typically, stopping times for SPC charts have the following forms:

$$\tau = \inf\{t > 0 : Z_t > a\} \quad \text{or} \quad \tau = \inf\{t > 0 : Z_t > a \text{ or } Z_t < b\},$$

where Z_t is a statistic generated by ξ_t, a and b are constants.
ARL can be defined formally by the following way

$$ARL = E_\infty(\tau) = T,$$

where T is a given parameter (usually large) and $E_\infty(.)$ is the expectation under the distribution "in-control" $F(x, \alpha_0)$.
Another characteristic of SPC is the "Average Delay (AD)" time which is the expectation of the alarm time under the assumption that the change-point occurs at a point $\theta = 1$, that is

$$AD = E_1(\tau),$$

where $E_1(.)$ is the expectation under the distribution "out-of-control" $F(x, \alpha)$.

Note

$$AD \leq \sup_{\theta} E_\theta(\tau - \theta + 1|\tau \geq \theta).$$

Of course, it would be desirable to optimise $\sup_{\theta} E_\theta(\tau - \theta + 1|\tau \geq \theta)$ but the latter quantity is usually difficult to calculate. One can expect that a SPC chart has a good performance if its AD is close to a minimal value. There are also many other criteria for optimality discussed in the literature (see e.g. Shiryaev, 1961, 1978, 1996; Lorden, 1971; Moustakides, 2008).

ARL and AD are the most popular measures used for comparing the performance of different control charts because they are comparatively easy to calculate. Numerical methods for the evaluation of ARL and AD have been discussed by many authors (see e.g. Barnard, 1959; Bissell, 1969; Woodall, 1983; Lucas and Saccucci, 1990; Gan, 1994; Srivastava and Wu, 1993, 1997; Frisén and Sonesson, 2006).

1.2 Research Objectives and Contributions

The overall objective of the thesis is related to the problem of using SPC charts for sequential detection of changes in distributions. We examine mathematical and statistical approaches for obtaining analytical approximations for the characteristics ARL and AD using EWMA charts.

The main aim of this thesis is to obtain analytical approximations for the characteristics of EWMA charts. We shall use a martingale relation between the expectations of alarm times and overshoots over a barrier for autoregressive processes firstly derived by Novikov (1990). Our approach combines the martingale technique with Monte Carlo simulations and it has been implemented using Mathematica®. Furthermore, a simple numerical algorithm for finding the optimal parameter values of EWMA chart has been established. The results of calculations of characteristics of EWMA chart are compared with other classical methods, such as the Markov Chain Approach (MCA) (see Brook and Evans, 1972; Lucas and Saccucci, 1990; Champ and Woodall, 1991), Integral Equations (IE) (Crowder, 1987; Srivastava and Wu, 1997) and Monte Carlo simulation (MC) (Roberts, 1959). In addition, we compare the performance of the EWMA chart with the performance of the CUSUM chart.

In the thesis, we also discuss the problem of detecting changes when the observations are from some non-Gaussian distributions such as Poisson and Bernoulli distributions. The martingale technique can be used to find the exact lower-bound for the expectation of the first alarm time for an EWMA chart and to compare the numerical results for EWMA with MC and other approaches.

Our contribution to the theory of SPC charts can be summarized as follows:

- We have developed a new analytical tool for determining the ARL and AD for an EWMA chart, including cases of Gaussian and some non-Gaussian distributions.
- We have suggested an algorithm for obtaining approximations for *optimal weights* and *alarm boundaries*.

Some results obtained using the suggested technique for Gaussian and some non-Gaussian distributions have been published (see Sukparungsee and Novikov, 2006, 2007).

1.3 Scope of Study

The scope of the study is:

1. *The martingale technique*

   Recall a process $\mathcal{M}_t$ is a martingale with respect to the filtration $\mathcal{F}_t$ if $E|\mathcal{M}_t| < \infty$ and

   $$E(\mathcal{M}_t|\mathcal{F}_s) = \mathcal{M}_s \quad (P - a.s.), \quad t \geq s.$$

   The application of martingales to study characteristics of stopping times is based on the martingale stopping theorem

   $$E(\mathcal{M}_\tau) = E(\mathcal{M}_0),$$

   which is valid for all bounded stopping times (τ) (see Chapter 3 for more details).

2. *The simulations and calculations*

   Monte Carlo simulations with 10^6 trials are carried out with Mathematica® and R software. Mathematica® is also used for some symbolic computations.

3. *Comparison of numerical results*

   The results from our approach are compared with other classical approaches to verify that the accuracy of our approach is as good as that of other approaches. Tables for optimal parameter values of EWMA designs are provided for observations drawn from Gaussian, Poisson and Bernoulli distributions.

1.4 Thesis Structure

The main body of the thesis is organized into the following chapters:

- In **Chapter 2** we review earlier work on the application of SPC charts to the detection of changes in distributions. The literature on Shewhart, CUSUM and EWMA charts and their properties are discussed.

- In **Chapter 3** we describe three standard methods for the numerical evaluation of ARL and AD: the Markov Chain Approach, Integral Equations and Monte Carlo simulation. In this chapter we also introduce a new analytical approximation, based on the so-called the "martingale technique".

- In **Chapter 4** we present the derivation of closed-form formulas for one and two-sided Gaussian EWMA charts using the martingale technique. We suggest a corrected approximation and describe the algorithms used for obtaining the numerical results. To illustrate an accuracy of the martingale approach, we also compare our numerical results with results obtained by MC, MCA and IE methods. We also compare the performance of an EWMA chart with the performance of a CUSUM chart.

- In **Chapter 5** we show the martingale technique can be applied to non-Gaussian distributions, in particular, to Poisson and Bernoulli distributions. We present the derivation of closed-form formulas for those distributions. As in Chapter 4, we compare the results from the martingale technique with the results from the other approaches. The performance of an EWMA chart is compared also with that of a CUSUM chart.

- In **Chapter 6** we summarize the main contributions of this thesis and suggest topics for further research.

- **Appendix A** contains codes developed in the Mathematica®.

- **Appendix B** presents a numerical algorithm for finding the optimal parameters of an EWMA chart.

Chapter 2

Statistical Process Control Charts for the Detection of Changes in Parameters

Statistical Process Control (SPC) charts are considered as the most important tools in quality control. The literature about SPC charts is extensive, especially about the problem of using SPC charts to detect a change-point in a process. In this chapter, we discuss some popular control charts and their basic properties and the criteria used to detect change-points in Section 2.1 .

In Section 2.2, we present the traditional control chart, the "Shewhart" chart, for independent univariate observations. The Cumulative Sum (CUSUM) chart and its main properties are discussed in Section 2.3. In Section 2.4 we describe the Exponentially Moving Weighted Average (EWMA) chart.

2.1 The Properties of Control Charts

It is assumed that a parameter value of independent observations of the process under control is to be sustained at some specified target value. However, in practice, this value can change at an unanticipated time, called the change-point time and usually denoted θ. In this thesis we suppose that the observations $\xi_1, \xi_2, \ldots$, are independent random variables with a distribution function $F(x, \alpha)$, where the parameter α has a value $\alpha = \alpha_0$ prior some change-point time $\theta \leq \infty$ (if $\theta = \infty$, there is no change). The parameter value α_0 corresponds to an "in-control" state of the process. After the change-point time θ, the new parameter value α corresponds to an "out-of-control" state where $\alpha \neq \alpha_0$. The parameter values for in-control state α_0 and out-of-control state α are assumed to be known. The in-control parameter α_0 could be measured in advance based on many observations. The out-of-control parameter α is usually unknown but it could be considered as a "target" constant which should be selected by practicers.

A major objective of SPC charts is to respond quickly to a parameter change, but not to respond (or, at least, to respond rarely) if the parameter does not change. These objectives are, however, in a conflict because sufficient amount of observations must be taken after the change-point to confirm that the parameter value has actually changed.

One of the important characteristics for SPC charts is the Average Run Length (ARL), which is the expectation of an alarm time τ required to detect a possible parameter change. Let $E_\theta(\tau)$ denote the expectation of τ under the assumption that the change-point occurs at point $\theta \leq \infty$. We set $\theta = \infty$ for the case when there is no change in parameter. Then, by definition

$$ARL \equiv E_\infty(\tau), \tag{2.1a}$$

and, usually, one requires to have equality

$$ARL = T, \tag{2.1b}$$

where T is a given number. Another important characteristics of SPC charts is the average time of delay that is the quantity

$$E_\theta(\tau - \theta + 1 | \tau \geq \theta). \tag{2.2}$$

Though Equation (2.2) depends on the parameter θ, some empirical studies (see e.g. Lucas and Saccucci, 1990) have suggested that this dependence is not essential. For this reason, and in order to compare our results with other methods, we have assumed that $\theta = 1$ in Equation (2.2), and call it Average Delay (AD): $AD = E_1(\tau)$.

They are many other criteria that have been used for optimality of SPC (see e.g. Shiryaev, 1961, 1978; Lorden, 1971). For example, an asymptotic theory was provided in Lorden (1971) and Shiryaev (1996) using a minimax approach. However, ARL and AD are still the most popular and commonly used characteristics for evaluating the performance of a control chart. These two criteria are used as the basis for the comparison of different charts throughout this thesis.

2.2 Shewhart Chart

Let $\xi_1, \xi_2, ..., \xi_t$ be independent and identically random variables (observations) with the mean (α_0) and the variance (σ^2).

The stopping time of a Shewhart chart is typically defined as a first passage time

$$\tau_A = \inf\{t > 0 : \xi_t > A_1 \quad \text{or} \quad \xi_t < A_2\}.$$

The levels A_1 and A_2 are usually called the upper control limit (UCL) and the lower control limit (LCL). Traditionally,

$$UCL = A_1 = \alpha_0 + L\sigma \quad \text{and} \quad LCL = A_2 = \alpha_0 - L\sigma,$$

where L is a constant (usually, $L = 3$ for the case of Gaussian distribution), see Figure 2.1.

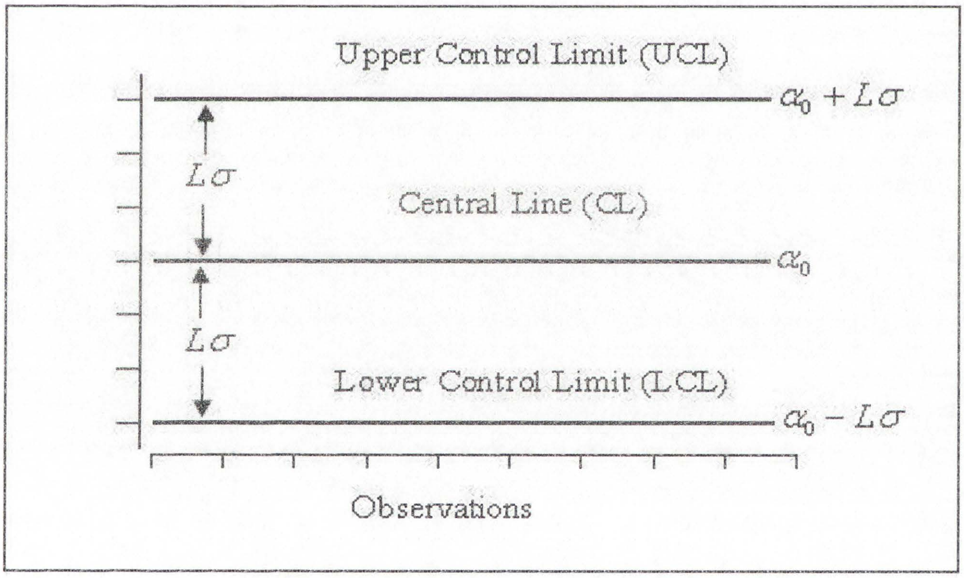


Figure 2.1: A typical Shewhart chart

In the case of Shewhart chart, ARL and AD can be easily calculated in terms of the error probabilities: the probability of type I error (P_I) (the test rejects "true" H_0: signal occurs when a point falls outside the control limit though there is no real change in parameter) and the probability of type II error (P_{II}) (the test rejects "true" H_1: no signal occurs when a point falls outside the control limit when there is a real change in parameter). The corresponding formulas are

$$ARL = E_\infty(\tau) = \frac{1}{P_I} \quad \text{and} \quad AD = E_1(\tau) = \frac{1}{(1 - P_{II})}.$$

2.3 CUSUM Chart

The Cumulative Sum (CUSUM) chart was first introduced by Page (1954). Denote the standard CUSUM defined by statistics Y_t with the following recursion:

$$Y_t = \max(Y_{t-1} + q(\xi_t), 0), \quad t = 1, 2, ..., \quad Y_0 = y, \tag{2.3}$$

where

$$q(x) = \log \frac{dF(x, \alpha)}{dF(x, \alpha_0)}$$

and it is assumed that the distribution $F(x, \alpha)$ is absolutely continuous with respect to the distribution $F(x, \alpha_0)$. It $F(x, \alpha)$ is a continuous distribution with a density function $f(x, \alpha)$ then $\frac{d}{dx}F(x, \alpha) = f(x, \alpha)$. One can find many other modifications of CUSUM charts in Hawkins and Olwell (1998).

Example 1. *If $\xi_t \sim Gaussian(\alpha, \sigma^2)$ then for all $x \in (-\infty, \infty)$,*

$$\begin{aligned} q(x) &= \log[\frac{f(x, \alpha)}{f(x, \alpha_0)}] \\ &= \frac{\alpha - \alpha_0}{\sigma^2}(x - \frac{\alpha + \alpha_0}{2}). \end{aligned}$$

Example 2. *If $\xi_t \sim Poisson(\alpha)$ then for $x = 0, 1, ...,$*

$$\begin{aligned} q(x) &= \log[\frac{f(x, \alpha)}{f(x, \alpha_0)}] \\ &= (\alpha_0 - \alpha_1) + x \log[\frac{\alpha}{\alpha_0}]. \end{aligned}$$

Example 3. *If $\xi_t \sim Bernoulli(\alpha)$, i.e. $P(\xi_t = 1) = \alpha = 1 - P(\xi_t = 0)$ then for $x = 0$ or 1,*

$$\begin{aligned} q(x) &= \log[\frac{f(x, \alpha)}{f(x, \alpha_0)}] \\ &= \log[(\frac{\alpha}{\alpha_0})^x (\frac{1 - \alpha}{1 - \alpha_0})^{1-x}] \\ &= x \log[(\frac{\alpha}{\alpha_0}) + (\frac{1 - \alpha}{1 - \alpha_0})] + (1 - x) \log[\frac{1 - \alpha}{1 - \alpha_0}]. \end{aligned}$$

The stopping time of a CUSUM chart is typically defined as a first passage time

$$\tau_B = \inf\{t > 0 : Y_t > B\},$$

where B is the control limit.

CUSUM charts have been discussed in many papers and books (see e.g. Page, 1961; Ewan, 1963; Brook and Evans, 1972; Hawkins, 1981; Gan, 1991; Hawkins, 1993; Woodall and Adams, 1993; Montgomery, 2005). Hawkins and Olwell (1998) provide a very accurate numerical approximation equation for evaluating ARL.

2.4 EWMA Chart

The Exponentially Weighted Moving Average (EWMA) chart was introduced by Roberts (1959). Crowder (1987); Lucas and Saccucci (1990); Srivastava and Wu (1993, 1997); Frisén (2003); Frisén and Sonesson (2006); Knoth (2007) have given detailed discussions and numerical comparisons for characteristics EWMA. The EWMA for the discrete time case is defined by statistics Z_t as the following recursion:

$$Z_t = (1-\lambda)Z_{t-1} + \lambda\xi_t, \quad t = 1, 2, \ldots . \tag{2.4}$$

The parameter $\lambda \in (0, 1)$ is a weight for previous observations and the initial value Z_0 is often assumed to be the target parameter value α_0, i.e. it is assumed that $Z_0 = \alpha_0$.

The EWMA statistic Z_t is a weighted average of all previous observations. A closed-form representation can be obtained for it by recursion of Equation (2.4). We first substitute for Z_{t-1} on the right hand side of Equation (2.4) to obtain

$$\begin{aligned} Z_t &= (1-\lambda)Z_{t-1} + \lambda\xi_t \\ &= (1-\lambda)[(1-\lambda)Z_{t-2} + \lambda\xi_{t-1}] + \lambda\xi_t \\ &= (1-\lambda)^2 Z_{t-2} + \lambda(1-\lambda)\xi_{t-1} + \lambda\xi_t. \end{aligned}$$

On continuing to substitute recursively for Z_{t-j}, $j = 2, 3, ..., t$, we obtain

$$Z_t = \lambda \sum_{k=0}^{t-1} (1-\lambda)^k \xi_{t-k} + (1-\lambda)^t Z_0. \tag{2.5}$$

EWMA is sometimes called a Geometric Moving Average (GMA). If the observations ξ_t

are independent random variables with mean α_0 and variance σ^2, then the expectation of Z_t is

$$E(Z_t) = \lambda \sum_{k=0}^{t-1} (1-\lambda)^k E(\xi_{t-k}) + (1-\lambda)^t Z_0$$
$$= \lambda \sum_{k=0}^{t-1} (1-\lambda)^k \alpha_0 + (1-\lambda)^t Z_0, \tag{2.6}$$

the variance of Z_t is

$$\sigma^2_{Z_t} = \lambda^2 \sum_{k=0}^{t-1} (1-\lambda)^{2k} \mathrm{Var}(\xi_{t-k})$$
$$= (\frac{\lambda}{2-\lambda})[1-(1-\lambda)^{2t}]\sigma^2, \tag{2.7}$$

where we used the equality

$$\sum_{k=0}^{t-1} (1-\lambda)^{2k} = \lambda[\frac{1-((1-\lambda)^2)^t}{1-(1-\lambda)^2}] = \frac{1-(1-\lambda)^{2t}}{\lambda(2-\lambda)}, \quad \lambda(2-\lambda) \neq 0.$$

As t increases ($t \to \infty$) then the term $[1-(1-\lambda)^{2t}]$ in Equation (2.7) approaches to one. This implies that the variance of Z_t will tend to the asymptotic value which is $(\frac{\lambda}{2-\lambda})\sigma^2$.

Therefore, the EWMA control limits can be taken as follows:

$$UCL = \alpha_0 + L\sigma\sqrt{\frac{\lambda}{2-\lambda}},$$
$$LCL = \alpha_0 - L\sigma\sqrt{\frac{\lambda}{2-\lambda}},$$

where L is a parameter to be chosen. Note that it is common to use $L = 3$ to detect the "out-of-control" state for the case of Gaussian observations. Set $H = L\sigma\sqrt{\frac{\lambda}{2-\lambda}}$ as a control limit. Note that when $\lambda = 1$, the above equation gives the same limits as the Shewhart chart. Without loss of generality, in case of Gaussian observations we will suppose that the mean value is zero ($\alpha_0 = 0$) and that the variance is one ($\sigma^2 = 1$).

If the anticipated shift in the mean value is positive (that is $\alpha > 0$), then we take the decision that the process is out-of-control when for the first time $Z_t > H$, that is *the stopping time of an one-sided EWMA chart* is

$$\tau_H = \inf\{t > 0 : Z_t > H\}. \tag{2.8}$$

If the anticipated shift in the mean could be positive and negative then it is natural to take the decision that the process is out-of-control when the first time $|Z_t| > H$ that is *the stopping time of a two-sided EWMA chart* is

$$\gamma_H = \inf\{t > 0 : |Z_t| > H\}. \tag{2.9}$$

In practice, it is desirable to find a value of the smoothing parameter λ and an out-of-control level H such that $AD = E_1(\tau)$ is minimal under constraints (2.1).

2.5 Discussion

It is well-known that the Shewhart chart is good for the detection of large changes, but it is not efficient for detecting small changes. This is because it only takes into account the last observation. CUSUM and EWMA charts take into account the previous observations and that is a reason why they are essentially more sensitive for the detection of small changes.

CUSUM charts are known to be asymptotic (as $T \to \infty$) optimal under a special minimax criterion (Lorden, 1971; Shiryaev, 1996). However, its performance can be inferior to EWMA for moderate values of α and T as indicated firstly by Lucas and Saccucci (1990). This result has been further verified for one-sided EWMA charts by Srivastava and Wu (1993, 1997) for Gaussian distribution and Sukparungsee and Novikov (2007) for some non-Gaussian distributions.

EWMA charts are inherently simpler than CUSUM charts and are also known to be more robust than CUSUM (see e.g. Lucas and Saccucci, 1990; Ergashev, 2003; Knoth, 2007) against occasional outlier and a series correlations in observations. EWMA charts are as simple and flexible as CUSUM charts, in addition they can also used to detect changes of parameters on either side simultaneously.

Chapter 3

Methods for Finding Analytical and Numerical Approximations for ARL and AD

Many authors (see Crowder, 1987; Yashchin, 1987; Lucas and Saccucci, 1990; Frisén, 1992; Srivastava and Wu, 1997) have extensively studied methods for evaluating of the performance of EWMA charts and for finding analytical and numerical approximations for the characteristics of EWMA charts. There are at least three standard methods that have been used for developing numerical approximations and for evaluating their performance: the Markov Chain Approach (MCA), Integral Equations (IE) and Monte Carlo simulations (MC).

In this chapter we develop a new analytical approach for evaluating performance based on a martingale technique. The chapter has four sections:

- Section 3.1 - explains the Markov Chain Approach (MCA)
- Section 3.2 - introduces the Integral Equations method (IE)
- Section 3.3 - discusses the Monte Carlo simulations approach (MC)
- Section 3.4 - describes the new analytical approach based on the martingale technique.

The discussion of the basic theory given in this chapter is based on the following papers: Liptser and Shiryaev (1977); Siegmund (1985); Basseville and Nikiforov (1993);

Shiryaev (1995); Gallagher (1996); Ross (1996); Borovkov (1999); Vardeman and Jobe (1999); Montgomery (2005).

3.1 The Markov Chain Approach

The Markov Chain Approach (MCA) is an important and popular technique for evaluating ARL. It was Brook and Evans (1972) who first studied the MCA, later Lucas and Saccucci (1990) developed the MCA for a continuous Markov Chain. According to this approach, to obtain a reasonable accuracy of approximations it is usually necessary to discretise the underlying process using a number of steps. As far as we know there are no theoretical results about a rate of convergence of this method, however its accuracy can be studied by direct comparison with Monte Carlo simulations.

Definition 1 (Markov processes). *A stochastic process X_t, $t = 1, 2, ...$ is a Markov process if for each*

$$t_1 \leq t_2 \leq ...t_n \leq t \leq t+s,$$

the conditional probability

$$P\{X_{t+s} = y | X_{t_1} = x_1, X_{t_2} = x_2, ..., X_{t_n} = x_n, X_t = x\} = P\{X_{t+s} = y | X_t = x\}.$$

For a Markov process, the probability for the process to be in a "future" state X_{t+s} is completely determined by the state X_t at the "present" time t.

3.1.1 The Markov Chain Approach

Consider a homogeneous Markov process with a finite number of states x_j, $j = 1, 2, ..., n, n+1$. The change from one state to another in one time step can be expressed by a transition probability matrix $\mathbf{P} = (P_{ij})$ whose elements are

$$P_{ij} := P(X_{t+1} = x_j | X_t = x_i). \tag{3.1}$$

Chains with a single absorbing state are of special interest to change-point detecting control charts. We assume that states x_j, $j = 1, 2, ..., n$ correspond to "in-control" states and x_{n+1} is an "out-of-control" state. If the chain starts in an in-control state x_i we set

τ_i as the first alarm time that the chain enters the out-of-control state,

$$\tau_i = \inf\{t > 0 : X_t = x_{n+1}\}, \quad X_0 = x_i, \quad i \leq n.$$

Since the values P_{ij} represent probabilities that the process will make a transition from state i to state j in one time step, and since the process must make a transition into some state, we have the following relations:

$$0 \leq P_{ij} \leq 1, \;\; i, j > 0; \qquad \sum_{j=1}^{n+1} P_{ij} = 1, \quad i = 1, 2, \ldots .$$

Let $\mathbf{P}$ denote a matrix of one-step transition probabilities P_{ij}, so that

$$\mathbf{P} = \begin{Vmatrix} P_{11} & P_{12} & P_{13} & \cdots \\ P_{21} & P_{22} & P_{23} & \cdots \\ \vdots & & & \\ P_{i1} & P_{i2} & P_{i3} & \cdots \\ \vdots & \vdots & \vdots & . \end{Vmatrix}$$

Since "in-control" states have indexes $1, 2, ..., n$ and "out-of-control" state is $n+1$, matrix $\mathbf{P}$ can be written in the following form:

$$\mathbf{P} = \begin{pmatrix} \mathbf{R} & (\mathbf{I}_n - \mathbf{R})\mathbf{1}_n \\ \mathbf{0}_n^T & 1 \end{pmatrix}$$

where,

$\mathbf{R}$ is the submatrix of transition probabilities (P_{ij}) for states $1, 2, ..., n$,

$\mathbf{I}_n$ is the unit $n * n$ matrix,

$\mathbf{1}_n$ is the unit $n * 1$ vector (all elements are 1),

$\mathbf{0}_n$ is zero $n * 1$ vector.

The matrix of transition probabilities for k steps contains the transition probabilities for moves from one state to another states in k steps,

$$\mathbf{P}^k = \begin{pmatrix} \mathbf{R}^k & (\mathbf{I}_n - \mathbf{R}^k)\mathbf{1}_n \\ \mathbf{0}_n^T & \mathbf{1} \end{pmatrix}$$

where $(\mathbf{I}_n - \mathbf{R}^k)\mathbf{1}_n$ is the vector of transition probabilities from states $i < n+1$ to the state $n+1$ in k steps. Hence,

$$
\begin{aligned}
P(\tau_i \leq k) &= \text{element}[(\mathbf{I}_n - \mathbf{R}^k)\mathbf{1}_n](i) \\
&= \text{sum of row (i)}(\mathbf{I}_n - \mathbf{R}^k) \\
&= \mathbf{p}_n^{(i)T}(\mathbf{I}_n - \mathbf{R}^k)\mathbf{1}_n, \qquad (3.2)
\end{aligned}
$$

where $\mathbf{p}_n^{(i)T}$ is the initial probability vector with 1 at i^{th} position and 0 otherwise. Then, Equation (3.2) is used to find the probability that the process starts in the in-control state x_i and first goes out-of-control state at time k:

$$
\begin{aligned}
P(\tau_i = k) &= P(\tau_i \leq k) - P(\tau_i \leq k-1) \\
&= \mathbf{p}_n^{(i)T}(\mathbf{I_n} - \mathbf{R}^k - (\mathbf{I_n} - \mathbf{R}^{k-1}))\mathbf{1}_n \\
&= \mathbf{p}_n^{(i)T}(\mathbf{R}^{k-1} - \mathbf{R}^k)\mathbf{1}_n. \qquad (3.3)
\end{aligned}
$$

We shall use the following result:

Lemma 3.1.1. *If* $\sum_{k=0}^{\infty} \mathbf{R}^k$ *converges absolutely, then* $\sum_{k=0}^{\infty} \mathbf{R}^k = (\mathbf{I}_n - \mathbf{R})^{-1}$.

Proof. If $\sum_{k=0}^{\infty} \mathbf{R}^k$ converges absolutely then we can multiply each terms of the series by another matrices and, in particular, we have the following relation

$$
\begin{aligned}
(\mathbf{I}_n - \mathbf{R}) \sum_{k=0}^{\infty} \mathbf{R}^k &= \sum_{k=0}^{\infty} (\mathbf{R}^k - \mathbf{R}^{k+1}) \\
&= (\mathbf{I} - \mathbf{R}) + (\mathbf{R} - \mathbf{R}^2) + \cdots \\
&= \mathbf{I}.
\end{aligned}
$$

Therefore, $\sum_{k=0}^{\infty} \mathbf{R}^k = (\mathbf{I}_n - \mathbf{R})^{-1}$. Remark, the series $\sum_{k=0}^{\infty} \mathbf{R}^k$ converges absolutely, e.g. under the condition $||R|| < 1$, where $||R||$ is the absolute value of the largest eigenvalue of $\mathbf{R}$. $\square$

Theorem 3.1.2. *The Average Run Length for the MCA n states with an initial state x_i is calculated as follows:*

$$ARL(n) = \mathbf{p}_n^{(i)T}(\mathbf{I}_n - \mathbf{R})^{-1}\mathbf{1}_n.$$

Proof. In order to proof Theorem (3.1.2), we use Equation (3.3) as follows:

$$\begin{aligned}
ARL(n) &= \sum_{k=0}^{\infty} kP(\tau_i = k) \\
&= \sum_{k=0}^{\infty} k\mathbf{p}_n^{(i)T}(\mathbf{R}^{k-1} - \mathbf{R}^k)\mathbf{1}_n \\
&= \sum_{k=0}^{\infty} \mathbf{p}_n^{(i)T}((k+1)\mathbf{R}^k - k\mathbf{R}^k)\mathbf{1}_n \\
&= \sum_{k=0}^{\infty} \mathbf{p}_n^{(i)T}\mathbf{R}^k\mathbf{1}_n \\
&= \mathbf{p}_n^{(i)T}(\mathbf{I}_n - \mathbf{R})^{-1}\mathbf{1}_n, \qquad (3.4)
\end{aligned}$$

where Lemma 3.1.1 is used to obtain the final equation. □

One may expect that when $n \to \infty$, $ARL(n)$ and $AD(n)$ converge to true values ARL and AD. However, we fail to find in literature any theoretical results of this type.

3.1.2 The Procedure for Calculation ARL and AD Using the Markov Chain Approach.

Assume that the allowed values for the variable being controlled are bounded by H_L from below and by H_U from above. The interval (H_L, H_U) is the in-control region and values outside this interval are the out-of-control region as shown in Figure 3.1. The procedure for obtaining two-sided ARL using a Markov Chain Approach has the following steps:

- Divide the interval between lower-bound and upper-bound for in-control parameter values (H_L, H_U) into n subintervals which are used as the n in-control states of a Markov Chain. Denote the midpoints of these subintervals as m_i, $i = 1, 2, .., n$

- Calculate a lower L_i and upper U_i bound of the i^{th} subinterval as follows:

$$L_i = H_L + \frac{(i-1)(H_U - H_L)}{n}, \qquad (3.5)$$

$$U_i = H_L + \frac{i(H_U - H_L)}{n}, \qquad (3.6)$$

and the midpoint m_i of the i^{th} subinterval is then

$$m_i = \frac{1}{2}(L_i + U_i) = H_L + \frac{(2i-1)(H_U - H_L)}{2n}. \qquad (3.7)$$

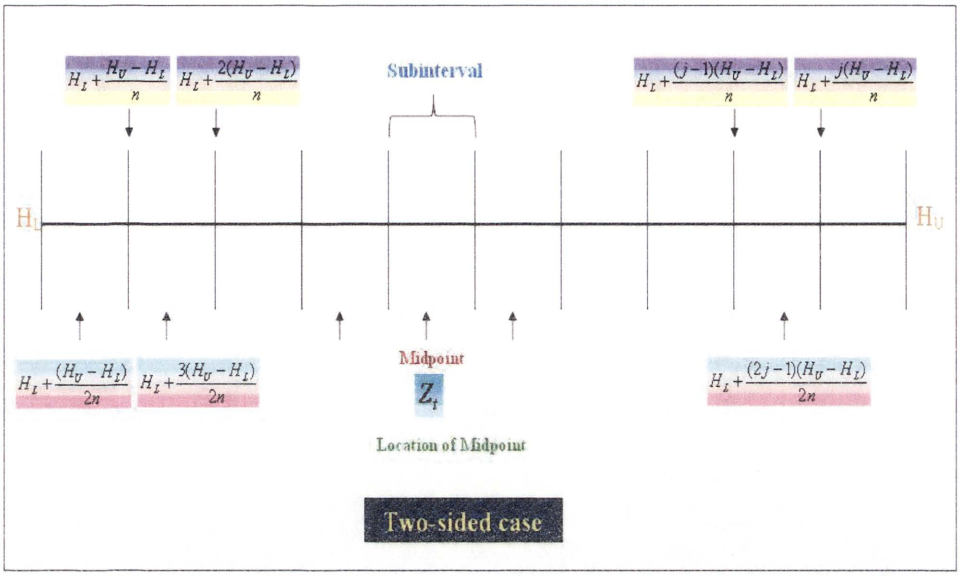


Figure 3.1: In-control region divided into n subintervals to form n in-control states of the Markov Chain

- Construct the matrix $\mathbf{R}$, where each element P_{ij} is the transition probability of going from i^{th} subinterval (i^{th} state of Markov Chain) to j^{th} subinterval in one step for $i, j = \{1, 2, .., n\}$ by substituting Equation (3.5), Equation (3.6) and Equation (3.7) into Equation (3.1). As an example, suppose that the EWMA statistic at time $t-1$, i.e., Z_{t-1}, is in the i^{th} subinterval. Then, according to the MCA we assume that Z_{t-1} is equal to the midpoint of the i^{th} subinterval and obtain the following transition probabilities

$$
\begin{aligned}
P_{ij} &= P(L_j < Z_t < U_j | Z_{t-1} = m_i) \\
&= P(H_L + \frac{(j-1)(H_U - H_L)}{n} < Z_t < H_L + \frac{j(H_U - H_L)}{n}) \\
&= P(H_L + \frac{(j-1)(H_U - H_L)}{n} < (1-\lambda)(H_L + \frac{(2i-1)(H_U - H_L)}{2n}) + \lambda(Z_t) \\
&\quad < H_L + \frac{j(H_U - H_L)}{n}) \\
&= P(H_L + \frac{H_U - H_L}{2n\lambda}(2(j-1) - (1-\lambda)(2i-1)) < Z_t \\
&\quad < H_L + \frac{H_U - H_L}{2n\lambda}(2j - (1-\lambda)(2i-1)) \\
P_{ij} &= F(H_L + \frac{H_U - H_L}{2n\lambda}(2j - (1-\lambda)(2i-1))) \\
&\quad - F(H_L + \frac{H_U - H_L}{2n\lambda}(2(j-1) - (1-\lambda)(2i-1))),
\end{aligned}
$$

where $F(\cdot)$ is the cumulative distribution function.

- Construct the vector $\boldsymbol{p}$ with length $n = [0, 0, 0, ..., 1, 0, ..., 0, 0]$. The number of states must be odd so that there is a unique middle value.

- Calculate $ARL(n)$ by using Equation (3.4).

Note that the approximation $ARL(n)$ for the case of one-sided EWMA can be calculated in a similar manner to the case of the two-sided EWMA by substituting H_L with zero. Then the result is

$$P_{ij} = F(\frac{H_U}{2n\lambda}(2j - (1-\lambda)(2i-1))) - F(\frac{H_U}{2n\lambda}(2(j-1) - (1-\lambda)(2i-1))).$$

3.2 Integral Equations

Crowder (1987) derived an Integral Equation for the ARL of an EWMA chart and showed that the ARL can be expressed in terms of a Fredholm Integral Equation of the second kind. Waldmann (1986) obtained an Integral Equation for the EWMA chart by the method of Page (1954). Srivastava and Wu (1993) followed a similar method to develop integral equations for a continuous-time model. Srivastava and Wu (1997) also showed that some adjustments had to be made to the Integral Equations for a discrete-time model to correct for the possible overshoot of the boundary. For the discrete-time model, approximations for ARL and AD can be found by solving the Integral Equations numerically.

Consider an EWMA chart where $\xi_1, \xi_2, ...$ are independent random variables with a continuous distribution with a probability density $f(y)$. Assume that the lower-bound and upper-bound for the variables ξ_t, $t = 1, 2, \ldots$ are $H_L = -H$ and $H_U = H$, correspondingly. Let $L(u)$ be the ARL of the EWMA chart with initial EWMA statistic $Z_0 = u$ (i.e, $L(u) = E_\infty(\tau_H)$). If an EWMA sequence begins at u, there are two possible cases where it will be after a single observation, ξ_1. First, if ξ_1 is in the out-of-control region, i.e. $((1-\lambda)u + \lambda\xi_1 < -H \quad \text{or} \quad (1-\lambda)u + \lambda\xi_1 > H)$, that means the signal is out-of-control and therefore the run length is 1.

Second, if ξ_1 is in the in-control region, i.e. $(-H < (1-\lambda)u + \lambda\xi_1 < H)$, then one observation will have been run and on average $L((1-\lambda)u + \lambda\xi_1)$ more observations will be produced before a signal is given. This can be expressed in the form

$$\begin{aligned} L(u) = {} & (1 - P[\frac{-H-(1-\lambda)u}{\lambda} < \xi_1 < \frac{H-(1-\lambda)u}{\lambda}]) \\ & + \int_{\frac{-H-(1-\lambda)u}{\lambda}}^{\frac{H-(1-\lambda)u}{\lambda}} (1 + L((1-\lambda)u + \lambda y)) f(y) dy \end{aligned}$$

$$L(u) = 1 + \int_{\frac{-H-(1-\lambda)u}{\lambda}}^{\frac{H-(1-\lambda)u}{\lambda}} L((1-\lambda)u + \lambda y) f(y) dy.$$

Setting $(1-\lambda)u + \lambda y = z$, we obtain

$$L(u) = 1 + \frac{1}{\lambda}\int_{-H}^{H} L(z) f(\frac{z-(1-\lambda)u}{\lambda}) dz, \tag{3.8}$$

where $f(\cdot)$ denotes the probability density function of ξ_t.

A similar approach (to that one used to obtain Equation (3.8)) can be used to derive an integral equation for an AD. It is assumed that the out-of-control signal occurs when $|Z_t - \alpha| > H$ for some constant H. Denote $L(u,\alpha)$ as the AD of the EWMA chart, given that the parameter of change is equal to α and that the initial EWMA is $Z_0 = u$, (i.e $L(u,\alpha) = E_1(\tau_H)$). The analog of Equation (3.8) is the Fredholm Integral Equation of the second kind (see for example, Wieringa, 1999).

3.3 Monte Carlo Simulations

Simulation studies can be used when exact analytical formulas are not available. In this thesis, the simulation technique has been used to examine the performance of the EWMA chart and to estimate boundary overshoots. The Mathematica® code for simulations for ARL and AD for both one-sided and two-sided EWMA charts are given in Appendix A. The code includes also the case of CUSUM.

The MC method is based on the Law of Large Numbers (LLN). The simplest version of LLN is the following:

Theorem 3.3.1. *(Weak Law of Large Numbers (WLLN)). Let $X_1, X_2, ..., X_n$ be independent random variables with finite expected value $E(X_i) = \mu$ and finite variance $V(X_i) = \sigma^2$. Let $S_n = X_1 + X_2 + ... + X_n$.*

Then for any $\varepsilon > 0$,

$$P(|\frac{S_n}{n} - \mu| \geq \varepsilon) \to 0 \quad as \quad n \to \infty,$$

equivalently,

$$P(|\frac{S_n}{n} - \mu| < \varepsilon) \to 1 \quad as \quad n \to \infty.$$

Let

$$\xi_n = \frac{\sum_{i=1}^{n}(X_i - \mu)}{\sqrt{n}\sigma}, \quad \Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} exp\{\frac{-x^2}{2}\}dx.$$

Theorem 3.3.2. *(Central Limit Theorem (CLT)). Let $X_1, X_2, ..., X_n$ be a sequence of i.i.d random variables such that $E(X_i) = \mu$, $E(X_i^2) < \infty$ and $Var(X_i) = \sigma^2 > 0$. Then*

$$\xi_n \xrightarrow{d} \xi \sim N(0,1)$$

or, equivalently, for all $x \in (-\infty, \infty)$

$$P(\xi_n \leq x) \to \Phi(x).$$

Suppose X is a random variable and we need to evaluate

$$J = E[g(X)],$$

where $g(X)$ is a given function. To estimate J we need to generate a sequence of independent random variables $X_1, X_2, ..., X_n$, such that $P(X_i \leq x) = P(X \leq x)$ and then according to WLLN we have

$$J_n := \frac{1}{n}\sum_{i=1}^{n} g(X_i) \xrightarrow{P} E[g(X)].$$

To estimate an accuracy of the approximation, we assume

$$Var(g(X)) := \sigma^2(g) < \infty,$$

and note

$$Var(J_n) = \frac{Var(g(X))}{n} = \frac{\sigma^2(g)}{n}.$$

Then, applying CLT we have the convergence $(J_n - J)\sqrt{n} \xrightarrow{d} N(0, \sigma^2(g))$. In particular, it implies that

$$|J_n - J| \leq 3\frac{\sigma(g)}{\sqrt{n}}$$

with probability approximately equal to 0.997... (for large n). The constant $\sigma^2(g)$ is usually unknown but it also can be estimated using WLLN:

$$\hat{\sigma}_n^2(g) := \frac{1}{n}\sum_{i=1}^{n} g^2(X_i) - (J_n)^2 \xrightarrow{P} E(g^2(X)) - J^2 = \sigma^2(g).$$

3.3.1 Control Variates

The method of control variates is one of the most widely used variance reduction techniques for MC. Suppose we want to evaluate $J = Eg(X)$, where

$$J_n = \frac{1}{n}\sum_{i=1}^{n} g(X_i).$$

This estimator for J is *consistent* that is $J_n \xrightarrow{P} E[g(X)]$, *unbiased* that is $EJ_n = E[g(X)]$ and it is *asymptotic normal* that is

$$P\{|\tilde{J}_n - J| < x\frac{\sigma(g)}{\sqrt{n}}\} \to 2\Phi(x) - 1,$$

where we suppose that $\sigma^2(g) = Var[g(X)] < \infty$.

A control variate is a function $q(X)$ such that the expectation $Q = Eq(X)$ is known. Consider the estimator $Q_n = \frac{1}{n}\sum\limits_{i=1}^{n} q(X_i)$ and another estimator for J in the form

$$\tilde{J}_n = J_n - a(Q_n - Q),$$

where a is a parameter which can be chosen to minimise the variance of $\tilde{J}_n$.
Set $\sigma^2(q) = Var[q(X)]$ and assume that $\sigma^2(q) < \infty$. Then by WLLN,

$$\tilde{J}_n = J_n - a(Q_n - Q) \xrightarrow{P} J = E[g(X)],$$
$$E(\tilde{J}_n) = J, \quad Var(\tilde{J}_n) = \frac{Var(g(X) - aq(X))}{n} = \frac{\sigma^2(g - aq)}{n}$$

and so by CLT

$$P\{|\tilde{J}_n - J| < x\frac{\sigma(g - aq)}{\sqrt{n}}\} \to 2\Phi(x) - 1,$$

where

$$\begin{aligned}\sigma^2(g - aq) &= Var(g(X) - aq(X)) \\ &= Var(g(X)) + a^2 Var(q(X)) - 2aCov(g(X), q(X)).\end{aligned}$$

Therefore, the estimator $\tilde{J}_n$ is also *consistent*, *unbiased* and *asymptotic normal* for any value a.

The variance $\sigma^2(g - aq)$ is a quadratic function of the parameter a and we may choose value a to minimise the variance $\sigma^2(g - aq)$. The optimal a is

$$a^* = \frac{Cov(g(X), q(X))}{\sigma^2(q)}$$

and the corresponding optimal variance is

$$\sigma^2(g - a^*q) = \sigma^2(g) - \frac{(Cov(g(X)), q(X))^2}{\sigma^2(q)} = \sigma^2(g)(1 - \rho^2_{g,q}),$$

where the correlation coefficient $\rho_{g,q}$ is

$$\rho_{g,q} = \frac{Cov(g(X), q(X))}{\sqrt{\sigma^2(g)\sigma^2(q)}}.$$

Note that if $|\rho_{g,q}|$ is near to 1 then the variance $\sigma^2(g - a^*q)$ may be much smaller than the variance $\sigma^2(q)$. That is why we can potentially reduce the variance of MC estimator in several times.

The quality of a control variate q depends on the correlation between g and q. In practice it is not likely that one would know the optimal value a^* but it can be estimated from Monte Carlo simulations. Using X_i we can evaluate $g_i = g(X_i)$ and $q_i = q(X_i)$, then

$$\begin{aligned}
\hat{\sigma}_q^2 &= \frac{1}{n}\sum_{i=1}^{n}(q_i - Q)^2, \\
\hat{J}_n &= \frac{1}{n}\sum_{i=1}^{n} g_i, \\
\hat{Cov}(g(X), q(X)) &= \frac{1}{n}\sum_{i=1}^{n}(g_i - \hat{J}_n)(q_i - Q), \\
\hat{a}^* &= \frac{\hat{Cov}(g(X), q(X))}{\hat{\sigma}_q^2}, \\
\tilde{J}_n &= \hat{J}_n - \hat{a}^*\frac{1}{n}\sum_{i=1}^{n}(q_i - Q)
\end{aligned}$$

Concerning other methods for variance reduction see e.g. Borovkov (2003).

3.4 The Martingale Approach

In this section we show how to exploit a martingale technique to obtain closed-form relations for the expectations of stopping times of first order autoregressive processes. We start with some general definitions and notation from stochastic analysis.

3.4.1 Martingales and Stopping times

We shall always assume that we have given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and there is a sequence of $\sigma-$algebras $\mathcal{F}_t \in \mathcal{F}$ such that

$$\mathcal{F}_s \in \mathcal{F}_t \quad \text{for} \quad t \geq s.$$

The sequence of $\sigma-$algebras $\{\mathcal{F}_t\}$ is called a filtration (or history, or information available to the moment $t = 0, 1, ...$). Typically, one may consider $\mathcal{F}_t$ as a collection of all (measurable) events generated by an underlying observed stochastic processes X_t, $t = 0, 1, ...$. For the latter case we use notation:

$$\mathcal{F}_t = \sigma\{X_0, ..., X_t\}.$$

A random process Y_t is **adapted** to $\mathcal{F}_t$ if Y_t is $\mathcal{F}_t$-measurable for any $t = 0, 1, ...$.

Definition 2. *A process $\mathcal{M}_t$, $0 \leq t < \infty$ is called a* **martingale** *with respect to filtration $\mathcal{F}_t$ if*

$$E(|\mathcal{M}_t|) < \infty, \tag{3.9}$$

and

$$E(\mathcal{M}_t|\mathcal{F}_{t-1}) = \mathcal{M}_{t-1}. \tag{3.10}$$

Definition 3 (Stopping time). *A stopping time τ is a nonnegative-integer-valued random variable such the event $\{\tau \leq t\} \in \mathcal{F}_t$ for any $t \geq 0$.*

Further it will be convenient to use notation $\mathrm{M}(\mathcal{F}_t, \mathbb{P})$ for the class of all martingales on the filtered probability space $(\Omega, \mathcal{F}_t, \mathcal{F}, \mathbb{P})$.

Proposition 3.4.1. *If $\mathcal{M}_t \in \mathrm{M}(\mathcal{F}_t, \mathbb{P})$ then*

$$E(\mathcal{M}_t) = E(\mathcal{M}_0), \tag{3.11}$$

$$E(\mathcal{M}_s) = E(E(\mathcal{M}_t|\mathcal{F}_s)) = E(\mathcal{M}_t) = \ldots = E(\mathcal{M}_0).$$

Theorem 3.4.2 (Martingale Stopping Theorem). *If $\mathcal{M}_t \in \mathrm{M}(\mathcal{F}_t, \mathbb{P})$, $t \geq 0$, then for any stopping time τ*

$$X_t := \mathcal{M}_{\min(\tau,t)} \in \mathrm{M}(\mathcal{F}_t, \mathbb{P})$$

and hence for any fixed $t \geq 0$

$$E(\mathcal{M}_{\min(\tau,t)}) = E(\mathcal{M}_0). \tag{3.12}$$

Further we describe some results from the papers of Novikov (1990) and Novikov and Kordzakhia (2007) on martingales and stopping times for first order autoregressive processes (AR(1)).

We define an AR(1) sequence as a solution of the recursive equation

$$Y_t = \beta Y_{t-1} + \eta_t, \quad t = 1, 2, ..., \quad Y_0 = y, \tag{3.13}$$

where η_t is a sequence of independent identically distributed random variables (*innovation*) and β are nonrandom constants, $0 < \beta < 1$.

Note that the equation for the EWMA statistic denoted by Z_t coincides with Equation (3.13) for Y_t if β is replaced by $1 - \lambda$ and the independent random variables η_t are replaced by $\lambda \xi_t$.

The solution of Equation (3.13) has the following representation for $t = 1, 2, ...,$

$$Y_t = \beta^t y + \sum_{k=0}^{t-1} \beta^k \eta_{t-k} \,. \tag{3.14}$$

Here we describe some martingales related to AR(1) sequences in the case when the innovation η_t has a distribution with a light right tail (see Propositions 3.4.4 and 3.4.5 below). In Theorem 3.4.10 we prove a martingale identity (the analog of classical Wald's identity for random walks) and show how to use it to obtain bounds for the expectation of the first passage time

$$\tau_a = \inf\{t > 0 : Y_t > a\}, \quad a \geq y,$$

where we assume $\inf\{\emptyset\} = \infty$ and so $\tau_a = \infty$ on the set $\{\sup_{t \geq 0} Y_t < a\}$.

Below we always consider martingales with respect to the natural filtration $\mathcal{F}_t = \sigma\{Y_0, Y_1, ..., Y_t\}$.

First, consider a martingale $\mathcal{M}_t$ of the form

$$\mathcal{M}_t = \beta^{vt} q_v(Y_t), \tag{3.15}$$

where a deterministic function $q_v(y)$ depends on a parameter v, the variable y takes values from the domain D of Y_t. Note that, typically, $D = (-\infty, \infty)$. Under the assumption that $\mathcal{M}_t$ has a finite expectation, by definition of martingales

$$E[\mathcal{M}_t|\mathcal{F}_{t-1}] = \mathcal{M}_{t-1} \quad - a.s.$$

which with Equation (3.15) is equivalent to the equation

$$\beta^{vt} E[q_v(\beta Y_{t-1} + \eta_t)|\mathcal{F}_{t-1}] = \beta^{v(t-1)} q_v(Y_{t-1}) \quad - a.s.$$

Here η_t is independent of $\mathcal{F}_{t-1}$ and Y_{t-1} may take any value from the domain D of Y_t. Therefore, the function $q_v(y)$ should be a solution of the equation

$$Eq_v(\beta y + \eta_1) = \beta^{-v} q_v(y), \quad y \in D. \tag{3.16}$$

Similar, if a martingale $\mathcal{M}_t$ has the form

$$\mathcal{M}_t = Q(Y_t) - t, \tag{3.17}$$

where $Q(y)$ is a deterministic function, then we obtain another equation

$$EQ(\beta y + \eta_1) = Q(y) + 1, \quad y \in D. \tag{3.18}$$

Martingales of the form (3.15) and (3.17) have already been discussed in Novikov (1990); Novikov and Ergashev (1993) under the assumptions

$$Ee^{u\eta_1} < \infty \quad \text{for} \quad 0 \leq u < \infty \tag{3.19}$$

and

$$E|\eta_1| < \infty. \tag{3.20}$$

Here we will also use the assumption (3.19) but will consider a relax condition (3.20) assuming only the existence of the logarithmic moment $E\log(1+|\eta_1|)$ (see Proposition 3.4.4 and 3.4.5 below) or moments of order $\delta > 0$ (see Proposition 3.4.6).

Denote the cumulant function of η_1 as follows

$$\psi(u) = \log Ee^{u\eta_1}, \quad 0 \leq u < \infty.$$

It is well known that if $\psi(u)$ is finite then it is a convex differentiable function for $u > 0$ (see e.g. Borovkov, 1973), $\psi(0) = 0$. In view of Equation (3.14) we have for any $u \in [0, \infty)$

$$Ee^{uY_t} = \exp\{\beta^t y + \sum_{k=0}^{t-1} \psi(\beta^k u)\}.$$

If $E\eta_t = m$ is finite then $\psi(u) = mu + o(u)$ as $u \to 0$. This fact implies that the partial sums $\sum_{k=0}^{t} \psi(\beta^k u)$ converge to a finite limit, say, $\phi(u)$ for any $u \geq 0$ as $t \to \infty$:

$$\sum_{k=0}^{t} \psi(\beta^k u) \to \phi(u) = \sum_{k=0}^{\infty} \psi(\beta^k u). \tag{3.21}$$

Note that under assumption (3.19) we may have[1] $E(\eta_t^-) = \infty$ or, equivalently, $\psi'(0) = -\infty$. Under the latter condition there exists $u_0 > 0$ such that $\psi(u) < 0$ for $u \in (0, u_0)$. It implies $\psi(\beta^k u) < 0$ for all $u \in (0, u_0)$ and for all $u \geq u_0$ and $k > \log(u/u_0)/\log(1/\beta)$. Therefore, the series $\sum_{k=0}^{\infty} \psi(\beta^k u)$ converges to a finite value or diverges to $-\infty$ for all $u > 0$. We now show that this series converges under the Vervaat condition

$$E \log(1 + |\eta_1|) < \infty. \tag{3.22}$$

Lemma 3.4.3. *Let conditions* (3.19) *and* (3.22) *hold. Then the function*

$$\phi(u) = \sum_{k=0}^{\infty} \psi(\beta^k u)$$

is differentiable for $u > 0$,

$$\phi(u) = \lim_{t\to\infty} \log Ee^{uY_t} = \log Ee^{u\Theta}, \quad \Theta \stackrel{d}{=} \sum_{k=0}^{\infty} \beta^k \eta_{k+1}$$

and

$$\phi(u) = \phi(\beta u) + \psi(u)\,, \; 0 \leq u < \infty. \tag{3.23}$$

Proof. Accordingly to the results of Vervaat (1979), under condition (3.22) the process Y_t converges in distribution as $t \to \infty$:

$$Y_t \xrightarrow{d} \Theta\,, \tag{3.24}$$

[1] $x^- = \max(-x, 0), x^+ = \max(x, 0)$

where Θ is a finite random variable. By Equation (3.14) we obtain

$$Y_t = \beta^{t}y + \sum_{k=0}^{t-1} \beta^k \eta_{k+1} \xrightarrow{d} \Theta,$$

where

$$\Theta \stackrel{d}{=} \sum_{k=0}^{\infty} \beta^k \eta_{k+1}. \tag{3.25}$$

Note that since $\sum_{k=0}^{\infty} \beta^k \eta_{k+1} = \beta \sum_{k=1}^{\infty} \beta^{k-1} \eta_{k+1} + \eta_1$ we have

$$\Theta \stackrel{d}{=} \beta\Theta + \eta_1, \tag{3.26}$$

where in the right-hand side (RHS) random variables Θ and η_1 are assumed to be independent.

Replacing η_k by η_k^-, we obtain from Equation (3.24) and (3.25) as $t \to \infty$

$$\sum_{k=0}^{t} \beta^k \eta_{k+1}^- \xrightarrow{d} \sum_{k=0}^{\infty} \beta^k \eta_{k+1}^-.$$

By the Lebesgue dominated convergence theorem this implies for any $u > 0$

$$Ee^{-u \sum_{k=0}^{t} \beta^k \eta_{k+1}^-} \to Ee^{-u \sum_{k=1}^{\infty} \beta^k \eta_k^-} > 0.$$

Since $\sum_{k=0}^{t-1} \beta^k \eta_{t-k} \stackrel{d}{=} \sum_{k=1}^{t} \beta^k \eta_k \geq - \sum_{k=1}^{t} \beta^k \eta_k^-$ we obtain that

$$\liminf_{t\to\infty} \sum_{k=0}^{t} \psi(\beta^k u) = \liminf_{t\to\infty} \log Ee^{u \sum_{k=0}^{t} \beta^k \eta_k} \geq \log Ee^{-u \sum_{k=0}^{\infty} \beta^k \eta_k^-} > -\infty$$

and, hence, Equation (3.21) holds under condition (3.22).

Thus, we have shown that for any $u \in [0, \infty)$ the function

$$\phi(u) = \lim_{t\to\infty} Ee^{uY_t} = \log Ee^{u\Theta}$$

is finite. It is a differentiable function because it is a cumulant function of a finite random variable. Furthermore, in view of Equation (3.26) $\phi(u)$ satisfies Equation (3.23). $\square$

If $E\eta_t = m$ is finite then $E\Theta = \frac{mu}{1-\beta}$ and as $u \to 0$

$$\phi(u) = \frac{mu}{1-\beta} + o(u). \tag{3.27}$$

We shall often use the following condition

$$\int_1^\infty e^{uy-\phi(u)}u^{v-1}du < \infty \tag{3.28}$$

with some real v (to be specified) and any $y \in D$.

A simple sufficient condition for validity of condition (3.28) is

$$\lim_{u\to\infty} \frac{\phi(u)}{u} = \infty. \tag{3.29}$$

Note that if the innovation η_t is bounded from above then Equation (3.29) does not hold (see Lemma 3.4.3 below) but still condition (3.28) could hold for $y \in D$.

Set

$$N_v(y) = \int_0^\infty e^{uy-\phi(u)}u^{v-1}du. \tag{3.30}$$

Proposition 3.4.4. *Let $v > 0$, conditions (3.19), (3.22) and condition (3.28) with $v > 0$ hold. Then*

$$\beta^{vt}N_v(Y_t) \quad \textit{is a martingale.} \tag{3.31}$$

Proof. The function $N_v(y)$ is finite due the imposed conditions. Now we are going to verify that Equation (3.16) holds for $q_v(y) = N_v(y)$. By Fubinni's theorem we have

$$EN_v(\beta y + \eta_1) = \int_0^\infty \exp\{u\beta y + \psi(u) - \phi(u)\}u^{v-1}du$$

and (with use of Equation (3.23))

$$= \int_0^\infty e^{u\beta y-\phi(\beta u)}u^{v-1}du = \beta^{-v}\int_0^\infty e^{uy-\phi(u)}u^{v-1}du$$
$$= \beta^{-v}N_v(y).$$

Thus, we have shown that the function $N_v(y)$ is a solution of Equation (3.16) and hence the process $\beta^{vt}N_v(Y_t)$ is a martingale. □

Set

$$H(y) = \frac{1}{\log(1/\beta)} \int_0^\infty u^{-1}(e^{uy} - 1)e^{-\phi(u)}du.$$

Proposition 3.4.5. *Let conditions* (3.19), (3.22) *and* (3.28) *with* $v = 0$ *hold. Then*

$$H(Y_t) - t \quad \text{is a martingale.} \tag{3.32}$$

Proof. The function $H(y)$ is finite due the imposed conditions. By Fubinni's theorem

$$E[H(\beta y + \eta_1)] = \frac{1}{\log(1/\beta)} \int_0^\infty u^{-1}(e^{u\beta y + \psi(u)} - 1)e^{-\phi(u)}du, \tag{3.33}$$

where the RHS is finite if for some $u_0 > 0$

$$\int_{u_0}^\infty u^{-1}|e^{u\beta y + \psi(u)} - 1|e^{-\phi(u)}du < \infty \tag{3.34}$$

and

$$\int_0^{u_0} u^{-1}|e^{u\beta y + \psi(u)} - 1|e^{-\phi(u)}du < \infty. \tag{3.35}$$

The integral $\int_{u_0}^\infty$ in Equation (3.34) is finite for any $u_0 > 0$ due to Equation (3.28) with $v = 0$; the integral $\int_0^{u_0}$ in Equation (3.35) is finite if and only if

$$\int_0^{u_0} u^{-1}|1 - e^{\psi(u)}|du < \infty. \tag{3.36}$$

If $E\eta_t = m$ is finite then in view of Equation (3.27), obviously, condition (3.36) holds. If $E\eta_t^- = \infty$ then there exists $u_0 > 0$ such that $\phi(u) \le 0$ for $u \in [0, u_0]$. This fact together with Lemma 3.4.3 lead to the following estimates

$$\begin{aligned}
\int_0^{u_0} u^{-1}|1 - e^{\psi(u)}|du &= \int_0^{u_0} u^{-1}(1 - e^{\psi(u)})du = E\int_0^{u_0} u^{-1}(1 - e^{u\eta_1})du \\
&\le EI\{\eta_1 \le 0\}\int_0^{-\eta_1 u_0} u^{-1}(1 - e^{-u})du \\
&\le P\{-1 \le \eta_1 \le 0\}\int_0^{u_0} u^{-1}(1 - e^{-u})du \\
&\quad + P\{\eta_1 < -1\}\Big(\int_0^{u_0} u^{-1}(1 - e^{-u})du + EI\{\eta_1 < -1\}\int_{u_0}^{-\eta_1 u_0} u^{-1}du\Big) \\
&\le \text{const.} + E\log(1 + |\eta_1|).
\end{aligned}$$

Thus, we have shown that integrals in Equation (3.34) and (3.35) are finite and therefore the RHS of Equation (3.33) can be now written (with use of Equation (3.23)) as follows:

$$E[H(\beta y+\eta_t)] = H(y) + \int_0^\infty u^{-1}\frac{(e^{u\beta y-\phi(\beta u)} - e^{uy-\phi(u)})}{\log(1/\beta)}du.$$

To satisfy Equation (3.18) we need to show that the last integral equals to 1. In fact, this type of integrals is well known; it is called the Frullani's integral. It equals really to 1 if it is absolutely convergent and the function $\phi(u)$ is continuous (the latter is, of course, true). The absolute convergence can be checked similarly to the verifications of Equation (3.34) and (3.35) as it had been done above. □

Proposition 3.4.6. *Let condition* (3.22) *hold,*

$$P\{\eta_1 = N\} = p = 1 - P\{\eta_1 \le 0\} > 0. \tag{3.37}$$

Then as $u \to \infty$

$$\phi(u) = \frac{u\,N}{1-\beta} + o(u).$$

Proof. We have

$$\psi(u) = \log(Ee^{u\eta_1}) = u\,N - g(u) \tag{3.38}$$

with

$$g(u) = -\log[EI\{\eta_1 \le 0\}\exp\{u(\eta_1 - N)\} + p] \ge 0.$$

Note

$$g'(u) = -\frac{EI\{\eta_1 \le 0\}(\eta_1 - N)\exp\{u(\eta_1 - N)\}}{EI\{\eta_1 \le 0\}\exp\{u(\eta_1 - N)\} + p} \to 0 \quad \text{as} \quad u \to \infty.$$

In view of Equation (3.38) we have

$$\phi(u) = \sum_{k=0}^{\infty}\psi(\beta^k u) = \frac{uN}{1-\beta} - \Delta(u), \tag{3.39}$$

where

$$\Delta(u) := \sum_{k=0}^{\infty} g(\beta^k u) \ge 0, \quad 0 \le u < \infty.$$

The function $\Delta(u)$ is differentiable and concave since by Lemma 3.4.3 the function $\phi(u)$ is differentiable and convex. Obviously,

$$\Delta(u) = \Delta(\beta u) + g(u), \quad 0 \le u < \infty. \tag{3.40}$$

From concavity of $\Delta(u)$ it follows that the derivative $\Delta'^{(u)}$ is non-increasing. It certainly has a lower-bound (as $\Delta(u) \geq 0$) and therefore there exists a finite limit

$$\lim_{u \to \infty} \Delta'(u) = A.$$

Applying L'Hospitale's rule we have

$$A = \lim_{u \to \infty} \frac{\Delta(u)}{u} = \lim_{u \to \infty} \frac{\Delta(\beta u) + g(u)}{u} = \beta A + \lim_{u \to \infty} \frac{g(u)}{u}.$$

As $\lim_{u \to \infty} \frac{g(u)}{u} = \lim_{u \to \infty} g'(u) = 0$, we obtain that $A = \beta A$ and therefore $A = 0$.
Proposition 3.4.6 is proved. □

In Novikov and Kordzakhia (2007), it is shown that under some mild assumptions on the left tail of η_1, the distribution of τ_a is exponentially bounded.

Theorem 3.4.7. *Let $a > 0$, conditions* (3.41) *hold with some $\delta \in (0, 1)$ and*

$$P\{\eta_1 > a(1 - \beta)\} > 0. \tag{3.41}$$

Then there exists $\alpha > 0$ such that

$$Ee^{\alpha \tau_a} < \infty.$$

Theorem 3.4.8. *Let $a > 0$, conditions* (3.22) *and* (3.41) *hold. Then*

$$E(\tau_a) < \infty.$$

Proof. We use Proposition 3.4.6 and the fact that the first passage time $\tilde{\tau}_a$ of AR(1) with a truncated innovation is greater than the original one (without truncation).

Let $N > a(1 - \beta) > 0$ such that $P\{\eta_1 > N\} > 0$ and let $\tilde{\tau}_a$ is the stopping time for AR(1) processes generated by the innovation with the property

$$P\{\tilde{\eta}_1 = N\} = p = 1 - P\{\tilde{\eta}_1 \leq 0\} > 0.$$

Then by Lemma 3.4.3 the corresponding cumulant function $\phi(u) = \frac{uN}{1-\beta} + o(u)$ as $u \to \infty$. By Proposition 3.4.5

$$E\tilde{\tau}_a \wedge t = \frac{1}{\log(1/\beta)} \int_0^\infty u^{-1}(Ee^{u\tilde{Y}_{\tilde{\tau}_a \wedge t}} - e^{uy})e^{-\phi(u)}du,$$

where $\tilde{Y}_{\tilde{\tau}_a \wedge t} \leq \beta a + N$. It implies

$$E\tilde{\tau}_a \wedge t \leq \frac{1}{\log(1/\beta)} \int_0^\infty u^{-1}(e^{u(\beta a+N)} - e^{uy})e^{-\frac{N}{1-\beta}(u)+o(u)}du = \text{const.} < \infty.$$

Lemma 3.4.9. (Fatou's Lemma) *Let $f_1, f_2, ...$ is a sequence of non-negative measurable functions defined on a measurable space (S, Σ, μ). Then the integral over S of the lower limit of f_t is less than or equal to the lower limit of the integral over S off f_t*

$$\int_S \liminf_{t\to\infty} f_t d\mu \leq \liminf_{t\to\infty} \int_S f_t d\mu.$$

To apply the lemma 3.4.9. We note that $\lim_{t\to\infty} (\tilde{\tau}_a \wedge t) = \lim_{t\to\infty} \tilde{\tau}_a$ and

$$E\tilde{\tau}_a \leq E \lim_{t\to\infty} (\tilde{\tau}_a \wedge t) \leq \lim_{t\to\infty} E(\tilde{\tau}_a \wedge t) \leq \text{const.} < \infty$$

The proof is completed. □

Now we prove a general martingale identity which can be used for derivation of bounds and asymptotic for $E(\tau_a)$.

Theorem 3.4.10. *Let conditions* (3.22) *and* (3.28) *with $v = 0$ hold. If $E(\tau_a) < \infty$ then*

$$E\tau_a = \frac{1}{\log(1/\beta)} \int_0^\infty u^{-1}(Ee^{uY_{\tau_a}} - e^{uy})e^{-\phi(u)}du. \tag{3.42}$$

Proof. By Proposition 3.4.5 and the Martingale Stopping Theorem (3.4.2) we have for any $t = 1, 2, ...,$

$$E(\tau_a \wedge t) = EH(Y_{\tau_a \wedge t}) - H(y) = \frac{1}{\log(1/\beta)} E \int_0^\infty u^{-1}(e^{uY_{\tau_a \wedge t}} - e^{uy})e^{-\phi(u)}du.$$

Since $\lim_{t\to\infty} E(\tau_a \wedge t) = E(\tau_a)$ and

$$\int_0^\infty u^{-1}(Ee^{uY_{\tau_a \wedge t}} - e^{uy})e^{-\phi(u)}du = EI(\tau_a \leq t) \int_0^\infty u^{-1}(e^{uY_{\tau_a}} - e^{uy})e^{-\phi(u)}du \tag{3.43}$$

$$+ EI(\tau_a > t) \int_0^\infty u^{-1}(e^{uY_t} - e^{uy})e^{-\phi(u)}du, \tag{3.44}$$

where the first term in RHS is a monotonic function of t. Therefore, we need only to show a convergence as $t \to \infty$ to zero for the latter integral term. Note that $Y_t \leq a$ on the set

$\{\tau_a > t\}$ and this implies the following upper-bound

$$EI(\tau_a > t) \int_0^\infty u^{-1}(e^{uY_t} - e^{uy})e^{-\phi(u)}du$$

$$\le P(\tau_a > t) \int_0^\infty u^{-1}(e^{ua} - e^{uy})e^{-\phi(u)}du \to 0$$

because $P(\tau_a > t) \to 0$ and the last integral is finite due to the imposed conditions.

To show that the lower-bound for the second integral term in the RHS of Equation (3.43) tends also to zero as $t \to \infty$, we note

$$EI(\tau_a > t) \int_0^\infty u^{-1}(e^{uy} - e^{uY_t})e^{-\phi(u)}du \le P(\tau_a > t) \int_0^\infty u^{-1}(e^{uy} - 1)e^{-\phi(u)}du$$
$$+ EI(\tau_a > t) \int_0^\infty u^{-1}(1 - e^{-uY_t^-})e^{-\phi(u)}du.$$

The first integral in the RHS converges, obviously, to zero.

Note that in view of Equation (3.14)

$$Y_t^- \le \beta^t x^- + \sum_{k=0}^{t-1} \beta^k \eta_{t-k}^- := Z_t, \quad t = 1, 2, ...,$$

where Z_t is a AR(1) process with innovation sequence η_k^-, $Z_0 = x^-$.

Since

$$\int_0^\infty u^{-1}(1 - e^{-uY_t^-})e^{-\phi(u)}du \le \int_1^\infty u^{-1}e^{-\phi(u)}du + \int_0^1 u^{-1}(1 - e^{-uZ_t})e^{-\phi(u)}du,$$

where $\int_1^\infty u^{-1}e^{-\phi(u)}du$ is finite by condition (3.28) with $v = 0$, we need only to verify that

$$\lim_{t\to\infty} EI(\tau_a > t) \int_0^1 u^{-1}(1 - e^{-uZ_t})du = 0.$$

Note

$$EI(\tau_a > t) \int_0^1 u^{-1}(1 - e^{-uZ_t})du = EI(\tau_a > t) \int_0^{Z_t} u^{-1}(1 - e^{-u})du$$

$$\le P(\tau_a > t, Z_t \le 1) \int_0^1 u^{-1}(1 - e^{-u})du + EI(\tau_a > t, Z_t > 1) \int_1^{Z_t+1} u^{-1}(1 - e^{-u})du.$$

The first term in the RHS here tends, obviously, to zero. For the second term we have

$$EI(\tau_a > t) \int_1^{Z_t+1} u^{-1}(1 - e^{-u})du \le EI(\tau_a > t) \log(Z_t + 1).$$

Since

$$\log(Z_t + 1) \leq \log(1 + y^- + \sum_{k=1}^{t} \eta_k^-) \leq \log(1 + y^-) + \sum_{k=1}^{t} \log(1 + \eta_k^-)$$

and by the Wald identity $E(\sum_{k=1}^{\tau_a} \log(1 + \eta_k^-)) = E(\log(1 + \eta_1^-))E\tau_a < \infty$, we obtain due to the Lebesgue dominated convergence theorem that as $t \to \infty$

$$EI(\tau_a > t)\log(Z_t + 1) \leq EI(\tau_a > t)\log(Z_{\tau_a} + 1) \to 0.$$

Now combining all estimates obtained above we complete the proof. $\square$

In Theorem 3.4.8 can be used for obtaining bounds and asymptotic of $E\tau_a$. Since the overshoot

$$\chi_a = Y_{\tau_a} - a$$

is always nonnegative, under the assumption of Theorem 3.4.10 we obtain the following lower-bound

$$E\tau_a \geq \frac{1}{\log(1/\beta)} \int_0^\infty u^{-1}(e^{ua} - e^{uy})e^{-\phi(u)}du. \tag{3.45}$$

The upper-bound can be obtained with use of truncation (if the original innovation is not bounded) of the innovation sequence $\{\eta_k\}$ from above by a constant, say, H. For the latter case, noting that $\tilde{Y}_{\tau_a} \leq \beta a + H$, we obtain that

$$E\tau_a \leq E\tilde{\tau}_a \leq \frac{1}{\log(1/\beta)} \int_0^\infty u^{-1}(e^{u(\beta a + H)} - e^{uy})e^{-\tilde{\phi}(u)}du,$$

where the function $\tilde{\phi}(u)$ is the corresponding cumulant of the limit distribution of the AR(1) sequence $\tilde{Y}_t$ with the truncated innovation.

We will demonstrate below that the lower-bound Equation (3.45) produces a quite reasonable numerical approximation when the parameter a is not large (see on Figure 4.1(a)). In any case, this lower-bound could be use to control accuracy of Monte Carlo approximations.

One also can use the martingale closed-form relation (3.42) to construct a control variable to reduce the variance of Monte Carlo simulation for τ_a. Indeed, the expectation of the of the random variable

$$\zeta = \tau_a - Q(Y_{\tau_a}),$$

where

$$Q(y) := \frac{1}{\log(1/\beta)} \int_0^\infty u^{-1}(e^{uy} - e^{uY_0})e^{-\phi(u)}du$$

is known because accordingly to Theorem 3.4.8 and 3.4.10

$$E(\zeta) = E(\tau_a) - \frac{1}{\log(1/\beta)} \int_0^\infty u^{-1}E(e^{uY_{\tau_a}} - e^{uY_0})e^{-\phi(u)}du = 0.$$

Besides, we can expect that $\zeta = \tau_a - Q(Y_{\tau_a})$ and τ_a are very correlated random variables because the term $Q(Y_{\tau_a}) = Q(a + \chi_H)$ where the overshoot χ_H seems to be not dependent on τ_a. At least we know that for random walks (i.e when $\beta = 1$) the overshoot χ_H and τ_a are asymptotically independent for $a \to \infty$ (see Siegmund, 1985). We plan to develop a methodology based on this consideration in a future research.

3.5 Discussion

Even though MCA, IE and MC are the popular standard approaches for determining the performance of EWMA charts, they face various limitations of calculation.

The MC is simple to program but it is usually very time consuming to run and difficult to use for optimisation. The MCA is considered as the most popular technique (Lucas and Saccucci, 1990). It is based on the use of matrix inversions for approximating a Markov Chain. However, there are no analytical results on the accuracy of this method. The IE is the most advanced method but it requires intensive programming or special software to implement, even for a Gaussian distribution.

In this chapter, new analytical approximations based on the martingale technique have been developed for evaluating the performance of an EWMA chart. This technique can be applied to Gaussian distribution and to non-Gaussian light-tailed distributions, all it will be shown in more detail in Chapters 4 and 5. Additionally, the closed-form formulas can be used not only to evaluate the performance of EWMA but also to find optimal design parameters of EWMA chart.

Chapter 4

EWMA Chart for Detection of a Change in a Gaussian Distribution

In this chapter, using the martingale technique, we derive analytical lower-bounds, closed-form expressions and approximations for characteristics of EWMA charts for the case of Gaussian distribution. Besides, we develop a new method for obtaining approximations for characteristics of EWMA which combines the martingale technique with Monte Carlo simulations.

This chapter has four sections. In Section 4.1 we derive closed-form formulas and analytical lower-bounds for ARL and AD for both one-sided and two-sided Gaussian EWMA charts. In Section 4.2 a correction term is introduced to the formulas to make the approximation more accurate. In Section 4.3 we present numerical results and make comparison of our results with other standard methods, such as the Markov Chain Approach, Integral Equations and Monte Carlo simulations. Some examples with tables of optimal parameter values for one-sided and two-sided Gaussian EWMA designs are presented in Section 4.4.

4.1 The Derivation of Closed-form Formulas for Gaussian EWMA

4.1.1 Expectation of first passage times for one-sided Gaussian EWMA

Here we use notations from Chapter 2 and apply some results from Chapter 3 to EWMA statistics.

Recall that the EWMA procedure is defined by the following recursion:

$$Z_t = (1-\lambda)Z_{t-1} + \lambda\xi_t, \quad t = 1, 2, ..., \quad Z_0 = z, \tag{4.1}$$

where $\lambda \in (0, 1)$ is a weighting factor.

Set

$$E_\theta e^{u\lambda\xi_t} := e^{u\psi_\theta(u)} < \infty \quad \text{for any } u \in (0, \infty),$$

where $\theta = \infty$ ("in-control" case for all time) or $\theta = 1$ ("out-of-control" case for all time) and let

$$\varphi_\theta(u) = \sum_{k=0}^{\infty} \psi_\theta((1-\lambda)^k u). \tag{4.2}$$

Now make the further assumptions that ξ_t has the Gaussian distribution, $\xi_t \sim N(\alpha, \sigma^2)$. Then we obtain

$$\varphi_\theta(u) = u\alpha + \frac{\lambda u^2\sigma^2}{(4-2\lambda)}.$$

Indeed, since for the case under consideration

$$\psi_\theta(u) = \alpha\lambda u + \frac{\lambda^2}{2}\sigma^2 u^2$$

then

$$\begin{aligned}
\varphi_\theta(u) &= \sum_{k=0}^{\infty} \psi_\theta((1-\lambda)^k u) \\
&= \sum_{k=0}^{\infty} [\alpha((1-\lambda)^k \lambda u) + \frac{1}{2}\sigma^2((1-\lambda)^{2k}\lambda^2 u^2)] \\
&= \frac{\alpha\lambda u}{1-(1-\lambda)} + \frac{1}{2}(\frac{\lambda^2\sigma^2 u^2}{1-(1-\lambda)^2}) \\
&= u\alpha + \frac{\lambda u^2\sigma^2}{(4-2\lambda)}.
\end{aligned}$$

It is usually supposed (without loss of generality) that $E_\infty(\xi_t) = \alpha_0 = 0$. Here we also use this assumption and therefore, we have

$$\varphi_\infty(u) = \frac{\lambda u^2\sigma^2}{(4-2\lambda)}, \quad \varphi_1(u) = u\alpha + \frac{\lambda u^2\sigma^2}{(4-2\lambda)}.$$

Consider now the case of one-sided Gaussian EWMA which is based on the first passage time

$$\tau_H = \inf\{t > 0 : Z_t > H\},$$

assuming that $H > z = 0$. Then, accordingly to Theorems 3.4.8 and 3.4.10 (conditions of these theorems hold trivially) the following identity hold:

$$E_\theta(\tau_H) = \frac{1}{|\ln(1-\lambda)|} \int_0^\infty u^{-1} E_\theta(e^{uZ_{\tau_H}} - 1)e^{-\varphi_\theta(u)} du. \tag{4.3}$$

In particular, it means that

$$ARL_1 = E_\infty(\tau_H) = \frac{1}{|\ln(1-\lambda)|} \int_0^\infty u^{-1} E_\infty(e^{uZ_{\tau_H}} - 1)e^{-\frac{\lambda u^2}{4-2\lambda}} du$$

and

$$AD_1 = E_1(\tau_H) = \frac{1}{|\ln(1-\lambda)|} \int_0^\infty u^{-1} E_1(e^{uZ_{\tau_H}} - 1)e^{-u\alpha-\frac{\lambda u^2}{4-2\lambda}} du.$$

Note that the overshoot

$$\chi_H = Z_{\tau_H} - H$$

is always a nonnegative random variable. Using the inequality

$$E_\theta(e^{uZ_{\tau_H}}) \geq e^{uH},$$

we obtain the following exact lower-bounds

$$ARL_1 \geq \frac{1}{|\ln(1-\lambda)|} \int_0^\infty u^{-1} (e^{uH} - 1)e^{-\frac{\lambda u^2}{4-2\lambda}} du \tag{4.4}$$

and

$$AD_1 \geq \frac{1}{|\ln(1-\lambda)|} \int_0^\infty u^{-1} (e^{uH} - 1)e^{-u\alpha-\frac{\lambda u^2}{4-2\lambda}} du. \tag{4.5}$$

Consider now the case of two-sided Gaussian EWMA chart, assuming again that $\alpha_0 = 0$ and $Z_0 = 0$, (these are typical assumptions). For two-sided case the alarm time is usually consider in the following form

$$\gamma_H = \inf\{t > 0 : |Z_t| > H\}.$$

To find proper closed-form relations for $ARL_2 = E_\infty(\gamma_H)$ and $AD_2 = E_1(\gamma_H)$ we note that accordingly to Proposition 3.4.5 the process

$$H(Z_t) - t = \frac{1}{|\ln(1-\lambda)|}\int_0^\infty u^{-1}(e^{uZ_t} - 1)e^{-u\alpha - \frac{\lambda u^2}{4-2\lambda}}du - t \in \mathrm{M}(\mathcal{F}_t, \mathbb{P}_\theta),$$

i.e. $H(Z_t) - t$ is a martingale with respect to the filtration $\mathcal{F}_t$ and the distribution $\mathbb{P}_\theta$.

Recall that we assume that $\alpha = 0$ in case $\theta = \infty$. Due to the symmetry of Gaussian distribution with zero mean (recall, we would like to find $ARL_2 = E_\infty(\gamma_H)$) we note the process

$$H(-Z_t) - t = \frac{1}{|\ln(1-\lambda)|}\int_0^\infty u^{-1}(e^{-uZ_t} - 1)e^{-\frac{\lambda u^2}{4-2\lambda}}du - t \in \mathrm{M}(\mathcal{F}_t, \mathbb{P}_\infty).$$

Since any linear combinations of martingales is a martingale we can conclude that also the process

$$\frac{H(Z_t) + H(-Z_t) - 2t}{2} = \frac{1}{|\ln(1-\lambda)|}\int_0^\infty u^{-1}(\cosh(uZ_t) - 1)e^{-\frac{\lambda u^2}{4-2\lambda}}du - t \in \mathrm{M}(\mathcal{F}_t, \mathbb{P}_\infty), \tag{4.6}$$

where $\cosh(x) = \frac{e^x + e^{-x}}{2}$.

Set

$$Q(z) = \frac{1}{|\ln(1-\lambda)|}\int_0^\infty u^{-1}(\cosh(uz) - 1)e^{-\frac{\lambda u^2}{4-2\lambda}}du.$$

Then in view of Equation (4.6) by the Martingale Stopping Theorem for any $t > 0$,

$$E_\infty(\min(\gamma_H, t)) = E_\infty Q(Z_{\min(\gamma_H, t)}).$$

Now due to Lemma Fatou and mononicity (or, repeating considerations which lead to Theorems 3.4.8 and 3.4.10) we can conclude that

$$\begin{aligned} ARL_2 &= E_\infty(\gamma_H) = E_\infty Q(Z_{\gamma_H}) \\ &= \frac{1}{|\ln(1-\lambda)|}\int_0^\infty u^{-1}E_\infty(\cosh(uZ_{\gamma_H}) - 1)e^{-\frac{\lambda u^2}{4-2\lambda}}du. \end{aligned} \tag{4.7}$$

Since

$$|Z_{\gamma_H}| > H,$$

we obtain the following lower-bound

$$ARL_2 \geq Q(H) = \frac{1}{|\ln(1-\lambda)|}\int_0^\infty u^{-1}(\cosh(uH)-1)e^{-\frac{\lambda u^2}{4-2\lambda}}du. \tag{4.8}$$

To find an approximation for AD_2 we still can use the fact that the processes

$$H(Z_t) - t = \frac{1}{|\ln(1-\lambda)|}\int_0^\infty u^{-1}(e^{uZ_t}-1)e^{-u\alpha-\frac{\lambda u^2}{4-2\lambda}}du - t \in \mathrm{M}(\mathcal{F}_t, \mathbb{P}_1),$$

which implies that

$$AD_2 = E_1(\gamma_H) = \frac{1}{|\ln(1-\lambda)|}\int_0^\infty u^{-1}E_1(e^{uZ_{\gamma_H}}-1)e^{-u\alpha-\frac{\lambda u^2}{4-2\lambda}}du.$$

Using the symbol of an indicator function $I\{A\}$ we obtain

$$E_1(e^{uZ_{\gamma_H}}-1) = E_1[I\{Z_{\gamma_H} > H\}(e^{uZ_{\gamma_H}}-1)] + E_1[I\{Z_{\gamma_H} < -H\}(e^{uZ_{\gamma_H}}-1)].$$

To study AD_2 we may assume (without loss of generality) the shift is positive, i.e. $E_1(\xi_1) = \alpha > 0$. Then the process Z_t has a positive drift up (to the upper level H) and therefore the probability of crossing the upper level $P(Z_{\gamma_H} > H)$ should be close to one and correspondingly the probability of crossing the lower-bound $P(Z_{\gamma_H} < -H)$ should be close to zero (at least, for not very small α). Beside, neglecting by the overshoots we have

$$I\{Z_{\gamma_H} > H\}(e^{uZ_{\gamma_H}}-1) \geq I\{Z_{\gamma_H} > H\}(e^{uH}-1)$$

and

$$I\{Z_{\gamma_H} < -H\}|(e^{uZ_{\gamma_H}}-1)| \leq I\{Z_{\gamma_H} < -H\}.$$

Note that the factor $(e^{uH}-1)$ could be very high for not too small uH. The presented considerations motivate the following approximation:

$$AD_2 \approx \frac{1}{|\ln(1-\lambda)|}\int_0^\infty u^{-1}E_1(e^{u(H+\chi_H)}-1)e^{-u\alpha-\frac{\lambda u^2}{4-2\lambda}}du,$$

where $\chi_H = Z_{\gamma_H} - H$ is the overshoot over the upper level H. Since the overshoot is nonnegative we get the approximate lower-bound

$$AD_2 \gtrapprox \frac{1}{|\ln(1-\lambda)|}\int_0^\infty u^{-1}(e^{uH}-1)e^{-u\alpha-\frac{\lambda u^2}{4-2\lambda}}du. \tag{4.9}$$

The closed-form formulas presented above can be expressed in the terms of standard special functions, for instance, *Hypergeometric and Gamma functions,* and easily computed numerically using the Mathematica® package. Alternatively, one could use numerical integration methods (Simpson rules, Gauss-Legendre quadratures etc.) which are also available in Mathematica®.

The closed-form formula for ARL_1 can be expressed in terms of the function

$$f[z,\rho] := \int_0^\infty \frac{(e^{uz}-1)}{u} e^{-\rho u^2} du, \quad \rho = \frac{\lambda}{4-2\lambda}. \tag{4.10}$$

With use of Mathematica® we found the following expressions for this function in terms of hypergeometric functions

$$\begin{aligned} f[z,\rho] &= \frac{1}{4(\rho)}(2\rho\pi \mathrm{Erfi}(\frac{z}{2\sqrt{\rho}}) + z^2 \mathrm{HypergeometricPDF}[\{1,1\},\{\frac{3}{2},2\},\frac{z^2}{4\rho}]) \\ &= \frac{1}{4(\rho)}(2\rho\pi \mathrm{Erfi}(\frac{z}{2\sqrt{\rho}}) + z^2\, {}_1F_1(\frac{3}{2},2;\frac{z^2}{4\rho}), \end{aligned} \tag{4.11}$$

where **Erfi(z)** is the imaginary error function **Erf(iz)/i** and ${}_1F_1$ is a hypergeometric function.

The closed-form formula of AD_1 of the one-sided EWMA depends on the sizes of the change in the parameter α. Let

$$G[\alpha,z,\rho] = \frac{1}{\ln|1-\lambda|}\int_0^\infty \frac{(e^{uz}-1)}{u} e^{-\alpha u-\rho u^2} du. \tag{4.12}$$

The integral term in Equation (4.12) can be rewritten as follows:

$$\begin{aligned} \int_0^\infty \frac{e^{uz}-1}{u} e^{-\alpha u-\rho u^2} du &= \int_0^\infty e^{-\rho u^2} \frac{(e^{uz}e^{-u\alpha}-1+1-e^{-u\alpha})}{u} du \\ &= \int_0^\infty \frac{e^{u(z-\alpha)}-1}{u} e^{-\rho u^2} du - \int_0^\infty \frac{(e^{-\alpha u}-1)}{u} e^{-\rho u^2} du \end{aligned}$$

and so

$$G[\alpha,z,\rho] = f[z-\alpha,\rho] - f[-\alpha,\rho].$$

The closed-form formula for ARL_2 can be expressed in terms of the function

$$ff[z,\rho] := \int_0^\infty u^{-1}(\cosh(uz)-1)e^{-\rho u^2} du = \frac{1}{2}(f[z,\rho]-f[-z,\rho]). \tag{4.13}$$

Below we find another representing for the function $ff[z,\rho]$ in terms of simple power series. We use the Taylor series

$$\cosh(x) - 1 = \sum_{n=1}^{\infty} \frac{x^{2n}}{(2n)!}.$$

Then substituting $x = uz$, we obtain

$$\frac{\cosh(uz) - 1}{u} = \frac{1}{u} \sum_{n=1}^{\infty} \frac{(uz)^{2n}}{(2n)!} = \sum_{n=1}^{\infty} \frac{z^{2n}}{(2n)!} u^{2n-1}.$$

Hence, the function $ff[z, \rho]$ can be written in the form:

$$\begin{aligned} ff[z, \rho] &= \int_0^{\infty} \sum_{n=1}^{\infty} \frac{z^{2n}}{(2n)!} u^{2n-1} e^{-\rho u^2} du \\ &= \sum_{n=1}^{\infty} \frac{z^{2n}}{(2n)!} \int_0^{\infty} u^{2n-1} e^{-\rho u^2} du. \end{aligned} \tag{4.14}$$

The last integral can be transformed into a Gamma function by the substitution $t = \rho u^2$:

$$\int_0^{\infty} u^{2n-1} e^{-\rho u^2} du = \frac{1}{2\rho^n} \int_0^{\infty} t^{n-1} e^{-t} dt = \frac{1}{2\rho^n} \Gamma(n) = \frac{(n-1)!}{2\rho^n}.$$

Consequently, we obtain

$$ff[z, \rho] = \sum_{n=1}^{\infty} \frac{z^{2n}(n-1)!}{(2n)!\rho^n}.$$

4.2 Corrected Approximation

The closed-form formulas include the overshoot $\chi_H = Z_{\tau_H} - H$ whose distribution is, generally speaking, unknown. Neglecting by overshoot, we obtained the explicit lower-bounds for ARL_1 and AD_1 which are easy to calculate but it may be not an accurate approximation. Here we provide some considerations about how to find a correction term to improve the accuracy.

Recall that the process EWMA is governed by the following equation,

$$Z_t = (1 - \lambda) Z_{t-1} + \lambda \xi_t, \quad Z_0 = z.$$

Setting $\tilde{Z}_t = \frac{Z_t}{\lambda}$ we obtain

$$\tilde{Z}_t = (1 - \lambda) \tilde{Z}_{t-1} + \xi_t, \quad \tilde{Z}_0 = z/\lambda \tag{4.15}$$

and, consequently, the alarm time is

$$\tau_H = \inf\{t > 0 : \tilde{Z}_t > \frac{H}{\lambda}\}.$$

If λ is close to zero (that is typical for applications in detection small changes) then $\tilde{Z}_t$ is close to a *random walk* $S_t = \sum_{k=1}^{t} \xi_k$, with i.i.d. increments ξ_t.

Based on these considerations we suggest to use the following approximation

$$\tilde{Z}_{\tau_H} \simeq \frac{H}{\lambda} + \tilde{\chi}_H,$$

where $\tilde{\chi}_H$ is the overshoot of the random walk $S_t = \sum_{k=1}^{t} \xi_k$ over level $\frac{H}{\lambda}$. This leads to the following approximation

$$Z_{\tau_H} = \lambda \tilde{Z}_{\tau_H} \simeq \lambda(\frac{H}{\lambda} + \tilde{\chi}_H) = H + \lambda \tilde{\chi}_H$$

and, correspondingly, for $u \geq 0$

$$E_\theta e^{uZ_{\tau_H}} \simeq E_\theta e^{u(H+\lambda\tilde{\chi}_H)} = e^{uH} E_\theta e^{\lambda u \tilde{\chi}_H}.$$

Note that due Jensen inequality $E_\theta e^{\lambda u \tilde{\chi}_H} \geq e^{\lambda u E_\theta \tilde{\chi}_H}$ and also due to Taylor expansion $e^x = 1 + x + o(x)$, $x \to 0$, we have for $\lambda \to 0$,

$$E_\theta e^{\lambda u \tilde{\chi}_H} = 1 + \lambda u E_\theta \tilde{\chi}_H + o(\lambda) = e^{\lambda u E_\theta \tilde{\chi}_H} + o(\lambda).$$

This leads to the next approximation

$$E_\theta e^{uZ_{\tau_H}} \simeq e^{u(H+\lambda E_\theta \tilde{\chi}_H)}.$$

For the cases when the level $\frac{H}{\lambda}$ is high (i.e. for small λ) we can use the following well-known result from theory of random walks (see e.g. Siegmund, 1985).

Theorem 4.2.1. *Let* $S_t = \sum_{k=1}^{t} \xi_k$, $\tau_b = \inf\{t > 0 : S_t > b\}$, *the distribution of* ξ_t *be non-lattice,* $E\xi_k = a \geq 0$, $0 < E|\xi_k|^2 < \infty$. *Then there exists limit*

$$\lim_{b \to \infty} E(S_{\tau_b} - b) = C.$$

The analytical calculation of the constant C could be a hard problem. Note that the constant C depends on $\alpha = E(\xi_1)$. Under the assumption $E|\xi_k|^3 < \infty$ the constant C can be represented in the following form

$$C = \frac{E(S^2_{\tau_+})}{2E(S_{\tau_+})}, \tag{4.16}$$

where $\tau_+ = \inf\{t > 0 : S_t > 0\}$, S_{τ_+} is the so-called positive ladder variable of a random walk S_t. If ξ_t is nonnegative then $S_{\tau_+} = \xi_1$.

Remark 1. *If the distribution of ξ_t is lattice (e.g. Poisson and Bernoulli distribution) then the statement of Theorem* (4.2.1) *still holds but Equation* (4.16) *is different (see Lorden, 1971; Siegmund, 1985).*

The monograph of Siegmund (1985) and the papers of Lotov (1996); Chang and Peres (1997) contain a lot of results concerning properties of limiting distribution of $P(S_{\tau_b} - b > x)$ as $b \to \infty$ for the case of Gaussian random walk. In particular, it is know that if $\xi_t \sim N(0,1)$ then

$$C = -\frac{\zeta(1/2)}{\sqrt{2\pi}} = 0.5826..., \tag{4.17}$$

where $\zeta(x)$ is the Riemann zeta function (see details in Siegmund, 1985; Chang and Peres, 1997).

Summarising the above considerations we suggest for the case $\xi_t \sim N(0,1)$ to use the as the formula

$$E_\theta e^{uZ_{\tau_H}} \simeq e^{u(H+C\lambda)}, \tag{4.18}$$

where $C = 0.5826$ for $\alpha = 0$ or when α is close to zero.

The closed-form formulas for one-sided Gaussian EWMA:

$$ARL_1 = E_\infty(\tau_H) \simeq \frac{1}{|\ln(1-\lambda)|}\int_0^\infty u^{-1}E_\infty(e^{u(H+C\lambda)} - 1)e^{-\frac{\lambda u^2}{4-2\lambda}}du, \tag{4.19}$$

$$AD_1 = E_1(\tau_H) \simeq \frac{1}{|\ln(1-\lambda)|}\int_0^\infty u^{-1}E_1(e^{u(H+C\lambda)} - 1)e^{-u\alpha-\frac{\lambda u^2}{4-2\lambda}}du. \tag{4.20}$$

The closed-form formulas for two-sided Gaussian EWMA:

$$ARL_2 = E_\infty(\gamma_H) \simeq \frac{1}{|\ln(1-\lambda)|}\int_0^\infty u^{-1}E_\infty(\cosh(u(H+C\lambda)) - 1)e^{-\frac{\lambda u^2}{4-2\lambda}}du, \tag{4.21}$$

$$AD_2 = E_1(\gamma_H) \simeq \frac{1}{|\ln(1-\lambda)|}\int_0^\infty u^{-1}E_1(e^{u(H+C\lambda)} - 1)e^{-u\alpha-\frac{\lambda u^2}{4-2\lambda}}du. \tag{4.22}$$

Note that similar approximations are often used in many other problems of sequential analysis, Siegmund (see e.g. 1985). The theoretical justification of such approximation is a very hard problem and, of course, the value $C = 0.5826$ should be used only as a "first approximation". A more accurate approximation can be obtained with Monte Carlo simulations and fitting with the non-linear least-square methods. The first approximation of the overshoot produces usually is a good approximation for small λ. A comparison of numerical results with lower-bounds, MC and first approximation with $C = 0.5826$ for Gaussian distribution in case $\lambda = 0.01$ is presented in Table 4.1 for one-sided EWMA case and Table 4.2 for two-sided EWMA case.

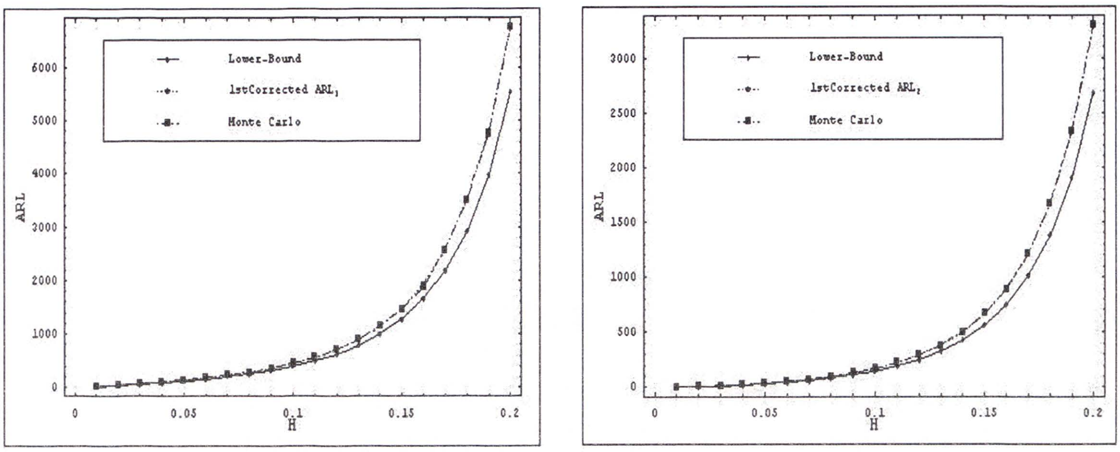
Table 4.1: Comparison of numerical results of ARL_1 between lower-bounds, first approximation and MC for one-sided Gaussian EWMA

H	Lower-bounds Eq. (4.4)	*First Approx.* (ARL_1)	*MC*	H	Lower-bounds Eq. (4.4)	*First Approx.* (ARL_1)	*MC*
.01	18.643	30.572	30.43±.02	.11	497.890	566.79	566.82±.18
.02	39.667	53.207	54.03±.03	.12	622.579	711.02	711.29±.22
.03	63.573	79.076	79.06±.04	.13	783.206	898.62	898.98±.28
.04	90.999	108.92	109.85±.05	.14	993.616	1146.83	1147.60±.36
.05	122.771	143.70	143.42±.06	.15	1274.05	1481.07	1482.73±.46
.06	159.961	184.67	185.47±.07	.16	1654.52	1939.40	1940.59±.60
.07	203.986	233.51	233.54±.08	.17	2180.26	2579.61	2576.85±.80
.08	256.735	292.47	291.21±.10	.18	2920.43	3490.93	3510.04±1.09
.09	320.757	364.59	363.38±.12	.19	3982.44	4813.22	4748.67±1.32
.10	399.536	454.08	456.97±.15	.20	5535.84	6769.30	6770.32±1.65

One could see that values of the first approximation are close to MC results for both one-sided and two-sided Gaussian EWMA chart. Plots of ARL_1 and ARL_2 are shown on Figure 4.1(a) and 4.1(b), respectively.

Table 4.2: Comparison of numerical results of ARL_2 between lower-bounds, first approximation and MC for two-sided Gaussian EWMA

H	Lower-bounds Eq. (4.8)	*First Approx.* (ARL_2)	*MC*	H	Lower-bounds Eq. (4.8)	*First Approx.* (ARL_2)	*MC*
.01	0.993	2.50	2.78±.006	.11	188.319	220.77	221.20±.06
.02	4.013	6.754	7.00±.002	.12	247.229	289.57	289.50±.08
.03	9.183	13.27	13.49±.003	.13	324.301	380.24	380.00±.11
.04	16.718	22.33	22.55±.006	.14	426.439	501.39	501.39±.15
.05	26.946	34.33	34.66±.009	.15	563.746	665.71	664.81±.20
.06	40.336	49.85	50.31±.01	.16	751.22	892.22	891.89±.26
.07	57.537	69.67	70.10±.02	.17	1011.46	1209.81	1208.41±.36
.08	79.445	94.86	95.24±.03	.18	1379.03	1663.09	1662.54±.50
.09	107.295	126.93	127.38±.03	.19	1907.63	2321.99	2326.09±.71
.10	142.793	167.93	168.61±.05	.20	2682.03	3297.95	3301.43±.94



(a) one-sided EWMA chart (b) two-sided EWMA chart

Figure 4.1: Comparison of lower-bounds, first approximations and simulations of ARL: Gaussian case

To approximate AD, we also use the value $C = 0.5826$ as a "first approximation" when λ is small (e.g. $\lambda = 0.01$ and 0.04). Some results of numerical calculation for the case of out-of-control parameter with $\alpha = 0.5$ are presented in Table 4.3 for one-sided case and Table 4.4 for two-sided case.

Table 4.3: Comparison of numerical results of AD_1 between lower-bounds, first approximation and MC for one-sided Gaussian EWMA

λ	H	Lower-bounds Eq. (4.5)	*First Approx.* (AD_1)	*MC*
	0.10	21.67	23.09	23.32±.003
0.01	0.15	34.52	36.12	36.41±.004
	0.20	49.20	51.06	51.27±.006
	0.10	5.020	6.34	6.62±.002
0.04	0.15	7.940	9.42	9.71±.002
	0.20	11.22	12.89	13.17±.003

Table 4.4: Comparison of numerical results of AD_2 between lower-bounds, first approximation and MC for two-sided Gaussian EWMA

λ	H	Lower-bounds Eq. (4.9)	*First Approx.* (AD_2)	*MC*
	0.10	21.67	23.09	23.37±.003
0.01	0.15	34.52	36.12	36.39±.004
	0.20	49.20	51.06	51.33±.006
	0.10	5.020	6.34	6.19±.001
0.04	0.15	7.940	9.42	9.56±.002
	0.20	11.22	12.89	13.15±.003

As mentioned earlier, to get a better accuracy results we will use a combination of MC and the analytical closed-form formulas. The algorithm for obtaining numerical charts of optimal set of parameters $\tilde{\lambda}$ and $\tilde{H}$ for approximating ARL and AD is discussed below.

4.2.1 Algorithm for Obtaining Numerical Approximations for ARL and AD

Recall that our goal is to find optimal parameters $(\tilde{\lambda}, \tilde{H})$ for EWMA under conditions

$$ARL \simeq T, \quad AD \to \min.$$

We have shown above that both ARL and AD for the one-sided and two-sided testing can be approximated with a function $Q_\theta(H + \lambda C)$ where $\theta = \infty$ (for "in-control" case) or $\theta = 1$ (for "out-of-control" case). The function $Q_\theta(z)$ when $\theta = \infty$ for the one-sided and two-sided case are given in Equation (4.19) and Equation (4.21), and when $\theta = 1$ for the one-sided and two-sided cases are given in Equation (4.20) and Equation (4.22). Now using the notation $Q_\theta(z)$ we can describe our algorithm (which could be used for Gaussian and non-Gaussian cases) for obtaining numerical approximation for ARL and AD.

The algorithm has four steps as follows:

1. *For given T and out-of-control parameter α use Equation (4.18) with the constant C from Equation (4.16) as an initial approximation. Find a "first approximation" (λ^*, H^*) for the optimal parameters that minimise AD^* (e.g. use Mathematica® command "FindRoot" as inverse function of $\lambda = \lambda(H)$ and "FindMinimun" as a minimal AD^*, see code on Appendix B).*

2. *Simulate EWMA with (λ^*, H^*) to find the corresponding "true" ARL.*

3. *Find an improved approximation for the constant $\tilde{C}$ from the equation $Q_\theta(H+\lambda\tilde{C}) = ARL$ with "ARL" from Step 2 (again use "FindRoot" for as inverse function $\tilde{C} = \tilde{C}(\lambda^*, H^*)$).*

4. *Repeat Step 1 with the value of the constant $\tilde{C}$ obtained from Step 3. We then obtain a "second approximation" $(\tilde{\lambda}, \tilde{H})$ for the set of optimal parameters under the constraint $ARL = T$.*

We illustrate here the steps discussed above for the following example. For example, we would like to design a two-sided Gaussian EWMA chart to detection a parameter change

$\alpha = 0.5$ for given $\alpha_0 = 0$ and $T = 500$. Then, according to algorithm described above we have the following steps

1. For given $T = 500$ and $\alpha = 0.5$, an initial approximation for (λ^*, H^*) is found from the closed-form formulas with the constant $C = 0.5826$ for the case of Gaussian observations (see code in Appendix B). The values obtained for the "first approximations" are: $(\lambda^* = 0.047025, H^* = 0.403263)$ and the minimal AD_2 is $AD_2^* = 28.4774$.

2. A simulation is carried out (with 10^6 repeats) for the pair of parameter values $(\lambda^* = 0.047025, H^* = 0.403263)$ from Step 1. From this simulation we obtain the "true" value $ARL_2 = 510.395$ corresponding to these parameter values.

3. The information from Steps 1 and 2 can be used to find an approximate value for the constant $\tilde{C}$ corresponding to the value $ARL_2 = 510.395$ from the equation $Q_\theta(H + \lambda\tilde{C}) = ARL_2$. The new value obtained is $\tilde{C} = 0.610832$.

4. Step 1 is repeated to find a second approximation for $(\tilde{\lambda}, \tilde{H})$ with the constant $\tilde{C} = 0.610832$ under the constraint $T = 500$. Then the values (0.047025,0.401954) are obtained as the "second approximation".

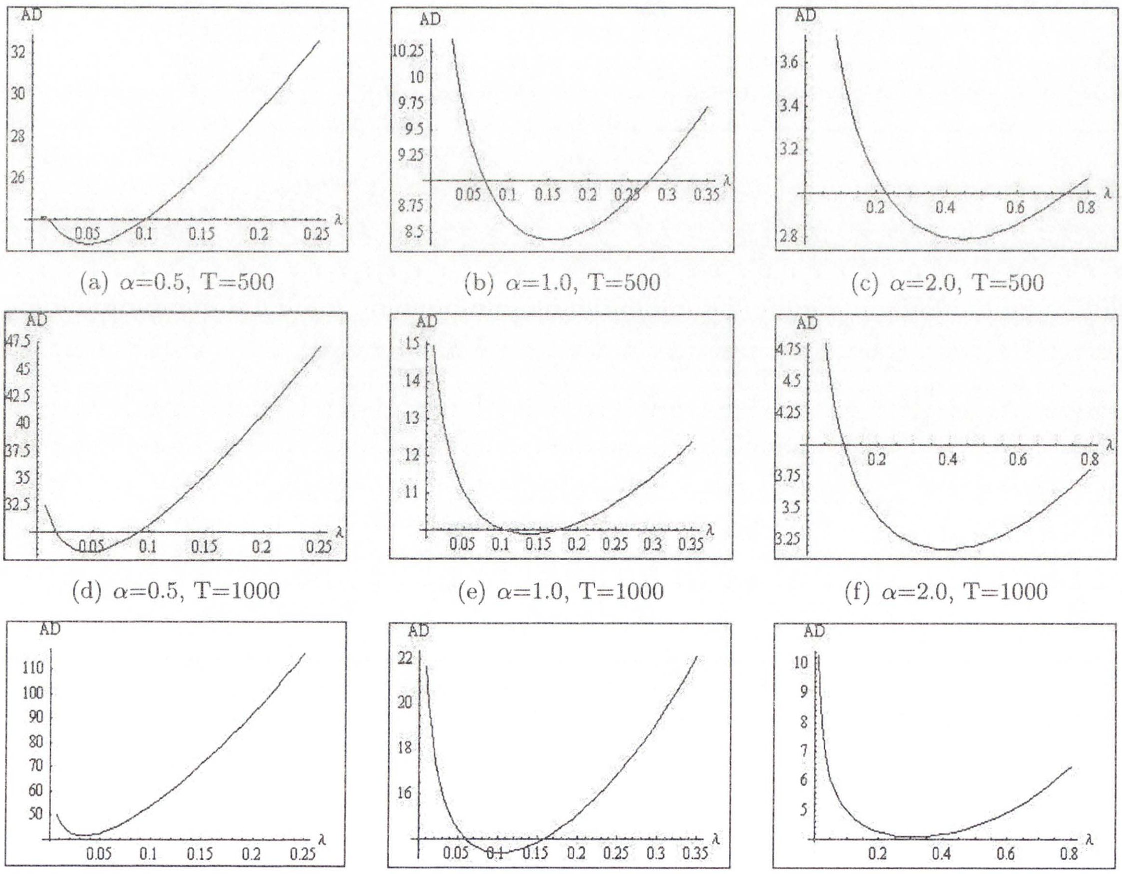
Finally, one could check the accuracy of the approximation with simulations. The simulation results obtained are $AD_2 = 28.7215$ and $ARL_2 = 500.922$. The percentage difference between the approximations and the simulations for the case under consideration is 0.85, less than 1%.

4.3 Numerical Results

4.3.1 Comparison of analytical approximations with simulations and other methodologies

Here we compare results obtained by our approach with those ones from the Markov Chain Approach, the Integral Equations approach and the Monte Carlo simulations. For the Markov Chain Approach we used the results of (Crowder (1987); Lucas and Saccucci (1990)), and for the Integral Equations approach the results of (Brook and Evans, 1972; Srivastava and Wu, 1997). We have carried out our own Monte Carlo simulations. Table 4.5 gives a comparison of the numerical results for the lower-bound from Equation (4.8), the closed-form formula from Equation (4.7) and the approximations of Crowder (1987) and Lucas and Saccucci (1990), and from the Monte Carlo simulations. The numbers in the table are percentage differences between values from the different methods. It can be seen that the percentage differences between our approximations and the Monte Carlo simulations do not exceed three percent (which seems to be reasonable for most applications). To find the constant C for our method we have always used our approximation combined with the Monte Carlo simulation. This corresponds to our suggestion of the value for the constant C for the case of a Gaussian distribution in Equation (4.16). For comparison, when $\lambda = 0.01$, $L = 2.0$, the obtained value for the constant $C = 0.583$ from Equation (4.16) compared with the value of $C = 0.5826$ for the case of a random walk Equation (4.17).

Table 4.6 shows a comparison of the optimal parameters $\widetilde{\lambda}$, $\widetilde{H}$ and AD_2^* for EWMA charts obtained from our approach with the optimal parameter values obtained by Lucas and Saccucci (1990) and Srivastava and Wu (1997). One of the main advantage of our approach is that we easily produces a curve for approximated values of AD and then a minimum point of AD can be found visually or numerically. These curves of approximated AD_2 for $\alpha = 0.5, 1.0$ and 2.0 and $T = 500, 1000$ and 5000 are shown in Figure 4.2.



(a) α=0.5, T=500 (b) α=1.0, T=500 (c) α=2.0, T=500

(d) α=0.5, T=1000 (e) α=1.0, T=1000 (f) α=2.0, T=1000

(g) α=0.5, T=5000 (h) α=1.0, T=5000 (i) α=2.0, T=5000

Figure 4.2: AD_2 for different magnitudes of change: two-sided Gaussian case

Table 4.5: Comparison of ARL_2 results from our closed-form formulas with other methods for two-sided Gaussian EWMA

λ	L	C	Lower-Bounds martingale using Eq. (4.8)	Closed-form by using Eq. (4.7)	Lucas& Saccucci 1990	Crowder 1987	Monte Carlo 10^6 trials	Percent Difference
			(1)	(2)	(3)	(4)	(5)	(2) and (5)
	1.0		59.28	**71.67**	-	60.11	71.9[0.06][a]	-0.32
0.01	2.0	0.583	447.91	**526.98**	-	453.13	527.02[0.49]	-0.01
	3.0		4236.14	**5282.0**	-	-	5288.46[5.14]	-0.12
	1.0		19.56	**27.07**	-	26.43	27.37[0.02]	-1.11
	2.0		147.79	**196.46**	-	192.16	196.46[0.18]	0.00
0.03	2.437	0.589	363.0	**499.21**	500	486.9	499.33[0.48]	-0.02
	2.989		1357.79	**1999.31**	2000	1939.4	2000.00[1.98]	-0.03
	3.00		1397.76	**2061.38**	-	-	2062.34[2.04]	-0.05
	1.0		11.61	**17.65**	-	17.12	17.89[0.02]	-1.36
0.05	2.0	0.597	87.76	**127.33**	-	123.75	127.36[0.12]	-0.02
	2.615		321.05	**500.29**	500	483.34	499.45[0.49]	0.17
	3.0		830.02	**1381.99**	-	1379.35	1379.39[1.36]	0.19
	1.0		8.21	**13.45**	-	13.31	13.70[0.01]	-1.86
0.07	2.0	0.604	62.03	**96.78**	-	-	96.78[0.09]	0.00
	2.015		63.89	**99.81**	100	98.81	99.83[0.09]	-0.02
	3.0		586.66	**1080.97**	-	1065.81	1075.61[1.06]	0.50
	1.0		3.67	**10.18**	-	10.10	10.43[0.01]	-2.46
	2.0		27.70	**73.18**	-	72.58	73.20[0.07]	-0.03
0.10	3.0	0.613	404.09	**848.40**	-	-	841.95[0.83]	0.76
	3.058		471.80	**1006.31**	1000	998.65	998.01[0.99]	0.82
	3.283		888.47	**2018.41**	2000	1993.13	1994.56[1.98]	1.18

[a] the standard deviation

Table 4.6: Comparison of optimal parameter values and AD_2^* for two-sided Gaussian EWMA

α	Methods	$T=500$				$T=1000$			
		C	λ	L	AD_2^*	C	λ	L	AD_2^*
0.5	Martingales	0.611	0.047	2.59	28.48	0.536	0.041	2.839	33.98
	MCA [a]		0.05	2.616	28.7		0.04	2.817	34.3
	IE [b]		0.04	2.614	28.4		0.03	2.845	33.7
	MC		0.047	2.59	28.72±.06		0.041	2.839	34.56±.06
1.0	Martingales	0.62	0.135	2.885	9.91	0.6	0.135	3.102	9.88
	MCA		0.13	2.883	10.2		0.11	3.087	11.7
	IE		0.11	2.981	9.60		0.09	3.194	11.1
	MC		0.135	2.885	10.2±.02		0.135	3.102	10.18±.02
2.0	Martingales	0.686	0.93	3.055	3.18	0.669	0.351	3.254	3.56
	MCA		0.36	3.046	3.51		0.33	3.247	3.90
	IE		0.34	3.141	2.9		0.31	3.355	3.30
	MC		0.393	3.055	3.52±.005		0.351	3.254	3.91±.006
α			$T=2000$				$T=5000$		
0.5	Martingales	0.586	0.035	3.037	39.70	0.546	0.029	3.311	47.53
	MCA		0.03	3.029	40.1		0.03	3.299	47.7
	IE		0.03	3.061	39.3		0.02	3.328	47.1
	MC		0.035	3.037	40.08±.07		0.029	3.311	48.21±.11
1.0	Martingales	0.607	0.106	3.294	13.21	0.578	0.092	3.553	14.92
	MCA		0.11	3.30	13.2		0.09	3.538	15.2
	IE		0.08	3.394	12.6		0.07	3.643	14.7
	MC		0.106	3.294	13.21±.02		0.092	3.553	15.27±.03
2.0	Martingales	0.655	0.317	3.445	4.30	0.626	0.282	3.696	4.48
	MCA		0.30	3.475	4.29		0.27	3.682	4.81
	IE		0.28	3.462	3.70		0.25	3.805	4.30
	MC		0.317	3.445	4.30±.006		0.282	3.696	4.85±.01

[a] Markov Chain Approach by Lucas & Saccucci (1990)
[b] Integro Equations method by Srivastava & Wu (1997)

4.3.2 Comparison of the performance of EWMA with CUSUM developed from Monte Carlo methods: Gaussian case

A comparison of the performance of EWMA and CUSUM charts is shown in Table 4.7. This table contains numerical values for the corresponding simulation results (with 10^6 sample trajectories) for the set of optimal parameters ($\tilde{\lambda} = 0.0454, \tilde{H} = 0.3924$) and $\tilde{C} = 0.5721$ when $T = 1000$ and a parameter change $\alpha = 1$. The table also includes ARL and AD for CUSUM for simulations (with 10^6 trials) and the boundary $A = 4.68$. Both CUSUM and EWMA are optimized for $T = 1000$ and $\alpha = 1$. Of course, at this point $AD_1 = 11.52$, EWMA is inferior than CUSUM but for values α less than $\alpha = 1$, AD of EWMA is better than CUSUM. Figure 4.3 shows that the performance of EWMA is superior to CUSUM for small changes while CUSUM performs better than EWMA for moderate changes.

Table 4.7: Comparison Monte Carlo simulations of ARL_1 and AD_1 for one-sided CUSUM and EWMA charts: Gaussian case

α	CUSUM	EWMA
	($A = 4.68$)	($\tilde{\lambda} = 0.0454, \tilde{H} = 0.3924$)
0	1010.79±.32	1013.41±.32
0.1	773.88±.14	280.67±.08
0.2	199.83±.04	113.86±.03
0.3	90.99±.02	61.22±.014
0.4	52.12±.009	39.56±.008
0.5	*34.23±.006*	*28.59±.005*
0.6	24.30±.004	22.18±.004
0.7	18.39±.003	18.02±.003
0.8	14.49±.002	15.19±.002
0.9	11.76±.002	13.12±.002
1.0	9.74±.002	11.52±.001
1.5	4.90±.001	7.26±.0006
2.0	3.09±.0004	5.34±.0004
3.0	1.70±.0003	3.59±.0002

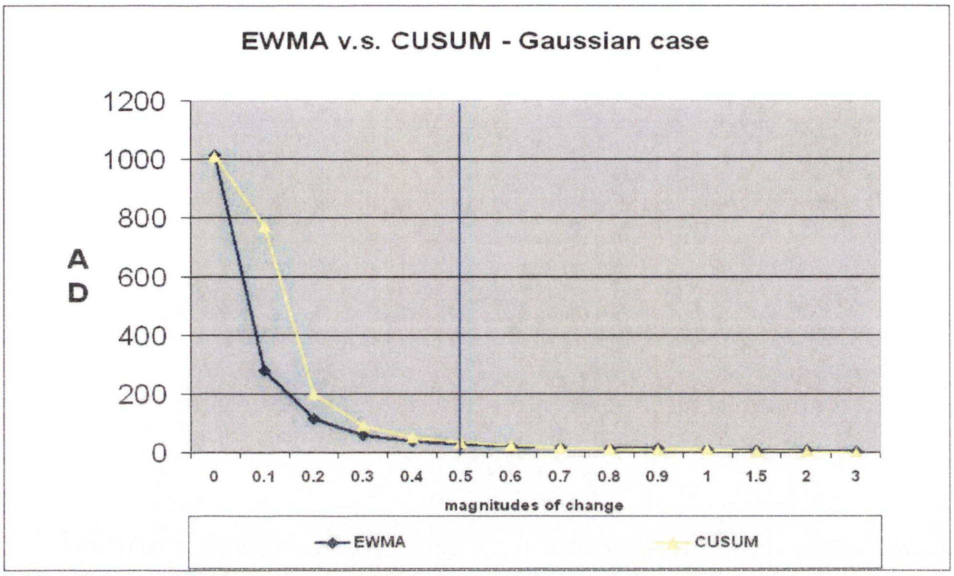


Figure 4.3: Comparison of AD_1 simulated by EWMA and by CUSUM for a one-sided Gaussian EWMA chart

4.4 Choices of Optimal Parameters of Gaussian EWMA Designs

In this section, we present numerical results for values of the optimal parameters $\widetilde{\lambda}$, $\widetilde{H}$ and AD^* of EWMA for both one-sided and two-sided cases in case $T = 500, 1000, 2000$ and 5000. Table 4.8 contains results for the one-sided Gaussian EWMA chart and in Table 4.9 for the two-sided Gaussian EWMA chart. The results are presented for a range of parameter changes α from 0.1 to 2.0. The λ values are in the range $0.01 < \lambda < 0.45$ and the desired in-control ARL value was set at $500, 1000, 2000$ and 5000.

Table 4.8: Optimal parameter values and AD_1^* of a one-sided Gaussian EWMA

T	α	$\widetilde{\lambda}$	$\widetilde{H}$	$\widetilde{C}$	AD_1^*	MC	T	$\widetilde{\lambda}$	$\widetilde{H}$	$\widetilde{C}$	AD_1^*	MC
	0.1	0.0095	0.1033	0.2469	132.13	138.8±.37		0.0014	0.0205	0.2517	151.92	156.07±.39
	0.2	0.0095	0.1082	0.1913	65.46	68.61±.20		0.0014	0.0205	0.2517	75.71	71.87±.14
	0.3	0.0217	0.2092	0.1587	43.35	46.59±.09		0.0180	0.2093	0.5748	56.73	57.23±.11
	0.4	0.0315	0.2744	0.1792	30.57	33.02±.06		0.0313	0.3065	0.5820	38.63	38.83±.07
	0.5	0.0454	0.383	0.234	22.85	24.93±.05		0.0454	0.3924	0.5721	28.19	28.62±.05
500	0.6	0.0682	0.4717	0.3115	17.79	19.37±.04	1000	0.0608	0.4722	0.5946	21.60	21.87±.04
	0.7	0.0877	0.5542	0.3699	14.30	15.49±.03		0.0775	0.5496	0.6048	17.15	17.47±.03
	0.8	0.1083	0.6308	0.4350	11.77	12.67±.02		0.0954	0.6257	0.6068	13.99	14.27±.02
	0.9	0.130	0.7056	0.4804	9.89	10.6±.01		0.1145	0.700	0.6176	11.66	11.99±.02
	1.0	0.153	0.777	0.5250	8.43	9.01±.01		0.135	0.7742	0.6227	9.88	10.17±.01
	1.5	0.283	1.1323	0.6450	4.48	4.82±.008		0.2505	1.1409	0.6186	5.14	5.52±.008
	2.0	0.446	1.523	0.709	2.79	3.15±.005		0.3925	1.5090	0.6840	3.17	3.52±.006
	0.1	0.0053	0.0996	0.6688	280.39	277.96±.06		0.0016	0.0503	0.4064	403.99	405.93±.74
	0.2	0.0079	0.1348	0.6454	122.30	121.72±.02		0.0077	0.1612	0.531	165.32	166.38±.3
	0.3	0.0174	0.2352	0.6214	71.20	71.05±.01		0.0149	0.2467	0.550	91.50	95.06±.16
	0.4	0.0280	0.3205	0.6082	47.14	47.12±.08		0.0236	0.3277	0.565	58.93	59.33±.1
	0.5	0.040	0.4010	0.6133	33.80	33.81±.06		0.0337	0.4067	0.564	41.49	41.91±.07
2000	0.6	0.0534	0.4780	0.6044	25.57	25.76±.04	5000	0.0452	0.4833	0.579	30.99	31.32±.05
	0.7	0.0681	0.5557	0.6093	20.11	20.31±.03		0.0578	0.5591	0.584	24.13	24.54±.04
	0.8	0.0840	0.6315	0.6020	16.28	16.55±.02		0.0717	0.6343	0.581	19.38	19.69±.04
	0.9	0.1010	0.7052	0.6080	13.48	13.79±.02		0.0865	0.7077	0.588	15.94	16.29±.03
	1.0	0.1191	0.7792	0.6047	11.37	11.96±.01		0.1024	0.7810	0.593	13.37	13.73±.03
	1.5	0.2238	1.1351	0.6422	5.82	6.12±.01		0.1952	1.1583	0.516	6.73	7.27±.02
	2.0	0.351	1.5010	0.6692	3.56	3.90±.005		0.3080	1.4990	0.636	4.08	4.45±.01

Table 4.9: Optimal parameter values and AD_2^* of a two-sided Gaussian EWMA

T	α	$\widetilde{\lambda}$	$\widetilde{H}$	$\widetilde{C}$	AD_2^*	MC	T	$\widetilde{\lambda}$	$\widetilde{H}$	$\widetilde{C}$	AD_2^*	MC
500	0.1	0.0039	0.0667	0.6151	216.73	218.61±.52	1000	0.0033	0.0722	0.6128	296.55	296.47±.65
	0.2	0.0114	0.1532	0.6290	98.26	98.15±.21		0.0099	0.1620	0.5919	126.07	127.09±.25
	0.3	0.0215	0.2399	0.6088	58.27	58.57±.12		0.0185	0.2473	0.5510	72.17	73.04±.14
	0.4	0.0334	0.3227	0.6107	39.22	39.40±.07		0.0288	0.3292	0.5408	47.51	48.30±.09
	0.5	0.0470	0.4020	0.6108	28.48	28.72±.06		0.0407	0.4089	0.5362	33.97	34.56±.06
	0.6	0.0621	0.4798	0.6085	21.76	21.95±.05		0.0539	0.4866	0.5428	25.66	26.16±.05
	0.7	0.0786	0.5557	0.6124	17.24	17.47±.04		0.0686	0.5618	0.5634	20.17	20.59±.04
	0.8	0.0964	0.6309	0.6109	14.05	14.27±.03		0.0844	0.6365	0.5727	16.32	16.70±.03
	0.9	0.1154	0.7038	0.6215	11.70	11.94±.03		0.1014	0.7096	0.5858	13.50	13.87±.02
	1.0	0.1354	0.7774	0.6197	9.91	10.2±.02		0.1194	0.7817	0.5994	11.39	11.73±.01
	1.5	0.2511	1.1362	0.6471	5.15	5.47±.02		0.2240	1.1378	0.6355	5.82	6.13±.009
	2.0	0.3929	1.5103	0.6859	3.17	3.52±.01		0.3512	1.5019	0.6690	3.56	3.91±.005
2000	0.1	0.0028	0.0785	0.5897	393.07	392.81±.78	5000	0.0024	0.0845	0.3914	542.85	543.8±1.04
	0.2	0.0085	0.1670	0.5104	156.91	155.82±.32		0.0069	0.1687	0.5174	201.17	201.77±.5
	0.3	0.0159	0.2492	0.5851	87.07	87.50±.15		0.0132	0.2504	0.5204	107.86	108.7±.26
	0.4	0.0249	0.3295	0.5809	56.25	56.66±.09		0.0208	0.3299	0.5309	68.27	69.06±.16
	0.5	0.0354	0.4077	0.5860	39.70	40.08±.07		0.0299	0.4077	0.5456	47.53	48.21±.11
	0.6	0.0472	0.4842	0.5950	29.71	30.04±.05		0.0402	0.4844	0.5503	35.21	35.79±.08
	0.7	0.0603	0.5597	0.5963	23.17	23.48±.04		0.0516	0.5602	0.5522	27.24	27.79±.06
	0.8	0.0746	0.6342	0.5993	18.64	18.97±.03		0.0641	0.6350	0.5562	21.77	22.23±.05
	0.9	0.0899	0.7079	0.6014	15.35	15.70±.02		0.0777	0.7085	0.5664	17.84	18.26±.04
	1.0	0.1063	0.7805	0.6066	13.21	13.21±.02		0.0922	0.7811	0.5777	14.92	15.27±.03
	1.5	0.2017	1.1377	0.6274	6.83	6.83±.01		0.1776	1.1394	0.6012	7.42	7.79±.02
	2.0	0.3176	1.4968	0.6554	4.30	4.30±.006		0.2818	1.4966	0.6255	4.47	4.84±.01

Chapter 5

An Adaptation and Expansion of the Martingale Technique to non-Gaussian Distributions

A common assumption used to determine statistical properties is that the individual observations are from a Gaussian distribution. However, in real applications the process characteristics are frequently non-Gaussian. A well-known example is data from processes where qualitative/attribute observations are made in quality control measurements. In these quality control measurements, the random variable ξ_t is the "number of defects produced" or "the fraction of defects per unit produced". If the number of defects is considered, then often the Poisson distribution is the appropriate distribution. Gan (1990) and Borror et al. (1998) have stressed the importance of the change-point problem for the Poisson distribution in their work on counting defects or nonconformities in manufacturing processes. Another important application has been discussed by Frisén (1992) who studied the number of incidences of adverse health events in patients under surveillance. Woodall (1997) has provided a listing of the control charts that have been used for attribute/count observations.

In this chapter, we present analytical and numerical approximations for the EWMA characteristics for observations from Poisson and Bernoulli distributions. We derive the closed-form formulas for ARL and AD using the martingale techniques. The numerical results obtained from the martingale formulas are compared with numerical results obtained from other approaches including Monte Carlo simulations. Typically, observations modelled by Poisson and Bernoulli distributions are tested in practice for one-sided alternatives, by this reason we do not examine two-sided charts.

5.1 EWMA Chart for Poisson Distribution

The performance of EWMA for a Poisson distribution has been studied by Gan (1990) and Borror et al. (1998). Gan (1990) studied the performance using the Markov Chain Approach and introduced three modified version of EWMA. Borror et al. (1998) studied the performance using the Markov Chain Approach and compared the results obtained with Monte Carlo simulations.

In this chapter using the martingale technique we derive closed-form represented for ARL and AD for Poisson EWMA chart. We assume that $\xi_1, \xi_2, ..., \xi_t$ are independent Poisson random variables that follow the change-point model:

$$\begin{aligned} \xi_t &\sim Poisson(\alpha_0), \quad t = 1, 2, ..., \theta - 1 \\ \xi_t &\sim Poisson(\alpha), \quad t = \theta, \theta + 1, ..., \qquad , \alpha > \alpha_0 \end{aligned}$$

5.1.1 Expectation of first passage times for one-sided Poisson EWMA

Our derivation of the approximations for ARL and AD is similar to the derivation of approximations for a Gaussian distribution given in Section 4.1. For non-Gaussian distributions, such as Poisson and Bernoulli, the functions $\varphi_\theta(u)$ defined in Equation (4.2), cannot be found in explicit form. However, the series given in Equation (5.2) is quickly convergent and so $E_\theta(\tau_H)$ in Equation (4.3) can be quickly calculated numerically with use of numerical integration (e.g. with the command *NIntegrate* in the Mathematica® software).

Recall, the EWMA statistic can be expressed in the recursive form

$$Z_t = (1 - \lambda) Z_{t-1} + \lambda \xi_t, \quad Z_0 = z,$$

and the stopping time

$$\tau_H = \inf\{t > 0 : Z_t > H\}.$$

Recall, we have the general closed-form presentation

$$E_\theta(\tau_H) = \frac{1}{|\ln(1 - \lambda)|} \int_0^\infty u^{-1} E_\theta(e^{uZ_{\tau_H}} - e^{uZ_0}) e^{-\varphi_\theta(u)} du, \tag{5.1}$$

where $\theta = \infty$ ("in-control" process) or $\theta = 1$ ("out-of-control" process) and the function

$$\varphi_\theta(u) = -u\alpha_0 + \sum_{k=0}^{\infty} \psi_\theta((1-\lambda)^k u). \tag{5.2}$$

If ξ_t are independent random variables with Poisson distribution with rate α then

$$\psi_\theta(u) = \log E_\theta(e^{u\lambda\xi_t}) = \log e^{\alpha(e^{u\lambda}-1)} = \alpha(e^{u\lambda}-1).$$

Without loss of generality, we assume that $E_\infty(\xi_t) = \alpha_0$. Then we obtain

$$\varphi_\infty(u) = -u\alpha_0 + \alpha_0 \sum_{k=0}^{\infty} (e^{u(1-\lambda)^k\lambda} - 1) \tag{5.3}$$

and

$$\varphi_1(u) = -u\alpha_0 + \alpha \sum_{k=0}^{\infty} (e^{u(1-\lambda)^k\lambda} - 1). \tag{5.4}$$

Substituting Equation (5.3) and (5.4) into Equation (5.1) we obtain closed-form expressions an one-sided ARL_1 and AD_1 Poisson EWMA. Without loss of generality, the initial value is usually assumed to be $Z_0 = 0$.

In particular, we obtain

$$ARL_1 = E_\infty(\tau_H) = \frac{1}{|\ln(1-\lambda)|} \int_0^\infty u^{-1} E_\infty(e^{uZ_{\tau_H}} - 1) e^{u\alpha_0 - \alpha_0 \sum\limits_{k=0}^{\infty} e^{u(1-\lambda)^k\lambda - 1}} du$$

and

$$AD_1 = E_1(\tau_H) = \frac{1}{|\ln(1-\lambda)|} \int_0^\infty u^{-1} E_1(e^{uZ_{\tau_H}} - 1) e^{u\alpha_0 - \alpha \sum\limits_{k=0}^{\infty} e^{u(1-\lambda)^k\lambda - 1}} du, \quad \alpha > \alpha_0.$$

Similar to the Gaussian case (Equation (4.19) and (4.20)) we recommend to use Equation (4.16) to find an initial approximation for the overshoot for the Poisson EWMA.

Note that if $\xi_t \sim Poisson(\alpha = 1)$, then

$$C = \frac{E(S^2_{\tau_+})}{2E(S_{\tau_+})} = \frac{E(\xi_t^2)}{2E(\xi_t)} = 1. \tag{5.5}$$

Then, we also use Equation (4.18) for obtaining closed-form formulas:

The closed-form formulas for one-sided Poisson EWMA:

$$ARL_1 = E_\infty(\tau_H) \simeq \frac{1}{|\ln(1-\lambda)|}\int_0^\infty u^{-1}E_\infty(e^{u(H+C\lambda)}-1)e^{u\alpha_0-\alpha_0\sum\limits_{k=0}^{\infty}e^{u(1-\lambda)^k\lambda-1}}du, \quad (5.6)$$

$$AD_1 = E_1(\tau_H) \simeq \frac{1}{|\ln(1-\lambda)|}\int_0^\infty u^{-1}E_1(e^{u(H+C\lambda)}-1)e^{u\alpha_0-\alpha\sum\limits_{k=0}^{\infty}e^{u(1-\lambda)^k\lambda-1}}du, \quad (5.7)$$

where $\alpha > \alpha_0$.

As discussed earlier, the overshoot is a nonnegative random variable $\chi_H = Z_{\tau_H} - H$. If the overshoot is neglected, we obtain the exact lower-bounds:

$$ARL_1 \geq \frac{1}{|\ln(1-\lambda)|}\int_0^\infty u^{-1}(e^{uH}-1)e^{u\alpha_0-\alpha_0\sum\limits_{k=0}^{\infty}e^{u(1-\lambda)^k\lambda-1}}du \quad (5.8)$$

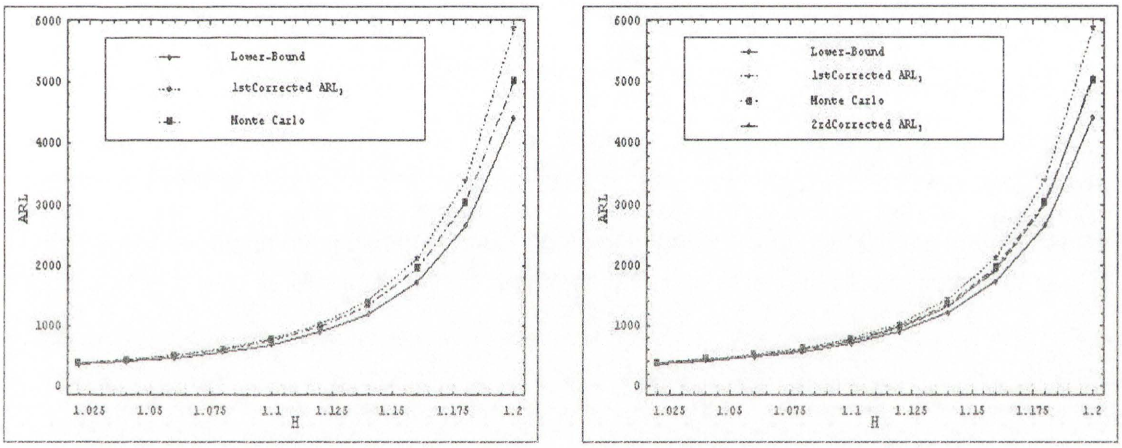
and

$$AD_1 \geq \frac{1}{|\ln(1-\lambda)|}\int_0^\infty u^{-1}(e^{uH}-1)e^{u\alpha_0-\alpha\sum\limits_{k=0}^{\infty}e^{u(1-\lambda)^k\lambda-1}}du. \quad (5.9)$$

On Table 5.1, we present the numerical results for ARL_1 when in-control parameter $\alpha_0 = 1$ and varieties of boundary H. Note that the "first approximation" (with an overshoot from Equation (5.5)) is not closed to the results of MC (see on Figure 5.1(a)). Our goal is to fit the analytical formula by using the nonlinear least-square method with MC results. For example, we use only three points (first, middle and last points) of MC results and fitting those points with analytical formula then we obtain the approximation of constant $C = 0.4912$. After improved an approximation of the overshoot, we obtain the "second approximation" which the numerical results get much better than the first approximation compared with the results of MC. In addition, the results are almost as good as with MC is demonstrated on Figure 5.1(b).

Table 5.1: Comparison of numerical results of ARL_1 between lower-bounds, first approximation, second approximation and MC for Poisson EWMA

H	Lower-bounds Eq. (5.8)	*First Approx.* (ARL_1)	*Second Approx.* $C = 0.4912$	*MC*
1.02	363.27	386.32	374.33	382.03±.05
1.04	412.61	442.85	427.02	436.64±.06
1.06	477.95	519.11	497.43	509.31±.08
1.08	567.88	626.30	595.32	609.08±.11
1.10	697.13	784.07	737.66	758.49±.15
1.12	892.14	1028.32	955.11	984.76±.22
1.14	1202.28	1427.70	1305.64	1344.01±.33
1.16	1724.06	2119.52	1903.86	1950.70±.51
1.18	2655.14	3391.66	2987.15	3034.67±.84
1.20	4419.93	5877.53	5071.40	5022.87±1.09



(a) before adjusted an overshoot (b) after adjusted an overshoot

Figure 5.1: Comparison of lower-bounds, approximations and simulations of ARL_1: Poisson case

5.2 Numerical Results

5.2.1 Comparison of analytical approximations with simulations and other methodologies

The numerical results for ARL and AD using the martingale approximations are compared with numerical results from the Markov Chain Approach and Monte Carlo simulations in Table 5.2. The percentage error between the approximations and the simulations is presented in the last column. Note that the errors decrease when L and λ are increased.

Table 5.2: Comparison of the numerical ARL_1 from the martingale approach with the numerical ARL_1 from the MCA and Monte Carlo simulations.

λ	L	C	Lower-Bound	Corrected Lower-Bound	ARL by MCA[a]	ARL by Monte Carlo[b]	Percentage Difference[c]
	1		198.37	216.82	256.66	235.65[0.27][d]	-7.99
0.01	2	0.3933	913.35	998.57	1035.57	1054.7[0.33]	-5.63
	3		6038.19	6847.70	6854.96	6841.09[2.16]	0.09
	1		63.25	80.76	86.25	86.12[0.03]	-6.56
0.03	2	0.6495	277.30	354.09	378.18	368.63[0.12]	-3.97
	3		1604.76	2237.54	2493.93	2235.61[0.70]	0.06
	1		36.71	53.04	55.02	54.65[0.02]	-2.93
0.05	2	0.7818	156.08	225.62	234.86	229.94[0.07]	-1.91
	3		834.69	1364.35	1455.8	1363.73[0.43]	0.06
	1		25.47	40.68	39.59	39.72[0.01]	2.4
0.07	2	0.8619	105.82	169.10	170.82	169.04[0.05]	0.04
	3		533.95	988.70	1030.92	988.77[0.31]	-0.007
	1		17.15	30.99	30.19	30.34[0.01]	2.14
0.10	2	0.9419	69.30	125.36	122.49	122.75[0.04]	2.12
	3		327.08	703.65	721.78	704.05[2.24]	-0.05

[a] n=500

[b] No. of rounds=10^6

[c] between martingale approach and Monte Carlo

[d] Standard deviations are shown in parentheses

5.2.2 Comparison of the performance of EWMA with CUSUM developed from Monte Carlo methods: Poisson case

We used the set of optimal parameter values found with the martingale approximations and have also compared the performance of EWMA and CUSUM using Monte Carlo simulations. Table 5.3 shows the numerical results for a optimal choice $\tilde{\lambda} = 0.0373$, $\tilde{H} = 0.3678$ and $\tilde{C} = 0.859$ with in-control parameter $\alpha_0 = 1$ and fixed $T = 1000$. Although CUSUM is usually considered as competitive with EWMA, the results in Table 5.3 show that CUSUM performance could be inferior to EWMA for moderate values of change in α as shown on Figure 5.2.

For given $T = 1000$ and in-control parameter $\alpha_0 = 1$, small changes can be detected better from an EWMA chart than from a CUSUM chart.

Table 5.3: Comparison Monte Carlo simulations of ARL_1 and AD_1 for one-sided CUSUM and EWMA charts: Poisson case

α	CUSUM	EWMA
	$(A = 4.11)$	$(\lambda = 0.0373, \tilde{H} = 0.3678)$
1.0	992.11±.31	983.08±.29
1.1	671.55±.13	259.58±.08
1.2	179.19±.04	110.76±.03
1.3	84.28±.02	62.29±.01
1.4	50.19±.01	41.74±.009
1.5	*34.05±.007*	*30.90±.006*
1.6	24.87±.005	24.34±.004
1.7	19.17±.004	20.01±.003
1.8	15.42±.003	17.07±.003
1.9	12.94±.003	14.77±.002
2.0	10.95±.002	13.10±.002
2.5	6.03±.001	8.37±.001
3.0	4.16±.0008	6.22±.001
4.0	2.59±.0005	4.22±.001

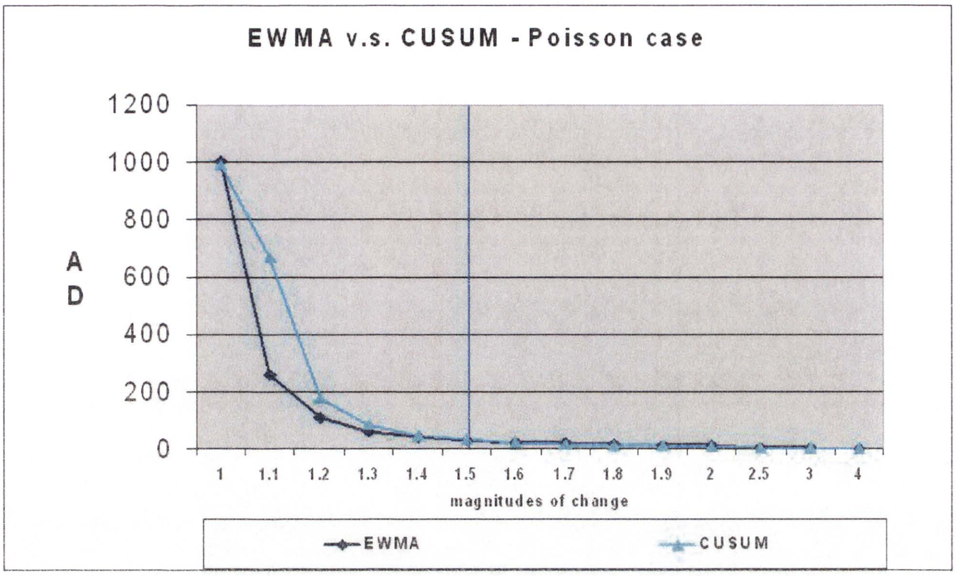


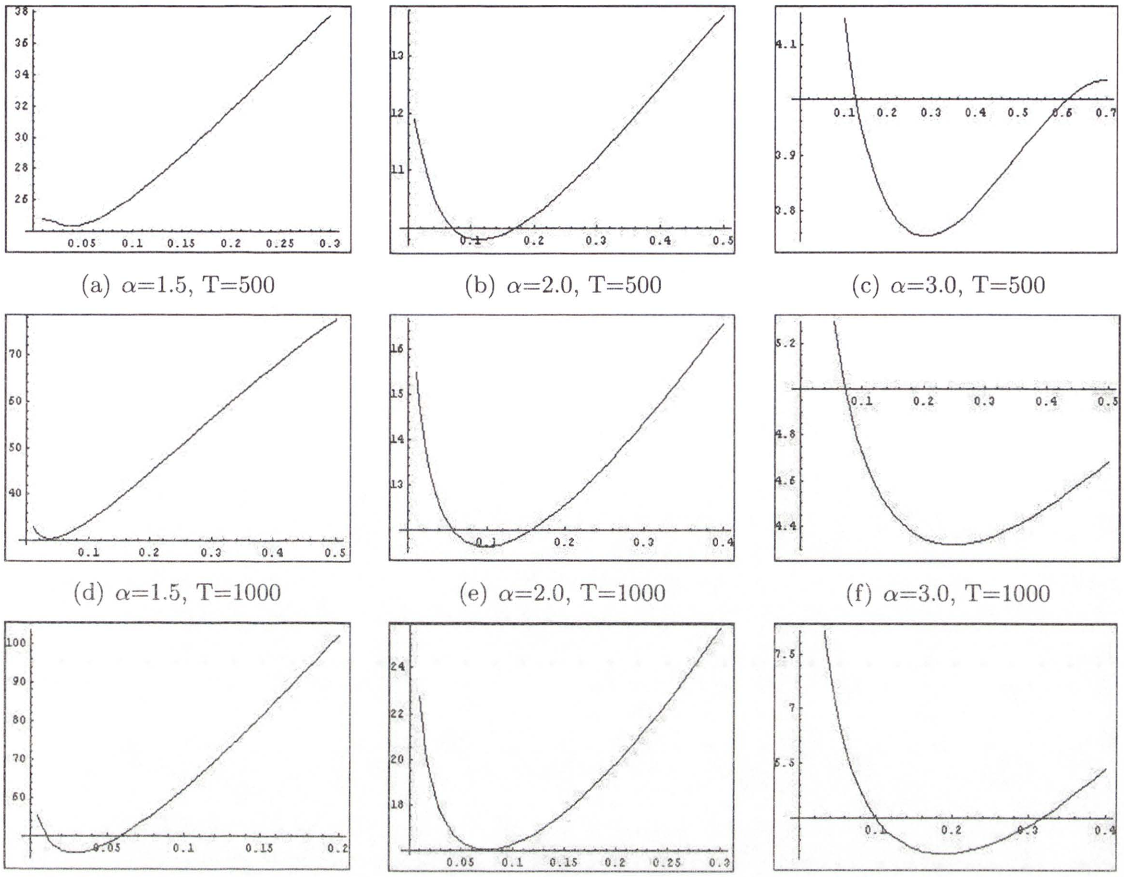
Figure 5.2: Comparison of AD_1 simulated by EWMA and by CUSUM for a one-sided Poisson EWMA chart

5.3 Choices of Optimal Parameters of Poisson EWMA Designs

Using the martingale approach, we can easily find the approximation for optimal parameters $(\tilde{\lambda}, \tilde{H})$ of EWMA designs. The parameter values for approximated optimal EWMA designs for values of $T = 500, 1000, 2000$ and 5000, magnitudes of change $\alpha = 1.1, 1.3, 1.5, 1.7, 1.9, 2.0, 2.5, 3.0$ and in-control parameter $\alpha = 1$ are shown in Table 5.4. This table gives values for the optimal pairs of parameters $(\tilde{\lambda}, \tilde{H})$, the corresponding corrected constant $\tilde{C}$ and the minimal AD^* for each size of changes. Figure 5.3 shows the curves of optimal AD when fixed in-control $T = 500, 1000$ and 5000 and magnitudes of change $\alpha = 1.5, 2$ and 3.

Table 5.4: Optimal parameter values and AD_1^* of a one-sided Poisson EWMA

ARL	α	$\tilde{\lambda}$	$\tilde{H}$	$\tilde{C}$	AD_1^*	MC
500	1.1	0.0096	0.102	0.7926	132.98	133.82±.036
	1.3	0.0213	0.204	0.8123	44.72	45.18±.009
	1.5	0.0385	0.323	0.8606	24.35	24.73±.005
	1.7	0.0677	0.489	0.9043	15.79	16.24±.003
	1.9	0.0973	0.634	0.9461	11.30	11.76±.002
	2.0	0.1124	0.703	0.9712	9.796	10.23±.002
	2.5	0.1930	1.113	1.0263	5.61	6.08±.001
	3.0	0.2863	1.371	1.2170	3.75	4.15±.0008
1000	1.1	0.0070	0.1024	0.6543	198.16	199.8±.051
	1.3	0.0164	0.2027	0.8353	58.84	59.25±.011
	1.5	0.0373	0.3678	0.8599	30.39	30.9±.006
	1.7	0.0610	0.5170	0.9069	19.17	19.26±.004
	1.9	0.0863	0.6562	0.9437	13.48	13.93±.003
	2.0	0.0995	0.7226	0.9704	11.62	12.02±.002
	2.5	0.1696	1.0395	1.0973	6.52	6.97±.001
	3.0	0.2480	1.3631	1.1847	4.32	4.17±.001
2000	1.1	0.0055	0.1048	0.1170	280.78	282.97±.068
	1.3	0.0152	0.2246	0.8231	74.44	74.83±.014
	1.5	0.0334	0.3842	0.8551	36.75	37.35±.007
	1.7	0.0538	0.5286	0.9032	22.68	23.17±.004
	1.9	0.0761	0.6654	0.9421	15.74	16.18±.003
	2.0	0.0878	0.7315	0.9627	13.50	13.92±.002
	2.5	0.1501	1.0448	1.0868	7.46	7.89±.001
	3.0	0.2187	1.3589	1.1588	4.90	5.56±.001
5000	1.1	0.0044	0.1128	0.7571	551.84	552.27± .087
	1.3	0.0132	0.2393	0.8013	96.40	97.3±.018
	1.5	0.0282	0.3918	0.8310	45.50	46.29±.008
	1.7	0.0456	0.5337	0.8867	27.46	27.89±.005
	1.9	0.0648	0.6698	0.9204	18.80	19.26±.003
	2.0	0.0750	0.7400	0.9529	16.04	16.49±.003
	2.5	0.1294	1.0481	1.0499	8.71	9.19±.002
	3.0	0.1887	1.3595	1.1052	5.67	6.13±.001



(a) α=1.5, T=500 (b) α=2.0, T=500 (c) α=3.0, T=500

(d) α=1.5, T=1000 (e) α=2.0, T=1000 (f) α=3.0, T=1000

(g) α=1.5, T=5000 (h) α=2.0, T=5000 (i) α=3.0, T=5000

Figure 5.3: AD_1 for different magnitudes of change: Poisson case

5.4 EWMA Chart for Bernoulli Distribution

5.4.1 The derivation of the closed-form formulas for the Bernoulli EWMA

In this section, we consider the EWMA procedure for a Bernoulli distribution. Somerville et al. (2002) have recommended to use the Bernoulli EWMA for detecting changes in binary attribute observations, i.e. observations with two results "failure" (0) or "success" (1) with probability of success (α).

Here we assume that $\xi_1, \xi_2, ..., \xi_t$ are independent Bernoulli random variables and

$$\begin{aligned}
\xi_t &\sim Bernoulli(\alpha_0), \quad t = 1, 2, ..., \theta - 1 \\
\xi_t &\sim Bernoulli(\alpha), \quad t = \theta, \theta + 1, ..., \quad \alpha > \alpha_0.
\end{aligned}$$

5.4.2 Expectation of first passage times for one-sided Bernoulli EWMA

If ξ_t are independent random variables with outcome 0 and 1 such that the probability of "success" $P(\xi_t = 1) = \alpha$ then

$$\psi_\theta(u) = \log(\alpha e^{u\lambda} + (1 - \alpha))$$

and

$$\varphi_\theta(u) = -u\alpha_0 + \log \prod_{k=0}^{\infty} (\alpha e^{(u(1-\lambda)^k \lambda)} + (1 - \alpha)), \tag{5.10}$$

where

$$\begin{aligned}
\varphi_\infty(u) &= -u\alpha_0 + \log \prod_{k=0}^{\infty} (\alpha_0 e^{(u(1-\lambda)^k \lambda)} + (1 - \alpha_0)), \\
\varphi_1(u) &= -u\alpha_0 + \log \prod_{k=0}^{\infty} (\alpha e^{(u(1-\lambda)^k \lambda)} + (1 - \alpha)).
\end{aligned}$$

Substituting those expressions into Equation (4.3), we obtain

$$E_\theta(\tau_H) = \frac{1}{|\ln(1-\lambda)|} \int_0^\infty \frac{E_\theta(e^{uZ_{\tau_H}} - 1)e^{u\alpha_0}}{u \prod_{k=0}^{\infty} (\alpha e^{(u(1-\lambda)^k \lambda)} + (1 - \alpha))} du. \tag{5.11}$$

In particular,

$$ARL_1 = E_\infty(\tau_H) = \frac{1}{|\ln(1-\lambda)|} \int_0^\infty \frac{E_\infty(e^{uZ_{\tau_H}} - 1)e^{u\alpha_0}}{u \prod_{k=0}^{\infty} (\alpha_0 e^{(u(1-\lambda)^k \lambda)} + (1 - \alpha_0))} du$$

and

$$AD_1 = E_1(\tau_H) = \frac{1}{|\ln(1-\lambda)|}\int_0^{\infty} \frac{E_1(e^{uZ_{\tau_H}}-1)e^{u\alpha_0}}{u\prod\limits_{k=0}^{\infty}(\alpha e^{(u(1-\lambda)^k\lambda)}+(1-\alpha))}du.$$

We also use considerations which lead to Equation (4.18) to obtain approximations for an overshoot. Note that if $\xi_t \sim Bernoulli(\alpha)$, then

$$C = \frac{E(S^2_{\tau_+})}{2E(S_{\tau_+})} = \frac{E(\xi_t^2)}{2E(\xi_t)} = \frac{\alpha}{2\alpha} = \frac{1}{2}. \tag{5.12}$$

The closed-form formulas for one-sided Bernoulli EWMA:

$$ARL_1 = E_{\infty}(\tau_H) \simeq \frac{1}{|\ln(1-\lambda)|}\int_0^{\infty} \frac{E_{\infty}(e^{u(H+C\lambda)}-1)e^{u\alpha_0}}{u\prod\limits_{k=0}^{\infty}(\alpha_0 e^{(u(1-\lambda)^k\lambda)}+(1-\alpha_0))}du \tag{5.13}$$

$$AD_1 = E_1(\tau_H) \simeq \frac{1}{|\ln(1-\lambda)|}\int_0^{\infty} \frac{E_1(e^{uZ_{\tau_H}}-1)e^{u\alpha_0}}{u\prod\limits_{k=0}^{\infty}(\alpha e^{(u(1-\lambda)^k\lambda)}+(1-\alpha))}du. \tag{5.14}$$

If the overshoot is neglected then we obtain exact lower-bounds:

$$ARL_1 \geq \frac{1}{|\ln(1-\lambda)|}\int_0^{\infty} \frac{(e^{uH}-1)e^{u\alpha_0}}{u\prod\limits_{k=0}^{\infty}(\alpha_0 e^{(u(1-\lambda)^k\lambda)}+(1-\alpha_0))}du \tag{5.15}$$

and

$$AD_1 \geq \frac{1}{|\ln(1-\lambda)|}\int_0^{\infty} \frac{(e^{uH}-1)e^{u\alpha_0}}{u\prod\limits_{k=0}^{\infty}(\alpha e^{(u(1-\lambda)^k\lambda)}+(1-\alpha))}du. \tag{5.16}$$

For the case of Bernoulli distribution, $\xi_t \leq 1$ and therefore we have also an upper-bound

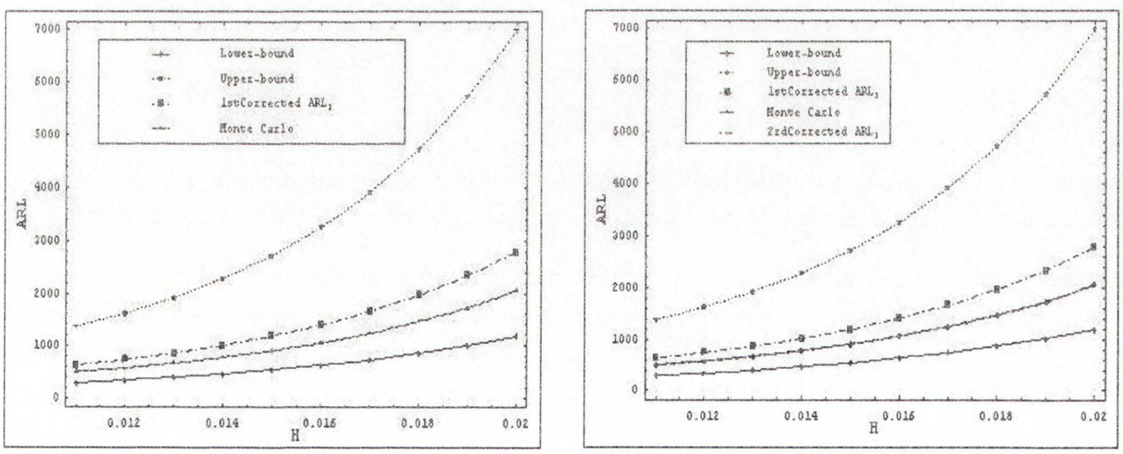
$$E_\theta(\tau_H) \leq Q_\theta((1-\lambda)H + \lambda\xi_t), \tag{5.17}$$

since $Z_t = (1-\lambda)Z_{t-1} + \lambda\xi_t$ and always $Z_{t-1} < H$ for $t = \tau_H$.

The accuracy of these lower-bounds and upper-bounds are reasonable for small λ, (i.e. $\lambda = 0.01$). The numerical results of approximations ARL_1, lower-bounds and upper-bounds are presented on Table 5.5 with in-control parameter $\alpha_0 = 0.01$ and including the "first approximation" with constant C from Equation (5.12) and MC. However, the accuracy of the first approximation differs significantly from MC. The closed-form formulas for ARL and AD presented above contain an overshoot. To get more accurate numerical approximations, we suggest to use a combination of MC and a martingale closed-from formula applying a nonlinear fitting as Poisson case. We then obtain the constant $C = 0.3279$ for evaluating a "second approximation" see Figure 5.4(a) and 5.4(b) below. The second approximation is more certainly accurate approximations.

Table 5.5: Comparison of numerical results of ARL_1 between lower-bounds, upper-bounds, first approximation, second approximation and MC for Bernoulli EWMA

H	Lower-bounds Eq. (5.15)	Upper-bounds Eq. (5.17)	*First Approx.* (ARL_1)	*Second Approx.* $C = 0.3279$	*MC*
0.011	288.22	1371.05	630.195	483.02	499.91±.17
0.012	338.39	1617.19	736.052	563.69	575.06±.19
0.013	396.05	1914.14	860.88	657.98	666.49±.22
0.014	462.64	2274.12	1008.84	768.76	779.63±.25
0.015	539.96	2712.58	1185.12	899.57	905.82±.29
0.016	630.20	3249.11	1396.21	1054.85	1056.05±.33
0.017	736.05	3908.62	1650.26	1240.11	1235.49±.39
0.018	860.88	4722.93	1957.52	1462.26	1465.13± .46
0.019	1008.84	5732.80	2330.96	1730.00	1718.92±.54
0.020	1185.12	6990.55	2787.02	2054.26	2049.68±.64



(a) before adjusted an overshoot (b) after adjusted an overshoot

Figure 5.4: Comparison of lower-bounds, upper-bounds, approximations and simulations of ARL_1: Bernoulli case

5.5 Numerical Results

5.5.1 Comparison of analytical approximations with Monte Carlo simulations

We have numerically calculated based on the closed-form formulas suggested above and compared these values with the results obtained from MC. The accuracy of the approach is confirmed by the simulations (see Table 5.6).

Table 5.6: Comparison of the approximations with Monte Carlo simulations for the Bernoulli EWMA

α	T = 500				
	$\tilde{\lambda}$	$\tilde{H}$	$\tilde{C}$	AD_1^*	MC
0.10	.0057	.0132	.3252	56.27	57.20±.001
0.15	.0241	.0545	.3132	30.39	30.77±.007
0.20	.0356	.0761	.3065	19.31	19.44±.004
0.30	.0834	.1539	.2985	10.62	11.31±.002
α	T = 1000				
0.10	.0057	.0191	.3187	83.25	83.97±.012
0.15	.0210	.0598	.2951	39.16	39.97±.009
0.20	.0396	.0993	.2913	23.97	23.65±.005
0.30	.0777	.1677	.2845	12.71	12.94±.002

5.5.2 Comparison of the performance of EWMA with CUSUM developed from Monte Carlo methods: Bernoulli case

We have also compared the EWMA characteristics with CUSUM for the Bernoulli distribution. Table 5.7 shows some numerical results for ARL and AD for a variety of out-of-control values α for the optimal parameter values of $\tilde{\lambda} = 0.0209$, $\tilde{H} = 0.0598$ and $\tilde{C} = 0.2951$ with in-control parameter value $\alpha_0 = 0.05$ and fixed $T = 1000$. Figure 5.5 shows a comparison of AD values for EWMA and CUSUM. The numerical results give similar results to those obtained for the Gaussian and Poisson distributions. The optimal EWMA charts perform slightly better than CUSUM for the detection of a small change in parameter.

Table 5.7: Comparison Monte Carlo simulations of ARL_1 and AD_1 for one-sided CUSUM and EWMA charts: Bernoulli case

α	CUSUM	EWMA
	$(A = 4.718)$	$(\tilde{\lambda} = 0.0209, \tilde{H} = 0.0598)$
0.05	1008.25±.31	1007.96±.32
0.06	486.18±.15	462.83±.14
0.08	178.18±.05	170.38±.05
0.10	96.73±.02	92.40±.02
0.12	63.93±.01	61.66±.01
0.14	47.50±.009	45.48±.01
0.15	*42.10±.008*	*40.08±.009*
0.16	37.69±.007	35.87±.007
0.18	31.34±.004	29.60±.006
0.20	26.72±.002	20.28±.003
0.30	15.34±.001	14.41±.002
0.40	10.62±.001	10.23±.001

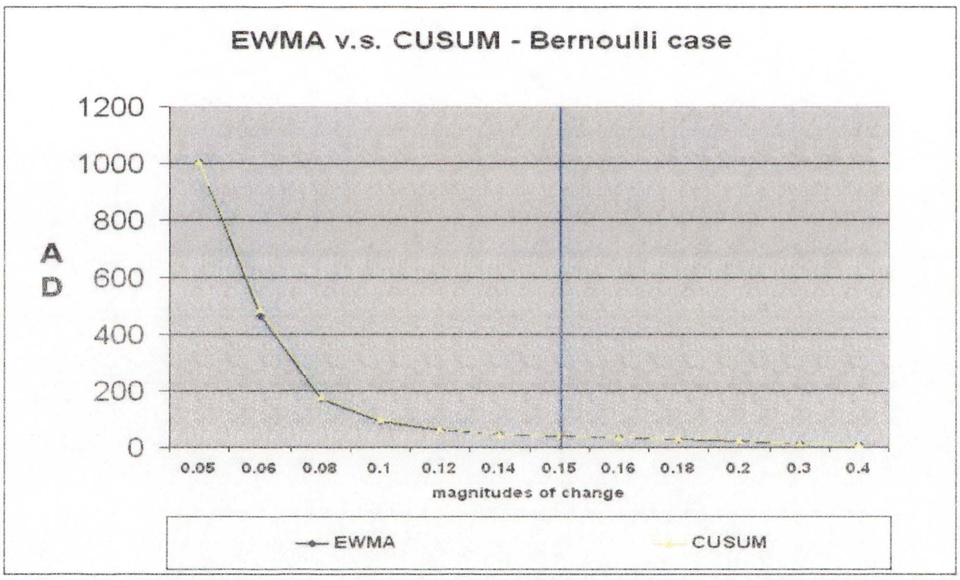


Figure 5.5: Comparison of AD_1 simulated by EWMA and by CUSUM for a one-sided Bernoulli EWMA chart

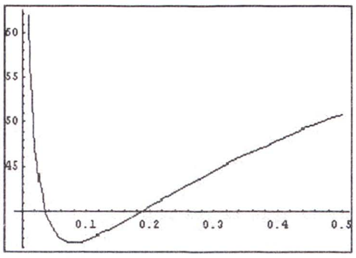
5.6 Choices of Optimal Parameters of Bernoulli EWMA Designs

Table 5.8 contains approximations for optimal values of parameters $(\tilde{\lambda}, \tilde{H})$ when observations are from a Bernoulli distribution. The values were calculated numerically for the one-sided EWMA case from Equation (5.13) and Equation (5.14). These optimal values were obtained by minimising AD values when fixed ARL values of 500, 1000, 2000 and 5000, in-control parameter $\alpha_0 = 0.05$ and the sizes of parameter change, $\alpha =$

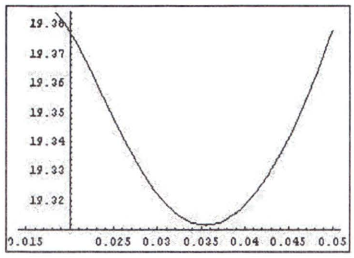
0.1, 0.15, 0.18, 0.22, 0.24, 0.25, 0.30. The numerical results from the martingale technique approximations are as good as the results from the Monte Carlo simulations. The suggested algorithms can be easily used to create curves of AD for a range of magnitudes of change as shown in Figure 5.6.

Table 5.8: Optimal parameter values and AD_1^* for a one-sided Bernoulli EWMA

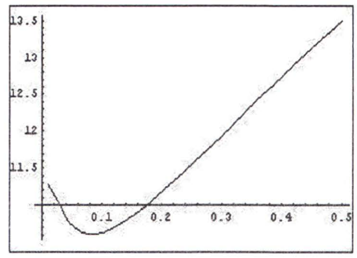
ARL	α	$\tilde{\lambda}$	$\tilde{H}$	$\tilde{C}$	AD_1^*	MC
500	0.10	0.0057	0.0132	0.3252	56.27	57.2±.014
	0.15	0.0073	0.0176	0.3225	28.60	29.01±.006
	0.18	0.0089	0.0216	0.3178	22.14	22.55±.005
	0.20	0.0356	0.0761	0.3065	19.31	19.44±.004
	0.22	0.0461	0.0947	0.3042	16.72	17.11±.004
	0.24	0.0557	0.1109	0.2991	14.69	15.22±.003
	0.25	0.0604	0.1185	0.2987	13.83	14.33±.003
	0.30	0.0834	0.1539	0.2985	10.62	11.31±.002
1000	0.10	0.0057	0.0191	0.3187	83.25	83.97±.012
	0.15	0.0209	0.0598	0.2951	39.16	40.09±.008
	0.18	0.0209	0.0845	0.2899	28.56	29.27±.006
	0.20	0.0323	0.0993	0.2913	23.97	23.65±.005
	0.22	0.047	0.1135	0.2897	20.54	21.05±.004
	0.24	0.0545	0.1274	0.2884	17.89	18.19±.004
	0.25	0.0583	0.1341	0.2908	16.79	17.03±.004
	0.30	0.0777	0.1677	0.2845	12.7	12.9±.002
2000	0.10	0.0022	0.0105	0.2854	110.72	112.35±.02
	0.15	0.0211	0.0706	0.2757	48.48	49.61±.009
	0.18	0.0301	0.0925	0.2752	34.66	35.33±.007
	0.20	0.0363	0.1064	0.2767	28.82	29.57±.006
	0.22	0.0427	0.1201	0.2778	24.51	25.15±.005
	0.24	0.0491	0.1334	0.2793	21.22	21.9±.004
	0.25	0.0524	0.1401	0.2775	19.85	20.47±.004
	0.30	0.0695	0.1725	0.2815	14.86	15.23±.003
5000	0.10	0.0069	0.037	0.2538	154.45	157.82±.03
	0.15	0.0183	0.0752	0.2383	61.44	63.5±.013
	0.18	0.0257	0.0964	0.2410	43.08	44.7±.009
	0.20	0.0310	0.101	0.2483	35.49	36.7±.007
	0.22	0.0364	0.1233	0.2529	29.96	31.06±.006
	0.24	0.0419	0.1366	0.2523	25.77	26.62±.005
	0.25	0.0447	0.1432	0.2522	24.04	24.81±.005
	0.30	0.0594	0.1751	0.2631	17.79	18.23±.003



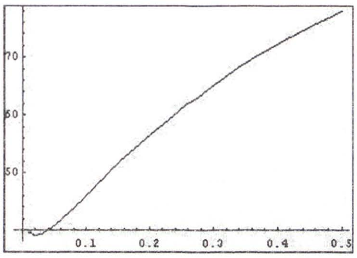
(a) α=0.15,T=500



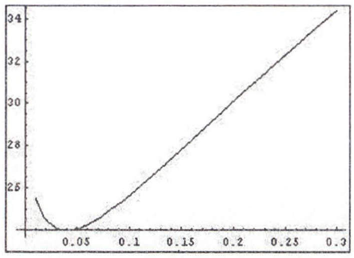
(b) α=0.20,T=500



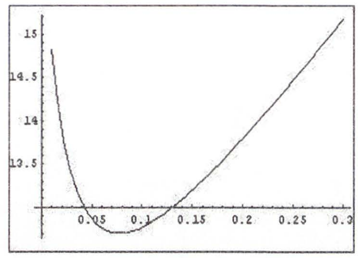
(c) α=0.30,T=500



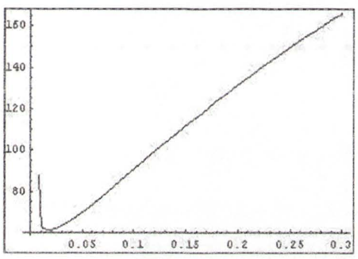
(d) α=0.15,T=1000



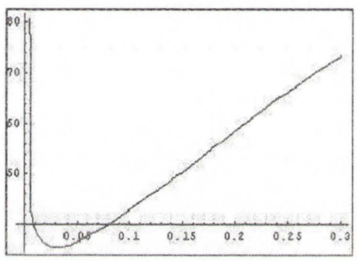
(e) α=0.20,T=1000



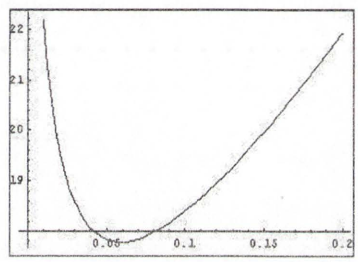
(f) α=0.30,T=1000



(g) α=0.15,T=5000



(h) α=0.20,T=5000



(i) α=0.30,T=5000

Figure 5.6: AD_1 for different magnitudes of change: Bernoulli case

Chapter 6

Conclusion and Recommendations for Further Research

6.1 Overall Conclusion

Statistical Process Control (SPC) charts are widely used to detect changes in parameters that show a process with a random component has gone from an "in-control" state to an "out-of-control" state. The control charts should not signal that a process is out-of-control when it is still in-control, but they should quickly signal when a process goes out-of-control. Two measures that are commonly used to analyse the performance of control charts are the Average Run Length (ARL) and the Average Delay (AD). The ARL is a measure of the average number of observations that will occur before an in-control process falsely gives an out-of-control signal. Therefore to reduce the number of false out-of-control signals a sufficiently large ARL is required. The AD is a measure of the average number of observations that will occur before an out-of-control process correctly gives an out-of-control signal. Therefore to reduce the time that the process is out-of-control, a small AD is required. The ARL and AD are therefore two conflicting criteria that must be balanced to give an optimal design of control chart.

In this thesis, we discussed SPC charts for cases of small and moderate changes in parameters. Two effective types of charts that have been used to detect small and moderate parameter changes are Exponentially Weighted Moving Average (EWMA) and Cumulative Sum (CUSUM) charts. We have demonstrated that the EWMA chart is easy to implement and calculate with a martingale technique. For EWMA charts until now we have no optimal theorem but we have shown by computer simulations that the performance of

EWMA chart is superior to a CUSUM chart.

In this thesis, a martingale technique has been used to obtain analytical approximations for ARL and AD for EWMA charts. We have shown that our martingale approximations are easy to program and calculate. Further, we compare an accuracy of our approximations with other approaches by using the percentage difference. The approximations also produce accurate results and reduce the computational time for finding optimal EWMA designs compared with other standard methods such as Markov Chain Approach (MCA), Integral Equations (IE) and Monte Carlo simulations methods. The limitation of the tradition methods have the some following crucial drawbacks: MCA, of course, requires many matrix inversions; IE requires a lot of programming and diverges under some conditions; MC requires a large numbers of trajectories, in particular for finding an optimal parameter values of EWMA designs. In addition, we have shown that the martingale approach can be adopted and expanded to non-Gaussian distributions such as the Poisson and Bernoulli distributions. We have also developed algorithms written in Mathematica® programs that we have used to obtain optimal numerical values for parameters for optimal EWMA designs for both Gaussian and some non-Gaussian distributions.

6.2 Recommendations for Further Research

In this thesis we have discussed some applications of the martingale method to the performance analysis of control charts. However, there are further applications that can be made. Some possible areas of future applications are as follows:

- The use of the martingale technique for determining the characteristics of EWMA charts when several variables may be monitored and even controlled. This is so-called Multivariate EWMA charts (MEWMA).

- In this thesis, we have assumed that observations are i.i.d. random variables. However, in real applications they could be serially-correlated observations such as in $AR(1), MA(1)$ etc. Serially-correlated data is always assumed in finance and insurance. The martingale technique could be used to analyse this serially-correlated data.

- We suggest to develop a software based on our analytical formulas in which the input parameters could be easily changed by the user, e.g., from the keyboard or through a GUI type dialogue box.

- The overshoot distribution is known only for the case of the Exponential distribution. The distribution of the overshoot has rarely been studied in the literature. Knowledge of the overshoot distribution would improve the accuracy of the overshoot correction to the values of the ARL and AD.

Appendix A

Mathematica® codes for Simulation of Sequentially Stochastic Processes

The appendix gives all of the Mathematica® codes used for simulation of the expectation of the first exit time of EWMA. The results from the Monte Carlo simulation are obtained by taking an average of the first exit times for a large number of trials.

Input requirement

In order to get the ARL and AD values from Monte Carlo and the EWMA martingale approximations the following input is required:

- smoothing parameter $\lambda \in (0, 1)$
- intensity parameters of the distribution of ξ_t given α_0 (in-control)
- shift parameter α, $\alpha > \alpha_0$
- initial value of Z_0 of process Z_t
- value of boundary, $H > Z_0$
- number of trials, n

Output

The output of the simulation program gives an expectation of the first exit time (or ARL) and an expectation of average delay time (AD) from the process Z_t for discrete finite times.

A.1 Gaussian Distribution

A.1.1 EWMA simulation

1. ARL for one-sided EWMA

   "EWMA-ARL1sided"

   ```
   <<Statistics`ContinuousDistributions`
   Clear[gaus, Z, j, λ, n]; n = 10^6; λ = 0.01; H = 1;
   gaus[_]:=Random[NormalDistribution[α0, 1]]; SeedRandom[]; t0 = TimeUsed[];
   X = 0; j = 0;
   Do[
   {For[i = 1; Z = 0, Z < H ,i++, Z = Z(1 − λ) + λgaus[0]],
   j = j + i − 1, X = X + (i − 1)^2}, {k, 1, n}];
   T = N[j/n]
   Sqrt[(X/n − T^2)/n]
   TimeUsed[] − t0
   ```

2. ARL for two-sided EWMA

   "EWMA-ARL2sided"

   ```
   Clear[gaus, Z, j, λ, n]; n = 10^6; λ = 0.01; H = 1;
   gaus[α0_]:=Random[NormalDistribution[α0, 1]]; SeedRandom[]; t0 = TimeUsed[];
   X = 0; j = 0;
   Do[
   {For[i = 1; Z = 0, Abs[Z] < H ,i++, Z = Z(1 − λ) + λgaus[0]],
   j = j + i − 1, X = X + (i − 1)^2}, {k, 1, n}];
   T = N[j/n]
   Sqrt[(X/n − T^2)/n]
   TimeUsed[] − t0
   ```

3. AD for one-sided and two-sided EWMA

   "EWMA-AD"

   ```
   Clear[gaus, Z, j, λ, n]; n = 10^6; λ = 0.01; H = 1;
   gaus[α_]:=Random[NormalDistribution[α, 1]]; SeedRandom[]; t0 = TimeUsed[];
   X = 0; j = 0;
   Do[
   {For[i = 1; Z = 0, Z < H ,i++, Z = Z(1 − λ) + λgaus[0.5]],
   ```

$j = j + i - 1, X = X + (i-1)\text{^}2\}, \{k, 1, n\}];$

$\text{AD} = N[j/n]$

$\text{Sqrt}[(X/n - \text{AD}\text{^}2)/n]$

$\text{TimeUsed}[] - \text{t0}$

A.1.2 CUSUM simulation

1. ARL for one-sided CUSUM

"CUSUM"

<<Statistics'ContinuousDistributions'

$\text{Clear}[\text{gaus}, X, Y, j]; A = 4.68; n = 10\text{^}6; \alpha_0 = 0; \alpha = 0.5;$

$\text{gaus}[\alpha_0\_]{:=}\text{Random}[\text{NormalDistribution}[\alpha_0]]; \text{SeedRandom}[]; \text{t0} = \text{TimeUsed}[];$

$X = 0; j = 0;$

Do[

$\{\text{For}[i = 1; Y = 0, Y < A\ , i{+}{+}, Y = \text{Max}[0, Y + (\frac{\alpha-\alpha_0}{\sigma^2})(\text{gaus} - (\frac{\alpha+\alpha_0}{2}))]],$

$j = j + i - 1, X = X + (i-1)\text{^}2\}, \{k, 1, n\}];$

$\text{T} = N[j/n]$

$\text{Sqrt}[(X/n - \text{T}\text{^}2)/n]$

$\text{TimeUsed}[] - \text{t0}$

2. AD for two-sided CUSUM

"CUSUM"

<<Statistics'ContinuousDistributions'

$\text{Clear}[\text{gaus}, X, Y, j]; A = 4.68; n = 10\text{^}6; \alpha_0 = 0; \alpha = 0.5;$

$\text{gaus}[\alpha\_]{:=}\text{Random}[\text{NormalDistribution}[\alpha]]; \text{SeedRandom}[]; \text{t0} = \text{TimeUsed}[];$

$X = 0; j = 0;$

Do[

$\{\text{For}[i = 1; Y = 0, Y < A\ , i{+}{+}, Y = \text{Max}[0, Y + (\frac{\alpha-\alpha_0}{\sigma^2})(\text{gaus} - (\frac{\alpha+\alpha_0}{2}))]],$

$j = j + i - 1, X = X + (i-1)\text{^}2\}, \{k, 1, n\}];$

$\text{AD} = N[j/n]$

$\text{Sqrt}[(X/n - \text{AD}\text{^}2)/n]$

$\text{TimeUsed}[] - \text{t0}$

A.2 Poisson Distribution

A.2.1 EWMA simulation

1. ARL for one-sided EWMA

   "EWMA-ARL1sided"

   <<Statistics'DiscreteDistributions'

   Clear[pois, Z, j, λ, n, T]; $n = 10$^6; $\lambda = 0.01$; $H = 1$;

   pois[α_0_]:=Random[PoissonDistribution[α_0]]; SeedRandom[]; t0 = TimeUsed[];

   $Z = 0; j = 0;$

   Do

   [{For[$i = 1$; $Z = 0, Z < H$,i++, $Z = Z(1 - \lambda) + \lambda$(pois[1] − 1)],

   $j = j + i - 1, X = X + (i - 1)$^2}, {$k$, 1, n}];

   T = $N[j/n]$

   Sqrt[(X/n − T^2)/n]

   TimeUsed[] − t0

2. AD for one-sided and two-sided EWMA

   "EWMA-AD"

   Clear[pois, Z, j, λ, n, T]; $n = 10$^6; $\lambda = 0.01$; $H = 1$;

   pois[α_]:=Random[PoissonDistribution[α]]; SeedRandom[]; t0 = TimeUsed[];

   $Z = 0; j = 0;$

   Do

   [{For[$i = 1$; $Z = 0, Z < H$,i++, $Z = Z(1 - \lambda) + \lambda$(pois[1.5] − 1)],

   $j = j + i - 1, X = X + (i - 1)$^2}, {$k$, 1, n}];

   AD = $N[j/n]$

   Sqrt[(X/n − AD^2)/n]

   TimeUsed[] − t0

A.2.2 CUSUM simulation

1. ARL for one-sided CUSUM

   "CUSUM"

   <<Statistics'DiscreteDistributions'

   Clear[pois, X, Y, j]; $b = 4.1$; $n = 10$^6; $\alpha_0 = 1$; $\alpha = 1.5$; cc = Log[α/α_0]; aa = $-\alpha + \alpha_0$;

   pois[α_0_]:=Random[PoissonDistribution[α_0]]; SeedRandom[]; t0 = TimeUsed[];

$X = 0; j = 0;$

Do[

{For[$i = 1; Y = 0, Y < b$,i++, $Y = $ Max[0, $Y + (-\alpha + \alpha_0)$ + pois(Log[α/α_0])]],

$j = j + i - 1, X = X + (i - 1)$^2}, {$k$, 1, n}];

T = N[j/n]

Sqrt[(X/n − T^2)/n]

TimeUsed[] − t0

2. AD for one-sided CUSUM

"CUSUM"

<<Statistics'DiscreteDistributions'

Clear[pois, X, Y, j]; $b = 4.1; n = 10$^6; $\alpha_0 = 1; \alpha = 1.5$; cc = Log[$\alpha/\alpha_0$]; aa = $-\alpha + \alpha_0$;

pois[α_]:=Random[PoissonDistribution[α]]; SeedRandom[]; t0 = TimeUsed[];

$X = 0; j = 0;$

Do[

{For[$i = 1; Y = 0, Y < b$,i++, $Y = $ Max[0, $Y + (-\alpha + \alpha_0)$ + pois(Log[α/α_0])]],

$j = j + i - 1, X = X + (i - 1)$^2}, {$k$, 1, n}];

AD = N[j/n]

Sqrt[(X/n − AD^2)/n]

TimeUsed[] − t0

A.3 Bernoulli Distribution

A.3.1 EWMA simulation

1. ARL for one-sided EWMA with

   "EWMA-ARL1sided"

   <<Statistics'DiscreteDistributions'

   Clear[ber, Z, j, λ, n, T]; $n = 10$^6; $\lambda = 0.01; H = 1; \alpha_0 = 0.05$;

   ber:=If[Random[] $> (1 - \alpha_0)$, 1, 0]; SeedRandom[]; t0 = TimeUsed[];

   $Z = 0; j = 0;$

   Do

   [{For[$i = 1; Z = 0, Z < H$,i++, $Z = Z(1 - \lambda) + \lambda$(ber $- \alpha_0$)],

   $j = j + i - 1, X = X + (i - 1)$^2}, {$k, 1, n$}];

   T = $N[j/n]$

   Sqrt[(X/n − T^2)/n]

   TimeUsed[] − t0

2. AD for one-sided and two-sided EWMA

   "EWMA-AD"

   Clear[ber, Z, j, λ, n, T]; $n = 10$^6; $\lambda = 0.01; H = 1; \alpha = 0.15$;

   ber:=If[Random[] $> (1 - \alpha)$, 1, 0]; SeedRandom[]; t0 = TimeUsed[];

   $Z = 0; j = 0;$

   Do

   [{For[$i = 1; Z = 0, Z < H$,i++, $Z = Z(1 - \lambda) + \lambda$(ber $- \alpha_0$)],

   $j = j + i - 1, X = X + (i - 1)$^2}, {$k, 1, n$}];

   AD = $N[j/n]$

   Sqrt[(X/n − AD^2)/n]

   TimeUsed[] − t0

A.3.2 CUSUM simulation

1. ARL for one-sided CUSUM

   "CUSUM"

   <<Statistics'DiscreteDistributions'

   Clear[ber, X, Y, j]; $b = 3.66$; $n = 10$^6; $\alpha_0 = 0.05; \alpha = 0.15$;

   ber:=If[Random[] $> (1 - \alpha_0)$, 1, 0]; SeedRandom[]; t0 = TimeUsed[];

$X = 0; j = 0;$

Do[

{For[$i = 1; Y = 0, Y < b$, i++, $Y =$ Max[0, Y + berLog[$(\frac{\alpha}{\alpha_0} + \frac{1-\alpha}{1-\alpha_0})$ + (1-ber)Log$\frac{1-\alpha}{1-\alpha_0}$]]],

$j = j + i - 1, X = X + (i - 1)$^2}, {$k, 1, n$}];

T = $N[j/n]$

Sqrt[$(X/n -$ T^2$)/n$]

TimeUsed[] − t0

2. AD for one-sided CUSUM

"CUSUM"

<<Statistics'DiscreteDistributions'

Clear[ber, X, Y, j]; $b = 3.66; n =$ 10^6; $\alpha_0 = 0.05; \alpha = 0.15;$

ber:=If[Random[] > $(1 - \alpha), 1, 0$]; SeedRandom[]; t0 = TimeUsed[];

$X = 0; j = 0;$

Do[

{For[$i = 1; Y = 0, Y < b$, i++, $Y =$ Max[0, Y + Log[ber$(\frac{\alpha}{\alpha_0} - \frac{1-\alpha}{1-\alpha_0}) + \frac{1-\alpha}{1-\alpha_0}$]]],

$j = j + i - 1, X = X + (i - 1)$^2}, {$k, 1, n$}];

AD = $N[j/n]$

Sqrt[$(X/n -$ AD^2$)/n$]

TimeUsed[] − t0

Appendix B

Mathematica® Codes for Calculation of Closed-form Formulas

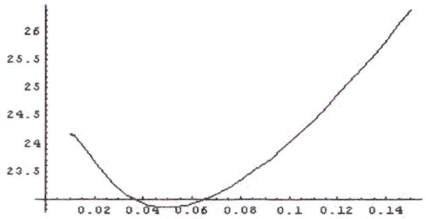
B.1 Procedure for obtaining the optimal parameter

These are the procedures for obtaining the optimal parameters (λ^*, H^*) from the martingale closed-form formulas for the Gaussian, Poisson and Bernoulli distributions.

B.1.1 Calculation codes for the case of one-side Gaussian EWMA chart

Step 1: **"Initial approximation"**

```
Clear[λ, b, c, α, T, QQ, FF, H]; T = 500; α0 = 0; α = 0.5; c = 0.583;
QQ[H_, λ_, c_, α_]:= − NIntegrate[(Exp[u(H + λc)] − 1)/u
Exp[−uα − (λu^2)/(4(1 − λ/2))], {u, 0, ∞}]/Log[1 − λ];
myinv[λ_]:=H/.(FindRoot[QQ[H, λ, c, 0] == T, {H, 0.1, 3}]);
FF[λ_]:= − NIntegrate[(Exp[u(myinv[λ] + λc)] − 1)/u
Exp[−αu − (λu^2)/(4(1 − λ/2))], {u, 0, ∞}]/Log[1 − λ];
Plot[FF[λ], {λ, 0.01, 0.3}]
FM = FindMinimum[FF[λ], {λ, 0.1, 1}]
AD = First[FM]
ww = λ/.Last[FM]
bb = myinv[ww]
```



Initial approximation

{22.8468, {$\lambda \to$ 0.0496229}}

22.8468

0.0496229

0.365274

Step 2: **<<Statistics'ContinuousDistributions'**

"Monte Carlo check 1"

Clear[gaus, Y, j, λ, n**];** n **= 10^5; alf = ww;**

gaus[m_]:=Random[NormalDistribution[α_0**, 1]]; SeedRandom[]; t0 = TimeUsed[];**

$X = 0; j = 0;$

Do[{For[$i = 1; Z = 0, Z <$ **bb ,** i**++,** $Y = Y(1 - \lambda) + \lambda$**gaus[0]],**

$j = j + i - 1, X = X + (i - 1)$**^2}, {**$k, 1, n$**}];**

newT = $N[j/n]$

TimeUsed[] – t0

Monte Carlo check 1

402.135

0.0148

1693.72

Step 3: **"Second approximation"**

newc = c**/.(FindRoot[QQ[bb, ww,** c**, 0] == newT, {**c**, .1, .7}])**

newH = H**/.(FindRoot[QQ[**H**, ww, newc, 0] ==** T**, {**H**, .1, 2}])**

0.240828

0.382254

"Corrected AD by closed-form martingale"

$\alpha = 0.5;$

$\lambda = \mathrm{ww}$;

$-\mathrm{NIntegrate}[(\mathrm{Exp}[u(\mathrm{newH} + \mathrm{wwnewc})] - 1)/u$

$\mathrm{Exp}[-\alpha u - \frac{\lambda u\text{^}2}{4(1-\lambda/2)}], \{u, 0.001, 1000\}]/\mathrm{Log}[1 - \lambda]$

Corrected AD

22.8468

Step 4: **"Final check MC for AD"**

$\mathrm{Clear}[\mathrm{gaus}, Z, j, \lambda, n]; n = 10\text{^}5; \lambda = \mathrm{ww}$;

$\mathrm{gaus}[\alpha\_]{:=}\mathrm{Random}[\mathrm{NormalDistribution}[\alpha, 1]]; \mathrm{SeedRandom}[]; \mathrm{t0} = \mathrm{TimeUsed}[]$;

$X = 0; j = 0$;

$\mathrm{Do}[\{\mathrm{For}[i = 1; Z = 0, Z < \mathrm{newH}\ , i{+}{+}, Z = Z(1 - \lambda) + \lambda \mathrm{gaus}[1]]$,

$j = j + i - 1, X = X + (i - 1)\text{^}2\}, \{k, 1, n\}]$;

$\mathrm{AD} = N[j/n]$

$\mathrm{Sqrt}[(X/n - \mathrm{AD}\text{^}2)/n]$

$\mathrm{TimeUsed}[] - \mathrm{t0}$

24.9532

0.055

93.813

"Final check MC for ARL"

$\mathrm{Clear}[\mathrm{gaus}, Z, j, \lambda, n]; n = 10\text{^}5; \lambda = \mathrm{ww}$;

$\mathrm{gaus}[\alpha_0\_]{:=}\mathrm{Random}[\mathrm{NormalDistribution}[\alpha_0, 1]]; \mathrm{SeedRandom}[]; \mathrm{t0} = \mathrm{TimeUsed}[]$;

$X = 0; j = 0$;

$\mathrm{Do}[\{\mathrm{For}[i = 1; Z = 0, Z < \mathrm{newH}\ , i{+}{+}, Z = Z(1 - \lambda) + \lambda \mathrm{gaus}[0]]$,

$j = j + i - 1, X = X + (i - 1)\text{^}2\}, \{k, 1, n\}]$;

$T = N[j/n]$

$\mathrm{Sqrt}[(X/n - T\text{^}2)/n]$

$\mathrm{TimeUsed}[] - \mathrm{t0}$

484.8

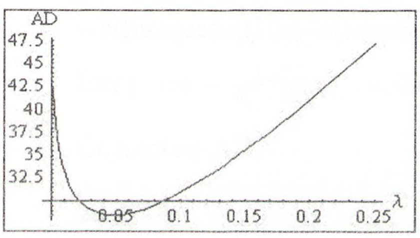
0.0156

1523.91

B.1.2 Calculation codes for the case of two-sided Gaussian EWMA chart

Step 1: "**Initial approximation**"

```
Clear[λ, b, c, α, T, QQ, FF, H]; T = 500; α0 = 0; α = 0.5; c = 0.583;
QQ[H_, λ_, c_, α_]:= − NIntegrate[(Cosh[u(H + λc)] − 1)/u
Exp[−uα − (λu^2)/(4(1 − λ/2))], {u, 0, ∞}]/Log[1 − λ];
myinv[λ_]:=H/.(FindRoot[QQ[H, λ, c, 0] == T, {H, 0.1, 3}]);
FF[λ_]:= − NIntegrate[(Exp[u(myinv[λ] + λc)] − 1)/u
Exp[−αu − (λu^2)/(4(1 − λ/2))], {u, 0, ∞}]/Log[1 − λ];
Plot[FF[λ], {λ, 0.01, 0.3}]
FM = FindMinimum[FF[λ], {λ, 0.1, 1}]
AD = First[FM]
ww = λ/.Last[FM]
bb = myinv[ww]
```



Initial approximation

{28.4774, {λ → 0.047025}}

28.4774

0.0470249

0.403263

Step 2: <<Statistics‘ContinuousDistributions‘

```
"Monte Carlo check 1"
Clear[gaus, Y, j, λ, n]; n = 10^5; alf = ww;
gaus[m_]:=Random[NormalDistribution[α0, 1]]; SeedRandom[]; t0 = TimeUsed[];
X = 0; j = 0;
Do[{For[i = 1; Z = 0, Abs[Z] < bb , i++, Y = Y(1 − λ) + λgaus[0]],
j = j + i − 1, X = X + (i − 1)^2}, {k, 1, n}];
newT = N[j/n]
TimeUsed[] − t0
```

Monte Carlo check 1

510.395

0.0148

1693.72

Step 3: **"Second approximation"**

newc = c/.(FindRoot[QQ[bb, ww, c, 0] == newT, {c, .1, .7}])

newH = H/.(FindRoot[QQ[H, ww, newc, 0] == T, {H, .1, 2}])

0.610832

0.401954

"Corrected AD by closed-form martingale"

$\alpha = 0.5$;

$\lambda =$ ww;

−NIntegrate[(Exp[u(newH + wwnewc)] − 1)/u

Exp[$-\alpha u - \frac{\lambda u\hat{}2}{4(1-\lambda/2)}$], {$u$, 0.001, 1000}]/Log[1 − λ]

Corrected AD

28.4774

Step 4: **"Final check MC for AD"**

Clear[gaus, Z, j, λ, n]; n = 10^5; λ = ww;

gaus[α_]:=Random[NormalDistribution[α, 1]]; SeedRandom[]; t0 = TimeUsed[];

$X = 0; j = 0$;

Do[{For[i = 1; Z = 0, Z < newH , i++, Z = Z(1 − λ) + λgaus[1]],

$j = j + i - 1$, $X = X + (i-1)$^2}, {k, 1, n}];

AD = N[j/n]

Sqrt[(X/n − AD^2)/n]

TimeUsed[] − t0

28.7215

0.055

93.813

"Final check MC for ARL"

Clear[gaus, Z, j, λ, n]; n = 10^5; λ = ww;

gaus[α_0_]:=Random[NormalDistribution[α_0, 1]]; SeedRandom[]; t0 = TimeUsed[];

$X = 0; j = 0;$

Do[{For[$i = 1; Z = 0$, Abs[Z] < newH , i++, $Z = Z(1 - \lambda) + \lambda$gaus[0]],

$j = j + i - 1, X = X + (i - 1)$^2}, {$k, 1, n$}];

$T = N[j/n]$

Sqrt[($X/n - T$^2)/n]

TimeUsed[] − t0

500.922

0.1559

1633.61

B.1.3 Calculation codes for the case of Poisson EWMA chart

Step 1: **"Initial approximation"**

Clear[$\lambda, b, C, \alpha_0, \alpha, T$, QQ, FF, H]; $T = 500; \alpha_0 = 1; \alpha = 1.5; C = 1;$

QQ[H_, λ_, C_, α_0_]:= − 1/Log[1 − a]NIntegrate[(Exp[$u(H + \lambda$C)] − 1)/u

Exp[$u - \alpha_0$Sum[(Exp[$u\lambda(1 - \lambda)$^k] − 1), {$k, 0, \infty$}]], {$u, 0, \infty$}]

myinv[λ_]:=H/.(FindRoot[QQ[H, λ, C, 1] == T, {H, 0.1, 5}]);

FF[λ_]:= − 1/Log[1 − λ]NIntegrate[(Exp[u(myinv[λ] + λC)] − 1)/u

Exp[$u - \alpha$ Sum[(Exp[$u\lambda(1 - \lambda)$^k] − 1), {$k, 0, \infty$}]], {$u, 0, \infty$}]

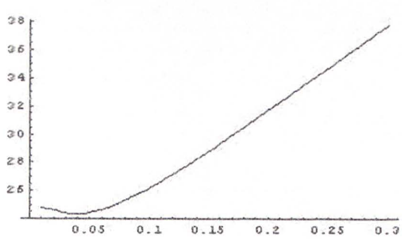
Plot[FF[λ], {λ, 0.01, 0.5}]

FM = FindMinimum[FF[λ], {λ, 0.01, 0.5}]

AD = First[FM]

ww = λ/.Last[FM]

bb = myinv[ww]



Initial approximation

{24.3544, {$a \to 0.0384861$}}

24.3544

0.0384861

0.317569

Step 2: **<<Statistics'DiscreteDistributions'**

"Monte Carlo check 1"

Clear[pois, Z, j, λ, n]; $n = 10$^6; $\lambda =$ ww;

pois[α_0_]:=Random[PoissonDistribution[α_0]]; SeedRandom[]; t0 = TimeUsed[];

$X = 0; j = 0$;

Do[{For[$i = 1; Z = 0, Z <$ bb , i++, $Z = Z(1-\lambda) + \lambda$(pois[1] $- 1$)],

$j = j + i - 1, X = X + (i-1)$^2}, {$k, 1, n$}];

newT = $N[j/n]$

TimeUsed[] − t0

Monte Carlo check 1

468.1903

0.0162

1693.72

Step 3: **"Second approximation"**

newC = C/.(FindRoot[QQ[bb, ww, C, 1] == newT, {C, .5, 2.0}])

newH = H/.(FindRoot[QQ[H, ww, newC, 1] == T, {H, 0.1, 0.5}])

Second approximation

0.860551

0.322935

"Corrected AD by closed-form martingale"

$\alpha = 1.5$;

$\lambda =$ ww;

$-1/$Log[$1 - \lambda$]NIntegrate[(Exp[u(newH + λnewC)] − 1)/u

Exp[$u - \alpha$Sum[(Exp[$u\lambda(1-\lambda)$^k] − 1), {$k, 0, \infty$}]], {$u, 0, \infty$}]

Corrected AD

24.3544

Step 4: **"Final check MC for AD"**

<<Statistics'DiscreteDistributions'

Clear[pois, Z, j, λ, n, AD]; $n = 10$^6; $\lambda =$ ww;

pois[α_]:=Random[PoissonDistribution[α]]; SeedRandom[]; t0 = TimeUsed[];

$X = 0; j = 0;$

$\text{Do}[\{\text{For}[i = 1; Z = 0, Z < \text{newH} , i{+}{+}, Z = Z(1 - \lambda) + \lambda(\text{pois}[1.5] - 1)],$

$j = j + i - 1, X = X + (i - 1)\hat{}2\}, \{k, 1, n\}];$

$\text{AD} = N[j/n]$

$\text{Sqrt}[((X/n) - (\text{AD}\hat{}2))/n]$

$\text{TimeUsed}[] - \text{t0}$

Final check MC for AD

24.72947

0.00553

153.578

"Final check MC for ARL"

$\text{Clear}[\text{pois}, Y, j, \text{alf}, n]; n = 10\hat{}6; \lambda = \text{ww};$

$\text{pois}[\alpha_0\_]{:=}\text{Random}[\text{PoissonDistribution}[\alpha_0]]; \text{SeedRandom}[]; \text{t0} = \text{TimeUsed}[];$

$X = 0; j = 0;$

$\text{Do}[\{\text{For}[i = 1; Z = 0, Z < \text{newH} , i{+}{+}, Z = Z(1 - \lambda) + \lambda(\text{pois}[1] - 1)],$

$j = j + i - 1, X = X + (i - 1)\hat{}2\}, \{k, 1, n\}];$

$T = N[j/n]$

$\text{Sqrt}[((X/n) - (T\hat{}2))/n]$

$\text{TimeUsed}[] - \text{t0}$

Final check MC for ARL

497.876

0.157827

8144.75

B.1.4 Calculation codes for the case of Bernoulli EWMA chart

Step 1: **"Initial approximation"**

$\text{Clear}[a, \lambda, C, \alpha_0, \alpha, T, \text{QQ}, \text{FF}, H]; T = 500; C = 0.5; \alpha_0 = 0.05; \alpha = 0.25;$

$\text{QQ}[\text{H}\_, \lambda\_, \text{C}\_, \alpha_0\_]{:=} - 1/\text{Log}[1 - \lambda]$

$(\text{NIntegrate}[(\text{Exp}[(H + \lambda C)u] - 1)\text{Exp}[u\alpha_0]/u$

$1/\prod_{k=0}^{\infty} (\alpha_0\text{Exp}[\lambda u(1 - \lambda)\hat{}k] + (1 - \alpha_0)), \{u, 0, 1\}]$

$+\text{NIntegrate}[(\text{Exp}[(H + \lambda C)u] - 1)\text{Exp}[u\alpha_0]/u$

$1/\prod_{k=0}^{\infty} (\alpha_0\text{Exp}[\lambda u(1 - \lambda)\hat{}k] + (1 - \alpha_0)), \{u, 1, \infty\}]);$

myinv[λ_]:=H/.(FindRoot[QQ[H, λ, C, α_0] == T, {H, 0.01, 0.2}]);

FF[λ_]:= − 1/Log[1 − λ]

(NIntegrate[(Exp[(myinv[λ] + λC)u] − 1)Exp[$u\alpha_0$]/u

$1/\prod_{k=0}^{\infty}$ (αExp[au(1 − a)^k] + (1 − α)), {u, 0.001, 1}]

+NIntegrate[(Exp[(myinv[λ] + λC)u] − 1) Exp[$u\alpha_0$]/u

$1/\prod_{k=0}^{\infty}$ (αExp[λu(1 − λ)^k] + (1 − α)), {u, 1, ∞}]);

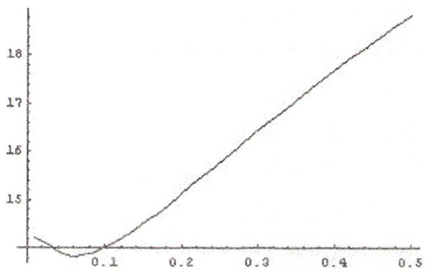
Plot[FF[λ], {λ, .01, .5}]

FM = FindMinimum[FF[λ], {λ, .01, .5}]

AD = First[FM]

ww = λ/.Last[FM]

bb = myinv[ww]



Initial approximation

{13.8289, {a → 0.0603573}}

13.8289

0.0603573

0.106329

Step 2: <<Statistics'DiscreteDistributions'

"Monte Carlo check 1"

Clear[ber, Z, j, λ, n]; n = 10^6; λ = ww;

ber:=If[Random[] > 0.95, 1, 0]; SeedRandom[]; t0 = TimeUsed[];

X = 0; j = 0;

Do[{For[i = 1; Z = 0, Z < bb , i++, Z = Z(1 − λ) + λ(ber − 0.05)],

j = j + i − 1, X = X + (i − 1)^2}, {k, 1, n}];

newT = N[j/n]

Sqrt[(X/n − newT^2)/n]

TimeUsed[] − t0

Monte Carlo check 1

329.843

0.104195

299.016

Step 3: **"Second approximation"**

$\mathbf{newC} = C/.(\mathbf{FindRoot}[\mathbf{QQ}[\mathbf{bb}, \mathbf{ww}, C, \alpha_0] == \mathbf{newT}, \{C, .1, .7\}])$

$\mathbf{newH} = H/.(\mathbf{FindRoot}[\mathbf{QQ}[H, \mathbf{ww}, \mathbf{newC}, \mathbf{p0}] == T, \{H, .01, 0.5\}])$

Second approximation

0.298388

0.118498

0.0603573

"Corrected AD by closed-form martingale"

$\alpha = 0.15; \lambda = \mathbf{ww};$

$-1/\mathbf{Log}[1 - a]$

$$(\mathbf{NIntegrate}[(\mathbf{Exp}[(H + \lambda\mathbf{newC})u] - 1)\mathbf{Exp}[u\alpha_0]/u$$
$$1/\prod_{k=0}^{\infty}(\alpha\mathbf{Exp}[\lambda u(1-\lambda)\hat{}k] + (1-\alpha)), \{u, 0.001, 1\}]$$
$$+\mathbf{NIntegrate}[(\mathbf{Exp}[(H + \lambda\mathbf{newC})u] - 1)\mathbf{Exp}[u\alpha_0]/u$$
$$1/\prod_{k=0}^{\infty}(\alpha\mathbf{Exp}[\lambda u(1-\lambda)\hat{}k] + (1-\alpha)), \{u, 1, \infty\}])$$

Corrected AD

13.8289

Step 4: **"Final check MC for AD"**

$\mathbf{Clear}[\mathbf{ber}, Y, j, \lambda, n]; n = 10\hat{}6; \lambda = \mathbf{ww};$

$\mathbf{ber}{:=}\mathbf{If}[\mathbf{Random}[] > 0.75, 1, 0]; \mathbf{SeedRandom}[]; \mathbf{t0} = \mathbf{TimeUsed}[];$

$X = 0; j = 0;$

$\mathbf{Do}[\{\mathbf{For}[i = 1; Z = 0, Z < \mathbf{newH}\ , i{+}{+}, Z = Z(1-\lambda) + \lambda(\mathbf{ber} - 0.05)],$

$j = j + i - 1, X = X + (i-1)\hat{}2\}, \{k, 1, n\}];$

$\mathbf{AD} = N[j/n]$

$\mathbf{Sqrt}[(X/n - \mathbf{AD}\hat{}2)/n]$

$\mathbf{TimeUsed}[] - \mathbf{t0}$

Final check MC for AD

14.4024

0.00303116

14.016

"Final check MC for ARL"

```
Clear[ber, Z, j, λ, n]; n = 10^6; λ = ww;
ber:=If[Random[] > 0.95, 1, 0]; SeedRandom[]; t0 = TimeUsed[];
X = 0; j = 0;
Do[{For[i = 1; Z = 0, Z < newH , i++, Z = Z(1 - λ) + λ(ber - 0.05)],
j = j + i - 1, X = X + (i - 1)^2}, {k, 1, n}];
T = N[j/n]
Sqrt[(X/n - T^2)/n]
TimeUsed[] - t0
```

Final check MC for ARL

494.776

0.156512

449.828

Bibliography

Anderson, E. (2002). Monitoring cyclical processes: A non-parametric approach. *Journal of Applied Statistics*, 29:973–990.

Barnard, G. A. (1959). Control chart and stochastic processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 23:149–153.

Basseville, M. and Nikiforov, I. V. (1993). *Detection of Abrupt Changes: Theory and Applications*. Prentice-Hall, Englewood Cliffs, NJ.

Bissell, A. F. (1969). CUSUM techniques for quality control (with discussion). *Applied Statistics*, 18:1–30.

Borovkov, A. A. (1973). *Stochastic Processes in Queueing Theory*. Applications of Mathematics, No.4, Springer-Verlag, New York, Berlin.

Borovkov, A. A. (1999). *Probability Theory*. Gordon and Breach, New York. English transl.

Borovkov, K. A. (2003). *Elements of Stochastic Modelling*. World Scientific, New Jersey-Singapore.

Borror, C. M., Champ, C. W., and Rigdon, S. E. (1998). Poisson EWMA control charts. *Journal of Quality Technology*, 30:352–361.

Brook, D. and Evans, D. A. (1972). An approach to probability distribution of cusum run length. *Biometrika*, 59:539–548.

Champ, C. W. and Woodall, W. H. (1991). A comparison of the markov chain and the integral equation approaches for evaluating the run length distribution of quality control charts. *Communications in Statistics: Simulation and Computation*, 20:191–204.

Chang, J. and Peres, Y. (1997). Ladder heights, Gaussian random walks, and the Riemann zeta function. *Annals of Probability*, 25:787–802.

Crowder, S. V. (1987). A simple method for studying run-length distributions of exponentially weighted moving average charts. *Technometrics*, 29:401–407.

Ergashev, B. A. (2003). On a CAPM monitoring based on the EWMA process control. Computing in Economics and Finance 2003 283, Society for Computational Economics.

Ewan, W. D. (1963). When and how to use Cu-Sum charts. *Technometrics*, 5:1–22.

Frisén, M. (1992). Evaluations of methods for statistical surveillance. *Statistics in Medicine*, 11:1489–1502.

Frisén, M. (2003). Statistical surveillance: Optimality and methods. *International Statistical Review*, 71:403–434.

Frisén, M. and Sonesson, C. (2006). Optimal surveillance based on exponentially weighted moving averages. *Sequential Analysis*, 25:379–403.

Gallagher, R. G. (1996). *Discrete Stochastic Processes.* Kluwer Academic, Boston.

Gan, F. F. (1990). Monitoring poisson observations using modified exponentially weighted moving average control charts. *Communications in Statistics-Simulation and Computational*, 19:103–124.

Gan, F. F. (1991). An optimal design of CUSUM quality control charts. *Journal of Quality Technology*, 23:279–286.

Gan, F. F. (1994). Design of optimal exponential CUSUM control charts. *Journal of Quality Technology*, 26:109–124.

Hawkins, D. M. (1981). A CUSUM for a scale parameter. *Journal of Quality Technology*, 13:228–235.

Hawkins, D. M. (1993). Cumulative Sum control charting: An underutilized SPC tool. *Quality Engineering*, 5:463–477.

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

Hillson, E. M., Reeves, J. H., and McMillan, C. A. (1998). A statistical signalling model for use in surveillance of adverse drug reaction data. *Journal of Applied Statistics*, 25:23–40.

Hunter, J. S. (1986). The exponentially weighted moving average. *Journal of Quality Technology*, 18:203–210.

Hutwagner, L., Thompson, W., Seeman, G. M., and Treadwell, T. (2003). The bioterrorism preparedness and response early aberration reporting system (EARS). *Journal of Urban Health*, 80:89–96.

Knoth, S. (2007). Accurate ARL calculation for EWMA control charts monitoring normal mean and variance simultaneously. *Sequential Analysis*, 26:251–263.

Liptser, R. S. and Shiryaev, A. N. (1977). *Statistics of Random Processes, I.General Theory; II.Applications.* Springer-Verlag, New York.

Lorden, G. (1971). Procedures for reacting to a change in distribution. *Annals of Mathematical Statistics*, 42:1897–1908.

Lotov, V. I. (1996). On some boundary crossing problems for gaussian random walks. *Annals of Probability*, 24:2154–2171.

Lucas, J. M. (1976). The design and use of V-mask control schemes. *Journal of Quality Technology*, 8:1–12.

Lucas, J. M. and Saccucci, M. S. (1990). Exponentially weighted moving average control schemes: Properties and enhancements (with discussion). *Technometrics*, 32:1–29.

Montgomery, D. C. (2005). *Introduction to Statistical Quality Control.* John Wiley, New York, 5th edition.

Moustakides, G. V. (1986). Optimal stopping times for detecting changes in distributions. *Annals of Statistics*, 14:1379–1387.

Moustakides, G. V. (2008). Sequential change detection revisited. *Annals of Statistics*, 36:787–807.

Novikov, A. A. (1990). On the first passage time of an autoregressive process over a level and application to a "disorder problem". *Theory of Probability and its Applications*, 35:269–279.

Novikov, A. A. and Ergashev, B. A. (1993). Limit theorems for the time of crossing a level by an autoregressive process. In *(translation) Proceedings of Steklov Mathematical Institute*, volume 202, pages 169–186.

Novikov, A. A. and Kordzakhia, N. (2007). Martingales and exit times of AR(1) processes. Research paper 205, QFRC, University of Technology, Sydney (UTS).

Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, 41:100–114.

Page, E. S. (1961). Cumulative Sum control charts. *Technometrics*, 3:1–9.

Pollak, M. (1985). Optimal stopping times for detecting changes in distributions. *Annals of Statistics*, 13:206–227.

Roberts, S. W. (1959). Control chart tests based on geometric moving average. *Technometrics*, 1:239–250.

Roberts, S. W. (1966). A comparison of some control chart procedures. *Technometrics*, 8:411–430.

Ross, S. M. (1996). *Stochastic Processes.* Wiley, New York, 2nd edition.

Shiryaev, A. N. (1961). The problem of the most rapid detection of a distribution in a stationary process. *Soviet Mathematics*, 2:795–799.

Shiryaev, A. N. (1978). *Optimal Stopping Rules.* Springer-Verlag, New York, 2rd edition.

Shiryaev, A. N. (1995). *Probability.* Springer-Verlag, New York, Inc., Secaucus, NJ, USA, 2nd edition. Translator-R.P. Boas.

Shiryaev, A. N. (1996). Minimax optimality of the method of cumulative sums (CUSUM) in the case of continuous time. *Russian Mathematical Surveys*, 51:750–751.

Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals.* Springer-Verlag, New York. 1st.

Sitter, R. R., Hanrahan, L. P., DeMets, D., and Anderson, H. A. (1990). A monitoring system to detect increased rates of cancer incidence. *American Journal of Epidemiology*, 13:123–130.

Somerville, S. E., Montgomery, D. C., and Runger, G. C. (2002). Filtering and smoothing methods for mixed particle count distributions. *International Journal of Production Research*, 40:2991–3013.

Srivastava, M. S. and Wu, Y. (1993). Comparison of EWMA, CUSUM and Shiryayev-Roberts procedure for detecting a shift in the mean. *Annals of Statistics*, 21:645–670.

Srivastava, M. S. and Wu, Y. (1997). Evaluation of optimum weights and average run lengths in EWMA control schemes. *Communications in Statistics: Theory and Methods*, 26:1253–1267.

Sukparungsee, S. and Novikov, A. A. (2006). On EWMA procedure for detection of a change in observations via martingale approach. *KMITL Science Journal*, 6:373–380.

Sukparungsee, S. and Novikov, A. A. (2007). Analytical approximations for average run lengths in EWMA charts in case of light-tailed distributions. In *International Conference of Mathematical Sciences*, Bangi-Putrajaya, Malaysia (ICMS 2007).

Vardeman, S. B. and Jobe, J. M. (1999). *Statistical Quality Assurance Methods for Engineers.* Wiley, New York.

Vervaat, W. (1979). On a stochastic difference equation and a representation of nonnegative infinitely divisible random variables. *Advanced in Applied Probabilities*, 11:750–783.

Waldmann, K. H. (1986). Bounds for the distribution of the run length of one-sided and two-sided CUSUM quality control schemes. *Technometrics*, 28:61–67.

Westgard, J., Groth, T., Aronsson, T., and de Verdier, C. (1977). Combined Shewhart-Cusum control chart for improved quality control in clinical chemistry. *Clinical Chemistry*, 23:1881–1887.

Wieringa, J. E. (1999). *Statistical process control for serially correlated data.* Phd thesis, University of Groningen, Groningen, the Netherlands.

Woodall, W. H. (1983). The distribution of the run length of one-sided CUSUM procedures for continuous random variables. *Technometrics*, 25:295–301.

Woodall, W. H. (1997). Control charts based on attribute data: Bibliography and review. *Journal of Quality Technology*, 29:172–183.

Woodall, W. H. and Adams, B. M. (1993). The statistical design of CUSUM charts. *Quality Engineering*, 5:559–570.

Yashchin, E. (1987). Some aspects of the theory of statistical control schemes. *IBM Journal of Research and Development*, 31:199–205.

Ye, N., Borror, C., and Zhang, Y. (2002). EWMA techniques for computer intrusion detection through anomalous changes in event intensity. *Quality and Reliability Engineering International*, 18:443–451.

Ye, N., Vilbert, S., and Chen, Q. (2003). Computer intrusion detection through EWMA for autocorrelated and uncorrelated data. *IEEE Transactions on Reliability*, 52:75–82.