UNIVERSITY OF ALBERTA

Inference Problems After CUSUM Tests

BY

KEYUE DING ©

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

IN

STATISTICS

DEPARTMENT OF MATHEMATICAL SCIENCES

EDMONTON, ALBERTA

FALL 1999

0-612-46830-5

UNIVERISTY OF ALBERTA
LIBRARY RELEASE FORM

NAME OF AUTHOR: Keyue Ding

TITLE OF THESIS: Inference Problems after CUSUM Tests

DEGREE: Doctor of Philosophy

YEAR THIS DEGREE GRANTED: 1999

Department of Biostatistics
University of Rochester
Medical Center
Rochester, N. Y 14642-8630

September 2, 1999

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **Inference Problems After CUSUM Tests** submitted by **Keyue Ding** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy** in **Statistics**.

Dr. Y. Wu (Supervisor)

Dr. P. M. Hooper

Dr. E. Gombay

Dr. D. P. Wiens

Dr. M. J. Zuo

Dr. M. Woodroofe (External Examiner)
University of Michigan

Date: Aug 30, 1999.

Abstract

This thesis systematically studies the inference problems after CUSUM test. It consists of four main chapters in which different aspects of the problems are discussed.

In Chapter 2, properties of minimum point of unbalanced two-sided random walk are investigated. These results not only can be used to study maximum likelihood estimator of change point, but also can be applied to research estimate of change point after CUSUM test. Chapter 3 gives a practical method for constructing the confidence intervals of change point after CUSUM test. The method can be implemented either by tracing back some zero points of the control chart before the detection point or by tracing back certain number of items produced before the stopping time. In Chapter 4, we study the biases of estimates of change point and change magnitude, and find that biases of both estimates are quite substantial. We therefore propose a bias correction method for practical use. In Chapter 5, we further investigate properties of estimate of change magnitude, and obtain the asymptotic distribution function of the estimate. Based on the asymptotic distribution function, we derive the confidence intervals of change magnitude.

ACKNOWLEDGMENTS

First and foremost I would like to express my sincere thanks and gratitude to my supervisor Dr. Yanhong Wu for his excellent guidance and advice throughout my graduate studies.

I like to acknowledge and thank the members of my graduate committee members: Dr. E. Gombay and Dr. P. M. Hooper, for their suggestions and stimulating discussions. Also, I wish to thank my examining committee members: Dr. D. P. Wiens, Dr. M. Woodroofe and Dr. M. Zuo for their time and valuable comments.

I wish to express my pecial thanks to Dr. A. N. Al-Hussaini and Dr. Karunamuni for their help and valuable time.

Many thanks and appreciation go to Mr. Yongkang Mao, Mrs. Huiling Mao and Mr. Xiaoming Sheng for their help, friendship and fun times we often had.

Finally, I would like to express my gratitude to University of Alberta for Ph. D. Scholarship support, to Dr. Sharma for Pundit RD Sharma Memorial Graduate Award and to Mrs. Whitney for Eoin L. Whitney Scholarship.

TO MY WIFE YAN GAO

AND

MY PARENTS

Contents

1 Introduction **1**

1.1 General Description and Motivation 1

1.2 The CUSUM Procedure 4

1.2.1 Definition of CUSUM Procedure 5

1.2.2 Representation in Form of Reflected Random Walk 7

1.3 Standard One Parameter Exponential Family 8

1.4 Statement of Problem 10

2 Unbalanced Two-Sided Random Walk **18**

2.1 Introduction 18

2.2 Notation and Preliminaries 19

2.3 Moments of Minimum Point of Two-sided Random Walk 24

2.3.1 Introduction 24

2.3.2 Second Order Expansion of $E(\nu_0)$ and $E|\nu_0|$ 26

2.3.3 Second Order Expansion for $E(\nu_0^2)$ 34

2.4 Simulation Results 40

2.4.1 Two-sided Normal Random Walk 40

2.4.2 Two-sided Non-negative Exponential Random Walk 43

3 Confidence Interval for Change Point **51**

3.1 Introduction 52

3.2 Method 54

3.3 Main Results and Their Intuitive Derivations 56

3.4 Local Second Order Approximations 61

3.5 Application 68

3.5.1 Detect Increase in Mean 69

3.5.2 Detect Increase in Variance 75

4 Biases of Change Point and Change Magnitude Estimations **85**

4.1 Introduction 86

4.2 Bias of Estimate of Change Point 89

4.3 Bias of Estimate of Change Magnitude 92

4.4 Application 104

4.4.1 Detect Increase in Mean 105

4.4.2 Detect Increase in Variance 110

4.5 Error Checking for (4.9) and (4.14) 114

5 Confidence Intervals for Change Magnitude **122**

5.1 Introduction 123

5.2 Asymptotic Distribution of $\hat{\theta}_1$ 125

5.3 Simulation Results 129

5.4 Moments of N_0 and $N_M + \sigma_M$ 132

6 Summary and Topics for Future Research **148**

List of Tables

2.1 $P(\tau_{m_0'} < \infty)$, $E\nu_0$ and $E|\nu_0|$ for $\theta_0 = -0.2$ 41

2.2 $P(\tau_{m_0'} < \infty)$, $E\nu_0$ and $E|\nu_0|$ for $\theta_0 = -0.25$ 42

2.3 $P(\tau_{m_0'} < \infty)$, $E\nu_0$ and $E|\nu_0|$ for $\theta_0 = -0.3$ 42

2.4 $P(\tau_{m_0'} < \infty)$, $E\nu_0$ and $E|\nu_0|$ for $\lambda_0 = 0.85$ 44

2.5 $P(\tau_{m_0'} < \infty)$, $E\nu_0$ and $E|\nu_0|$ for $\lambda_0 = 0.8$ 45

2.6 $P(\tau_{m_0'} < \infty)$, $E\nu_0$ and $E|\nu_0|$ for $\lambda_0 = 0.75$ 45

3.1 Non-coverage Probability for $\theta_0 = \delta/2 = -0.2$ 72

3.2 Non-coverage Probability for $\theta_0 = -0.25$ 73

3.3 Non-coverage Probability for $\theta_0 = -0.30$ 73

3.4 $E^\nu[N - L_s | N > \nu]$ for $\theta_0 = -0.2$ 74

3.5 $E^\nu[N - L_s | N > \nu]$ for $\theta_0 = -0.25$ 74

3.6 $E^\nu[N - L_s | N > \nu]$ for $\theta_0 = -0.3$ 75

3.7 Non-coverage Probability for $\epsilon_0 = 0.2$ 79

3.8 Non-coverage Probability for $\epsilon_0 = 0.25$ 80

3.9 Non-coverage Probability for $\epsilon_0 = 0.3$ 81

3.10 $E^\nu[N - L_s|N > \nu]$ for $\epsilon_0 = 0.2$ 81

3.11 $E^\nu[N - L_s|N > \nu]$ for $\epsilon_0 = 0.25$ 82

3.12 $E^\nu[N - L_s|N > \nu]$ for $\epsilon_0 = 0.3$ 82

4.1 Comparison of Approximated and Simulated Bias (Normal). 109

4.2 Comparison of Approximated and Simulated Bias (Exponential). . . . 113

5.1 Coverage Probabilities of 90% Confidence Intervals 131

5.2 Coverage Probabilities of 95% Confidence Intervals 131

5.3 Coverage Probabilities of 98% Confidence Intervals 132

Chapter 1

Introduction

1.1 General Description and Motivation

In many practical and experimental situations. certain statistical properties of an observed phenomenon which evolves in time may change abruptly at some unknown point(s). The detection and characterization of such a change are key problems of interest in many scientific fields. Typical examples can be found in quality control, clinical trials, speech signals recognition and etc. In statistical literature, such problems are the so-called *change point* problems. During the last few decades, extensive researches have been carried out in this field. For some recent reviews, we refer to Siegmund (1985), Basseville and Nikiforov (1993), Ghosh and Sen (1991), Lai (1995) and the references therein.

One of the earliest applications of change point detection lies in the area of quality control, or continuous production monitoring. In this typical application, on-line quality control procedures are used to reach a decision as each new observation is collected. Consider a production process which may be *in control* (products meet customer's requirements) or *out of control* (products do not meet customer's requirements). In such a situation , changes happen when the manufacturing process deviates from the in control conditions and enters the out of control state. From both safety and quality points of view, it is necessary to detect the changes and stop the process for inspection and repair as quickly as possible after assignable causes are identified. Statistical control procedures have long been used in on-line quality control to detect changes which indicate the trend of deterioration of product quality. As a simple model, we can assume that when the process being observed is in control, the observations are iid with a specific probability distribution. When the process is out of control, this distribution changes. If a parametric approach is adopted, the changes in the distributions are reflected in changes of parameters associated with the probability distributions. For example, if we assume that the quality characteristic process X_t is characterized as a normal process. Under normal operating conditions, observations comply with a normal distribution with mean μ_0 and standard deviation σ_0, where μ_0 is the target value and σ_0 reflects the permissive variation of quality. Any shift from the target value μ_0 or increase in the standard deviation results in poor quality. So two basic types of changes could happen in the parameters:

1: *Shift from the target value (the reference mean value)* μ_0 *toward* $\mu(\mu \neq \mu_0$ *with constant standard deviation. This type of change is described as system error.*

2: *Increase in the standard deviation from* σ_0 *to* $(1+\epsilon)\sigma_0$ *(*$\epsilon > 0$*) with constant mean. This type of change is called as random error.*

Composite changes can also happen, but can be broken up into the two basic types. The problems now become how to design a statistical decision function and a decision rule that can detect these changes as quickly as possible after changes occur.

To monitor production quality, groups of random samples of size, say m, are usually taken at regular time intervals, and monitoring procedures based on either sample mean $\bar{X}_t$ or sample variance s_t^2 (sample range R_t) are plotted with appropriate control limit(s). The Shewhart, CUSUM and EWMA control charts are the most commonly used ones in practice; see Montgomery (1996) for tutorial overview of these control charts. The process is stopped at the first time when the control charts fall outside of control region(s). Then, we may try to identify the sources of disorders (changes) and isolate the assignable sources. For this, we may use the information provided by control charts until the alarm time to estimate certain process parameters after change, such as the mean, standard deviation, and so forth. These estimates may then be used to adjust the system and reset the monitoring procedure. The object of this thesis is to consider the inference problems on the change point and change magnitude after changes have been detected. Most of investigations will be focused on the corresponding problems after the sequential CUSUM test.

1.2 The CUSUM Procedure

The CUSUM procedure is a sequential statistical procedure which is used to detect a change in a probability distribution. It was proposed by Page (1954) and is commonly used in situations where data (observations) are being monitored sequentially, and its application to the on-line quality control is the most prominent example. In a typical application, we would be observing a sequence of random variables, usually representing a certain quality characteristic of a manufacturing process, sequentially, and want to stop the process as soon as possible if the distribution of the random variables shift from the original in-control distribution to the undesirable out-of-control distribution, and then inspect and adjust the system.

From statistical point of view, the data (observations) are a realization of a random process. Because of this random behavior, large fluctuations can occur in the data even if the process is in control, and these fluctuations result in *false alarms*, which are the cases when control chart fall out side of control region. On the other hand, again because of randomness, a decision rule may not be able to detect the change immediately after it occurs. So the best solution is a quick detection of change(s) with as few false alarms as possible. In practice, the optimal solution is basically a tradeoff between quick detection and few false alarms. The CUSUM procedure provides a solution to such a kind of problems.

1.2.1 Definition of CUSUM Procedure

We first consider the simple case. Suppose that $X_1, X_2, \ldots, X_\nu$ are iid r.v. with distribution function F_{θ_0}, and $X_{\nu+1}, X_{\nu+2}, \ldots$ are iid with distribution function F_{θ_1}, where F_{θ_0} and F_{θ_1} are known while ν is unknown. ν is the so called change point. We want to stop the process as soon as possible after ν. For simplicity, suppose that f_{θ_0} and f_{θ_1} are density functions of F_{θ_0} and F_{θ_1} with respect to some reference measure. Then at time n, the log likelihood ratio statistic under the hypothesis that the distribution has changed by time n (i.e. hypothesis $\nu < n$) with respect to the hypothesis that change has not occurred (i.e. hypothesis $\nu > n$) is

$$\begin{aligned}
\Lambda_n &= \max_{0 \le \nu \le n} \ln\left\{\frac{\prod_{i=1}^{\nu} f_{\theta_0}(X_i) \prod_{i=\nu+1}^{n} f_{\theta_1}(X_i)}{\prod_{i=1}^{n} f_{\theta_0}(X_i)}\right\} \\
&= \max_{0 \le \nu \le n} \ln \prod_{i=\nu+1}^{n} \frac{f_{\theta_1}(X_i)}{f_{\theta_0}(X_i)} \\
&= \max_{0 \le \nu \le n} (\tilde{S}_n - \tilde{S}_\nu)
\end{aligned}$$

$$= \tilde{S}_n - \min_{0 \le \nu \le n} \tilde{S}_\nu$$

$$= \tilde{S}_n - \tilde{m}_n, \qquad (1.1)$$

where

$$\tilde{S}_k = \sum_{i=1}^{k} \ln\left(\frac{f_{\theta_1}(X_i)}{f_{\theta_0}(X_i)}\right), \quad \tilde{S}_0 = 0,$$

and

$$\tilde{m}_k = \min_{0 \le i \le k} \tilde{S}_i$$

Thus, we would like to stop the process and reject the null hypothesis at the first time when Λ_n exceeds some specified threshold, say b, i.e. at time

$$\tilde{N}_b = \inf\{n : \Lambda_n > b\}. \qquad (1.2)$$

This is the CUSUM procedure.

The key property of the log likelihood ratio $\tilde{S}_n$ is that it has a negative drift before change, and a positive drift after change. Therefore, the relevant information, as far as the change is concerned, lies in the difference between the value of log likelihood ratio and its current minimum value; and it is clear that the detection rule is nothing but a comparison between the cumulative sum $\tilde{S}_n$ and an adaptive threshold $\tilde{m}_n + b$. Because of $\tilde{m}_n$, this threshold not only varies on-line, but also keeps complete information of the past observations. This provides the basis for the optimal properties of CUSUM procedure.

1.2.2 Representation in Form of Reflected Random Walk

The CUSUM procedure can be represented in form of reflected random walks. In the problem described in the previous subsection, when $X_1, X_2, \ldots$ are iid with distribution F_{θ_0}, the process $\{\tilde{S}_k, k \geq 0\}$, being the sum of iid random variables, is called a random walk. The corresponding reflected random walk $\{W_k, k \geq 0\}$ with reflecting barrier at 0 is defined by

$$W_0 = 0, W_n = \max\{0, W_{n-1} + \ln(\frac{f_{\theta_1}(X_n)}{f_{\theta_0}(X_n)})\}, \quad for \quad n > 0. \tag{1.3}$$

The CUSUM procedure can be equivalently defined as stopping at the first time

$$N_b' = \inf\{n, W_n > b\}. \tag{1.4}$$

The relationship between the reflected random walk $\{W_n, n \geq 0\}$ and the underlying random walk $\{\tilde{S}_n, n \geq 0\}$ is as follows. The increment between W_{n-1} and W_n is the same as that between $\tilde{S}_{n-1}$ and $\tilde{S}_n$ when the increment $\ln(\frac{f_{\theta_1}(X_n)}{f_{\theta_0}(X_n)})$ is not causing W_n to be negative. Otherwise, i.e. adding $\ln(\frac{f_{\theta_1}(X_n)}{f_{\theta_0}(X_n)})$ to W_{n-1} would cause W_n to be negative, then the process is reflected at 0, by taking $W_n = 0$ instead of of taking the full increment $\ln(\frac{f_{\theta_1}(X_n)}{f_{\theta_0}(X_n)})$. Reflected random walks and their related theory play a central role in studying CUSUM procedure (Chang (1992), Siegmund (1985)). Details will be given in Chapter 3.

1.3 Standard One Parameter Exponential Family

To describe the problems concerned in the thesis under a unified framework, we introduce the concept of *standard one parameter exponential family*. From now on, we will restrict our attention to this case. In this subsection, we give some notations and facts related to the standard one parameter exponential family. It should be pointed out that similar material is also available from Siegmund (1985) and Chang (1992).

A family of distribution $\{F_\theta, \theta \in \Theta\}$ indexed by the parameter θ is called a standard one parameter exponential family if the distribution F_θ is of the form

$$dF_\theta(x) = \exp(\theta x - \psi(\theta))dF_0(x), \tag{1.5}$$

where Θ contains an interval with 0 in it, and the function ψ is normalized so that $\psi(0) = \psi'(0) = 0$, $\psi''(0) = 1$.

It can be easily verified that

$$\psi'(\theta) = \int x dF_\theta(x), \psi''(\theta) = \int (x - \psi'(\theta))^2 dF_\theta(x),$$

and $\psi'(\theta) <, =, or > 0$ according to $\theta <, =, or > 0$.

We will generally assume that the observations $X_1, X_2, \ldots$ are iid random variables having distribution F_θ which belongs to the standard one parameter exponential family, and the notations P_θ, E_θ will denote the corresponding probability and expectation. So $E_\theta(X_1) = \psi'(\theta)$.

The function ψ, the cumulant generating function of F_0, is convex, and its Taylor expansion at $\theta = 0$ is

$$\psi(\theta) = \frac{1}{2}\theta^2 + \frac{\gamma}{6}\theta^3 + \frac{\kappa}{24}\theta^4 + O(\theta^5),$$

where $\gamma = E_0(X_1^3)$ and $\kappa = E_0(X_1^4) - 3$. From this, $\mu = \psi'(\theta)$ has the expansion

$$\mu = \theta + \frac{\gamma}{2}\theta^2 + \frac{\kappa}{6}\theta^3 + O(\theta^4).$$

For small $\theta \neq 0$, there corresponds exactly one $\tilde{\theta} \neq 0$, necessarily opposite sign, for which $\psi(\theta) = \psi(\tilde{\theta})$. For $\theta_0 < 0 < \theta_1$, we conveniently set

$$\Delta_i = \theta_i - \tilde{\theta}_i, for \quad i = 0, 1.$$

Note that $\Delta_0 < 0 < \Delta_1$.

Suppose that, in (1.3), f_{θ_0} and f_{θ_1} are density functions of F_{θ_0} and F_{θ_1} which belong to the above defined exponential family and $\theta_0 < 0 < \theta_1$ satisfy $\psi(\theta_0) = \psi(\theta_1)$. Then the CUSUM procedure, after rescaling the parameters, can be written as stopping at the first time

$$\tau_d = \inf\{n : S_n - m_n > d\},$$

where

$$S_0 = 0, S_k = \sum_{i=1}^{k} X_i, for \quad k > 0, and \quad m_n = \min_{0 \le k \le n} S_k.$$

Or, equivalently, rewrite as making an alarm at time

$$N = \inf\{n : T_n > b\}, \tag{1.6}$$

where

$$T_0 = 0, T_n = \max\{0, T_{n-1} + X_n\}, \quad for \quad n > 0.$$

Of particular important examples of the above defined standard exponential families are:

(**1**) The normal family, where $\psi(\theta) = \frac{1}{2}\theta^2$ and F_0 is the distribution function of standard normal.

(**2**) The shifted non-negative exponential family, where $\psi(\theta) = -\theta - ln(1-\theta)$ and $f_0(x) = e^{-(x+1)}I_{(x>-1)}$. If the density function of usual non-negative exponential family is $\frac{1}{\lambda}e^{-\frac{x}{\lambda}}$, then the transformation of variable to the standardized variable is $X - 1$ and $\theta = 1 - \frac{1}{\lambda}$.

(**3**) The $Gamma(\alpha, \beta)$ family with α being fixed, where $\psi(\theta) = -\sqrt{\alpha}\theta - \alpha ln(1 - \frac{\theta}{\sqrt{\alpha}})$ and $f_0(x) = \frac{\alpha^{\alpha/2}}{\Gamma(\alpha)}(x+\sqrt{\alpha})^{\alpha-1}e^{-\sqrt{\alpha}(x+\sqrt{\alpha})}I_{(x>-\sqrt{\alpha})}$. If the density function of usual $Gamma(\alpha, \beta)$ family is $\frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-\frac{x}{\beta}}$, then the transformation of variable to the standardized variable is $\frac{X-\alpha}{\sqrt{\alpha}}$ and $\theta = \sqrt{\alpha}(1 - \frac{1}{\beta})$.

1.4 Statement of Problem

Ever since it was introduced, the CUSUM procedure has attracted heavy attention of both theoretical and applied statisticians, and its properties and operating characteristics have been extensively studied. Page (1954, 1955, 1957, 1961, 1963) showed that CUSUM procedure performs better in detecting small and moderate changes than other commonly used procedures, such as, Shewhart chart and EWMA chart. Barnard (1959) called CUSUM procedure "a fundamental change in the classical procedure". Roberts (1966), Pollak and Siegmund (1985), Srivastava and Wu (1993) compared it with other commonly used procedures, such as EWMA and Shiryayev-Roberts procedures. Lorden (1971), who first gave the definition (1.1) of the procedure, proved that the procedure has an asymptotic optimality in terms of the average conditional delay in detection time. The asymptotic optimality is developed on the basis that CUSUM procedure is a sequence of *Wald Sequential Probability Ratio Tests* (SPRT). Moustakides(1986) showed that Lorden's asymptotic optimality of CUSUM procedure also holds in a non-asymptotic sense. In the literature, a commonly used performance measure of control procedures is the *average run length* (ARL) of samples. However, the exclusive use of the ARL has been criticized, and use of percentage points of the run length distribution has been recommended (Bissell, 1969). To evaluate the operating characteristics, many techniques have been developed both numerically and approximately. For example, Brook and Evans (1972) gave the algorithm to calculate ARLs based on the Markov structure of the CUSUM process, and related numerical algorithms were also given by Woodall (1983), and Waldmann (1986). These algorithms have been developed to the point that, given a computer and enough time, we can essentially obtain the true probability distribution of ARLs at a desired value. Siegmund (1985) provided a very accurate approximation formula for calculating ARL's.

In spite of the extraordinary accomplishments listed above on the CUSUM procedures, those researches are mainly focused on the design problems and the operating characteristics, while the inference problems after CUSUM test received less attention. More precisely, there is no systematic investigation on the estimations of the change-point and change magnitude after the detection. However, the estimations and construction of confidence intervals for the change-point and change magnitude are a rather crucial issue in practice. For example, in quality control, after the change has been detected, we want to stop the process for inspection and repair. Apparently, we want to decide how many items should be inspected to guarantee the uniformity of the quality. That means, we need to estimate the change point and construct confidence intervals for the change point. From process adjustment and control point of view, we would also like to estimate the change magnitude as a reference to adjust the system for the next stage. Hinkley (1971) gave a natural way of estimating change-point ν by $\hat{\nu} = \max\{n < N : T_n = 0\}$, and in the case of detecting change in the mean of a process, estimating process mean after change by $\hat{\mu}_1 = T_N/(N - \hat{\nu})$. Simulation studies were conducted without conditioning on $N > \nu$. The results show that negative bias for $\hat{\nu}$ appears mainly due to the false alarm possibility; and positive biases for $\hat{\mu}$ occurs due to the overshoot at the boundary crossing time. He also pointed out the necessity of further investigation. However, in considering these problems , conditioning on $N > \nu$ is very crucial, otherwise the estimations are meaningless. On the other hand, simulation cannot provide any of usual sorts of insights which one might hope could be provided by an analytic approximation. Theoretical results are also necessary for developing simple methods for bias correction. The main goal of this thesis is to solve the above questions by some analytic approximations.

The thesis is organized in the following manner. In each chapter, different aspects of the problem are investigated. Chapter 2 presents some fundamental results that are necessary for a full development of subsequent chapters. Much of notation is defined, and some preliminary results available from literature are given. Then properties of the so-called *two-sided random walk* are investigated, and these results build the basis for studying inference problems after CUSUM test.

Chapter 3 considers the problem of constructing confidence intervals for change point based on the information contained in the CUSUM control chart. A practical method for constructing confidence intervals of change point is proposed by tracing back a certain number of zero points of the control chart. The method is implemented by approximating the non-coverage probability. Furthermore, the average lengths of confidence intervals are derived, and comparison of our approximations with the simulation results is also conducted.

Chapter 4 turns to studying biases of the estimates of change point and change magnitude. Under certain conditions, the second order approximations for the biases of change point and change magnitude are obtained. Simulation studies are also performed to check the accuracy of our approximations. Finally, a simple bias correction method is thus proposed for practical use.

In Chapter 5, the estimate of change magnitude is further investigated. At first, its asymptotic distribution function is derived. Then based on the asymptotic distribution function, the confidence intervals for the change magnitude are obtained.

Finally, in Chapter 6, major results are summarized, possible extensions are discussed and some topics for future research are suggested.

REFERENCES

Barnard, G. A. (1959). Control charts and stochastic process. *J. R. Statist. Soc. Ser. B*, **16**, 151-174.

Basseville, M., Nikiforov, I. (1993). Detection of abrupt changes. Englewood Cliffs, NJ: Prentice-Hall.

Bissell, A. F. (1969). CUSUM techniques for quality control(with discussion). *Applied Statistics*, **18**, 1-30.

Brook, D., Evans, D. (1972). An approach to the probability distribution of cusum run length. *Biometrika*, **59**, 539-549.

Chang, J.T.(1992). Moments of the first ladder height of random walks with small drift. *Ann. Appl. Probab.* **2**, 714-738.

Ghosh, B. K., Sen, P. K. (1991). Handbook of sequential analysis. Marcel Dekker. Inc. New York.

Hinkley, D. V.(1971). Inference about the change point from cumulative sum tests. *Boimetrika*, **58**, 509-523.

Lai, T. L.(1995). Sequential changepoint detection in quality control and dynamical systems. *J. R. Statist. Soc. Ser. B*, **57**, 613-658.

Lorden, G.(1971). Procedures for reacting to a change in distribution. *Ann. Statist.* **6**, 1897-1908.

Montgomery, D. C. (1996). Introduction to statistical quality control. 3rd ed., New York: Wiley.

Moustakides, G. V.(1986). Optimal stopping times for detecting changes in distribution. *Ann. Statist.* , **14**, 1379-1387.

Page, E. S.(1954). Continuous inspection schemes.*Biometrika*, **41**, 100-114.

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

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

Page, E. S.(1961). Cumulative sum charts. *Technometrics*, **3**, 1-9.

Page, E. S.(1963). Controlling the standard deviation by CUSUM and warning lines. *Technometrics*, **5**, 307-315.

Pollak, M. and Siegmund, D.(1985). A diffusion process and its applications to detecting a change in the drift of Brownian motion. *Biometrika*, **72**, 267-280.

Pollak, M. and Siegmund, D.(1986). Convergence of quasi-stationary to stationary distribution for stochastically monotone Markov processes. *J. Appl. Probab.* , **23**, 215-220.

Roberts, S. M.(1966). A comparison of some control chart procedures. *Technometrics*, **8**, 411-430.

Siegmund, D.(1985). *Sequential Analysis: Tests and Confidence Intervals*. Springer, New York.

Srivastava, M. S. and Wu, Y.(1993). Comparison of EWMA, CUSUM and Shiryayev-Roberts procedures for detecting a shift in the mean. *Ann. Statist.*, **21**, 645-670.

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.

Woodall, W. H. (1983). The distribution of the run length of one-sided cusum procedures for continuous random variables. *Technometrics*, **25**, 295-301.

Chapter 2

Unbalanced Two-Sided Random Walk

2.1 Introduction

The purpose of this chapter is to set the stage for subsequent chapters. It contains background materials and fundamental theory that are necessary for full development of the upcoming chapters.

In Section 2.2, we first introduce some notations which will be used throughout the thesis, and then present some preliminary results available from literature. For the preliminary results, we first give two results about the local expansions of the moments of ladder's height and the correlation between ladder epoch and ladder's height, then we list two results which are regarding the convergence rates of overshoot and the correlation between the crossing time and the overshoot. Finally, we introduce the so-called *Wald's Likelihood Ratio Identity*, which is then used to derive the convergence rate for the boundary crossing probability.

In Section 2.3, properties of two-sided random walk are investigated. These results will be applied as essential techniques in later chapters to considering inference problems after CUSUM test; besides, they can also be used to study properties of maximum likelihood estimator for change point; and in term of random walk, these results have their own independent interest as well.

Finally, in Section 2.4, simulation results are given to show the accuracy of approximations obtained in Section 2.3.

2.2 Notation and Preliminaries

In this section, we introduce some notations and preliminaries that will be used throughout the thesis.

Suppose that $X_1, X_2, \ldots$ are independent and identically distributed (iid) random variables. As usual, let $S_0 = 0$ and $S_k = \sum_{i=1}^{k} X_i$ for $k > 0$. The process $\{S_k, k \geq 0\}$ is called a random walk. The ladder epoch τ_+ and the descending ladder epoch τ_- are defined as

$$\tau_+ = \inf\{k > 0 : \; S_k > 0\}, \tag{2.1}$$

and

$$\tau_- = \inf\{k > 0 : \ S_k \le 0\}, \tag{2.2}$$

and the ladder height is S_{τ_+} (S_{τ_-}). More generally, define the first passage time at boundary x as

$$\tau_x = \inf\{k > 0 : \ S_k \le x\}, \quad for \quad x < 0, \tag{2.3}$$

$$\tau_x = \inf\{k > 0 : \ S_k > x\}, \quad for \quad x > 0, \tag{2.4}$$

and denote the overshoot at x as

$$R_x = S_{\tau_x} - x. \tag{2.5}$$

Now let $\{F_\theta, \theta \in \Theta\}$ be a standard one parameter exponential family as described in Section 1.3, and $X_1, X_2, \ldots$ be iid with distribution function F_θ for some θ in the interior of Θ. Define the characteristic function $\phi_\theta(t)$ by

$$\phi_\theta(t) = \int_{-\infty}^{\infty} e^{itx} F_\theta(dx).$$

Stone (1965) defined that the distribution F_θ is *strongly non-lattice* if

$$\lim_{|t|\to\infty} \sup |\phi_\theta(t)| < 1.$$

With above defined notations, we now present some results that will be used extensively throughout the thesis. The first result gives a local expansion on the moments of ladder height, which was first discussed by Siegmund (1979), and further studied by Chang (1992).

Theorem 2.1 [Chang (1992)]: *Suppose that $\{F_\theta : \theta \in \Theta\}$ is a standard one parameter exponential family, with F_0 being strongly non-lattice. Then for any $k > 0$, as $0 < \theta \to 0$ $(0 > \theta \to 0)$, we have*

$$E_\theta S_{\tau_+}^k = E_0 S_{\tau_+}^k + \frac{k}{k+1}(E_0 S_{\tau_+}^{k+1})\theta + O(\theta^2).$$

$$(E_\theta S_{\tau_-}^k = E_0 S_{\tau_-}^k + \frac{k}{k+1}(E_0 S_{\tau_-}^{k+1})\theta + O(\theta^2).)$$

The second result is concerning the correlation between ladder epoch and ladder height.

Theorem 2.2 [Chang (1992)]: *Under the conditions of Theorem 2.1, for any $k > 0$, as $0 < \theta \to 0$ $(0 > \theta \to 0)$, we have*

$$\mu E_\theta(\tau_+ S_{\tau_+}^k) = \frac{1}{k+1} E_0 S_{\tau_+}^{k+1} + O(\theta).$$

$$(\mu E_\theta(\tau_- S_{\tau_-}^k) = \frac{1}{k+1} E_0 S_{\tau_-}^{k+1} + O(\theta).)$$

where $\mu = \psi'(\theta) = E_\theta X_1$.

The next two results are related to the convergence rate of the renewal function. Stone (1965) showed that the convergence rate for the renewal function is exponentially fast. Siegmund (1979) further showed that in the standard one parameter exponential family, the result holds uniformly in a neighborhood of 0 if the baseline distribution function is strongly non-lattice.

Because of the exponential convergence rate, under any distribution P_θ for $\theta \geq 0$, R_x approaches a limiting distribution as $x \to \infty$. Let R_∞ denote a random variable with this limiting distribution. The moments of R_∞ are given by

$$\rho_+^k(\theta) = E_\theta(R_\infty^k) = \frac{E_\theta(S_{\tau_+}^{k+1})}{(k+1)E_\theta(S_{\tau_+})} \tag{2.6}$$

for $k > 0$. For convenience, let $\rho_+(\theta) = \rho_+^1(\theta)$, $\rho_+ = \rho_+(0)$. Analogously, we define $\rho_-^k(\theta)$, $\rho_-(\theta)$ and ρ_- for $\theta < 0$.

Using the exponential convergence rate of the renew function, Chang (1992) derived the convergence rate for the distribution function of the overshoot R_x.

Theorem 2.3 [Chang (1992)]: *Under the conditions of Theorem 2.1, there exist positive θ^*, r and C such that*

$$|P_\theta(R_x < y) - P_\theta(R_\infty < y)| \leq Ce^{-r(x+y)},$$

$$|E_\theta(e^{-\Delta R_x}) - E_\theta(e^{-\Delta R_\infty})| \leq C\Delta e^{-rx},$$

uniformly for $\theta \in [0, \theta^*]$, $\Delta = \theta - \tilde{\theta}$ and $x, y \geq 0$.

A relation that will be used in this paper is

$$(E_0 S_{\tau_-})(E_0 S_{\tau_+}) = -\frac{1}{2}, \tag{2.7}$$

which hold for any standard one parameter exponential family. This result can be derived by differentiating the Wiener-Hopf factorization

$$(1 - E_0 e^{itS_{\tau_+}})(1 - E_0 e^{itS_{\tau_-}}) = 1 - E_0 e^{itX_1} \tag{2.8}$$

twice, and then letting $t = 0$.

Next result is a modified version of Theorem 3.1 of Chang (1992) which generalizes a result of Lai and Siegmund (1979) about the limit of the covariance between the crossing time and the overshoot.

Theorem 2.4 [Chang (1992)]: *There exist positive θ^*, r and C such that uniformly for $\theta \in [0, \theta^*]$,*

$$|\mu Cov_\theta(\tau_x, R_x) - \int_0^\infty (E_\theta R_x - E_\theta R_\infty) P_\theta(\tau_{-x} = \infty) dx| \leq Ce^{-rx}.$$

Another fundamental tool we will use repeatly is Wald's Likelihood Ratio Identity. We will use it in the following form. Let $X_1, X_2, \ldots$ be as above, $\Im_n$ be the sigma field generated by $X_1, \ldots, X_n$, τ be a stopping time adapted to $\{\Im_n\}$, and suppose Y is measurable with respect to $\Im_\tau$. Then we have

Theorem 2.5 [Wald's Likelihood Ratio Identity]: *For any θ, $\theta' \in \Theta$, and a $\Im_\tau$ measurable variable Y,*

$$E_\theta(Y; \tau < \infty) = E_{\theta'}[Ye^{(\theta-\theta')S_\tau - \tau(\psi(\theta)-\psi(\theta'))}; \tau < \infty].$$

For the proof of theorem 2.5, we refer to Siegmund (1985).

Finally, using Theorem 2.5 and 2.3, we derive the convergence rate for the boundary crossing probability $P_\theta(\tau_{-x} < \infty)$. Related results were discussed by Carlsson (1983), Klass (1983) and Wu (1999).

Theorem 2.6: *There exist $\theta^* > 0$ and positive constants C and r such that uniformly for $\theta \in [0, \theta^*]$,*

$$|P_\theta(\tau_{-x} < \infty)/e^{-\Delta x} E_{\tilde{\theta}} e^{\Delta R_{-\infty}} - 1| \leq C \Delta e^{-rx},$$

where $\tilde{\theta} < 0$ satisfies $\psi(\theta) = \psi(\tilde{\theta})$, and $\Delta = \theta - \tilde{\theta}$.

2.3 Moments of Minimum Point of Two-sided Random Walk

2.3.1 Introduction

Let $\{X_i\}$ be independent random variables with distribution function F_{θ_0} for $i < 0$ and F_{θ_1} for $i > 0$, where $\theta_0 < 0 < \theta_1$, and F_{θ_0} and F_{θ_1} belong to a standard one parameter exponential family as defined in Section 1.3. The two-sided random walk is defined as

$$Z_n = \begin{cases} \sum_{i=1}^n X_i & for \quad n > 0 \\ 0 & for \quad n = 0 \\ -\sum_{i=n}^{-1} X_i & for \quad n < 0, \end{cases} \tag{2.9}$$

and denote the minimum point ν_0 of Z_n as

$$Z_{\nu_0} = \min_{-\infty < n < \infty} Z_n. \tag{2.10}$$

Then ν_0 is the maximum likelihood estimate of change point. Thus the behavior of ν_0 is of particular interest in estimating and making inference about the change point. For details of how to relate the two-sided random walk to the maximum likelihood estimator of the change point, please refer to Hinkley (1970).

In this section, we give second order expansions for the first two moments of ν_0 under the condition that θ_0 and θ_1 approach zero at the same order. These results can be used to study the bias and variance of the maximum likelihood estimator for the change point. Wu (1999) considered similar problems for the case that $\{X_i\}$ are from a normal population with different means in each side of random walk. which is a special case of the question discussed in this section. The fundamental tool we will use is the theorems given in section 2.2.

For the convenience of presentation, we introduce some standard notations which are used throughout this section. Denote

$$m_0 = \inf_{-\infty<n<0} Z_n = -\sup_{-\infty<n<0}(-Z_n) = -M_0; \quad and \quad m_1 = \inf_{0<n<\infty} Z_n,$$

and

$$m_0' = \min(0, m_0), \quad M_0' = \max(0, M_0) \quad and \quad m_1' = \min(0, m_1).$$

Further, we denote ν_2 as the minimum point of $S_n = Z_n$ for $n > 0$, i.e. $Z_{\nu_2} = m_1$. Finally, we use $P_{\theta_1}(.)$ to represent the probability measure associated with the random walk $\{S_n\}$ for $n > 0$, and $P(.)$ the probability measure associated with $\{Z_n\}$ for $-\infty < n < \infty$ when there is no confusion.

In the next two subsections, we give the second order expansions for the first and second moments of ν_0.

2.3.2 Second Order Expansion of $E(\nu_0)$ and $E|\nu_0|$

First write

$$E(\nu_0) = E[\nu_0; \nu_0 < 0] + E[\nu_0; \nu_0 > 0]$$

$$= E[\nu_0; m_0 < m_1'] + E[\nu_0; m_1 < m_0']. \quad (2.11)$$

As the two terms on the right hand side of (2.11) have similar structures, we shall only give the detailed derivation for $E[\nu_0; m_1 < m_0']$.

From the strong Markov property of $S_n = \sum_{i=1}^{n} X_i$ for $n > 0$, we can write

$$E[\nu_0; m_1 < m_0'] = E[\nu_0; \tau_{m_0'} < \infty]$$

$$= E[\tau_{m_0'}; \tau_{m_0'} < \infty] + E[\nu_2]P(\tau_{m_0'} < \infty), \quad (2.12)$$

where we assume that $\tau_0 = \tau_-$ when $m_0' = 0$ as in the following discussion.

We first give the local expansions of some quantities related to the ladder epoch $\tau_+(\tau_-)$.

Lemma 2.1: *As* $\theta_0, \theta_1 \to 0$,

$$P_{\theta_1}(\tau_- = \infty) = -\Delta_1 E_0 S_{\tau_-} e^{\frac{\Delta_1}{2}\rho_-} + O(\Delta_1^3).$$

$$P_{\theta_0}(\tau_+ = \infty) = -\Delta_0 E_0 S_{\tau_+} e^{\frac{1}{2}\Delta_0 \rho_+} + O(\Delta_0^3).$$

$$E_{\theta_1}[\tau_-;\tau_- < \infty] = \frac{1}{\tilde{\mu}_1} E_0 S_{\tau_-} e^{\frac{\Delta_1}{2}\rho_-} + O(\Delta_1),$$

$$E_{\theta_0}[\tau_+;\tau_+ < \infty] = \frac{1}{\tilde{\mu}_0} E_0 S_{\tau_+} e^{\frac{\Delta_0}{2}\rho_+} + O(\Delta_0),$$

where $\tilde{\mu}_1 = \psi'(\tilde{\theta}_1)$ *and* $\tilde{\mu}_o = \psi'(\tilde{\theta}_0)$.

Proof: From Wald's Likelihood Ratio Identity, as $\theta_1 \to 0$, we have

$$P_{\theta_1}(\tau_- < \infty) = E_{\tilde{\theta}_1} e^{\Delta_1 S_{\tau_-}}$$

$$= 1 + \Delta_1 E_{\tilde{\theta}_1} S_{\tau_-} + \frac{\Delta_1^2}{2} E_{\tilde{\theta}_1} S_{\tau_-}^2 + O(\Delta_1^3)$$

$$= 1 + \Delta_1 (E_0 S_{\tau_-} + \frac{\tilde{\theta}_1}{2} E_0 S_{\tau_-}^2) + \frac{\Delta_1^2}{2} E_0 S_{\tau_-}^2 + O(\Delta_1^3)$$

$$= 1 + \Delta_1 E_0 S_{\tau_-} (1 + \frac{\Delta_1}{4} \frac{E_0 S_{\tau_+}^2}{E_0 S_{\tau_+}}) + O(\Delta_1^3)$$

$$= 1 + \Delta_1 E_0 S_{\tau_-} e^{\frac{\Delta_1}{2}\rho_-} + O(\Delta_1^3),$$

where in the third equation from last, we used Theorem 2.1 and the fact that $\tilde{\theta}_1 = -\theta_1 + O(\Delta_1^2)$. This completes the proof of the first equation of the lemma.

To prove the third equation of this lemma, again by Wald's Likelihood Ratio Identity and Theorem 2.2, we have

$$E_{\theta_1}(\tau_-, \tau_- < \infty) = E_{\tilde{\theta}_1}(\tau_- e^{\Delta_1 S_{\tau_-}})$$

$$= E_{\tilde{\theta}_1}(\tau_- + \Delta_1 \tau_- S_{\tau_-}) + O(\Delta_1)$$

$$= \frac{1}{\tilde{\mu}_1} E_{\tilde{\theta}_1} S_{\tau_-} + \frac{1}{2\tilde{\mu}_1} \Delta_1 E_0 S_{\tau_-}^2 + O(\Delta_1)$$

$$= \frac{1}{\tilde{\mu}_1}(E_0 S_{\tau_-} + \frac{\tilde{\theta}_1}{2} E_0 S_{\tau_-}^2) + \frac{\Delta_1}{2} E_0 S_{\tau_-}^2 + O(\Delta_1)$$

$$= \frac{1}{\tilde{\mu}_1} E_0 S_{\tau_-} e^{\frac{1}{2}\Delta_1 \rho_-} + O(\Delta_1).$$

Similarly, we can prove the other two equations.

Next three lemmas give the second order expansions for the three quantities appeared on the right hand side of (2.12) respectively.

Lemma 2.2: *As* $0 < \theta_1 \to 0$,

$$E_{\theta_1}(\nu_2) = \frac{E_{\theta_1}[\tau_-; \tau_- < \infty]}{P_{\theta_1}(\tau_- = \infty)} = -\frac{1}{\tilde{\mu}_1 \Delta_1} + O(1).$$

Proof: We only have to note that ν_2 is equivalent in distribution to $\sum_{i=1}^{K-1} \tau_-^{(i)}$, where $\tau_-^{(i)}$'s are iid r.v., which have the same distribution as that of $[\tau_- | \tau_- < \infty]$, and K is an independent geometric r.v. with $p = P_{\theta_1}(\tau_- = \infty)$.

Lemma 2.3: *As* $\theta_1, |\theta_0| \to 0$ *at the same order,*

$$P(\tau_{m_0'} < \infty) = \tilde{p}_0 + O(\Delta_1^3).$$

where

$$\tilde{p}_0 = C_1 e^{(C_2/C_1)\Delta_0 \Delta_1}, \tag{2.13}$$

with

$$C_1 = \frac{-\Delta_0}{\Delta_1 - \Delta_0} e^{\Delta_1(\rho_+ + \rho_-)},$$

$$C_2 = \frac{1}{2}\rho_+^2 + \frac{1}{2} + E_0 S_{\tau_+} \rho_- - \frac{1}{2}(\rho_+^2 - r_1) + \frac{1}{2}(\rho_-^2 - r_0) - C_0,$$

and

$$C_0 = \int_0^\infty E_0(R_{-x} - \rho_-) dE_0(R_x - \rho_+),$$

$$r_1 = \frac{E_0(S_{\tau_+}^3)}{3E_0(S_{\tau_+})} - \rho_+^2, r_0 = \frac{E_0(S_{\tau_-}^3)}{3E_0(S_{\tau_-})} - \rho_-^2.$$

Proof: Conditioning on whether $m_0' = 0$ or not, we can write

$$P(\tau_{m_0'} < \infty)$$

$$= P_{\tilde{\theta}_1}(\tau_- < \infty)P_{\tilde{\theta}_0}(\tau_+ = \infty) + P_{\tilde{\theta}_1}(\tau_{-M_0} < \infty; M_0 > 0). \tag{2.14}$$

The first term of (2.14) can be evaluated by results of Lemma 2.1. For the second term, first from Appendix 1, we have

$$E_{\tilde{\theta}_1} e^{\Delta_1 R_{-\infty}} = e^{\Delta_1 \rho_-} + O(\Delta_1^3). \tag{2.15}$$

Combining (2.15) and Theorem 2.6, we obtain that

$$P_{\theta_1}(\tau_{-x} < \infty) = e^{-\Delta_1(x+\rho_-)}(1 + O(\Delta_1^3) + O(\Delta_1 e^{-rx})). \tag{2.16}$$

uniformly for small $\theta_1 > 0$ and some $r > 0$ as $x \to \infty$. Using (2.16) and Wald's Likelihood Ratio Identity, we can write

$$\begin{aligned}
P_{\theta_1}(\tau_{-M_0} < \infty; M_0 > 0) &= -\int_0^\infty P_{\theta_1}(\tau_{-x} < \infty) dP_{\theta_0}(M_0 > x) \\
&= -\int_0^\infty P_{\theta_1}(\tau_{-x} < \infty) dP_{\theta_0}(\tau_x < \infty) \\
&= -\int_0^\infty E_{\tilde{\theta}_1}(e^{\Delta_1 S_{\tau_{-x}}}) d(E_{\tilde{\theta}_0}(e^{\Delta_0 S_{\tau_x}}))
\end{aligned}$$

$$\begin{aligned}
= \quad & -\textstyle\int_0^\infty e^{-\Delta_1(x-\rho_-)} de^{\Delta_0(x+\rho_+)} \\
& -\textstyle\int_0^\infty e^{-\Delta_1(x-\rho_-)} d[e^{\Delta_0(x+\rho_+)}(E_{\tilde{\theta}_0} e^{\Delta_0(R_x-\rho_+)} - 1)] \\
& -\textstyle\int_0^\infty e^{-\Delta_1(x-\rho_-)} [E_{\tilde{\theta}_1}(e^{\Delta_1(R_{-x}-\rho_-)} - 1)] de^{\Delta_0(x+\rho_+)}
\end{aligned}$$

$$-\int_0^\infty e^{-\Delta_1(x-\rho_-)}[E_{\tilde{\theta}_1} e^{-\Delta_1(R_{-x}-\rho_-)} - 1]d[P_{\theta_0}(\tau_x < \infty) - e^{\Delta_0(x+\rho_+)}]. \qquad (2.17)$$

The first term of (2.17) can be calculated directly as

$$-\frac{\Delta_0}{\Delta_1 - \Delta_0} e^{\Delta_1\rho_- + \Delta_0\rho_+}.$$

For the third term, we use the fact that $E_0 R_{-x} - \rho_-$ is exponentially bounded as $x \to \infty$ from the strong renewal theorem and therefore is integrable. By the fact that

$$E_{\tilde{\theta}_1}(e^{\Delta_1(R_{-x}-\rho_-)} - 1) = \Delta_1 E_0(R_{-x} - \rho_-) + O(\Delta_1^2).$$

The third term of (2.17) is approximately

$$-\Delta_0\Delta_1 \int_0^\infty (E_0 R_{-x} - \rho_-)dx + O(\Delta_1^3).$$

Similarly, by integrating by part and note that $R_0 = S_{\tau_+}$, the second term is equal to

$$e^{\Delta_1\rho_-}[P_{\theta_0}(\tau_+ < \infty) - e^{\Delta_0\rho_+}] - \Delta_0\Delta_1 \int_0^\infty (E_0 R_x - \rho_+)dx + O(\Delta_1^3).$$

For the fourth term of (2.17), from Appendix 2, we know that

$$\frac{d}{dx}(P_{\theta_0}(\tau_x < \infty) - e^{\Delta_0(x+\rho_+)}) = \Delta_0 \frac{d}{dx} E_0 R_x + O(\Delta_0^2). \qquad (2.18)$$

Thus, the fourth term of (2.17) equals to

$$-\Delta_0\Delta_1 \int_0^\infty (E_0 R_{-x} - \rho_-)d(E_0 R_x - \rho_+) + O(\Delta_1^3).$$

We complete the proof of Lemma 2.3 by the following facts:

$$\int_0^\infty (E_0 R_x - \rho_+)dx = \frac{1}{2}(\rho_+^2 - r_1), \tag{2.19}$$

$$\int_0^\infty (E_0 R_{-x} - \rho_-)dx = -\frac{1}{2}(\rho_-^2 - r_0). \tag{2.20}$$

Proof of (2.19) and (2.20):

Let M_0 be defined as in Section 2.2.1 with $X_i's$ have distribution F_{θ_0}. Then

$$E_{\theta_0}(M_0) = \int_0^\infty P_{\theta_0}(M_0 > x)dx$$

$$= \int_0^\infty P_{\theta_0}(\tau_x < \infty)dx$$

$$= \int_0^\infty e^{\Delta_0(x+\rho_+)}dx + \int_0^\infty e^{\Delta_0(x+\rho_+)}(E_{\tilde{\theta}_0} e^{\Delta_0(R_x-\rho_+)} - 1)dx + O(\Delta_0^3)$$

$$= -\frac{1}{\Delta_0}e^{\Delta_0\rho_+} + \Delta_0 \int_0^\infty (E_0 R_x - \rho_+)dx + O(\Delta_0^2).$$

On the other hand. by Theorem 1 of Siegmund (1979),

$$E_{\theta_0}(M_0) = -\frac{1}{\Delta_0} - \rho_+ - \frac{\Delta_0}{2}r_1 + O(\Delta_0^2).$$

So, from the two expressions of $E_{\theta_0}(M_0)$, we have

$$\int_0^\infty (E_0 R_x - \rho_+)dx = \frac{1}{2}(\rho_+^2 - r_1) + O(\Delta_0).$$

Since $\int_0^\infty (E_0 R_x - \rho_+)dx$ is free of Δ_0, it is indeed equal to $\frac{1}{2}(\rho_+^2 - r_1)$. This completes the proof of (2.19).

Similarly, we can show (2.20).

By combining the above approximations and some algebraic simplifications, we get the desired result.

Lemma 2.4: *As* $|\theta_0|$ *and* $\theta_1 \to 0$ *at the same order,*

$$E[\tau_{m_0'};\tau_{m_0'} < \infty] = \frac{\Delta_0}{\tilde{\mu}_1(\Delta_1 - \Delta_0)^2} e^{\Delta_0(\rho_+ + \rho_-)} + O(1).$$

Proof: First we write

$$E[\tau_{m_0'}, \tau_{m_0'} < \infty]$$

$$= E_{\theta_1}(\tau_-, \tau_- < \infty) P_{\theta_0}(\tau_+ = \infty) + E[\tau_{-M_0}, \tau_{-M_0} < \infty, M_0 > 0]. \tag{2.21}$$

The first term of (2.21) can be evaluated by using Lemma 2.1. For the second term, from Wald's Likelihood Ratio Identity, we have

$$E_{\theta_0}[E_{\theta_1}(\tau_{-M_0}, \tau_{-M_0} < \infty | M_0 > 0)] = E_{\theta_0}[E_{\tilde{\theta}_1}(\tau_{-M_0} e^{\Delta_1 S_{\tau_{-M_0}}})]$$

$$= E_{\theta_0}[E_{\tilde{\theta}_1}(\tau_{-M_0} e^{-\Delta_1(M_0 - \rho_-)})] + E_{\theta_0}[e^{-\Delta_1 M_0} E_{\tilde{\theta}_1}(\tau_{-M_0}(e^{-\Delta_1(R_{-M_0} - \rho_-)} - 1))]$$

$$= E_{\theta_0}[E_{\tilde{\theta}_1}(\tau_{-M_0} e^{-\Delta_1(M_0 - \rho_-)})] + O(1). \tag{2.22}$$

We will give the proof of (2.22) in Appendix 3. On the other hand,

$$E_{\theta_0}[E_{\tilde{\theta}_1}(\tau_{-M_0} e^{-\Delta_1(M_0 - \rho_-)})]$$

$$= \frac{1}{\tilde{\mu}_1} e^{\Delta_1 \rho_-} E_{\theta_0}[e^{-\Delta_1 M_0} E_{\tilde{\theta}_1} S_{\tau_{-M}}]$$

$$= \frac{1}{\tilde{\mu}_1} e^{\Delta_1 \rho_-} E_{\theta_0}[e^{-\Delta_1 M_0}(-M_0 + \rho_-)](1 + O(\Delta_1^2))$$

$$= \frac{1}{\tilde{\mu}_1} e^{\Delta_1 \rho_-} (\rho_- E_{\theta_0}[e^{-\Delta_1 M_0}] - E_{\theta_0}[M_0 e^{-\Delta_1 M_0}])(1 + O(\Delta_1^2)). \tag{2.23}$$

To approximate (2.23), we first note that M_0 is equivalent in distribution to $\sum_{i=1}^{K-1} S_{\tau_+}^{(i)}$, where $S_{\tau_+}^{(i)}$, for $i > 0$, are iid r.v.s, which have the same distribution as that of $[S_{\tau_+} | \tau_+ < \infty]$, and K is an independent Geometric random variable with $p = P_{\theta_0}(\tau_+ = \infty)$. Thus

$$E_{\theta_0}[e^{-\Delta_1 M_0}] = E_{\theta_0}[\sum_{k=2}^{\infty} e^{-\Delta_1 \sum_{i=1}^{k-1} S_{\tau_+^{(i)}}} (1 - P_{\theta_0}(\tau_+ = \infty))^{k-1} P_{\theta_0}(\tau_+ = \infty)]$$

$$= P_{\theta_0}(\tau_+ = \infty) \frac{E_{\theta_0}[e^{-\Delta_1 S_{\tau_+}}, \tau_+ < \infty]}{1 - E_{\theta_0}[e^{-\Delta_1 S_{\tau_+}}, \tau_+ < \infty]}$$

$$= -\frac{\Delta_0}{\Delta_1 - \Delta_0} + O(\Delta_0).$$

and

$$E_{\theta_0}[M_0 e^{-\Delta_1 M_0}]) = E_{\theta_0}[\sum_{k=2}^{\infty}(\sum_{i=1}^{k-1} S_{\tau_+^{(i)}} e^{-\Delta_1 \sum_{i=1}^{k-1} S_{\tau_+^{(i)}}} (1 - P_{\theta_0}(\tau_+ = \infty))^{k-1} P_{\theta_0}(\tau_+ = \infty)]$$

$$= P_{\theta_0}(\tau_+ = \infty)] \frac{E_{\theta_0}[S_{\tau_+} e^{-\Delta_1 S_{\tau_+}}, \tau_+ < \infty]}{(1 - E_{\theta_0}[e^{-\Delta_1 S_{\tau_+}}, \tau_+ < \infty])^2}$$

$$= -\frac{\Delta_0}{(\Delta_1 - \Delta_0)^2} e^{\Delta_0 \rho_+} + O(\Delta_0). \tag{2.24}$$

The proof is completed by combining the above approximations.

Combining the results of Lemma 2.1 to 2.4, the following result is obtained.

Theorem 2.7: *As $|\theta_0|$ and θ approach zero at the same order,*

$$E[\nu_0] = \frac{\Delta_0}{\tilde{\mu}_1(\Delta_1 - \Delta_0)^2} e^{\Delta_0(\rho_+ + \rho_-)} + \frac{\Delta_0}{\tilde{\mu}_1 \Delta_1(\Delta_1 - \Delta_0)} e^{\Delta_1(\rho_+ + \rho_-)}$$

$$- \frac{\Delta_1}{\tilde{\mu}_0(\Delta_1 - \Delta_0)^2} e^{\Delta_1(\rho_+ + \rho_-)} + \frac{\Delta_1}{\tilde{\mu}_0 \Delta_0(\Delta_1 - \Delta_0)} e^{\Delta_0(\rho_+ + \rho_-)} + O(1),$$

and

$$E|\nu_0| = \frac{\Delta_0}{\tilde{\mu}_1(\Delta_1 - \Delta_0)^2} e^{\Delta_0(\rho_+ + \rho_-)} + \frac{\Delta_0}{\tilde{\mu}_1 \Delta_1(\Delta_1 - \Delta_0)} e^{\Delta_1(\rho_+ + \rho_-)}$$

$$+\frac{\Delta_1}{\tilde{\mu}_0(\Delta_1-\Delta_0)^2}e^{\Delta_1(\rho_+ +\rho_-)}-\frac{\Delta_1}{\tilde{\mu}_0\Delta_0(\Delta_1-\Delta_0)}e^{\Delta_0(\rho_+ +\rho_-)}+O(1).$$

2.3.3 Second Order Expansion for $E(\nu_0^2)$

In this subsection, we further explore the techniques used in the last section and derive the second order expansion for $E(\nu_0^2)$. As the techniques are basically the same, we only provide the main steps of proofs. The basic assumptions are the same as in the last subsection, i.e. that θ, $|\theta_0| \to 0$ at the same order. Same notations will be used. First, we write

$$\begin{aligned} E[\nu_0^2] &= E[\nu_0^2;\nu_0<0]+E[\nu_0^2;\nu_0>0] \\ &= E[\nu_0^2;m_0<m_1']+E[\nu_0^2;m_1<m_0']. \end{aligned} \tag{2.25}$$

As the two terms on the right hand side of (2.25) have similar structures, we shall only give the detailed derivation for $E[\nu_0^2;m_1<m_0']$. For $n>0$, we can write

$$\begin{aligned} E[\nu_0^2;m_0<m_1'] &= E[(\tau_{m_0'}+\nu_2)^2;\tau_{m_0'}<\infty] \\ &= E[\tau_{m_0'}^2;\tau_{m_0'}<\infty]+2E[\tau_{m_0'}\nu_2;\tau_{m_0'}<\infty]+E[\nu_2^2;\tau_{m_0'}<\infty] \\ &= E[\tau_{m_0'}^2;\tau_{m_0'}<\infty]+2E[\nu_2]E[\tau_{m_0'};\tau_{m_0'}<\infty]+E[\nu_2^2]P(\tau_{m_0'}<\infty). \end{aligned} \tag{2.26}$$

Therefore, there are five quantities to be evaluated separately. Approximations for $E(\nu_2)$, $P(\tau_{m_0'}<\infty)$ and $E[\tau_{m_0'};\tau_{m_0'}<\infty]$ have already been given in the last subsection. The expansions for the other two will be given after three lemmas.

Lemma 2.5: *For any stopping time N adapted to $\Im_n = \sigma\{X_1, \ldots, X_n\}$, if $E_\theta N^2 < \infty$ and $E_\theta X_1^2 < \infty$,*

$$E_\theta N^2 = \frac{1}{\mu^2}[\psi''(\theta)E_\theta N + 2\mu E_\theta(NS_N) - E_\theta S_N^2],$$

where $S_n = \sum_{i=1}^n X_i$, $X_i's$ are iid with distribution function F_θ which belongs to the standard one parameter exponential family defined in Section 1.3, and $\mu = \psi'(\theta)$.

Proof: It is easy to check that $\{(S_n - n\mu)^2 - n\psi''(\theta), \Im_n\}$ is a martingale. Under the condition that $E_\theta N^2 < \infty$ and $E_\theta X_1^2 < \infty$, we know that

$$E_\theta[|(S_N - N\mu)^2 - N\psi''(\theta)|] < \infty.$$

Therefore, we have

$$E_\theta[(S_N - N\mu)^2 - N\psi''(\theta)] = 0.$$

Solving the last equation for $E_\theta N^2$, we get the desired result.

The next lemma can be derived from Theorem 2.1, 2.2 and the fact that $\theta_1 = \frac{1}{2}\Delta_1 + O(\Delta_1^2)$, $\theta_0 = \frac{1}{2}\Delta_0 + O(\Delta_0^2)$.

Lemma 2.6: *As $\theta_1, \theta_0 \to 0$,*

$$\mu_1 E_{\theta_1}[\tau_+ S_{\tau_+}] = \frac{1}{2}E_0 S_{\tau_+}^2 + O(\Delta_1) = \rho_+ E_0 S_{\tau_+} + O(\Delta_1);$$

$$\mu_0 E_{\theta_0}[\tau_- S_{\tau_-}] = \frac{1}{2}E_0 S_{\tau_-}^2 + O(\Delta_0) = \rho_- E_0 S_{\tau_-} + O(\Delta_0);$$

$$E_{\theta_1} S_{\tau_+} = E_0 S_{\tau_+} e^{\frac{\Delta_1}{2}\rho_+} + O(\Delta_1^2),$$

$$E_{\theta_0} S_{\tau_-} = E_0 S_{\tau_-} e^{\frac{\Delta_0}{2}\rho_-} + O(\Delta_0^2).$$

Lemma 2.7:

$$E_{\theta_1}[\tau_-^2; \tau_- < \infty] = \frac{\psi''(\tilde{\theta}_1)}{\tilde{\mu}_1^3} E_0 S_{\tau_-} e^{\frac{\Delta_1}{2}\rho_-} + O(\frac{1}{\Delta_1}).$$

Proof: From Wald's likelihood ratio identity, we have

$$E_{\theta_1}[\tau_-^2; \tau_- < \infty] = E_{\tilde{\theta}_1}[\tau_-^2 e^{\Delta_1 S_{\tau_-}}]$$

$$= E_{\tilde{\theta}_1}[\tau_-^2] + \Delta_1 E_{\tilde{\theta}_1}[\tau_-^2 S_{\tau_-}] + O(\frac{1}{\Delta_1}). \tag{2.27}$$

From Lemma 2.5, we have

$$E_{\tilde{\theta}_1}[\tau_-^2] = \frac{1}{\tilde{\mu}_1^2}[E_{\tilde{\theta}_1}\tau_-\psi''(\tilde{\theta}_1) + 2\tilde{\mu}_1 E_{\tilde{\theta}_1}(\tau_- S_{\tau_-}) - E_{\tilde{\theta}_1} S_{\tau_-}^2]$$

$$= \frac{1}{\tilde{\mu}_1^2}[\frac{1}{\tilde{\mu}_1} E_{\tilde{\theta}_1} S_{\tau_-}\psi''(\tilde{\theta}_1) + E_0 S_{\tau_-}^2 - E_0 S_{\tau_-}^2 + O(\Delta_1)]$$

$$= \frac{\psi''(\tilde{\theta}_1)}{\tilde{\mu}_1^3} E_0 S_{\tau_-} e^{-\frac{\Delta_1}{2}\rho_-} + O(\frac{1}{\Delta_1}), \tag{2.28}$$

where in the last equation, we used lemma 2.6, and in the second to the last equation, we used Theorem 2.1.

To approximate the second term in (2.27), let $h(\theta) = E_\theta S_{\tau_-}$ for $\theta < 0$. Then

$$h''(\theta) = E_\theta[-\psi''(\theta)\tau_- S_{\tau_-} + S_{\tau_-}(S_{\tau_-} - \mu\tau_-)^2].$$

On the other hand, by Theorem 4.2 of Chang (1992), we have $h''(\theta) = \frac{1}{3}E_0 S_{\tau_-}^3 - \alpha^{(1)}$, and $\alpha^{(1)}$ is a constant defined in (4.1) of Chang (1992). Therefore, we have

$$E_{\tilde{\theta}_1}[\tau_-^2 S_{\tau_-}] = \frac{1}{\tilde{\mu}_1}[2\tilde{\mu}_1 E_{\tilde{\theta}_1}[\tau_- S_{\tau_-}^2] - E_{\tilde{\theta}_1}[S_{\tau_-}^3] + E_{\tilde{\theta}_1}[\tau_- S_{\tau_-}]\psi''(\tilde{\theta}_1) + O(\frac{1}{\Delta_1^2})$$

$$= \frac{1}{2\tilde{\mu}_1^3} E_0 S_{\tau_-}^2 \psi''(\tilde{\theta}_1) + O(\frac{1}{\Delta_1^2}). \qquad (2.29)$$

Insert (2.28) and (2.29) into (2.27), we complete the proof of lemma 2.7.

By using the same argument as in the proof for Lemma 2.2, we obtain the following result.

Lemma 2.8:

$$E_\theta[\nu_2^2] = \frac{E_\theta[\tau_-^2; \tau_- < \infty]}{P_\theta(\tau_- = \infty)} + 2(\frac{E_\theta[\tau_-; \tau_- < \infty]}{P_\theta(\tau_- = \infty)})^2$$

$$= -\frac{\psi''(\tilde{\theta}_1)}{\Delta_1 \tilde{\mu}_1^3} + \frac{2}{\Delta_1^2 \tilde{\mu}_1^2} + O(\frac{1}{\Delta_1^2}).$$

Lemma 2.9: *As θ_1 and $\theta_0 \to 0$,*

$$E[\tau_{m_0'}^2; \tau_{m_0'} < \infty] = \frac{\psi''(\tilde{\theta}_1)}{\tilde{\mu}_1^3} \frac{\Delta_0}{(\Delta_1 - \Delta_0)^2} e^{\Delta_0(\rho_+ + \rho_-)} - \frac{2\Delta_0}{\tilde{\mu}_1^2(\Delta_1 - \Delta_0)^3} e^{-\Delta_1\rho_+ + \Delta_0\rho_-} + O(\frac{1}{\Delta_0^2}).$$

Proof: We first write

$$E[\tau_{m_0'}^2; \tau_{m_0'} < \infty]$$

$$= E_{\theta_1}[\tau_-^2; \tau_- < \infty] P_{\theta_0}(\tau_+ = \infty) + E[\tau_{-M_0}^2; \tau_{-M_0} < \infty, M_0 > 0]. \qquad (2.30)$$

The first term of (2.30) can be shown to be the order of $O(\frac{1}{\Delta_1^2})$. Again by Wald's Likelihood Ratio Identity, we can write the second term of (2.30) as

$$E_{\theta_0}[E_{\theta_1}(\tau_{-M_0}^2; \tau_{-M_0} < \infty | M_0 > 0)]$$

$$= E_{\theta_0}[E_{\tilde{\theta}_1}(\tau_{-M_0}^2 e^{-\Delta_1(M_0 - R_{-M_0})} | M_0 > 0)]$$

$$= E_{\theta_0}[e^{-\Delta_1(M_0-\rho_-)}E_{\tilde{\theta}_1}(\tau^2_{-M_0}|M_0 > 0)]$$

$$+E_{\theta_0}[E_{\tilde{\theta}_1}(\tau^2_{-M_0}e^{-\Delta_1(M_0-\rho_-)}(e^{\Delta_1(R_{-M_0}-\rho_-)} - 1)|M_0 > 0)]. \qquad (2.31)$$

From Appendix 4, we know that the second term in (2.31) is the order of $O(\frac{1}{\Delta_1^2})$. So we only need to approximate the first term on the right hand side of (2.31).

First, we note that

$$E_{\tilde{\theta}_1}(\tau^2_{-M_0}|M_0) = \frac{1}{\tilde{\mu}_1^2}[\psi''(\tilde{\theta}_1)E_{\tilde{\theta}_1}(\tau_{-M_0}|M_0) + 2\tilde{\mu}_1 E_{\tilde{\theta}_1}(\tau_{-M_0}S_{\tau_{-M_0}}|M_0) - E_{\tilde{\theta}_1}(S^2_{\tau_{-M_0}}|M_0)]$$

$$= \frac{1}{\tilde{\mu}_1^3}\psi''(\tilde{\theta}_1)(-M_0+\rho_-)+\frac{1}{\tilde{\mu}_1^3}\psi''(\tilde{\theta}_1)(E_{\tilde{\theta}_1}(R_{-M_0}|M_0)-\rho_-(\tilde{\theta}_1))+\frac{M_0^2}{\tilde{\mu}_1^2}-\frac{2\rho_-}{\tilde{\mu}_1^2}M_0+O(\frac{1}{\Delta_1^2}).$$

Insert the above approximation into the first term on the right hand side of (2.31), we have

$$E_{\theta_0}[e^{-\Delta_1(M_0-\rho_-)}E_{\tilde{\theta}_1}(\tau^2_{-M_0}|M_0)] = \frac{1}{\tilde{\mu}_1^3}\psi''(\tilde{\theta}_1)E_{\theta_0}(-M_0e^{-\Delta_1(M_0-\rho_-)})$$

$$+\frac{\rho_-}{\tilde{\mu}_1^3}\psi''(\tilde{\theta}_1)E_{\theta_0}e^{-\Delta_1(M_0-\rho_-)} + \frac{\rho_-}{\tilde{\mu}_1^3}\psi''(\tilde{\theta}_1)E_{\theta_0}[e^{-\Delta_1(M_0-\rho_-)}(E_{\tilde{\theta}_1}(R_{-M_0}|M_0) - \rho_-(\tilde{\theta}_1)))]$$

$$+\frac{1}{\tilde{\mu}_1^2}E_{\theta_0}[M_0^2e^{-\Delta_1(M_0-\rho_-)}] - \frac{2\rho_-}{\tilde{\mu}_1^2}E_{\theta_0}[M_0e^{-\Delta_1(M_0-\rho_-)}] + O(\frac{1}{\Delta_1^2}).$$

By the technique used in proving Lemma 2.3, we have

$$E_{\theta_0}[e^{-\Delta_1(M_0-\rho_-)}(E_{\tilde{\theta}_1}(R_{-M_0}|M_0) - \rho_-(\tilde{\theta}_1)))]$$

$$= \int_0^\infty e^{-\Delta_1(x-\rho_-)}E_{\tilde{\theta}_1}(R_{-x} - \rho_-(\tilde{\theta}_1))dP_{\theta_0}(\tau_x < \infty)$$

$$= \Delta_0\int_0^\infty e^{-\Delta_1(x-\rho_-)}e^{\Delta_0(x+\rho_+)}E_{\tilde{\theta}_1}(R_{-x} - \rho_-(\tilde{\theta}_1))dx$$

$$+\Delta_0^2 \int_0^\infty e^{-\Delta_1(x-\rho_-)} e^{\Delta_0(x+\rho_+)} E_{\tilde{\theta}_1}(R_{-x} - \rho_-(\tilde{\theta}_1)) d(E_0 R_x - \rho_+) + O(\Delta_0^2)$$

$$= O(\Delta_0).$$

Now, with previous results available, we only need to approximate $E_{\theta_0} M_0^2 e^{-\Delta_1 M_0}$. In fact,

$$E_{\theta_0} M_0^2 e^{-\Delta_1 M_0} = E_{\theta_0}[\sum_{k=2}^{\infty}(\sum_{i=1}^{k-1} S_{\tau_+^{(i)}})^2 e^{-\Delta_1 \sum_{i=1}^{k-1} S_{\tau_+^{(i)}}} (1 - P^{\theta_0}(\tau_+ = \infty))^{k-1} P^{\theta_0}(\tau_+ = \infty)]$$

$$= \sum_{k=3}^{\infty}(k-1)(k-2)[E^{\theta_0}(S_{\tau_+} e^{-\Delta_1 S_{\tau_+}}, \tau_+ < \infty)]^2 [E^{\theta_0}(e^{-\Delta_1 S_{\tau_+}}, \tau_+ < \infty)]^{k-3} P^{\theta_0}(\tau_+ = \infty)$$

$$+ \sum_{k=2}^{\infty}(k-1)[E^{\theta_0}(S_{\tau_+}^2 e^{-\Delta_1 S_{\tau_+}}, \tau_+ < \infty)][E^{\theta_0}(e^{-\Delta_1 S_{\tau_+}}, \tau_+ < \infty)]^{(k-2)} P^{\theta_0}(\tau_+ = \infty)$$

$$= \frac{2[E^{\theta_0}(S_{\tau_+} e^{-\Delta_1 S_{\tau_+}}, \tau_+ < \infty)]^2 P^{\theta_0}(\tau_+ = \infty)}{[1 - E^{\theta_0}(e^{-\Delta_1 S_{\tau_+}}, \tau_+ < \infty)]^3} + \frac{E^{\theta_0}(S_{\tau_+}^2 e^{-\Delta_1 S_{\tau_+}}, \tau_+ < \infty) P^{\theta_0}(\tau_+ = \infty)}{[1 - E^{\theta_0}(e^{-\Delta_1 S_{\tau_+}}, \tau_+ < \infty)]^2}$$

$$= -\frac{2\Delta_0}{(\Delta_1 - \Delta_0)^3} e^{\Delta_0 \rho_+ - (\Delta_1 - \Delta_0)\rho_+} + O(1). \tag{2.32}$$

The proof of Lemma 2.9 is completed by combining the approximations of the quantities involved and some simplifications.

Combining the results of Lemmas 2.1-2.9, the following result is obtained.

Theorem 2.8: *As θ_1 and $\theta_0 \to 0$,*

$$E(\nu_0^2) = \frac{\psi''(\tilde{\theta}_1)}{\tilde{\mu}_1^3} \frac{\Delta_0}{(\Delta_1 - \Delta_0)^2} e^{\Delta_0(\rho_+ + \rho_-)} - \frac{2\Delta_0}{\tilde{\mu}_1^2 (\Delta_1 - \Delta_0)^3} e^{-\Delta_1 \rho_+ + \Delta_0 \rho_-}$$

$$- \frac{2\Delta_0}{\tilde{\mu}_1^2 \Delta_1 (\Delta_1 - \Delta_0)^2} e^{\Delta_0(\rho_+ + \rho_-)} + \frac{\psi''(\tilde{\theta}_1)\Delta_0}{\tilde{\mu}_1^3 \Delta_1 (\Delta_1 - \Delta_0)} e^{\Delta_1(\rho_+ + \rho_-)}$$

$$- \frac{2\Delta_0}{\tilde{\mu}_1^2 \Delta_1^2 (\Delta_1 - \Delta_0)} e^{\Delta_1(\rho_+ + \rho_-)} - \frac{2\Delta_1}{\tilde{\mu}_0^2 \Delta_0 (\Delta_1 - \Delta_0)^2} e^{\Delta_1(\rho_+ + \rho_-)}$$

$$-\frac{\psi''(\tilde{\theta}_0)\Delta_1}{\tilde{\mu}_0^3\Delta_0(\Delta_1-\Delta_0)}e^{\Delta_0(\rho_++\rho_-)}+\frac{2\Delta_1}{\tilde{\mu}_0^2\Delta_0^2(\Delta_1-\Delta_0)}e^{\Delta_0(\rho_++\rho_-)}$$
$$+\frac{\psi''(\tilde{\theta}_0)}{\tilde{\mu}_0^3}\frac{\Delta_1}{(\Delta_1-\Delta_0)^2}e^{\Delta_1(\rho_++\rho_-)}+\frac{2\Delta_1}{\tilde{\mu}_0^2(\Delta_1-\Delta_0)^3}e^{\Delta_1\rho_+-\Delta_0\rho_-}+O(\frac{1}{\Delta_0^2}).$$

2.4 Simulation Results

In this section, simulation results are presented to check the accuracy of the results obtained in Section 2.3. Attention is focused on values of $P(\tau_{m_0'} < \infty)$, $E\nu_0$ and $E|\nu_0|$, which will play an important role in the subsequent chapters. To save space, we only give results for normal and non-negative exponential cases.

2.4.1 Two-sided Normal Random Walk

In this subsection, we assume that $X_n's$ are independent normal random variables with variance 1 and mean $\theta_0 < 0$ for $n < 0$, and mean $\theta_1 > 0$ for $n > 0$.

In this case, $\psi(\theta) = \frac{1}{2}\theta^2$; $E_0S_{\tau_+} = -E_0S_{\tau_-} = \frac{1}{\sqrt{2}}$; $\rho_+ = -\rho_- \approx 0.583$, which will be denoted as ρ; $r_0 = r_1 = \frac{1}{4}$ and $C_0 = \frac{1}{2}(\frac{1}{\sqrt{2}} - \rho_+)^2$. Substitute the related quantities into Lemma 2.3 and Theorem 2.7, we have

Corollary 2.1: *As* $\theta_1, |\theta_0| \to 0$ *at the same order,*

$$P(\tau_{m_0'} < \infty) = -\frac{\theta_0}{\theta_1-\theta_0}e^{-\theta_1(\theta_1-\theta_0)} + O(\theta_0^3).$$

Table 2.1: $P(\tau_{m_0'} < \infty)$, $E\nu_0$ and $E|\nu_0|$ for $\theta_0 = -0.2$

θ_1	$P(\tau_{m_0'} < \infty)$	$E\nu_0$	$E\|\nu_0\|$
0.2	0.4616	0	18.75
	0.4568	-0.43(0.27)	18.21
0.25	0.3972	-4.5	15.56
	0.4068	-3.55(0.23)	14.78
0.3	0.3443	-6.94	14.06
	0.3509	-6.10(0.21)	13.06

$$E\nu_0 = \frac{1}{2\theta_1^2} - \frac{1}{2\theta_0^2} + O(1),$$

and

$$E|\nu_0| = \frac{1}{2\theta_1^2} + \frac{1}{2\theta_0^2} - \frac{1}{(\theta_1 - \theta_0)^2} + O(1).$$

These results have been obtained by Wu (1999).

Table 2.1, 2.2 and 2.3 give comparisons of the simulated values with approximated values for $P(\tau_{m_0'} < \infty)$, $E\nu_0$ and $E|\nu_0|$. The simulated values are based on 10000 replications. For each fixed θ_0, we check three values of θ_1 around $-\theta_0$. In each cell, the top number is the approximated value while the bottom number is the simulation value, and the value given in the parenthesis is standard error for the simulated value.

From these tables, we can see that our approximations are very accurate.

Table 2.2: $P(\tau_{m_0'} < \infty)$, $E\nu_0$ and $E|\nu_0|$ for $\theta_0 = -0.25$

θ_1	$P(\tau_{m_0'} < \infty)$	$E\nu_0$	$E\|\nu_0\|$
0.25	0.4412	0	12
	0.4552	-0.02(0.19)	11.73
0.3	0.3854	-2.44	10.25
	0.4012	-2.32(0.17)	10.07
0.4	0.2966	-4.88	8.76
	0.3154	-4.81(0.15)	8.58

Table 2.3: $P(\tau_{m_0'} < \infty)$, $E\nu_0$ and $E|\nu_0|$ for $\theta_0 = -0.3$

θ_1	$P(\tau_{m_0'} < \infty)$	$E\nu_0$	$E\|\nu_0\|$
0.3	0.4176	0	8.33
	0.4428	0.17(0.14)	8.21
0.35	0.3676	-1.47	7.27
	0.3826	-1.53(0.12)	7.16
0.4	0.3239	-2.43	6.64
	0.3426	-2.45(0.11)	6.47

2.4.2 Two-sided Non-negative Exponential Random Walk

In this subsection, we assume that $Y_n's$ are independent non-negative exponential random variables with mean $\lambda_0 < 1$ for $n < 0$, and mean $\lambda_1 > 1$ for $n > 0$. Let $X_n = Y_n - 1$, then $X_n's$ belong to the standard one parameter exponential family.

In this case, $\psi(\theta) = -\theta - \ln(1-\theta)$.

$$\theta_0 = 1 - \frac{1}{\lambda_0}, \theta_1 = 1 - \frac{1}{\lambda_1}.$$

$$\mu_0 = -1 + \lambda_0, \mu_1 = -1 + \lambda_1.$$

$$\rho_+ = 1, r_1 = 1, \rho_- = -\frac{1}{3}, r_0 = \frac{1}{18} \quad and \quad C_0 = 0.$$

Substitute the above corresponding values into Lemma 2.3 and Theorem 2.7. following results are obtained.

Corollary 2.2: *As $\theta_1, |\theta_0| \to 0$ at the same order,*

$$P(\tau_{m_0'} < \infty) = C_1 e^{\frac{25}{36C_1}\Delta_0\Delta_1} + O(\Delta_0^3),$$

with $C_1 = -\frac{\Delta_0}{\Delta_1 - \Delta_0} e^{\frac{2}{3}\Delta_1}$.

$$E\nu_0 = \frac{\Delta_0}{\tilde{\mu}_1(\Delta_1 - \Delta_0)^2} e^{\frac{2}{3}\Delta_0} + \frac{\Delta_0}{\tilde{\mu}_1\Delta_1(\Delta_1 - \Delta_0)} e^{\frac{2}{3}\Delta_1}$$

$$-\frac{\Delta_1}{\tilde{\mu}_0(\Delta_1 - \Delta_0)^2} e^{\frac{2}{3}\Delta_1} + \frac{\Delta_1}{\tilde{\mu}_0\Delta_0(\Delta_1 - \Delta_0)} e^{\frac{2}{3}\Delta_0} + O(1),$$

and

$$E|\nu_0| = \frac{\Delta_0}{\tilde{\mu}_1(\Delta_1 - \Delta_0)^2} e^{\frac{2}{3}\Delta_0} + \frac{\Delta_0}{\tilde{\mu}_1\Delta_1(\Delta_1 - \Delta_0)} e^{\frac{2}{3}\Delta_1}$$

Table 2.4: $P(\tau_{m_0'} < \infty)$, $E\nu_0$ and $E|\nu_0|$ for $\lambda_0 = 0.85$

λ_1	$P(\tau_{m_0'} < \infty)$	$E\nu_0$	$E\|\nu_0\|$
1.2	0.5379	2.95	26.18
	0.5546	3.16(0.32)	25.84
1.25	0.4904	-2.78	22.06
	0.5036	-1.36(0.28)	19.25
1.3	0.4527	-6.08	19.97
	0.4614	-4.27(0.26)	16.74

$$+\frac{\Delta_1}{\tilde{\mu}_0(\Delta_1 - \Delta_0)^2} e^{\frac{2}{3}\Delta_1} - \frac{\Delta_1}{\tilde{\mu}_0 \Delta_0 (\Delta_1 - \Delta_0)} e^{\frac{2}{3}\Delta_0} + O(1).$$

Similar to the last subsection, Table 2.4, 2.5 and 2.6 present comparisons of simulated values with approximated values for $P(\tau_{m_0'} < \infty)$, $E\nu_0$ and $E|\nu_0|$. The simulated values are based on 10000 replications. For each λ_0, we check three values of λ_1, which give values of θ_1 close to $\tilde{\theta}_0$ after transformation. As in Table 2.1-2.3, in each cell, the top number is the approximated value while the bottom number is the simulation value, and the value given in the parenthesis is standard error for the simulated value.

From Table 2.4-2.6, again, we see that our approximations are very accurate.

Table 2.5: $P(\tau_{m_0'} < \infty)$, $E\nu_0$ and $E|\nu_0|$ for $\lambda_0 = 0.8$

λ_1	$P(\tau_{m_0'} < \infty)$	$E\nu_0$	$E\|\nu_0\|$
1.2	0.6127	11.10	21.03
	0.6283	8.90(0.27)	18.39
1.3	0.5268	2.13	13.69
	0.5382	1.83(0.25)	12.35
1.4	0.4679	-1.37	11.24
	0.4726	-1.79(0.17)	10.27

Table 2.6: $P(\tau_{m_0'} < \infty)$, $E\nu_0$ and $E|\nu_0|$ for $\lambda_0 = 0.75$

λ_1	$P(\tau_{m_0'} < \infty)$	$E\nu_0$	$E\|\nu_0\|$
1.4	0.5195	2.27	8.66
	0.5326	1.34(0.13)	7.78
1.5	0.4760	0.48	7.38
	0.4896	0.04(0.11)	6.46
1.6	0.4434	-0.59	6.71
	0.4363	-1.24(0.11)	5.92

APPENDIX 1: *Proof of (2.15)*

Using Taylor expansion and the fact that all moments of $R_{-\infty}$ exist, we have

$$
\begin{aligned}
E_{\tilde{\theta}_1} e^{\Delta_1 R_{-\infty}} &= 1 + \Delta_1 E_{\tilde{\theta}_1} R_{-\infty} + \frac{1}{2}\Delta_1^2 E_{\tilde{\theta}_1} R_{-\infty}^2 + O(\Delta_1^3) \\
&= 1 + \Delta_1 \frac{E_{\tilde{\theta}_1} S_{\tau_-}^2}{2E_{\tilde{\theta}_1} S_{\tau_-}} + \frac{1}{2}\Delta_1^2 \frac{E_{\tilde{\theta}_1} S_{\tau_-}^3}{3E_{\tilde{\theta}_1} S_{\tau_-}} + O(\Delta_1^3) \\
&= 1 + \Delta_1 \frac{E_0 S_{\tau_-}^2 + \frac{2}{3} E_0 S_{\tau_-}^3 \tilde{\theta}_1}{2(E_0 S_{\tau_-} + \frac{1}{2} E_0 S_{\tau_-}^2 \tilde{\theta}_1)} + \frac{1}{2}\Delta_1^2 \frac{E_0 S_{\tau_-}^3}{3E_0 S_{\tau_-}} + O(\Delta_1^3) \\
&= 1 + \Delta_1 \frac{E_0 S_{\tau_-}^2 - \frac{1}{3} E_0 S_{\tau_-}^3 \Delta_1}{2(E_0 S_{\tau_-} - \frac{1}{4} E_0 S_{\tau_-}^2 \Delta_1)} + \frac{1}{2}\Delta_1^2 (\rho_-^2 + r_0) + O(\Delta_1^3) \\
&= 1 + \Delta_1 \frac{\rho_- - \frac{1}{2}(\rho_-^2 + r_0)\Delta_1}{1 - \frac{1}{2}\rho_- \Delta_1} + \frac{1}{2}\Delta_1^2 (\rho_-^2 + r_0) + O(\Delta_1^3) \\
&= 1 + \Delta_1 \rho_- + \frac{1}{2}\Delta_1^2 \rho_-^2 + O(\Delta_1^3) \\
&= e^{\Delta_1 \rho_-} + O(\Delta_1^3).
\end{aligned}
$$

APPENDIX 2: *Proof of (2.18)*

We first write

$$
\begin{aligned}
\frac{d}{dx}(P_{\theta_0}(\tau_x < \infty) - e^{\Delta_0(x+\rho_+)}) &= \frac{d}{dx}(e^{\Delta_0(x+\rho_+)}(E_{\tilde{\theta}_0}(e^{\Delta_0(R_x-\rho_+)} - 1))) \\
= \Delta_0 e^{\Delta_0(x+\rho_+)} E_{\tilde{\theta}_0}(e^{\Delta_0(R_x-\rho_+)} - 1) &+ e^{\Delta_0(x+\rho_+)} \frac{d}{dx}(E_{\tilde{\theta}_0}(e^{\Delta_0(R_x-\rho_+)} - 1)) \\
&= e^{\Delta_0(x+\rho_+)} \frac{d}{dx}(E_{\tilde{\theta}_0}(e^{\Delta_0(R_x-\rho_+)} - 1)) + O(\Delta_0^2).
\end{aligned}
$$

For a given $\Delta x > 0$, conditional on $\{R_x < \Delta x\}$ or $\{R_x \geq \Delta x\}$, we have

$$E_{\tilde{\theta}_0} e^{\Delta_0 R_{x+\Delta x}} = E_{\tilde{\theta}_0}[e^{\Delta_0 R_{x+\Delta x}}; R_x \geq \Delta x] + E_{\tilde{\theta}_0}[e^{\Delta_0 R_{x+\Delta x}}; R_x < \Delta x]$$

$$= E_{\tilde{\theta}_0}[e^{\Delta_0 R_x}; R_x \geq \Delta x] + E_{\tilde{\theta}_0}[e^{\Delta_0 R_x} | R_x < \Delta x] E_{\tilde{\theta}_0}[e^{\Delta_0 R'_{(\Delta x - R_x)}}; R_x < \Delta x]$$

where R'_x is the overshoot at the boundary x for another independent copy of $\{S_n\}$. Thus, we have

$$E_{\tilde{\theta}_0}[e^{\Delta_0 R_{x+\Delta x}} - e^{\Delta_0 R_x}] = E_{\tilde{\theta}_0}[e^{\Delta_0 R_{x+\Delta x}} - e^{\Delta_0 R_x}; R_x < \Delta x]$$

$$= E_{\tilde{\theta}_0}[e^{\Delta_0 R_x} | R_x < \Delta x] E_{\tilde{\theta}_0}[(e^{\Delta_0 R'_{(\Delta x - R_x)}} - 1); R_x < \Delta x]$$

$$= \Delta_0 E_{\tilde{\theta}_0} R_0 f_{R_x}(0) \Delta x + O(\Delta_0^2 \Delta x) + O(\Delta_0 (\Delta x)^2)$$

$$= \Delta_0 E_0 R_0 f_{R_x}(0) \Delta x + O(\Delta_0^2 \Delta x) + O(\Delta_0 (\Delta x)^2).$$

Therefore

$$\frac{d}{dx}(P_{\theta_0}(\tau_x < \infty) - e^{\Delta_0 (x+\rho_+)}) = \Delta_0 E_0 R_0 f_{R_x}(0) + O(\Delta_0^2).$$

In the same way, we can prove that

$$\frac{d}{dx} E_0 R_x = E_0 R_0 f_{R_x}(0).$$

This completes the proof of (2.18).

APPENDIX 3: *Proof of (2.22)*

Since the technique is similar to that used in proving lemma 2.6, we only give the main steps.

Since $E_{\theta_0}[E_{\tilde{\theta}_1}[\tau_{-M}e^{-\Delta_1(M-\rho_-)}(e^{\Delta_1(R_{-M}-\rho_-)}-1)|M]]$ can be written as

$$-\int_0^\infty E_{\tilde{\theta}_1}[\tau_{-x}e^{-\Delta_1(x-\rho_-)}(e^{\Delta_1(R_{-x}-\rho_-)}-1)]dP_{\theta_0}(\tau_x<\infty)$$

$$=-\int_0^\infty e^{-\Delta_1(x-\rho_-)}E_{\tilde{\theta}_1}[\tau_{-x}(e^{\Delta_1(R_{-x}-\rho_-)}-1)]dP_{\theta_0}(\tau_x<\infty)$$

$$=-\Delta_1\int_0^\infty e^{-\Delta_1(x-\rho_-)}E_{\tilde{\theta}_1}[\tau_{-x}(R_{-x}-E_{\tilde{\theta}_1}R_{-\infty})]dP_{\theta_0}(\tau_x<\infty)$$

$$-O(\Delta_1^2)\int_0^\infty e^{-\Delta_1(x-\rho_-)}E_{\tilde{\theta}_1}\tau_{-x}dP_{\theta_0}(\tau_x<\infty) \qquad (2.33)$$

By Theorem 2.3 and the dominated convergence theorem, we know that the first term in the last equation is equal to

$$-\frac{\Delta_1}{\tilde{\mu}_1}\int_0^\infty E_{\tilde{\theta}_1}(R_{-x}-E_{\tilde{\theta}_1}R_{-\infty})P_{\tilde{\theta}_1}(M<x)dx\int_0^\infty e^{-\Delta_1(x-\rho_-)}dP_{\theta_0}(\tau_x<\infty)$$

$$-O(\frac{\Delta_1}{\tilde{\mu}_1})\int_0^\infty e^{-\Delta_1(x-\rho_-)}e^{-rx}dP_{\theta_0}(\tau_x<\infty)$$

$$=O(\frac{\Delta_0\Delta_1}{(\Delta_1-\Delta_0)\tilde{\mu}_1})+O(\frac{\Delta_1}{\tilde{\mu}_1})=O(1).$$

Similarly, we can show that the second term of (2.33) is equal to

$$O(\frac{\Delta_1^2}{\tilde{\mu}_1}\int_0^\infty xe^{-\Delta_1 x}de^{\Delta_0 x})=O(\frac{\Delta_1^2\Delta_0}{(\Delta_1-\Delta_0)^2\tilde{\mu}_1})=O(1).$$

This completes the proof of (2.22).

APPENDIX 4: *Approximate the second term of (2.31)*

Since it is similar to the proof of (2.22), only main steps are given.

$$E_{\theta_0}[E_{\tilde{\theta}_1}(\tau^2_{-M_0}e^{-\Delta_1(M_0-\rho_-)}(e^{\Delta_1(R_{-M_0}-\rho_-)}-1)|M_0>0)]$$

$$=\int_0^\infty e^{-\Delta_1(x-\rho_-)}E_{\tilde{\theta}_1}[\tau^2_{-x}(e^{\Delta_1(R_{-x}-\rho_-)}-1)]dP_{\theta_0}(\tau_x<\infty)$$

$$=\Delta_1\int_0^\infty e^{-\Delta_1(x-\rho_-)}E_{\tilde{\theta}_1}[\tau^2_{-x}(R_{-x}-\rho_-(\tilde{\theta}_1))]dP_{\theta_0}(\tau_x<\infty)$$

$$+O(\Delta_1^2)\int_0^\infty e^{-\Delta_1(x-\rho_-)}E_{\tilde{\theta}_1}\tau^2_{-x}dP_{\theta_0}(\tau_x<\infty)$$

$$=O(\frac{1}{\Delta_1^2}).$$

where in the last equation, we use a fact that $E_{\tilde{\theta}_1}(\tau^2_{m_0'})=O(\frac{1}{\Delta_1^4})$.

REFERENCES

Carlsson, H. (1983). Remainder term estimates of the renewal function. *Ann. Probab.*, **11**, 143-157.

Chang, J. T. (1992). On moments of the first ladder height of random walks with small drift. *Ann. Appl. Probab.*, **2**, 714-738.

Hinkley, D. V. (1971). Inference about the change-point in a sequence of random variables. *Biometrika*, **57** 1-17.

Klass, T. (1983). On the maximum of a random walk with small negative drift. *Ann. Probab.*, **11**, 491-505.

Lai, T. L. and Siegmund, D. (1979). A non-linear renewal theory with applications to sequential analysis, II. *Ann Probab.*, **7**, 60-76.

Siegmund, D. (1979). Corrected diffusion approximations in certain random walk problems. *Adv. Appl. Probab.*, **11**, 701-719.

Siegmund, D. (1985). *Sequential Analysis.* New York: Springer.

Stone, C. (1965). On moment generating functions and renewal theory. *Ann. Math. Statist.*, **36**, 1298-1301.

Wu, Y. (1999). Second order expansions for the moments of minimum point of an unbalanced two-sided normal random walk. *Ann. Inst. Statist. Math.*, **51**, 187-200.

Chapter 3

Confidence Interval for Change Point

The goal of this chapter is to construct confidence interval for the change point after CUSUM test. This chapter is organized as follows. Section 3.1 gives the general description of the problem. The method of constructing the confidence interval for the change point is proposed in Section 3.2. Section 3.3 presents the main results and their intuitive derivations. The rigorous proofs are given in Section 3.4. Finally. in Section 3.5, we apply the results to the cases of detecting changes in the mean or variance of a normal process, and compare them with the simulation results.

3.1 Introduction

Suppose that $\{X_k\}$ are independent random variables with distribution function F_{θ_0} for $k \leq \nu$ and F_{θ_1} for $k > \nu$, where $\theta_0 < 0 < \theta_1$, F_{θ_0} and F_{θ_1} belong to a standard one parameter exponential family as defined in section 1.3. and ν is the change point. We also assume that the distribution function F_0 is strongly non-lattice.

For sequential detection of the change point, Page(1954) proposed the CUSUM procedure where an alarm is made at

$$N = \inf\{n > 0 : \; T_n > d\}, \tag{3.1}$$

where T_n is the CUSUM process

$$T_n = \max(0, T_{n-1} + X_n), \quad with \quad T_0 = 0, \tag{3.2}$$

and d is the control limit which is chosen to give a specified ARL_0. Details of how to change the classical CUSUM procedure to the current version of CUSUM procedure will be given in the section 3.5.

The optimality of the CUSUM procedure has been studied by Lorden (1971) and Moustakides (1986), and the comparison with other procedures can be seen in Pollak and Siegmund (1985), Roberts (1966), Srivastava and Wu (1993). But so far, less attention has been paid to the inference problem after CUSUM test. In this chapter, we propose a method for constructing a lower confidence interval of the change point ν after the sequential CUSUM detection. From quality control point of view, this is a rather critical issue. In order to decide how long the process has been out of control before the detection, we need to estimate the change point. It is well known that when $\psi(\theta_0) = \psi(\theta_1)$, the maximum likelihood estimator for ν is

$$\hat{\nu} = \max\{n < N : \ T_n = 0\}, \tag{3.3}$$

i.e. the first zero point of T_n counting backward from the detection time N. Note that $\hat{\nu}$ can be computed recursively by memorizing the last zero point before the alarming time N.

To construct a confidence interval for ν, the basic difficulty is the memory problem. As the change usually occurs quite far away from the starting pcint, it is simply impossible to restore all the data until the alarming time. Motivated from the maximum likelihood estimator, a simple method of constructing lower confidence interval for the change point is proposed as follows:

Instead of tracking only the last zero point of T_n before detection, we keep tracking the last s zero points before the current detection time. By properly choosing the value of s, we can construct a confidence interval for ν formed from the s-th last zero point until the detection time to guarantee that it has the given coverage probability, say $1 - \alpha$.

An obvious advantage of this method is that all the related quantities can be computed recursively with fixed memory.

By assuming that ν and d approach infinity (which is true in most practical cases), we are able to find the asymptotic coverage probabilities and the average length of confidence intervals conditioning on the change being detected. By further assuming that θ_0 and θ_1 are small, we give the second order expansions for the quantities involved in calculating the coverage probabilities and the lengths of confidence intervals. The assumption is purely technical in order to give simple formulae for the related quantities, and also for checking the accuracy of the asymptotic results.

3.2 Method

The basic assumption here is that ν and d approach infinity, which is satisfied in most practical cases as the change usually occurs quite far away from the beginning and the average in control run length is quite large.

The following notations are standard in our discussion. Denote

$$S_n = S_0 + \sum_{i=1}^{n} X_i, \quad where \;\; S_0 \; issomeconstant.$$

Let

$$M = \sup_{0 \le k < \infty} S_n, \quad when \quad S_0 = 0.$$

Also, let

$$N_x = \inf\{n : \; S_n \le 0 \;\; or \quad > d | S_0 = x\}, \quad for \quad x \ge 0,$$

the two-sided boundary crossing time.

For notational convenience, let $P^\nu(.) = P^\nu_{\theta_0,\theta_1}(.)$ denote the probability when the change point is ν, $P_{\theta_1}(.) = P^0(.)$ when the change occurs at zero, and $P_{\theta_0}(.) = P^\infty(.)$ when there is no change occurring. Similarly, we denote $E^\nu(.)$, $E_{\theta_1}(.)$ and $E_{\theta_0}(.)$ as the corresponding expectations.

Now the method of constructing the confidence interval for ν conditioning on $N > \nu$ is formally described as follows:

Let L_k denote the k-th last zero point of $T_n = 0$ counted backward from $\hat{\nu}$ for $k = 1, 2, \ldots$ and let $L_0 = \hat{\nu}$. Then $[L_s, N)$ is a $1-\alpha$-level confidence interval if we choose s satisfying $s = \inf\{k \geq 0 : P^\nu(\nu < L_k | N > \nu) \leq \alpha\}$.

The advantage of this method is that one can calculate all related quantities by using the parallel computing technique for any given value of s due to the renewal property of T_n.

Also, it is noted that this is actually a conservative method. To obtain more accurate lower confidence limit L^*, one can interpolate between L_s and L_{s-1}.

Now we turn to the problem of calculating the non-coverage probability

$$p_s = P^\nu(\nu < L_s | N > \nu). \tag{3.4}$$

Firstly, it is noted that, conditioning on the value of T_ν given $N > \nu$, the zero points of T_n for $n > \nu$ will behave like a defective delayed renewal process and the total number of zero points after the first zero point follows a geometric distribution with terminating probability $P_{\theta_1}(S_{N_0} > d)$.

Secondly, we note that the event $\{\nu < L_s\}$ occurs if and only if the number of zero points after ν is larger than or equal to $s+1$. Thus, we have

$$p_s = P_{\theta_1}(S_{N_{T_\nu}} \le 0 | N > \nu)[P_{\theta_1}(S_{N_0} \le 0)]^s, \tag{3.5}$$

where T_ν is independent of $\{S_n - T_\nu, n > \nu\}$. In Section 3.3, the asymptotic values of p_s as well as the average length of $E[N - L_s | N > \nu]$ will be derived.

3.3 Main Results and Their Intuitive Derivations

In this section, the basic assumptions are both ν and d approach infinity and $\theta_0 < 0 < \theta_1$. From Pollak and Siegmund (1986), as ν and $d \to \infty$,

$$P_{\theta_0}(T_\nu \in dx | N > \nu) \to P_{\theta_0}(M \in dx). \tag{3.6}$$

Thus,

$$P_{\theta_1}(S_{N_{T_\nu}} \le 0 | N > \nu) \to p_0 = P_{\theta_1}(\tau_{-M} < \infty), \tag{3.7}$$

where M is the maximum value of another independent copy of $\{S_n\}$ with drift μ_0. Also, it is obvious that

$$P_{\theta_1}(S_{N_0} \le 0) \to p = P_{\theta_1}(\tau_- < \infty). \tag{3.8}$$

Therefore, we have the following theorem.

Theorem 3.1: *As* $d, \nu \to \infty$,

$$p_s \to p_0 p^s.$$

To force the non-coverage probability less than α, *s is thus to be taken as* $s = \lceil \ln(\alpha/p_0)/\ln p \rceil$, *where* $\lceil x \rceil$ *represents the smallest integer which is no less than* x.

Once the confidence interval is given, the next natural question is to evaluate its expected length. From quality control point of view, this tells us the number of items we have to inspect in order to guarantee the uniformity of the quality.

By conditioning on $N > \nu$ and the location of the change point ν, we can calculate the average length as follows.

(i) With probability $(1 - p_0)$, the event $\{\nu > \hat{\nu}\}$ occurs and the corresponding average length is

$$sE_{\theta_0}(N_0|S_{N_0} \leq 0) + E_{\theta_0}[\nu - \hat{\nu}|\nu > \hat{\nu}] + E_{\theta_1}[N_{T_\nu}|S_{N_{T_\nu}} > d].$$

Firstly, it is noted that, as $d \to \infty$,

$$E_{\theta_0}(N_0|S_{N_0} \leq 0) = E_{\theta_0}(\tau_-) + o(1).$$

Secondly, we denote

$$S'_n = S_\nu - S_{\nu-n}, \qquad for \qquad 0 \leq n \leq \nu.$$

Then $\hat{\nu} - \nu$ is actually the maximum point of S'_n. In fact, from the definition of $\hat{\nu}$, we know that $T_{\hat{\nu}} = 0$ and $S_{\hat{\nu}} = \min_{0 \leq k \leq \nu} S_k$. Therefore

$$T_\nu = S_\nu - \min_{0 \leq k \leq \nu} S_k.$$

Denote σ_x as the maximum point of S'_n with maximum value x. The following asymptotic result holds.

$$sE_{\theta_0}(N_0|S_{N_0} \leq 0) + E_{\theta_0}[\nu - \hat{\nu}|\nu > \hat{\nu}] + E_{\theta_1}[N_{T_\nu}|S_{N_{T_\nu}} > d]$$

$$= sE_{\theta_0}(\tau_-) + E_{\theta_1}[N_M|S_{N_M} > d] + E[\sigma_M|\tau_{-M} = \infty] + o(1),$$

where the following two facts are used:

(1) The event $\{\nu > \hat{\nu}\}$ is asymptotically equivalent to the event $\{\tau_{-M} = \infty\}$.

(2) $\nu - \hat{\nu}$ is asymptotically equivalent in distribution to the maximum point σ_M of another independent copy of $\{S_n\}$ with the maximum value M if we look at T_n backward starting from ν.

Both facts can be derived from (3.6).

(ii) With probability $p_0 p^s$, the event $\{\nu < L_s\}$ occurs and the average length is

$$sE_{\theta_1}[N_0|S_{N_0} \leq 0] + E_{\theta_1}[N_0|S_{N_0} > d]$$

$$= sE_{\theta_1}[\tau_-|\tau_- < \infty] + E_{\theta_1}[N_0|S_{N_0} > d] + o(1).$$

(iii) With probability $(1-p)p^{k-1}p_0$, the event $\{L_k < \nu \leq L_{k-1}\}$ occurs and the corresponding average length is

$$E_{\theta_0}[\nu - L_k|L_k < \nu \leq L_{k-1}] + E_{\theta_1}[N_{T_\nu}|S_{N_{T_\nu}} \leq 0] + E_{\theta_1}[N_0|S_{N_0} > d]$$

$$+(k-1)E_{\theta_1}[N_0|S_{N_0} \leq 0] + (s-k)E_{\theta_0}[N_0|S_{N_0} \leq 0]$$

$$= E[\sigma_M + \tau_{-M}|\tau_{-M} < \infty] + E_{\theta_1}[N_0|S_{N_0} > d] + (k-1)E_{\theta_1}[\tau_-|\tau_- < \infty] + (s-k)E_{\theta_0}(\tau_-),$$

for $k = 1, ..., s$, where again we use the fact that event $\{S_{N_{T_\nu}} \le 0\}$ is asymptotically equivalent to the event $\{\tau_{-M} < \infty\}$.

The only two quantities depending on d are $E_{\theta_1}[N_0 | S_{N_0} > d]$ and $E_{\theta_1}[N_M | S_{N_M} > d]$ which can be calculated as follows:

First, we note that

$$E_{\theta_1}[N_0; S_{N_0} > d] = E_{\theta_1}[N_0] - E_{\theta_1}[N_0; S_{N_0} \le 0]$$

$$= \frac{1}{\mu_1} E_{\theta_1}[S_{N_0}] - E_{\theta_1}[\tau_-; \tau_- < \infty] + o(1)$$

$$= \frac{1}{\mu_1}(E_{\theta_1}[S_{N_0}; S_{N_0} > d] + E_{\theta_1}[S_{N_0}; S_{N_0} \le 0])$$

$$- E_{\theta 1}[\tau_-; \tau_- < \infty] + o(1)$$

$$= \frac{1}{\mu_1}((1-p) E_{\theta 1}[S_{N_0} | S_{N_0} > d] + E_{\theta_1}[S_{\tau_-}; \tau_- < \infty])$$

$$- E_{\theta_1}[\tau_-; \tau_- < \infty] + o(1)$$

$$= \frac{1}{\mu_1}((1-p)(d + E_{\theta_1} R_\infty) + E_{\theta_1}[S_{\tau_-}; \tau_- < \infty])$$

$$- E_{\theta_1}[\tau_-; \tau_- < \infty] + o(1). \tag{3.9}$$

Similarly,

$$E_{\theta_1}[N_M; S_{N_M} > d] = E_{\theta_1}[N_M] - E_{\theta_1}[N_M; S_{N_M} \le 0]$$

$$= \frac{1}{\mu_1}[E_{\theta_1}(S_{N_M}) - E_{\theta_0}(M)] - E_{\theta_1}[\tau_{-M}; \tau_{-M} < \infty] + o(1)$$

$$= \frac{1}{\mu_1}[(1-p_0)(d + E_{\theta_1} R_\infty) + E_{\theta_1}[S_{\tau_{-M}} + M; \tau_{-M} < \infty]$$

$$-E_{\theta_0}[M]] - E_{\theta_1}[\tau_{-M}; \tau_{-M} < \infty] + o(1). \tag{3.10}$$

Summing up the above results, we get

Theorem 3.2: *As $\nu, d \to \infty$, the average length of the confidence interval $[L_s, N)$ conditioning on $N > \nu$ is asymptotically equal to*

$$E_\nu[N - L_s|N > \nu] = \frac{p_0}{1-p} E_{\theta_1}[N_0; S_{N_0} > d] + E_{\theta_1}[N_M; S_{N_M} > d]$$

$$+E_{\theta_1}[\sigma_M; \tau_{-M} = \infty] + (1 - p^s) E[\tau_{-M} + \sigma_M; \tau_{-M} < \infty]$$

$$+\frac{p_0(1-p^s)}{(1-p)}(E_{\theta_1}[\tau_-; \tau_- < \infty] - E_{\theta_0}[\tau_-]) + sE_{\theta_0}[\tau_-] + o(1), \tag{3.11}$$

where $E_{\theta_1}[N_0; S_{N_0} > d]$ and $E_{\theta_1}[N_M; S_{N_M} > d]$ are given by (3.9) and (3.10) respectively.

Remark: For fixed values of θ_0 and θ_1, it can be seen from Theorem 2 that the first order of the average length of the confidence interval is approximately $(d + E_{\theta_1} R_\infty)/(\mu_1)$.

The case that $s = 0$ is of particular interest as it gives the average length from the estimator $\hat{\nu}$ to the alarm time N.

The difficulty to apply these results is that there are no explicit expressions for the related quantities. In the next section, we shall give certain local second order expansions for the non-coverage probabilities and the average lengths of the confidence intervals when both θ_0 and θ_1 approach 0, and $\theta_0 d \to \infty$.

3.4 Local Second Order Approximations

In this section, second order approximations are developed under the following assumption:

(A): *Both θ_0 and θ_1 approach zero at the same order, and $|\theta_0|^{1+\gamma}d \to \infty$ for some* $\gamma > 0$.

The assumption is purely technical, and is satisfied in most practical situations. As we shall see, the approximations are quite accurate and can be easily used for designing the confidence interval for given significance level, say, α.

The techniques used here are those developed in Chapter 2. The following lemma gives the results concerning the convergence rate of (3.6).

Lemma 3.1: *As $\nu \to \infty$, we have*

$$P^{\nu}(T_\nu > x | N > \nu) = P_{\theta_0}(M > x) + O(e^{\Delta_0 d}).$$

Proof: First we write

$$P^{\nu}(T_\nu > x|N > \nu) = \frac{P_{\theta_0}(T_\nu > x, N > \nu)}{P^{\nu}(N > \nu)}$$

$$= \frac{P_{\theta_0}(T_0 = 0, 0 \le T_1 \le d, \ldots, 0 \le T_{\nu-1} \le d, x < T_\nu \le d)}{P_{\theta_0}(T_0 = 0, 0 \le T_1 \le d, \ldots, 0 \le T_\nu \le d)}. \tag{3.12}$$

Let J be the number of zero points of $\{T_n, 0 \le n < \nu\}$, and $\{L_j, 1 \le j \le J\}$ be the time n when $T_n = 0$ counting from $n = 0$. Then we have

$$P_{\theta_0}(T_0 = 0, 0 \le T_1 \le d, \ldots, 0 \le T_{\nu-1} \le d, x < T_\nu \le d)$$

$$= \sum_{j=1}^{\nu-1} [P_{\theta_0}(S_{N_0} \leq 0)]^{j-1} P_{\theta_0}(T_{L_j} = 0, 0 < T_{L_j+1} \leq d, ..., 0 < T_{\nu-1} \leq d, x < T_\nu \leq d).$$

If we look at $\{T_{L_j}, T_{L_j+1}, ..., T_\nu\}$ backward, then it is a random walk starting from T_ν with increment $-X_i$, where $X_i's$ are iid with distribution function F_{θ_0}. Thus by Theorem 1 (or (1)) of Siegmund (1976), we have

$$P_{\theta_0}(T_{L_j} = 0, 0 < T_{L_j+1} \leq d, ..., 0 \leq T_{\nu-1} \leq d, x < T_\nu \leq d) = P_{\theta_0}(-S_{N_x} \leq 0). \quad (3.13)$$

as $\nu \to \infty$.

Similarly, we have

$$P_{\theta_0}(T_0 = 0, 0 \leq T_1 \leq d, ..., 0 \leq T_\nu \leq d) = \sum_{j=1}^{\nu-1} [P_{\theta_0}(S_{N_0} \leq 0)]^{j-1} (1 - P_{\theta_0}(-S_{N_d} \leq 0)). \quad (3.14)$$

as $\nu \to \infty$.

By the total probability law, we have

$$P_{\theta_0}(\tau_x < \infty) = P_{\theta_0}(\tau_x < \infty, -S_{N_x} \leq 0) + P_{\theta_0}(\tau_x < \infty, -S_{N_x} > d)$$

$$= P_{\theta_0}(-S_{N_x} \leq 0) + P_{\theta_0}(\tau_{S_{N_x}} < \infty | S_{N_x} > d) P_{\theta_0}(-S_{N_x} > d).$$

By Theorem 2.6,

$$P_{\theta_0}(\tau_{S_{N_x}} < \infty | S_{N_x} > d) = O(e^{\Delta_0 d}).$$

Thus

$$P_{\theta_0}(-S_{N_x} \leq 0) = P_{\theta_0}(\tau_x < \infty) + O(e^{\Delta_0 d}).$$

Similarly, we can show that

$$P_{\theta_0}(-S_{N_d} \le 0) = O(e^{\Delta_0 d}).$$

Combining results obtained and the fact that $P_{\theta_0}(\tau_x < \infty) = P_{\theta_0}(M > x)$, Lemma 3.1 is proved.

Combining Lemma 3.1, (3.7) and Lemma 2.3, we have

Lemma 3.2: *Under the assumption of (A),*

$$p_0 = \tilde{p}_0 + O(\Delta_0^3) + O(e^{\Delta_0 d}),$$

where $\tilde{p}_0$ is the quantity defined in (2.13).

Next lemma gives approximation for p.

Lemma 3.3: *Under the assumption of (A),*

$$p = P_{\theta_1}(S_{N_0} \le 0) = 1 + \Delta_1 E_0 S_{\tau_-} e^{\frac{\Delta_1}{2}\rho_-} + O(\Delta_1^3) + O(e^{-\Delta_1 d}).$$

Proof: By the total probability law, we have

$$P_{\theta_1}(\tau_- < \infty) = P_{\theta_1}(\tau_- < \infty, S_{N_0} \le 0) + P_{\theta_1}(\tau_- < \infty, S_{N_0} > d)$$

$$= P_{\theta_1}(S_{N_0} \le 0) + P_{\theta_1}(\tau_{-S_{N_0}} < \infty | S_{N_0} > d) P_{\theta_1}(S_{N_0} > d).$$

Thus

$$P_{\theta_1}(S_{N_0} \le 0) = P_{\theta_1}(\tau_- < \infty) - P_{\theta_1}(\tau_{-S_{N_0}} < \infty | S_{N_0} > d) P_{\theta_1}(S_{N_0} > d). \quad (3.15)$$

From Theorem 2.6, it is easy to see that

$$P_{\theta_1}(\tau_{-S_{N_0}} < \infty | S_{N_0} > d) = O(e^{-\Delta_1 d}).$$

Then the proof is completed by applying Lemma 2.1.

Applying above approximations to Theorem 3.1, we have

Theorem 3.3: *Under the assumption of (A), as ν approach infinity, we have*

$$p_s \approx \tilde{p}_0(1 + \Delta_1 E_0 S_{\tau_-} e^{\frac{\Delta_1}{2}\rho_-})^s,$$

where $\tilde{p}_0$ is given in Lemma 2.3.

To approximate the average length of the confidence interval, by Theorem 3.2, we only need to approximate the follow items:

1. $E_{\theta_0}[\tau_-]$;
2. $E_{\theta_1}[\tau_-; \tau_- < \infty]$;
3. $E[\tau_{-M} + \sigma_M; \tau_{-M} < \infty]$;
4. $E[\sigma_M; \tau_{-M} = \infty]$;
5. $E_{\theta_1}[N_0; S_{N_0} > d]$;
6. $E_{\theta_1}[N_M; S_{N_M} > d]$.

By Wald's Identity and Theorem 2.1,

$$E_{\theta_0}[\tau_-] = \frac{1}{\mu_0} E_0 S_{\tau_-} e^{\frac{\Delta_0}{2}\rho_-} + O(\Delta_0). \tag{3.16}$$

The approximation for $E_{\theta_1}[\tau_-; \tau_- < \infty]$ has been given in Lemma 2.1.

From Lemma 2.4, we know that

$$E[\tau_{-M}; \tau_{-M} < \infty] = \frac{\Delta_0}{\tilde{\mu}_1(\Delta_1 - \Delta_0)^2} e^{\Delta_0(\rho_+ + \rho_-)} + O(1). \quad (3.17)$$

In the following, we give a series of lemmas which contain approximations for the rest of items.

Lemma 3.4: *As θ_0 and $\theta_1 \to 0$,*

$$E_{\theta_0}[\sigma_M] = -\frac{1}{\tilde{\mu}_0\Delta_0} + O(1),$$

$$E[\sigma_M; \tau_{-M} < \infty] = -\frac{\Delta_0}{\tilde{\mu}_0(\Delta_1 - \Delta_0)^2} e^{\Delta_1(\rho_+ + \rho_-)} + O(1).$$

$$E[\sigma_M; \tau_{-M} = \infty] = -\frac{1}{\tilde{\mu}_0\Delta_0} + \frac{\Delta_0}{\tilde{\mu}_0(\Delta_1 - \Delta_0)^2} e^{\Delta_1(\rho_+ + \rho_-)} + O(1).$$

Proof: Let $\tau_+^{(i)}$, for $i > 0$, be iid r.v.s. which have the same distribution function as that of $(\tau_+ | \tau_+ < \infty)$, and K be a Geometric random variable with $p = P_{\theta_0}(\tau_+ = \infty)$, which is independent of all variables involved. Then σ_M is equivalent in distribution to $\sum_{i=1}^{K-1} \tau_+^{(i)}$. Thus

$$E_{\theta_0}(\sigma_M) = \sum_{k=2}^{\infty}\sum_{i=1}^{k-1} E_{\theta_0}(\tau_+^{(i)}(1 - P_{\theta_0}(\tau_+ = \infty))^{k-1} P_{\theta_0}(\tau_+ = \infty)$$

$$= \sum_{k=2}^{\infty}(k-1)E_{\theta_0}(\tau_+, \tau_+ < \infty)P_{\theta_0}(\tau_+ = \infty)$$

$$= \frac{E_{\theta_0}(\tau_+, \tau_+ < \infty)}{P_{\theta_0}(\tau_+ = \infty)}.$$

From Lemma 2.1, we have

$$E_{\theta_0}(\tau_+, \tau_+ < \infty) = \frac{1}{\tilde{\mu}_0} E_0(S_{\tau_+}) e^{\frac{1}{2}\Delta_0 \rho_+} + O(\Delta_0),$$

and

$$P_{\theta_0}(\tau_+ = \infty) = -\Delta_0 E_0(S_{\tau_+}) e^{\frac{1}{2}\Delta_0 \rho_+} + O(\Delta_0^3).$$

So, we have

$$E_{\theta_0}(\sigma_M) = -\frac{1}{\tilde{\mu}_0 \Delta_0} + O(1).$$

To prove the second result of Lemma 3.4. first note that M is equivalent in distribution to $\sum_{i=1}^{K-1} S_{\tau_+}^{(i)}$, where $S_{\tau_+}^{(i)}$. for $i > 0$. are iid r.v.s. which have the same distribution function as that of $(S_{\tau_+} | \tau_+ < \infty)$. Similar to the proof of Lemma 2.4. we have

$$E[\sigma_M, \tau_M < \infty] = E[\sigma_M P_{\theta_1}(\tau_{-M} < \infty | M)]$$

$$= E_{\theta_0}[\sigma_M e^{-\Delta_1 (M - \rho_-)}](1 + O(\Delta_1^2))$$

$$= \sum_{k=2}^{\infty} E_{\theta_0}[(\sum_{i=1}^{k-1} \tau_+^{(i)}) e^{-\Delta_1 \sum_{i=1}^{k-1} S_{\tau_+^{(i)}}}](1 - P_{\theta_0}(\tau_+ = \infty)) P_{\theta_0}(\tau_+ = \infty) e^{\Delta_1 \rho_-}(1 + O(\Delta_1^2))$$

$$= \sum_{k=2}^{\infty} (k-1) E^{\theta_0}(\tau_+ e^{-\Delta_1 S_{\tau_+}}, \tau_+ < \infty)[E_{\theta_0}(\epsilon^{-\Delta_1 S_{\tau_+}}, \tau_+ < \infty)]^{k-2} P_{\theta_0}(\tau_+ = \infty) e^{\Delta_1 \rho_-}$$

$$(1 + O(\Delta_1^2))$$

$$= P_{\theta_0}(\tau_+ = \infty) e^{\Delta_1 \rho_-} E_{\theta_0}(\tau_+ e^{-\Delta_1 S_{\tau_+}}, \tau_+ < \infty) \frac{1}{(1 - E_{\theta_0}(e^{-\Delta_1 S_{\tau_+}}, \tau_+ < \infty))^2}(1 + O(\Delta_1^2)).$$

By similar technique used in the proof of Lemma 2.1, we have

$$E_{\theta_0}(\tau_+ e^{-\Delta_1 S_{\tau_+}}, \tau_+ < \infty) = \frac{1}{\tilde{\mu}_0} E_0(S_{\tau_+})(1 - \frac{1}{2}\Delta_0 \rho_+ - (\Delta_1 - \Delta_0)\rho_+) + O(\Delta_1),$$

and

$$E_{\theta_0}(e^{-\Delta_1 S_{\tau_+}}, \tau_+ < \infty) = 1 - (\Delta_1 - \Delta_0)E_0(S_{\tau_+})(1 - \frac{1}{2}\Delta_0\rho_+ - (\Delta_1 - \Delta_0)\rho_+) + O(\Delta_1^3).$$

So

$$E[\sigma_M, \tau_M < \infty] = -\frac{\Delta_0}{\tilde{\mu}_0(\Delta_1 - \Delta_0)^2} e^{\Delta_1(\rho_+ + \rho_-)} + O(1).$$

By the relationship of

$$E[\sigma_M, \tau_M = \infty] = E[\sigma_M] - E[\sigma_M, \tau_M < \infty],$$

we finish the proof of Lemma 3.4.

Lemma 3.5 : *As* $\theta_1 \to 0$.

$$E_{\theta_1}[N_0; S_{N_0} > d] = -\frac{d + \rho_+}{\mu_1}\Delta_1 E_0 S_{\tau_-} e^{\frac{1}{2}\Delta_1\rho_-} + \frac{1}{\mu_1} E_0 S_{\tau_-} e^{\frac{3}{2}\Delta_1\rho_-} - \frac{1}{\tilde{\mu}_1} E_0 S_{\tau_-} e^{\frac{1}{2}\Delta_1\rho_-} + O(\Delta_1).$$

Proof: Firstly, we note that

$$E_{\theta_1} R_\infty = \rho_+(\theta_1) = \rho_+ + O(\theta_1).$$

By similar technique used in proving Lemma 2.1, we know that

$$E_{\theta_1}[S_{\tau_-}; \tau_- < \infty] = E_{\tilde{\theta}_1} S_{\tau_-} e^{\Delta_1 S_{\tau_-}}$$

$$= E_{\tilde{\theta}_1}[S_{\tau_-} + \Delta_1 S_{\tau_-}^2] + O(\Delta_1^2)$$

$$= E_0 S_{\tau_-} e^{\frac{3}{2}\Delta_1\rho_-} + O(\Delta_1^2). \tag{3.18}$$

The proof of Lemma 3.5 is completed by combining (3.9), Lemma 2.1 and the above approximations.

Lemma 3.6: *As θ_0 and $\theta_1 \to 0$,*

$$E[N_M; S_{N_M} > d] = \frac{1}{\mu_1}[(d+\rho_+)(1-\tilde{p}_0)+\rho_-\tilde{p}_0+\frac{1}{\Delta_0}+\rho_+] - E[\tau_{-M}; \tau_{-M} < \infty] + O(1).$$

Proof: By the same technique used in proving Lemma 2.4, we have

$$E_{\theta_1}[S_{\tau_{-M}} + M; \tau_{-M} < \infty] = \rho_-\tilde{p}_0 + O(\Delta_1).$$

Combining (3.10) and the fact that

$$E_{\theta_0}M = -\frac{1}{\Delta_0} - \rho_+ - \frac{\Delta_0}{2}r_1 + O(\Delta_0^2).$$

Lemma 3.6 is proved.

3.5 Application

This section demonstrates how to use results obtained in the practical situation, and conducts simulation studies to check the accuracy of the approximations given in Section 3.3 and 3.4.

In the field of quality control, we might be monitoring some characteristic of a manufacturing process. In many practical situations, the quality characteristic process X_t is assumed to be normally distributed with mean μ_0 and standard deviation σ_0, where μ_0 is the target value and σ_0 reflects the variation of quality. Any shift from the target value μ_0 or increase in the process variation results in poor quality, and we want to detect the change as soon as possible. To monitor the quality characteristic process, groups of random samples of size, say m, are usually taken at some regular time intervals, and CUSUM charts based on either sample mean $\bar{X}_t$ or sample variance s_t^2 are plotted with the appropriate control limit(s) respectively. We would like to stop the process for inspection and repair when some points of the CUSUM chart fall out side of control limit(s). Since false alarms are often costly, we therefore assume the in control Average Run Length (ARL) of samples to be very large which ensures that the assumption (A) is true.

In the following, we will apply our results to detecting changes in either the mean or the standard deviation of a normal process. For simplicity, we only consider detecting the increase in mean or variance.

3.5.1 Detect Increase in Mean

Without lose of generality, we assume that the observed process $\bar{X}_t$ comes from a normal population with mean 0 and standard deviation 1 when the process is in control, and with mean μ and standard deviation 1 when the process is out of control. Then the CUSUM procedure is defined as making alarm at

$$N = \inf\{n > 0 : \ T_n > d\},$$

where T_n is the CUSUM process

$$T_n = \max(0, T_{n-1} + Y_n), \quad with \quad T_0 = 0,$$

where $Y_n = \bar{X}_n - \frac{\delta}{2}$, and d is the control limit with reference value δ. Usually, the reference value δ is the change magnitude which we are interested in detecting quickly, and serves as an preliminary estimate of μ, the true change magnitude.

In this case, $\psi(\theta) = \frac{1}{2}\theta^2$; $\theta_0 = -\frac{\delta}{2}$, $\theta_1 = \mu - \frac{\delta}{2}$; $\mu_0 = -\tilde{\mu}_0 = -\frac{\delta}{2}$, $\mu_1 = -\tilde{\mu}_1 = \mu - \frac{\delta}{2}$; $E_0 S_{\tau_+} = -E_0 S_{\tau_-} = \frac{1}{\sqrt{2}}$; $\rho_+ = -\rho_- \approx 0.583$, which will be denoted as ρ; $r_0 = r_1 = \frac{1}{4}$ and $C_0 = \frac{1}{2}(\frac{1}{\sqrt{2}} - \rho_+)^2$. Substitute the related quantities into Theorem 3.3 and 3.2, we have

Corollary 3.1: *Under Condition (A), we have*

$$\tilde{p}_0 = \frac{\delta}{2\mu} e^{-\mu(\mu-\delta/2)}. \tag{3.19}$$

$$p = 1 - \sqrt{2}(\mu - \frac{\delta}{2}) e^{-(\mu-\delta/2)\rho} + O(\delta^3). \tag{3.20}$$

and thus we have

$$p_s \approx \frac{\delta}{2\mu} e^{-\mu(\mu-\delta/2)} (1 - \sqrt{2}(\mu - \frac{\delta}{2}) e^{-(\mu-\delta/2)\rho})^s.$$

At $\mu = \delta$,

$$p_s \approx \frac{1}{2} e^{-\delta^2/2} (1 - \frac{\delta}{\sqrt{2}} e^{-\delta\rho/2})^s.$$

Corollary 3.2: *Under Condition (A), we have*

$$E^\nu[N - L_s | N > \nu] = \frac{1}{\mu - \delta/2}(d + 2\rho) - \frac{1}{\delta(\mu - \delta/2)} - \frac{\tilde{p}_0}{(\mu - \delta/2)^2} + \frac{2}{\delta^2} - \frac{p^s}{2\mu(\mu - \delta/2)}$$
$$+ \frac{\tilde{p}_0(1 - p^s)}{(1 - p)} (E_{\mu-\delta/2}[\tau_-; \tau_- < \infty] - E_{-\delta/2}[\tau_-]) + sE_{-\delta/2}[\tau_-] + O(1),$$

where

$$E_{\mu-\delta/2}[\tau_-;\tau_-<\infty]=\frac{1}{\sqrt{2}(\mu-\delta/2)}e^{-(\mu-\delta/2)\rho}+O(\delta), \tag{3.21}$$

$$E_{-\delta/2}[\tau_-]=\frac{\sqrt{2}}{\delta}e^{\frac{1}{2}\delta\rho}+O(\delta). \tag{3.22}$$

At $\mu=\delta$,

$$E^{\nu}[N-L_s|N>\nu]=\frac{2(d+2\rho)}{\delta}-\frac{4\tilde{p}_0}{\delta^2}-\frac{p^s}{\delta^2}$$

$$+sE_{-\delta/2}[\tau_-]+\frac{\tilde{p}_0(1-p^s)}{(1-p)}(E_{\delta/2}[\tau_-;\tau_-<\infty]-E_{-\delta/2}[\tau_-])+O(1).$$

To check the accuracy of the approximations, we conduct a simulation study. For average in-control run length $ARL_0=E_{-\delta/2}N=1000$, we take $\delta=0.4, 0.5$, and 0.6. The design for the control limit d is given by using the approximation (2.57) of Siegmund (1985):

$$ARL_0\approx\frac{2}{\delta^2}(e^{\delta(d+2\rho)}-1-\delta(d+2\rho)).$$

The accuracy of this approximation has been discussed by Wu (1994). From this formula, the values of d are given by 9.96, 8.59 and 7.56 corresponding to $\delta=0.4, 0.5$ and 0.6, respectively. The change point $\nu=100$. Comparisons of theoretical values with simulation values of non-coverage probabilities are given in Table 3.1, 3.2 and 3.3, where $\theta_0=-\delta/2$ and $\theta_1=\mu-\delta/2$. In each cell, the top number is the approximated value while the bottom number is the simulated value.

Table 3.1: Non-coverage Probability for $\theta_0 = \delta/2 = -0.2$

$\theta_1 \backslash s$	0	1	2	3	4	5	6	7	8
0.2	0.4616	0.3454	0.2584	0.1934	0.1443	0.1083	0.0810	0.0606	0.0454
	0.4602	0.3414	0.2505	0.1843	0.1385	0.1048	0.0784	0.0574	0.0425
0.25	0.3972	0.2758	0.1915	0.1330	0.0923	0.0641	0.0445	0.0309	0.0215
	0.4025	0.2808	0.1928	0.1296	0.0917	0.0606	0.0436	0.0317	0.0237
0.3	0.3443	0.2217	0.1427	0.0919	0.0591	0.0381	0.0245	0.0158	0.0102
	0.3525	0.2284	0.1499	0.0968	0.0614	0.0409	0.0261	0.0165	0.0113

Comparisons of theoretical values with simulation values for the length of $E^{\nu}[N - L_s]$ are presented in table 3.4, 3.5 and 3.6, where $\theta_0 = -\delta/2$, $\theta = \mu - \delta/2$. Values for s are chosen to be 0, which gives length corresponding to $E^{\nu}[N - \hat{\nu} | N > \nu]$, and other two numbers, which give lengths corresponding to 90% and 95% confidence intervals. For the third row, the top numbers are the approximated values of $E^{\nu}[N - L_s | N > \nu]$, while the middle numbers are the simulation values, and the the bottom numbers given in the parenthesis are standard errors for the simulated values.

From these tables, we can see that approximations obtained in this chapter are surprisingly accurate.

Table 3.2: Non-coverage Probability for $\theta_0 = -0.25$

$\theta_1 \backslash s$	0	1	2	3	4	5	6	7	8
0.25	0.4412 0.4528	0.3064 0.3117	0.2128 0.2148	0.1477 0.1479	0.1026 0.1045	0.0712 0.0720	0.0495 0.0503	0.0344 0.0347	0.0239 0.0251
0.3	0.3854 0.4012	0.2481 0.2541	0.1598 0.1636	0.1028 0.1041	0.0662 0.0654	0.0426 0.0442	0.0274 0.0281	0.0176 0.0180	
0.4	0.2966 0.3623	0.1637 0.2210	0.0904 0.1298	0.0499 0.0789	0.0275 0.0465	0.0152 0.0265			

Table 3.3: Non-coverage Probability for $\theta_0 = -0.30$

$\theta_1 \backslash s$	0	1	2	3	4	5	6	7
0.3	0.4176 0.4347	0.2689 0.2803	0.1731 0.1820	0.1115 0.1171	0.0718 0.0751	0.0462 0.0477	0.0297 0.0303	0.0192 0.0201
0.35	0.3676 0.3873	0.2193 0.2291	0.1308 0.1403	0.0780 0.0842	0.0465 0.0488	0.0277 0.0307	0.0165 0.0182	
0.4	0.3239 0.3472	0.1788 0.1967	0.0987 0.1081	0.0545 0.0576	0.0301 0.0297	0.0166 0.0167		

Table 3.4: $E^{\nu}[N - L_s|N > \nu]$ for $\theta_0 = -0.2$

θ_1	0.2	0.2	0.2	0.25	0.25	0.25	0.3	0.3	0.3
s	0	6	8	0	4	6	0	3	5
	37.84	65.58	73.89	36.21	53.98	62.23	34.09	47.04	55.21
	38.16	64.24	72.12	36.28	53.46	61.20	33.78	46.51	53.37
	(0.25)	(0.31)	(0.36)	(0.24)	(0.28)	(0.31)	(0.23)	(0.26)	(0.28)

Table 3.5: $E^{\nu}[N - L_s|N > \nu]$ for $\theta_0 = -0.25$

θ_1	0.25	0.25	0.25	0.3	0.3	0.3	0.4	0.4	0.4
s	0	5	6	0	4	5	0	3	4
	27.04	46.68	50.09	26.54	40.98	44.35	23.61	34.00	37.32
	28.57	46.58	49.81	25.33	40.87	43.96	23.39	33.41	36.77
	(0.22)	(0.26)	(0.28)	(0.21)	(0.25)	(0.26)	(0.21)	(0.24)	(0.25)

Table 3.6: $E^{\nu}[N - L_s|N > \nu]$ for $\theta_0 = -0.3$

θ_1	0.3	0.3	0.3	0.35	0.35	0.35	0.4	0.4	0.4
s	0	4	5	0	3	4	0	2	4
	21.64	34.36	37.28	20.50	29.82	32.72	19.37	25.52	31.30
	21.56	34.17	37.33	20.52	29.97	32.70	18.99	25.20	30.89
	(0.20)	(0.25)	(0.26)	(0.19)	(0.22)	(0.24)	(0.19)	(0.21)	(0.23)

3.5.2 Detect Increase in Variance

Without lose of generality, we assume that the observed process s_t^2 comes from a population of $\chi^2(p)$ when the process is in control, and from a population of $(1+\epsilon)^2\chi^2(p)(\epsilon > 0)$ when the process is out of control. Then the CUSUM procedure is defined as making alarm at

$$N = \inf\{n > 0 : \ T_n > d\},$$

where T_n is the CUSUM process

$$T_n = \max(0, T_{n-1} + Y_n), \qquad with \quad T_0 = 0,$$

and

$$Y_n = \frac{s_n^2[(1+\epsilon_0)^2 - 1]}{2\sqrt{2p}(1+\epsilon_0)^2 \ln(1+\epsilon_0)} - \sqrt{\frac{p}{2}},$$

and d is the control limit with reference value ϵ_0 which is the relative increase change magnitude we are interested in detecting quickly. The ϵ_0 also serves as an preliminary estimate of ϵ, the true relative increase change magnitude.

In this case,

$$\psi(\theta) = -\sqrt{\frac{p}{2}}\theta - \frac{p}{2}\ln(1 - \sqrt{\frac{2}{p}}\theta),$$
$$\theta_0 = \sqrt{\frac{p}{2}}(1 - \frac{2(1+\epsilon_0)^2\ln(1+\epsilon_0)}{(1+\epsilon_0)^2 - 1}),$$
$$\theta_1 = \sqrt{\frac{p}{2}}(1 - \frac{2(1+\epsilon_0)^2\ln(1+\epsilon_0)}{(1+\epsilon)^2[(1+\epsilon_0)^2 - 1]}),$$
$$\mu_0 = \sqrt{\frac{p}{2}}(\frac{(1+\epsilon_0)^2 - 1}{2(1+\epsilon_0)^2\ln(1+\epsilon_0)} - 1),$$
$$\mu_1 = \sqrt{\frac{p}{2}}(\frac{(1+\epsilon)^2[(1+\epsilon_0)^2 - 1]}{2(1+\epsilon_0)^2\ln(1+\epsilon_0)} - 1).$$

For simplicity, we only give results of a special case, i.e. $p = 2$. For $\epsilon_0 = 0.2, 0.25$ and 0.3, we choose the average in-control run length $ARL_0 = E_{\theta_0}N = 1000$, and $\nu = 100$. The control limit d is given by using the approximation (10.17) of Siegmund (1985):

$$ARL_0 \approx \frac{1}{\Delta_0\mu_0}(e^{-\Delta_0(d+\rho_+-\rho_-)} - 1 + \Delta_0(d + \rho_+ - \rho_-))$$

From this formula, we obtain that $d = 10.083, 8.794$, and 7.823 corresponding to $\epsilon_0 = 0.2, 0.25$, and 0.3 respectively. Simulation results are based on 10000 replications.

It is noted that, when $p = 2$, S_{τ_+} is a exponential random variable on $(0, \infty)$, and S_{τ_-} is uniform distributed on $(-1, 0)$. So, we have

$$\rho_+ = 1, r_1 = 1, \rho_- = -\frac{1}{3}, r_0 = \frac{1}{18} \quad and \quad C_0 = 0.$$

Substituting the above corresponding values into Theorem 3.2 and 3.3, The following results are obtained.

Corollary 3.3: *Under condition (A), we have*

$$\tilde{p}_0 = C_1 e^{\frac{25}{36C_1}\Delta_0\Delta_1}. \tag{3.23}$$

with $C_1 = -\frac{\Delta_0}{\Delta_1-\Delta_0}e^{\frac{2}{3}\Delta_1}$*;*

$$p = 1 - \frac{1}{2}\Delta_1 e^{-\frac{1}{6}\Delta_1} + O(\Delta_1^3). \tag{3.24}$$

and thus we have

$$p_s \approx C_1 e^{\frac{25}{36C_1}\Delta_0\Delta_1}(1 - \frac{1}{2}\Delta_1 e^{-\frac{1}{6}\Delta_1})^s.$$

Corollary 3.4: *Under condition (A), we have*

$$E^\nu[N - L_s|N > \nu] = \frac{d+2}{\mu_1} - \tilde{p}_0(\frac{1}{\Delta_1\mu_1} - \frac{1}{\Delta_1\tilde{\mu}_1}) + \frac{1}{\Delta_0\mu_1}$$

$$-\frac{1}{\Delta_0\tilde{\mu}_0} + \frac{\Delta_0}{\tilde{\mu}_1(\Delta_1-\Delta_0)^2}e^{2\Delta_0/3} - p^s(\frac{\Delta_0}{\tilde{\mu}_1(\Delta_1-\Delta_0)^2}e^{2\Delta_0/3} - \frac{\Delta_0}{\tilde{\mu}_0(\Delta_1-\Delta_0)^2}e^{2\Delta_1/3})$$

$$+\frac{\tilde{p}_0(1-p^s)}{(1-p)}(E_{\theta_1}[\tau_-;\tau_- < \infty] - E_{\theta_0}[\tau_-]) + sE_{\theta_0}[\tau_-] + o(1).$$

where

$$E_{\theta_1}[\tau_-;\tau_- < \infty] = -\frac{1}{2\tilde{\mu}_1}e^{-\Delta_1/6} + O(\Delta_1). \tag{3.25}$$

$$E_{\theta_0}[\tau_-] = -\frac{1}{2\mu_0}e^{-\Delta_0/6} + O(\Delta_0). \tag{3.26}$$

Table 3.7, 3.8, and 3.9 give the comparisons of the simulated values with the approximated values of the non-coverage probabilities for $s = 0, 1, \ldots,$ in some case up to 15. The top number in each cell is the approximated value, while the bottom number is the simulated value.

Again, similar to the last subsection, comparison of theoretical values with simulation values for the lengths of $E^{\nu}[N - L_s]$ are presented in table 3.10, 3.11 and 3.12. ϵ_0 are chosen as 0.2, 0.25, and 0.3. Values for s are chosen to be 0, which gives length corresponding to $E^{\nu}[N - \hat{\nu}|N > \nu]$, and other two numbers, which give lengths corresponding to 90% or 95% confidence intervals. For the third row, the top number is the approximated value of $E^{\nu}[N - L_s|N > \nu]$, while the middle number is the simulated value, and the the bottom number given in the parenthesis is standard error for the simulated value.

Once again, we find that the approximations obtained in this chapter are surprisingly accurate.

Table 3.7: Non-coverage Probability for $\epsilon_0 = 0.2$

$\epsilon\backslash s$	0	1	2	3	4	5	6	7	8
0.2	0.5516	0.4570	0.3786	0.3136	0.2598	0.2152	0.1783	0.1477	0.1224
	0.5627	0.4664	0.3814	0.3158	0.2644	0.2211	0.1834	0.1506	0.1268
0.25	0.4666	0.3560	0.2716	0.2072	0.1581	0.1206	0.0920	0.0702	0.0536
	0.4723	0.3672	0.2843	0.2158	0.1646	0.1266	0.0962	0.0720	0.0554
0.3	0.4084	0.2877	0.2027	0.1428	0.1006	0.0709	0.0499	0.0352	0.0248
	0.4189	0.2963	0.2126	0.1504	0.1041	0.0733	0.0502	0.0346	0.0226
$\epsilon\backslash s$	9	10	11	12	13	14	15		
0.2	0.1014	0.0840	0.0696	0.0576	0.0477	0.0396	0.0328		
	0.1035	0.0848	0.0712	0.0588	0.0472	0.0395	0.0328		
0.25	0.0409	0.0312	0.0238	0.0181	0.0138	0.0106	0.0081		
	0.0431	0.0327	0.0252	0.0188	0.0151	0.0113	0.0091		

Table 3.8: Non-coverage Probability for $\epsilon_0 = 0.25$

$\epsilon \backslash s$	0	1	2	3	4	5	6	7	8
0.25	0.5482	0.4347	0.3446	0.2732	0.2166	0.1718	0.1361	0.1080	0.0856
	0.5578	0.4426	0.3524	0.2820	0.2213	0.1765	0.1385	0.1100	0.0889
0.3	0.4789	0.3507	0.2568	0.1881	0.1377	0.1009	0.0739	0.0541	0.0396
	0.4904	0.3637	0.2640	0.1930	0.1434	0.1067	0.0786	0.0553	0.0401
0.35	0.4288	0.2908	0.1971	0.1337	0.0906	0.0614	0.0417	0.0282	0.0192
	0.4302	0.2916	0.1971	0.1353	0.0906	0.0634	0.0450	0.0310	0.0195
$\epsilon \backslash s$	9	10	11	12	13	14	15		
0.25	0.0679	0.0538	0.0427	0.0338	0.0268	0.0213	0.0169		
	0.0709	0.0550	0.0449	0.0353	0.0284	0.0225	0.0188		

Table 3.9: Non-coverage Probability for $\epsilon_0 = 0.3$

$\epsilon\backslash s$	0	1	2	3	4	5	6	7	8
0.3	0.5418	0.4116	0.3126	0.2375	0.1804	0.1370	0.1041	0.0791	0.0601
	0.5511	0.4144	0.3123	0.2386	0.1831	0.1337	0.1036	0.0804	0.0607
0.35	0.4840	0.3405	0.2395	0.1685	0.1185	0.0834	0.0588	0.0413	0.0290
	0.4926	0.3489	0.2469	0.1734	0.1220	0.0876	0.0610	0.0445	0.0319
0.4	0.4408	0.2878	0.1879	0.1227	0.0801	0.0523	0.0342	0.0223	0.0146
	0.4418	0.2848	0.1898	0.1231	0.0822	0.0539	0.0335	0.0231	0.0160
$\epsilon\backslash s$	9	10	11	12	13	14	15		
0.3	0.0456	0.0347	0.0263	0.0200	0.0152	0.0115	0.0088		
	0.0459	0.0341	0.0260	0.0198	0.0158	0.0117	0.0092		

Table 3.10: $E^{\nu}[N - L_s | N > \nu]$ for $\epsilon_0 = 0.2$

ϵ	0.2	0.2	0.2	0.25	0.25	0.25	0.3	0.3	0.3
s	0	10	13	0	6	9	0	4	6
	34.21	72.38	82.63	31.50	53.23	63.34	28.26	42.43	49.17
	34.46	70.24	79.12	30.89	50.82	60.26	27.62	40.78	47.04
	(0.31)	(0.35)	(0.36)	(0.29)	(0.32)	(0.33)	(0.28)	(0.30)	(0.31)

Table 3.11: $E^{\nu}[N - L_s|N > \nu]$ for $\epsilon_0 = 0.25$

ϵ	0.25	0.25	0.25	0.3	0.3	0.3	0.35	0.35	0.35
s	0	8	11	0	5	8	0	4	6
	25.28	50.97	59.64	22.95	38.41	47.02	20.62	32.69	38.40
	25.61	49.86	58.12	23.05	38.14	46.48	20.28	32.48	37.95
	(0.24)	(0.31)	(0.33)	(0.24)	(0.26)	(0.28)	(0.21)	(0.25)	(0.26)

Table 3.12: $E^{\nu}[N - L_s|N > \nu]$ for $\epsilon_0 = 0.3$

ϵ	0.3	0.3	0.3	0.35	0.35	0.35	0.4	0.4	0.4
s	0	6	9	0	5	7	0	4	6
	19.55	36.54	44.22	17.66	31.22	36.27	15.92	26.57	29.09
	19.97	36.69	44.10	18.10	31.42	36.42	16.53	26.94	31.90
	(0.20)	(0.24)	(0.26)	(0.19)	(0.23)	(0.24)	(0.18)	(0.21)	(0.23)

REFERENCES

Lorden, G. (1971). Procedures for reacting to a change in distribution. *Ann. Statist.*, **6**, 1897-1908.

Moustakides, G. V. (1986). Optimal stopping times for detecting changes in distribution. *Ann. Statist.*, **14**, 1379-1387.

Page, E. S. (1954). Continuous inspection schemes.*Biometrika*, **41**, 100-114.

Pollak, M. and Siegmund, D. (1985). A diffusion process and its applications to detecting a change in the drift of Brownian motion. *Biometrika*, **72**, 267-280.

Pollak, M. and Siegmund, D. (1986). Convergence of quasi-stationary to stationary distribution for stochastically monotone Markov processes. *J. Appl. Probab.*, **23**, 215-220.

Roberts, S. M. (1966). A comparison of some control chart procedures. *Technometrics*, **8**, 411-430.

Siegmund, D. (1976. The equivalence of absorbing and reflecting barrier problems for stochastically monotone Markov processes. *Ann. Prob.*, **4**, 914-924.

Siegmund, D. (1979). Corrected diffusion approximations in certain random walk problems. *Adv. Appl. Prob.*, **11**, 701-719.

Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals.* Springer, New York.

Srivastava, M. S. and Wu, Y. (1993). Comparison of EWMA, CUSUM and Shiryayev-Roberts procedures for detecting a shift in the mean. *Ann. Statist.*, **21**, 645-670.

Wu, Y. (1994). Design of control charts for detecting the change point. *Change Point Problems*, IMS Lecture Notes– Monograph Series, **23**, 330-345.

Chapter 4

Biases of Change Point and Change Magnitude Estimations

This chapter discusses the biases of estimates of change point and change magnitude after CUSUM test. It is well known that estimation after sequential test usually gives certain bias (Cox (1952), Siegmund (1978), Whitehead (1986), Woodroofe (1992)). There is no exception here, and we find that the biases are quite substantial for both estimates.

Section 4.1 gives the general description of the problem and defines the estimation of change point and change magnitude. Section 4.2 discusses the bias of estimate of change point. While the bias of estimate of change magnitude is studied in Section 4.3. In Section 4.4, theoretical results are applied to the case of monitoring the changes in the mean or variance of a normal process, and simulation studies are also conducted there to check the accuracy of theoretical results.

4.1 Introduction

As in Chapter 3, we assume that $\{X_k\}$ are independent random variables with distribution function F_{θ_0} for $k \leq \nu$ and F_{θ_1} for $k > \nu$, where $\theta_0 < 0 < \theta_1$. F_{θ_0} and F_{θ_1} belong to a standard one parameter exponential family as defined in section 1.3, and ν is the change point. It is also assumed that the baseline distribution function, F_0, is strongly non-lattice, i.e. $\lim_{|\lambda|\to\infty} \sup |E_0(\exp(i\lambda X_1))| < 1$.

For sequential detection of the change point, Page (1954) proposed the CUSUM procedure where an alarm is made at

$$N = \inf\{n > 0 : \ T_n > d\},$$

where T_n is the CUSUM process

$$T_n = \max(0, T_{n-1} + X_n), \quad with \quad T_0 = 0,$$

and d is a prescribed constant corresponding to a given ARL_0. Details about how to change the classical CUSUM procedure to the current version of CUSUM procedure will be given in the section of application.

The optimality of the CUSUM procedure has been studied by Lorden (1971) and

Moustakides (1986). Comparison with other procedures can be seen in Pollak and Siegmund (1985), Roberts (1966), Srivastava and Wu (1993).

However, the estimation of change point and change magnitude of the parameter after detection received less attention in the literature. From practical point of view, this is a rather critical issue. For example, in quality control, after change has been detected, we want to stop the process for inspection and repair. From process adjustment and control point of view, we would like to estimate the change magnitude as an initial estimation for the next stage.

It is well known that when $\psi(\theta_0) = \psi(\theta_1)$ the maximum likelihood estimator for ν is

$$\hat{\nu} = \max\{n < N : \ T_n = 0\}, \tag{4.1}$$

i.e. the first zero point of T_n counting backward from the detection time N. However. as θ_1 is usually unknown. we would be interested in estimating θ_1 and ν simultaneously. A natural way is still to use $\hat{\nu}$ as the estimator of ν, and estimate θ_1 by first estimating $\mu_1 = \psi'(\theta_1)$ as

$$\hat{\mu}_1 = T_N/(N - \hat{\nu}), \tag{4.2}$$

and then solve the equation $\hat{\mu}_1 = \psi'(\hat{\theta}_1)$ for $\hat{\theta}_1$ to obtain an estimate of change magnitude θ_1. In order to investigate the properties of $\hat{\theta}_1$, we have to study the properties of $\hat{\nu}$ and $\hat{\mu}_1$ first.

In this chapter, conditioning on $N > \nu$, the biases of $\hat{\nu}$ and $\hat{\mu}_1$ are studied by analytic approximations. The main results, i.e the second order approximations for the biases are developed under the following assumption:

(B): *Both* θ_0 *and* θ_1 *approach zero at the same order, and* ν *and* d *tend to infinity such that for some* $0 < \gamma < 1$, $d|\theta_0|^{1+\gamma} \to \infty$, *and* $d|\theta_0|^2 \to 0$.

The following notations are used in our discussion. Denote

$$S_n = S_0 + \sum_{i=1}^{n} X_i,$$

and

$$M = \sup_{0 \le k < \infty} S_n, \qquad with \qquad S_0 = 0.$$

Let

$$\tau_x = \inf\{n : \ S_n > x | S_0 = 0\}, \quad for \ \ x > 0; \qquad and$$

$$= \inf\{n : S_n \le x | S_0 = 0\} \quad for \ \ x < 0.$$

be the boundary crossing time and

$$R_x = S_{\tau_x} - x$$

be the overshoot at the boundary x, and

$$\tau_+ = \inf\{n : \ S_n > 0 | S_0 = 0\}, \quad and \quad \tau_- = \inf\{n : \ n > 0, S_n \le 0 | S_0 = 0\}$$

be the ladder epochs. Also, denote

$$N_x = \inf\{n : \ n > 0, S_n \le 0 \ \ or \ \ > d | S_0 = x\}, \qquad for \qquad x \ge 0$$

as the two-sided boundary crossing time.

For notational convenience, let $P^{\nu}(.) = P^{\nu}_{\theta_0,\theta_1}(.)$ denote the probability when the change point is ν, $P_{\theta_1}(.) = P^0(.)$ when the change occurs at zero, and $P_{\theta_0}(.) = P^{\infty}(.)$ when there is no change occurring. $E^{\nu}(.)$, $E_{\theta_1}(.)$ and $E_{\theta_0}(.)$ denote the corresponding expectations.

4.2 Bias of Estimate of Change Point

First we write

$$E[\hat{\nu} - \nu | N > \nu] = E[\hat{\nu} - \nu; \hat{\nu} > \nu | N > \nu] - E[\nu - \hat{\nu}; \hat{\nu} < \nu | N > \nu]. \quad (4.3)$$

From the renewal property of T_n at its zero points,

$$E[\hat{\nu} - \nu; \hat{\nu} > \nu | N > \nu]$$

$$= E[\hat{\nu} - \nu; S_{N_{T_\nu}} \leq 0]$$

$$= E_{\theta_1}[N_{T_\nu}; S_{N_{T_\nu}} \leq 0] + P_{\theta_1}(S_{N_{T_\nu}} \leq 0) E_{\theta_1}(\nu_2), \quad (4.4)$$

where ν_2 is the number of samples from the last zero point of S_n to the point where S_n goes back to zero after ν.

From (3.6), (3.7), Lemma 3.1 and Lemma 3.2, we obtain that

$$P_{\theta_1}(S_{N_{T_\nu}} \leq 0) = P_{\theta_1}(\tau_{-M} < \infty) + O(e^{\Delta_0 d}) = \tilde{p}_0 + O(\Delta_1^3) + O(e^{\Delta_0 d}). \quad (4.5)$$

where $\tilde{p}_0$ is the value defined in (2.13).

At the same time, from (3.6), (3.7) and Lemma 2.4, the approximation for the first term of the right hand side of (4.4) is obtained in the next lemma.

Lemma 4.1: *Under Condition (B),*

$$E_{\theta_1}[N_{T_\nu}; S_{N_{T_\nu}} \le 0] = E_{\theta_1}[\tau_{-M}; \tau_{-M} < \infty] + o(1)$$

$$= \frac{\Delta_0}{\tilde{\mu}_1(\Delta_1 - \Delta_0)^2} e^{\Delta_0(\rho_+ + \rho_-)} + O(1). \tag{4.6}$$

From (3.8) and Lemma 3.2, the following result can be derived.

Lemma 4.2: *Under Condition (B),*

$$E_{\theta_1}(\nu_2) = \frac{E_{\theta_1}[N_0, S_{N_0} \le 0]}{P_{\theta_1}(S_{N_0} > d)};$$

$$= \frac{E_{\theta_1}[\tau_-; \tau_- < \infty]}{P_{\theta_1}(\tau_- = \infty)} + o(1);$$

$$= \frac{1}{\tilde{\mu}_1 \Delta_1} + O(1).$$

Summarizing Lemmas 4.1-4.2 and (4.5), the next result follows.

Lemma 4.3: *Under Condition (B),*

$$E[\hat{\nu} - \nu; \hat{\nu} > \nu | N > \nu] = \frac{\Delta_0}{\tilde{\mu}_1(\Delta_1 - \Delta_0)^2} e^{\Delta_0(\rho_+ + \rho_-)} + \tilde{p}_0 \frac{1}{\tilde{\mu}_1 \Delta_1} + O(1).$$

To evaluate the second term of (4.3), under Condition (B), using Lemma 3.1, we have

$$P_{\theta_0}(T_\nu < x | N > \nu) = P_{\theta_0}(M < x) + O(e^{\Delta_0 d}),$$

and

$$P_{\theta_1}(S_{N_M} > d) = P_{\theta_1}(\tau_{-M} = \infty) + O(\Delta_0^3) + O(e^{\Delta_0 d}).$$

On the other hand, by looking at $\{S_k\}$ backward with starting point ν, we see that $\hat{\nu} - \nu$ is actually the maximum point of $S'_n = S_\nu - S_{\nu-n}$ for $0 \le n \le \nu$ with drift μ_0. In fact, from the definition of $\{T_n\}$, we know that $S_{\hat{\nu}} = \min_{0 \le k \le \nu} S_k$. Suppose $T_\nu = S_\nu - \min_{0 \le k \le \nu} S_k = x$. Then,

$$\max_{0 \le n \le \nu} S'_n = \max_{0 \le k \le \nu} (S_\nu - S_k) = x = S_\nu - S_{\hat{\nu}}.$$

If denote σ_x as the maximum point of S'_n with maximum value x, then under Condition (B), the second term of (4.3) is equal to

$$E_{\theta_0}[\sigma_M P_\theta(\tau_{-M} = \infty)] + O(1).$$

Using results given in Lemma 3.4, we obtain the following approximation for $E[\nu - \hat{\nu}; \hat{\nu} \le \nu | N > \nu]$.

Lemma 4.4: *Under Condition (B),*

$$E[\nu - \hat{\nu}; \hat{\nu} \le \nu | N > \nu] = -\frac{1}{\Delta_0 \tilde{\mu}_0} + \frac{\Delta_0}{\tilde{\mu}_0(\Delta_1 - \Delta_0)^2} e^{\Delta_1(\rho_+ + \rho_-)} + O(1).$$

Finally, the bias for the estimation of change point is obtained by summarizing the results of Lemma 4.3 and Lemma 4.4.

Theorem 4.1: *Under Condition (B),*

$$E[\hat{\nu} - \nu | N > \nu] = \frac{\Delta_0}{\tilde{\mu}_1(\Delta_1 - \Delta_0)^2} e^{\Delta_0(\rho_+ + \rho_-)} - \tilde{p}_0 \frac{1}{\tilde{\mu}_1 \Delta_1} + \frac{1}{\Delta_0 \tilde{\mu}_0}$$
$$- \frac{\Delta_0}{\tilde{\mu}_0(\Delta_1 - \Delta_0)^2} e^{\Delta_1(\rho_+ + \rho_-)} + O(1).$$

REMARK: From Theorem 4.1, we see that, even in the case $\psi(\theta_1) = \psi(\theta_0)$, the bias of $\hat{\nu}$ is not negligible and is positive (negative) for small θ_1 and θ_0 according to $\rho_+ + \rho_- > 0 (< 0)$. This gives us another typical example to show the effect of the sequential sampling rule. However, this bias becomes less significant for larger θ_0's. A detailed simulation study is reported in Section 4.4.

4.3 Bias of Estimate of Change Magnitude

It is well known that estimation after sequential testing usually gives certain bias, see Cox (1952) and Siegmund (1978) for initial studies. The estimation studied here provides another example. We find that the bias of $\hat{\mu}$ is quite significant and bias correction is definitely necessary.

First we write

$$E[\hat{\mu}_1 | N > \nu] = E[\frac{T_N}{N - \hat{\nu}}; \nu > \hat{\nu} | N > \nu] + E[\frac{T_N}{N - \hat{\nu}}; \nu \leq \hat{\nu} | N > \nu]. \tag{4.7}$$

Note, conditioning on $\{\nu \le \hat{\nu}\}$, $\{T_n, \hat{\nu} \le n \le N | N > \nu\}$ behaves stochastically equivalent to a random walk $\{S_n : 0 \le n \le N_0 | S_{N_0} > d\}$. Thus, from Lemma 3.1, we have

$$E[\frac{T_N}{N-\hat{\nu}}; \nu \le \hat{\nu} | N > \nu]$$

$$= E_{\theta_1}[\frac{S_{N_0}}{N_0} | S_{N_0} > d] P_{\theta_1}(S_{N_{T_\nu}} \le 0)$$

$$= (P_{\theta_1}(\tau_{-M} < \infty) + O(e^{\Delta_0 d}) + O(\Delta_1^3)) E_{\theta_1}[\frac{S_{N_0}}{N_0} | S_{N_0} > d].$$

By a similar analysis which leads to Lemma 4.4, we obtain that

$$E[\frac{T_N}{N-\hat{\nu}}; \nu > \hat{\nu} | N > \nu]$$

$$= E[\frac{T_N}{N-\nu+\nu-\hat{\nu}}; \nu > \hat{\nu}, S_{N_{T_\nu}} > d]$$

$$= E[\frac{S_{N_M}}{N_M+\sigma_M} | S_{N_M} > d](P_{\theta_1}(\tau_{-M} = \infty) + O(\Delta_0^3) + O(e^{\Delta_0 d})).$$

where σ_M is given in Lemma 3.4.

Now, the question becomes to approximate $E_{\theta_1}[\frac{S_{N_0}}{N_0} | S_{N_0} > d]$ and $E[\frac{S_{N_M}}{N_M+\sigma_M} | S_{N_M} > d]$.

To find the approximations, we shall use the following Taylor series expansion for $f(x, y) = \frac{x}{y}$:

$$f(x,y) = \frac{x_0}{y_0} + \frac{1}{y_0}(x - x_0) - \frac{x_0}{y_0^2}(y - y_0) + \frac{x_0}{y_0^3}(y - y_0)^2 - \frac{1}{y_0^2}(x - x_0)(y - y_0)$$
$$-\frac{x^*}{y^{*4}}(y - y_0)^3 + \frac{1}{y^{*3}}(y - y_0)^2(x - x_0), \qquad (4.8)$$

where $|x^* - x_0| \le |x - x_0|$ and $|y^* - y_0| \le |y - y_0|$.

Substitute $x = \frac{(S_{N_0}|S_{N_0}>d)}{E[N_0|S_{N_0}>d]}$, $y = \frac{(N_0|S_{N_0}>d)}{E[N_0|S_{N_0}>d]}$, $x_0 = E[x]$ and $y_0 = E[y]$ into (4.8), we get

$$E_\theta[\frac{S_{N_0}}{N_0}|S_{N_0} > d] \approx \frac{E[S_{N_0}|S_{N_0} > d]}{E[N_0|S_{N_0} > d]} + \frac{E[S_{N_0}|S_{N_0} > d]}{(E[N_0|S_{N_0} > d])^3} Var(N_0|S_{N_0} > d)$$

$$-\frac{1}{(E[N_0|S_{N_0} > d])^2} Cov(N_0, S_{N_0}|S_{N_0} > d). \tag{4.9}$$

It will be shown in Section 4.5 that the error of this approximation is at the order of $o(\frac{1}{d+\rho})$.

Approximations for each term on the right hand side of (4.9) will be given in the following several lemmas.

Lemma 4.5: *Under Condition (B),*

$$E[S_{N_0}|S_{N_0} > d] = d + \rho_+ + o(1).$$

This result is obvious (Siegmund (1985), (10.22)), and is listed here for easy reference.

Lemma 4.6: *Under Condition (B),*

$$E_{\theta_1}[N_0|S_{N_0} > d] = \frac{d+\rho_+}{\mu_1} - \frac{1}{\mu_1\Delta_1} + \frac{1}{\tilde{\mu}_1\Delta_1} + O(\frac{1}{\Delta_1}).$$

Proof: From (3.9),

$$E_{\theta_1}[N_0, S_{N_0} > d] = \frac{1}{\mu_1}((d+\rho_+)P_{\theta_1}(S_{N_0} > d) + E_{\theta_1}[S_{\tau_-}; \tau_- < \infty])$$

$$-E_{\theta_1}[\tau_-; \tau_- < \infty] + o(1).$$

Combining it with (3.15), Lemma 2.1, and omitting the lower order terms,

$$E_{\theta_1}[N_0, S_{N_0} > d] = \frac{1}{\mu_1}((d+\rho_+)P_{\theta_1}(S_{N_0} > d) + \frac{1}{\mu_1}E_0 S_{\tau_-} - \frac{1}{\tilde{\mu}_1}E_0 S_{\tau_-} + O(1).$$

Note that under Condition (B),

$$P_{\theta_1}(S_{N_0} > d) = P_{\theta_1}(\tau_- = \infty) + O(e^{-\Delta_1 d})$$

$$= -\Delta_1 E_0 S_{\tau_-} e^{\frac{\Delta_1}{2}\rho_-} + O(\Delta_1^3),$$

by Lemma 2.1. The proof of Lemma 4.6 is completed by some simplifications.

Lemma 4.7: *Under Condition (B),*

$$Var_{\theta_1}(N_0 | S_{N_0} > d) = \frac{d+\rho_+}{\mu_1^3}\psi''(\theta_1)$$

$$-\frac{1}{\mu_1^3\Delta_1}\psi''(\theta_1) + \frac{1}{\tilde{\mu}_1^3\Delta_1}\psi''(\tilde{\theta}_1) - (\frac{1}{\mu_1\Delta_1} - \frac{1}{\tilde{\mu}_1\Delta_1})^2 + O(\frac{1}{\Delta_1^3}).$$

Proof: First we write

$$Var_{\theta_1}(N_0 | S_{N_0} > d) = E_{\theta_1}[N_0^2 | S_{N_0} > d] - (E_{\theta_1}[N_0 | S_{N_0} > d])^2.$$

Note that

$$E_{\theta_1}[N_0^2; S_{N_0} > d] = E_{\theta_1}[N_0^2] - E_{\theta_1}[\tau_-^2; \tau_- < \infty] + O(1). \qquad (4.10)$$

By Lemma 2.5,

$$E_{\theta_1}[N_0^2] = \frac{1}{\mu_1^2}[\psi''(\theta_1)E_{\theta_1}N_0 + 2\mu_1 E_{\theta_1}(N_0 S_{N_0}) - E_{\theta_1}S_{N_0}^2]$$

$$= \frac{1}{\mu_1^2}[\psi''(\theta_1)E_{\theta_1}[N_0 | S_{N_0} > d]P_{\theta_1}(S_{N_0} > d) + \psi''(\theta_1)E_{\theta_1}(\tau_-, \tau_- < \infty) +$$

$$2\mu_1(d+\rho_+)E_{\theta_1}[N_0|S_{N_0}>d]P_{\theta_1}(S_{N_0}>d)+2\mu_1 E_{\theta_1}(S_{\tau_-}\tau_-,\tau_-<\infty)-$$

$$(d+\rho_+)^2P_{\theta_1}(S_{N_0}>d)-E_{\theta_1}(S^2_{\tau_-},\tau_-<\infty)]+O(d+\rho_+)$$

$$=[\frac{(d+\rho_+)^2}{\mu_1^2}+\frac{d+\rho_+}{\mu_1^3}(-\frac{2\mu_1}{\Delta_1}+\frac{2\mu_1^2}{\tilde{\mu}_1\Delta_1}+\psi''(\theta_1))-\frac{1}{\mu_1^3\Delta_1}\psi''(\theta_1)]P_{\theta_1}(S_{N_0}>d)+O(\frac{1}{\Delta_1^2}). \tag{4.11}$$

On the other hand, by Lemma 2.7,

$$E_{\theta_1}[\tau_-^2;\tau_-<\infty]=\frac{\psi''(\tilde{\theta}_1)}{\tilde{\mu}_1^3}E_0S_{\tau_-}+O(\frac{1}{\Delta_1^2}). \tag{4.12}$$

Combining (4.11) and (4.12), we have

$$E_{\theta_1}[N_0^2|S_{N_0}>d]=\frac{(d+\rho_+)^2}{\mu_1^2}+\frac{d+\rho_+}{\mu_1^3}(\psi''(\theta_1)-\frac{2\mu_1}{\Delta_1}+\frac{2\mu_1^2}{\tilde{\mu}_1\Delta_1})$$
$$-\frac{1}{\mu_1^3\Delta_1}\psi''(\theta_1)+\frac{1}{\tilde{\mu}_1^3\Delta_1}\psi''(\tilde{\theta}_1)+O(\frac{1}{\Delta_1^3}). \tag{4.13}$$

Using results obtained in Lemma 4.6, Lemma 4.7 is proved.

We now show that

$$\frac{1}{(E[N_0|S_{N_0}>d])^2}Cov(N_0,S_{N_0}|S_{N_0}>d)=o(\frac{1}{d+\rho_+}).$$

By Cauchy-Schwartz inequality, we have

$$|Cov(N_0,S_{N_0}|S_{N_0}>d)|\le(Var(N_0|S_{N_0}>d))^{1/2}(Var(S_{N_0}|S_{N_0}>d))^{1/2}.$$

Since $Var(S_{N_0}|S_{N_0}>d)$ approaches a constant, then

$$\frac{1}{(E[N_0|S_{N_0}>d])^2}Cov(N_0,S_{N_0}|S_{N_0}>d)=O((\frac{\mu_1}{d+\rho_+})^2(\frac{d+\rho_+}{\mu_1^3})^{1/2})=o(\frac{\mu_1}{d+\rho_+}).$$

Combining results of Lemmas 4.5-4.7, the following result is obtained after some simplifications.

Lemma 4.8: *Under Condition (B),*

$$E[\frac{S_{N_0}}{N_0}|S_{N_0} > d] = \mu_1 + \frac{1 + \psi''(\theta_1)}{d} + o(\frac{1}{d + \rho_+}).$$

By similar arguments leading to (4.9), we obtain that

$$E[\frac{S_{N_M}}{N_M + \sigma_M}|S_{N_M} > d] = \frac{E[S_{N_M}|S_{N_M} > d]}{E[N_M + \sigma_M|S_{N_M} > d]}$$

$$+\frac{E[S_{N_M}|S_{N_M} > d]}{(E[N_M + \sigma_M|S_{N_M} > d])^3} Var(N_M + \sigma_M|S_{N_M} > d) + o(\frac{1}{d + \rho_+}). \qquad (4.14)$$

Again, we shall show that the approximation is really at the order of $o(1/(d + \rho_+))$ in Section 4.5. The following four lemmas give the corresponding results given in Lemmas 4.5-4.8.

Lemma 4.9: *Under Condition (B),*

$$E[S_{N_M}|S_{N_M} > d] = d + \rho_+ + o(1).$$

Lemma 4.10: *Under Condition (B),*

$$E[N_M|S_{N_M} > d] = \frac{d + \rho_+}{\mu_1} + \frac{1}{\mu_1 \Delta_0} - \frac{1}{\mu_1 \Delta_1} - \frac{\Delta_0}{\tilde{\mu}_1 \Delta_1 (\Delta_1 - \Delta_0)} + O(\frac{1}{\Delta_1}).$$

Proof: By Lemma 4.2, we have

$$E[N_M; S_{N_M} > d] = E[N_M] - E[\tau_{-M}; \tau_{-M} < \infty] + o(1)$$

$$= \frac{1}{\mu_1}[ES_{N_M} - E(M)] - E[\tau_{-M}; \tau_{-M} < \infty] + o(1)$$

$$= \frac{1}{\mu_1}[(d + \rho_+)P_{\theta_1}(S_{N_M} > d) + \frac{1}{\Delta_0}] - \frac{\Delta_0}{\tilde{\mu}_1(\Delta_1 - \Delta_0)^2} + O(\frac{1}{\Delta_1})$$

Combining (3.7) with Lemma 2.4, we obtain that

$$E[N_M | S_{N_M} > d] = \frac{E[N_M; S_{N_M} > d]}{P(S_{N_M} > d)}$$

$$= \frac{d + \rho_+}{\mu_1} + \frac{\Delta_1 - \Delta_0}{\mu_1 \Delta_1 \Delta_0} - \frac{\Delta_0}{\tilde{\mu}_1 \Delta_1 (\Delta_1 - \Delta_0)} + O(\frac{1}{\Delta_1}),$$

which is the desired result.

From Lemma 3.4, the following result follows.

Lemma 4.11: *Under Condition (B).*

$$E[\sigma_M | S_{N_M} > d] = -\frac{1}{\tilde{\mu}_0 \Delta_0} + \frac{1}{\tilde{\mu}_0 \Delta_1} + \frac{\Delta_0}{\tilde{\mu}_0 \Delta_1 (\Delta_1 - \Delta_0)} + O(\frac{1}{\Delta_1}).$$

Lemma 4.12: *Under Condition (B).*

$$Var(N_M + \sigma_M | S_{N_M} > d) = \frac{d + \rho_+}{\mu_1^3} \psi''(\theta_1) + O(\frac{1}{\Delta_1^4}).$$

Proof: The idea in the proof is similar to that used in proving Lemma 4.8, and thus some details are omitted.

First note that

$$Var(N_M + \sigma_M | S_{N_M} > d) = E[(N_M + \sigma_M)^2 | S_{N_M} > d] - (E[N_M + \sigma_M | S_{N_M} > d])^2. \tag{4.15}$$

From Lemma 4.10 and 4.11, we only need to calculate $E[(N_M+\sigma_M)^2|S_{N_M}>d]$. We write

$$E[(N_M+\sigma_M)^2|S_{N_M}>d] = E[N_M^2|S_{N_M}>d]+2E[N_M\sigma_M|S_{N_M}>d]+E[\sigma_M^2|S_{N_M}>d]. \tag{4.16}$$

We will find approximation for each term on the right side of (4.16).

First, it is easy to verify that

$$\begin{aligned} E[\sigma_M^2] &= \sum_{k=2}^{\infty} E_{\theta_0}(\sum_{i=1}^{k-1}\tau_i^{(+)})^2(1-p)^{(k-1)}p \\ &= \frac{E_{\theta_0}(\tau_+^2,\tau_+<\infty)}{p}+\frac{2(E_{\theta_0}(\tau_+,\tau_+<\infty))^2}{p^2} \\ &= -\frac{\psi''(\tilde{\theta}_0)}{\tilde{\mu}_0^3\Delta_0}+\frac{2}{\tilde{\mu}_0^2\Delta_0^2}+O(\frac{1}{\Delta_0^3}), \end{aligned} \tag{4.17}$$

where $p = P_{\theta_0}(\tau_+=\infty)$.

Similarly,

$$\begin{aligned} E[\sigma_M^2, S_{N_M}\le 0] &= E[\sigma_M^2 e^{-\Delta_1 M}]+O(\frac{1}{\Delta_0^3}) \\ = p\frac{E_{\theta_0}(\tau_+^2e^{-\Delta_1 S_{\tau_+}},\tau_+<\infty)}{(1-E_{\theta_0}(e^{-\Delta_1 S_{\tau_+}},\tau_+<\infty))^2} &+ p\frac{2(E_{\theta_0}(\tau_+e^{-\Delta_1 S_{\tau_+}},\tau_+<\infty))^2}{(1-E_{\theta_0}(e^{-\Delta_1 S_{\tau_+}},\tau_+<\infty))^3}+O(\frac{1}{\Delta_0^3}) \\ &= -\frac{\Delta_0\psi''(\tilde{\theta}_0)}{\tilde{\mu}_0^3(\Delta_1-\Delta_0)^2}-\frac{2\Delta_0}{\tilde{\mu}_0^2(\Delta_1-\Delta_0)^3}+O(\frac{1}{\Delta_0^3}) \end{aligned} \tag{4.18}$$

Combining (4.17) with (4.18), we have

$$\begin{aligned} E[\sigma_M^2|S_{N_M}>d] &= \frac{2}{\tilde{\mu}_0^2\Delta_0^2}-\frac{2}{\tilde{\mu}_0^2\Delta_0\Delta_1}+\frac{2\Delta_0}{\tilde{\mu}_0^2\Delta_1(\Delta_1-\Delta_0)^2} \\ &-\frac{\psi''(\tilde{\theta}_0)}{\tilde{\mu}_0^3\Delta_0}+\frac{\psi''(\tilde{\theta}_0)}{\tilde{\mu}_0^3\Delta_1}+\frac{\Delta_0\psi''(\tilde{\theta}_0)}{\tilde{\mu}_0^3\Delta_1(\Delta_1-\Delta_0)}+O(\frac{1}{\Delta_1^3}). \end{aligned} \tag{4.19}$$

Second,

$$E[N_M\sigma_M] = E_{\theta_0}[\sigma_M E_{\theta_1}(N_M|M)]$$

$$= \frac{1}{\mu_1}[ES_{N_M}\sigma_M - E(\sigma_M M)]$$

$$= \frac{d+\rho_+}{\mu_1}E[\sigma_M, S_{N_M} > d] - \frac{1}{\mu_1}E(\sigma_M M) + O(\frac{1}{\Delta_0^3})$$

$$= \frac{d+\rho_+}{\mu_1}E[\sigma_M|S_{N_M} > d]P_{\theta_1}(S_{N_M} > d) - \frac{1}{\mu_1}E(\sigma_M M) + O(\frac{1}{\Delta_0^3}). \tag{4.20}$$

and

$$E(\sigma_M M) = \sum_{k=2}^{\infty} E_{\theta_0}(\sum_{i=1}^{k-1}\tau_i^{(+)}\sum_{j=1}^{(k-1)} S_{\tau_j}^{(+)})(1-p)^{(k-1)}p$$

$$= \frac{E_{\theta_0}(\tau_+ S_{\tau_+}, \tau_+ < \infty)}{p} + \frac{2E_{\theta_0}(\tau_+, \tau_+ < \infty)E_{\theta_0}(S_{\tau_+}, \tau_+ < \infty)}{p^2}$$

$$= \frac{2}{\tilde{\mu}_0\Delta_0^2} + O(\frac{1}{\Delta_0^2}). \tag{4.21}$$

Since

$$E[N_M\sigma_M, S_{N_M} \leq 0] = E[\sigma_M\tau_{-M}, \tau_{-M} < \infty] + O(\frac{1}{\Delta_0^2})$$

$$= -\frac{1}{\tilde{\mu}_1}E[\sigma_M M e^{-\Delta_1 M}] + O(\frac{1}{\Delta_0^3})$$

$$= -\frac{1}{\tilde{\mu}_1}\sum_{k=2}^{\infty} E_{\theta_0}(\sum_{i=1}^{k-1}\tau_i^{(+)}\sum_{j=1}^{(k-1)} S_{\tau_j}^{(+)}e^{-\Delta_1\sum_{j=1}^{(k-1)}S_{\tau_j}^{(+)}})(1-p)^{(k-1)}p + O(\frac{1}{\Delta_0^3})$$

$$= -\frac{p}{\tilde{\mu}_1}[\frac{2E_{\theta_0}(\tau_+ e^{-\Delta_1 S_{\tau_+}}, \tau_+ < \infty)E_{\theta_0}(S_{\tau_+}e^{-\Delta_1 S_{\tau_+}}, \tau_+ < \infty)}{(1 - E_{\theta_0}(e^{-\Delta_1 S_{\tau_+}}, \tau_+ < \infty)^3}]$$

$$\frac{E_{\theta_0}(\tau_+ S_{\tau_+}e^{-\Delta_1 S_{\tau_+}}, \tau_+ < \infty)}{(1 - E_{\theta_0}(e^{-\Delta_1 S_{\tau_+}}, \tau_+ < \infty)^2} + O(\frac{1}{\Delta_0^3})$$

$$= \frac{2\Delta_0}{\tilde{\mu}_0\tilde{\mu}_1(\Delta_1 - \Delta_0)^3} + O(\frac{1}{\Delta_0^3}). \tag{4.22}$$

Then, from (4.20) and (4.22), we have

$$E[N_M\sigma_M|S_{N_M} > d] = \frac{d+\rho_+}{\mu_1}\left(-\frac{1}{\tilde{\mu}_0\Delta_0} + \frac{1}{\tilde{\mu}_0\Delta_1} + \frac{\Delta_0}{\tilde{\mu}_0\Delta_1(\Delta_1-\Delta_0)}\right)$$

$$-\frac{2}{\tilde{\mu}_0\mu_1\Delta_0^2} + \frac{2}{\tilde{\mu}_0\mu_1\Delta_0\Delta_1} - \frac{2\Delta_0}{\tilde{\mu}_0\tilde{\mu}_1\Delta_1(\Delta_1-\Delta_0)^2} + O\left(\frac{1}{\Delta_0^3}\right). \tag{4.23}$$

Finally, we settle down to evaluate $E[N_M^2|S_{N_M} > d]$. Since

$$E[N_M^2; S_{N_M} > d] = E[N_M^2] - E[N_M^2; S_{N_M} \leq 0]$$

$$= E[N_M^2] - E[\tau_{-M}^2; \tau_{-M} < \infty] + O\left(\frac{1}{\Delta_0^2}\right).$$

From Lemma 2.9, we only need to find $E[N_M^2]$.

Given $M = x$, *for* $0 < x < d$, $\{(S_{N_x} - x - \mu_1 N_x)^2 - N_x\psi''(\theta_1), \Im_{N_x}\}$ is a martingale. Similar to Lemma 2.5, we have

$$E[N_x^2] = \frac{1}{\mu_1^2}(\psi''(\theta_1)E[N_x] + 2\mu_1 E[N_x(S_{N_x} - x)] - E(S_{N_x} - x)^2).$$

Thus

$$E[N_M^2] = \frac{1}{\mu_1^2}(\psi''(\theta_1)E[N_M] + 2\mu_1 E[N_M(S_{N_M} - M)] - E(S_{N_M} - M)^2)$$

$$= \frac{1}{\mu_1^2}(\psi''(\theta_1)E[N_M] + 2\mu_1 E[N_M S_{N_M}] - 2\mu_1 E[N_M M] - E[S_{N_M}^2] + 2E[S_{N_M} M] - E[M^2]).$$

Since

$$E[N_M M] = E[ME(N_M|M)] = \frac{1}{\mu_1}E[M(S_{N_M} - M)]$$

$$= \frac{1}{\mu_1}E[MS_{N_M}] - \frac{1}{\mu_1}EM^2,$$

then

$$E[N_M^2] = \frac{1}{\mu_1^2}(\psi''(\theta_1)E[N_M] + 2\mu_1 E[N_M S_{N_M}] - E[S_{N_M}^2] + E[M^2]). \tag{4.24}$$

The remain question is now to find approximations for terms on right hand side of (4.24). For this, first note

$$E[N_M S_{N_M}] = (d+\rho_+)E[N_M|S_{N_M} > d]P_{\theta_1}(S_{N_M} > d) + E[N_M S_{N_M}; S_{N_M} \leq 0]$$

$$= (d+\rho_+)E[N_M|S_{N_M} > d]P_{\theta_1}(S_{N_M} > d) + O(\frac{1}{\Delta_1^2}).$$

At the same time, by similar argument leading to (2.32), we have

$$E_{\theta_0} M^2 = \sum_{k=2}^{\infty} E_{\theta_0}(\sum_{i=1}^{(k-1)} S_{\tau_i^+})^2 (1-p)^{(k-1)} p$$

$$= \sum_{k=2}^{\infty}(k-1)E_{\theta_0}(S_{\tau_+}^2, \tau_+ < \infty)(1-p)^{(k-2)}p$$

$$+ \sum_{k=2}^{\infty}(k-1)(k-2)(E_{\theta_0}(S_{\tau_+}, \tau_+ < \infty))^2(1-p)^{(k-3)}p$$

$$= \frac{2(E_{\theta_0}(S_{\tau_+}, \tau_+ < \infty))^2}{p^2} + \frac{E_{\theta_0}(S_{\tau_+}^2, \tau_+ < \infty)}{p}$$

$$= \frac{2}{\Delta_0^2} + O(\frac{1}{\Delta_0}). \tag{4.25}$$

Finally,

$$E[S_{N_M}^2] = E[(d+R_{d-M})^2]P_{\theta_1}(S_{N_M} > d) + O(1) = (d+\rho_+)^2 P_{\theta_1}(S_{N_M} > d) + O(de^{\Delta_0 d}). \tag{4.26}$$

Combining (4.24), (4.25), (4.26), Lemma 2.9 and Lemma 4.10, we have

$$E[N_M^2|S_{N_M} > d] = \frac{(d+\rho_+)^2}{\mu_1^2} + \frac{d+\rho_+}{\mu_1^3}(\psi''(\theta_1) + \frac{2\mu_1}{\Delta_0} - \frac{2\mu_1}{\Delta_1} - \frac{2\mu_1^2\Delta_0}{\tilde{\mu}_1\Delta_1(\Delta_1 - \Delta_0)})$$

$$+\frac{2}{\mu_1^2\Delta_0^2}-\frac{2}{\mu_1^2\Delta_0\Delta_1}+\frac{\psi''(\theta_1)}{\mu_1^3\Delta_0}-\frac{\psi''(\theta_1)}{\mu_1^3\Delta_1}-\frac{\Delta_0\psi''(\tilde{\theta}_1)}{\tilde{\mu}_1^3\Delta_1(\Delta_1-\Delta_0)}+\frac{2\Delta_0}{\tilde{\mu}_1^2\Delta_1(\Delta_1-\Delta_0)^2}+O(\frac{1}{\Delta_0^3}). \tag{4.27}$$

From (4.19). (4.23), (4.27), Lemma 4.10 and 4.11, we obtain that

$$Var(N_M+\sigma_M|S_{N_M}>d)=\frac{d+\rho_+}{\mu_1^3}\psi''(\theta_1)+O(\frac{1}{\Delta_1^4}).$$

This completes the proof of Lemma 4.12.

Summarizing the results of Lemmas 4.9-4.12, the following result is obtained after some simplifications.

Lemma 4.13: *Under Condition (B),*

$$E[\frac{S_{N_M}}{N_M+\sigma_M}|S_{N_M}>d]=\mu_1+\frac{1}{d}(1+\psi''(\theta_1)-\frac{\Delta_1^2}{2\Delta_0^2})+o(\frac{1}{d+\rho_+}).$$

Summing up the results from Lemma 4.8 and Lemma 4.13, finally we obtain another main result of this chapter.

Theorem 4.2: *Under Condition (B),*

$$E[\hat{\mu}_1-\mu_1|N>\nu]=\frac{1}{d}(1+\psi''(\theta_1)-\frac{\Delta_1^3}{2\Delta_0^2(\Delta_1-\Delta_0)})+o(\frac{1}{d+\rho_+}). \tag{4.28}$$

In particular, when $\psi(\theta_1)=\psi(\theta_0)$,

$$E[\hat{\mu}_1-\mu_1|N>\nu]=\frac{1}{d}(\frac{3}{4}+\psi''(\theta_1))+o(\frac{1}{d+\rho_+}). \tag{4.29}$$

4.4 Application

In this section, we demonstrate how to use our theory in the practical situation. and conduct simulation studies to check the accuracy of the approximations given in Section 4.2 and 4.3.

In the field of quality control, we might be monitoring some characteristic of a manufacturing process, and in many practical situations. the quality characteristic process X_t is assumed to be normally distributed with mean μ_0 and standard deviation σ_0, where μ_0 is the target value and σ_0 reflects the variation of quality. Any shift from the target value μ_0 or increase in the process variation results in poor quality. and we want to detect the change as soon as possible. To monitor the quality characteristic process, groups of random samples of size. say m, are usually taken at some regular time intervals. Meanwhile CUSUM charts based on either sample mean $\bar{X}_t$ or sample variance s_t^2 are plotted with appropriate control limit(s). We would like to stop the process for inspection and repair when some point of the CUSUM chart falls out-side of control limit. Since false alarms are often costly, we therefore assume the in control Average Run Length(ARL) of samples to be very large.

In the following, we will apply our results to detect changes in mean or standard deviation respectively. For simplicity, we only consider detecting the increase in mean or variance.

4.4.1 Detect Increase in Mean

Without lose of generality, we assume that the observed process $\bar{X}_t$ comes from a normal population with mean 0 and standard deviation 1 when the process is in control, and with mean μ and standard deviation 1 when the process is out of control. Then the CUSUM procedure is defined as making alarm at

$$N = \inf\{n > 0 : \; T_n > d\},$$

where T_n is the CUSUM process

$$T_n = \max(0, T_{n-1} + Y_n), \quad with \quad T_0 = 0,$$

where $Y_n = \bar{X}_n - \frac{\delta}{2}$, and d is the control limit with reference value δ. Usually, the reference value δ is the change magnitude which we are interested in detecting quickly. δ also serves as a preliminary estimate of μ, the true change magnitude.

In this case,

$$\psi(\theta) = \frac{1}{2}\theta^2, \quad \theta_0 = -\frac{\delta}{2}, \quad \theta_1 = \mu - \frac{\delta}{2},$$

$$\mu_0 = -\tilde{\mu}_0 = -\frac{\delta}{2}, \quad \mu_1 = -\tilde{\mu}_1 = \mu - \frac{\delta}{2},$$

$$E_0 S_{\tau_+} = -E_0 S_{\tau_-} = \frac{1}{\sqrt{2}}, \quad \rho_+ = -\rho_- \approx 0.583,$$

$$r_0 = r_1 = \frac{1}{4} \quad and \quad C_0 = \frac{1}{2}(\frac{1}{\sqrt{2}} - \rho_+)^2.$$

Insert the related quantities into Theorem 4.1 and 4.2, we have

Corollary 4.1: Under condition (B),

$$E[\hat{\nu} - \nu] | N > \nu] = \frac{1}{2(\mu - \delta/2)^2} - \frac{2}{\delta^2} + O(1).$$

when $\mu = \delta$,

$$E[\hat{\nu} - \nu] | N > \nu] = O(1).$$

Corollary 4.2: Under condition (B),

$$E[\hat{\mu}_1 - \mu_1 | N > \nu] = \frac{1}{d}(2 - \frac{2(\mu - \delta/2)^3}{\mu\delta^2}) + o(\frac{1}{d + \rho_+}).$$

when $\mu = \delta$,

$$E[\hat{\mu}_1 - \mu_1 | N > \nu] = \frac{7}{4d} + o(\frac{1}{d + \rho_+}).$$

In the following, simulation results are presented to check the accuracy of our approximations. Let $\theta_0 = -\delta/2$ and $\theta_1 = \mu - \delta/2$. For average in-control run length $ARL_0 = E_{-\delta/2}N = 1000$, we choose $\theta_0 = -0.2, -0.25$, and -0.3, the corresponding control limit d is determined by using Siegmund's approximation (Siegmund (1985), (2.57)) which gives $d = 9.96, 8.82$ and 7.72 respectively. All simulations are replicated for 10000 times. The change point is taken at $\nu = 100$ which guarantees that it has little effect on the theoretical approximation. In Table 4.1, we give the comparisons of simulated and approximated biases for $\hat{\nu}$ and $\hat{\mu}_1$. In each cell, the top number is the approximation value while the bottom number is the simulated value, and its standard error is given in the parentheses.

From this comparison, we can draw the following several conclusions. First, the bias of $\hat{\nu}$ is not negligible for small θ_0's and becomes less significant for larger θ_0's particularly at $\theta_0 = -\theta_1$. Second, the variance of $\hat{\nu}$ is very large and thus makes the estimation less reliable. Third, the bias of $\hat{\mu}_1$ is definitely not negligible and a correction is thus necessary. Also the bias at $\theta_0 = -\theta_1$ is relatively stable which confirms our theoretical approximation which is free of θ_0. When $\theta_1 > -\theta_0$, the bias of $\hat{\mu}_1$ becomes smaller, but reduces less dramatically. Fourth, the variance of $\hat{\mu}_1$ is also relatively large. Fifth, the comparison shows that the our approximations are generally quite accurate and can be used in practice. Thus the following bias correction method is proposed.

First substitute the initial estimation of $\hat{\theta}_1$ for θ_1 into the approximations of the biases of $\hat{\nu}$ and $\hat{\theta}_1$ given in Corollary 4.1 and 4.2 respectively, and obtain the estimated biases. Then let

$$\tilde{\nu} = \hat{\nu} - (est.bias(\hat{\nu}));$$

and

$$\tilde{\theta}_1 = \hat{\theta}_1/(1 + (est.bias(\hat{\theta}_1))/\hat{\theta}).$$

Although this method slightly over-corrects the bias of $\hat{\nu}$ to the left and under-corrects the bias of $\hat{\theta}$, the formula is very simple and easy to be implemented for practical purposes.

Finally, let us say more about the assumption **(B)**. Under this assumption, the effect of ν is basically negligible in both biases. Also, the effect of d is negligible for the approximation of $E[\hat{\nu}-\nu|N > \nu]$. However, in the approximation of $E[\hat{\mu}_1-\mu_1|N > \nu]$. the expansion is very weak as we only assume that $d\theta^{1+\gamma} \to \infty$ for some $\gamma > 0$ and $d\theta^2 \to 0$. This shows that the approximation works well for θ neither too large nor too small. The simulation results confirm this point.

Table 4.1: Comparison of Approximated and Simulated Bias (Normal).

$\theta_0(d)$	θ_1	Bias($\hat{\nu}$)	Bias($\hat{\theta}_1$)
-0.2	0.2	0	0.1757
(9.96)		-0.816(0.28)	0.1755(0.0022)
	0.25	-4.5	0.1628
		-3.99(0.28)	0.1533(0.0023)
	0.30	-6.94	0.1431
		-6.05 (0.24)	0.1252(0.0024)
-0.25	0.25	0	0.1975
(8.86)		-0.131(0.21)	0.1938(0.0026)
	0.30	-2.44	0.1865
		-2.47(0.18)	0.1685(0.0026)
	0.40	-4.875	0.1498
		-3.77(0.16)	0.1463(0.0027)
-0.3	0.3	0	0.2267
(7.72)		-0.28(0.14)	0.2096(0.0028)
	0.35	-1.47	0.2166
		-1.41(0.12)	0.1876(0.0029)
	0.4	-2.43	0.2025
		-2.44(0.11)	0.1616(0.0029)

4.4.2 Detect Increase in Variance

Without lose of generality, we assume that the observed process s_t^2 comes from a population of $\chi^2(p)$ when the process is in control, and from a population of $(1+\epsilon)^2\chi^2(p)(\epsilon > 0)$ when the process is out of control. Then the CUSUM procedure is defined as making alarm at

$$N = \inf\{n > 0 : \ T_n > d\},$$

where T_n is the CUSUM process

$$T_n = \max(0, T_{n-1} + Y_n), \quad with \quad T_0 = 0,$$

$Y_n = \frac{s_n^2[(1+\epsilon_0)^2-1]}{2\sqrt{2p}(1+\epsilon_0)^2 \ln(1+\epsilon_0)} - \sqrt{\frac{p}{2}}$, and d is the control limit with reference value ϵ_0, the relative increase change magnitude we are interested in detecting quickly. ϵ_0 also serves as a preliminary estimate of ϵ, the true relative increase change magnitude.

In this case,

$$\psi(\theta) = -\sqrt{\frac{p}{2}}\theta - \frac{p}{2}\ln(1 - \sqrt{\frac{2}{p}}\theta),$$

$$\theta_0 = \sqrt{\frac{p}{2}}(1 - \frac{2(1+\epsilon_0)^2\ln(1+\epsilon_0)}{(1+\epsilon_0)^2 - 1}), \quad \theta_1 = \sqrt{\frac{p}{2}}(1 - \frac{2(1+\epsilon_0)^2\ln(1+\epsilon_0)}{(1+\epsilon)^2[(1+\epsilon_0)^2-1]}),$$

$$\mu_0 = \sqrt{\frac{p}{2}}(\frac{(1+\epsilon_0)^2-1}{2(1+\epsilon_0)^2\ln(1+\epsilon_0)} - 1), \quad \mu_1 = \sqrt{\frac{p}{2}}(\frac{(1+\epsilon)^2[(1+\epsilon_0)^2-1]}{2(1+\epsilon_0)^2\ln(1+\epsilon_0)} - 1).$$

For simplicity, we only give results of a special case, i.e. $p = 2$. For $\epsilon_0 = 0.2$, 0.25, and 0.3, we choose the average in-control run length $ARL_0 = E_{\theta_0}N = 10000$, and

$\nu = 100$. Simulation results are based on 10000 replications. The control limit d is given by using the approximation (10.17) of Siegmund (1985):

$$ARL_0 \approx \frac{1}{\Delta_0 \mu_0}(e^{-\Delta_0(d+\rho_+-\rho_-)} - 1 + \Delta_0(d + \rho_+ - \rho_-))$$

From this formula, we obtain that $d = 10.083$, 8.794, and 7.823 corresponding to $\epsilon_0 = 0.2$, 0.25, and 0.3.

Substituting

$$\rho_+ = 1, \quad r_1 = 1, \quad \rho_- = -\frac{1}{3}, \quad r_0 = \frac{1}{18} \quad and \quad C_0 = 0$$

into Theorem 4.1 and 4.2, following results are obtained.

Corollary 4.3: *Under condition (B), for $p = 2$, we have*

$$E[\hat{\nu} - \nu | N > \nu] = \frac{\Delta_0}{\tilde{\mu}_1(\Delta_1 - \Delta_0)^2} e^{\frac{2}{3}\Delta_0} - \tilde{p}_0 \frac{1}{\tilde{\mu}_1 \Delta_1} + \frac{1}{\Delta_0 \tilde{\mu}_0} - \frac{\Delta_0}{\tilde{\mu}_0(\Delta_1 - \Delta_0)^2} e^{\frac{2}{3}\Delta_1} + O(1). \tag{4.30}$$

where

$$\tilde{p}_0 = C_1 e^{\frac{25}{36 C_1}\Delta_0 \Delta_1},$$

with $C_1 = -\frac{\Delta_0}{\Delta_1 - \Delta_0} e^{\frac{2}{3}\Delta_1}$.

Corollary 4.4: *Under condition (B), for $p = 2$, we have*

$$E[\hat{\mu}_1 - \mu_1 | N > \nu] = \frac{1}{d + \rho_+}(1 + \frac{(1+\epsilon)^4[(1+\epsilon_0)^2 - 1]^2}{(1+\epsilon_0)^4[\ln(1+\epsilon_0)^2]^2} - \frac{\Delta_1^3}{2\Delta_0^2(\Delta_1 - \Delta_0)}) + o(\frac{1}{d + \rho_+}). \tag{4.31}$$

In particular, when $\psi(\theta_1) = \psi(\theta_0)$*, i.e.* $\epsilon_0 = \epsilon$*,*

$$E[\hat{\mu}_1 - \mu_1 | N > \nu] = \frac{1}{d+\rho_+}\left(\frac{3}{4} + \frac{[(1+\epsilon_0)^2 - 1]^2}{[\ln(1+\epsilon_0)^2]^2}\right) + o\left(\frac{1}{d+\rho_+}\right).$$

Table 4.2 gives the comparison of the simulated values with the approximated values for the biases of $\hat{\nu}$ and $\hat{\mu}_1$. The order of the numbers are arranged in the same way as in Table 4.1.

From Table 4.2, we can draw the following conclusions. First, the approximations of the bias of $\hat{\nu}$ is generally very precise. Second, the bias of $\hat{\mu}_1$ is very large, and therefore, bias correction is definitely required. This is the most important conclusion of this chapter. Third, approximation for the bias of $\hat{\mu}_1$ is systematically less than the simulated bias. This may be due to the following reasons. (1): As mentioned in Section 4.4.1, the expansion of $E[\hat{\mu}_1 - \mu_1 | N > \nu]$ is very weak. (2): The approximation only uses information contained in the first two moments of the population distribution, and reflects nothing about the skewness and kurtosis of the population distribution. Therefore, when the population distribution is skewed, the approximation can not be very precise.

Table 4.2: Comparison of Approximated and Simulated Bias (Exponential).

$\epsilon_0(d)$	ϵ_1	Bias($\hat{\nu}$)	Bias($\hat{\mu}_1$)
0.2 (10.083)	0.2	3.61 3.79(0.304)	0.2188 0.3372(0.0055)
	0.25	-4.27 -3.08(0.24)	0.2108 0.3019(0.0063)
	0.3	-7.33 -6.08(0.21)	0.1935 0.2791(0.0070)
0.25 (8.794)	0.25	2.80 2.34(0.23)	0.2659 0.4042(0.0068)
	0.3	-1.67 -1.17(0.18)	0.2683 0.3834(0.0074)
	0.35	-3.71 -3.59(0.17)	0.2660 0.3662(0.0080)
0.3 (7.823)	0.3	2.26 2.05(0.17)	0.3169 0.4866(0.0087)
	0.35	-0.51 -0.42(0.15)	0.3281 0.4522(0.0087)
	0.4	-1.93 -2.07(0.13)	0.3375 0.4557(0.0098)

4.5 Error Checking for (4.9) and (4.14)

To show that the remain terms of (4.9) are at the order of $o(\frac{1}{d+\rho})$, we first note that, conditioning on the event $\{S_{N_0} > d\}$,

$$\frac{S_{N_0}}{E_{\theta_1}[N_0|S_{N_0} > d]} - \frac{E_{\theta_1}[S_{N_0}|S_{N_0} > d]}{E_{\theta_1}[N_0|S_{N_0} > d]} \to 0 \ \textit{in probability as} \ d \to \infty.$$

and

$$\frac{N_0}{E_{\theta_1}[N_0|S_{N_0} > d]} \to 1 \ \textit{in probability as} \ d \to \infty.$$

Thus, for the third order terms in Taylor expansion, the coefficient of $E_{\theta_1}[(\frac{S_{N_0}}{E[N_0|S_{N_0}>d]} - \frac{E_{\theta_1}[S_{N_0}|S_{N_0}>d]}{E_{\theta_1}[N_0|S_{N_0}>d]})(\frac{N_0}{E_{\theta_1}[N_0|S_{N_0}>d]} - 1)^2|S_{N_0} > d]$ converges to 1 in probability, and the coefficient of $E_{\theta_1}[(\frac{N_0}{E_{\theta_1}[N_0|S_{N_0}>d]} - 1)^3|S_{N_0} > d]$ is at the order of $\frac{E_{\theta_1}[S_{N_0}|S_{N_0}>d]}{E_{\theta_1}[N_0|S_{N_0}>d]} = O(\mu_1)$ in probability. Thus, we only need to show that

$$E_{\theta_1}[(\frac{S_{N_0}}{E[N_0|S_{N_0} > d]} - \frac{E_{\theta_1}[S_{N_0}|S_{N_0} > d]}{E_{\theta_1}[N_0|S_{N_0} > d]})(\frac{N_0}{E_{\theta_1}[N_0|S_{N_0} > d]} - 1)^2|S_{N_0} > d] = o(\frac{1}{d+\rho}), \tag{4.32}$$

and

$$E_{\theta_1}[(\frac{N_0}{E_{\theta_1}[N_0|S_{N_0} > d]} - 1)^3|S_{N_0} > d] = o(\frac{1}{\mu_1(d+\rho)}). \tag{4.33}$$

We now begin to prove (4.32) and (4.33).

Lemma 4.14: *Under Condition (B), we have*

$$E[(N_0 - E[N_0|S_{N_0} > d])^3|S_{N_0} > d] = 3\psi''(\theta_1)\frac{d+\rho_+}{\mu_1^5} + O(\frac{(d+\rho_+)^2}{\mu_1^3}) + O(\frac{1}{\mu_1^6}), \tag{4.34}$$

and thus

$$E_{\theta_1}[(\frac{N_0}{E_{\theta_1}[N_0|S_{N_0} > d]} - 1)^3|S_{N_0} > d] = o(\frac{1}{\mu_1(d+\rho)}),$$

and

$$E_{\theta_1}[(\frac{S_{N_0}}{E[N_0|S_{N_0} > d]} - \frac{E_{\theta_1}[S_{N_0}|S_{N_0} > d]}{E_{\theta_1}[N_0|S_{N_0} > d]})(\frac{N_0}{E_{\theta_1}[N_0|S_{N_0} > d]} - 1)^2|S_{N_0} > d] = o(\frac{1}{d+\rho}).$$

Proof: First we write

$$E[(N_0 - E[N_0|S_{N_0} > d])^3|S_{N_0} > d] = E[N_0^3|S_{N_0} > d]$$

$$-3E[N_0|S_{N_0} > d]E[N_0^2|S_{N_0} > d] + 2(E[N_0|S_{N_0} > d])^3,$$

and note that

$$E[N_0^3; S_{N_0} > d] = E[N_0^3] - E[N_0^3; S_{N_0} \leq 0]$$

$$= EN_0^3 - E[\tau_-^3; \tau_- < \infty] + O(1).$$

To evaluate the $E[N_0^3]$, note that

$$\int e^{\theta S_n - n\psi(\theta)} dF_0(x_1)...dF_0(x_n) = 1. \tag{4.35}$$

Differentiate both sides of equation (4.35) three times with respect to (w.r.t) θ, we know that

$$\{(S_n - n\mu_1)^3 - 3n\psi''(\theta_1)(S_n - n\mu_1) - n\psi^{(3)}(\theta_1), \Im_n\} \tag{4.36}$$

is a martingale under P_{θ_1}, where $\psi^{(3)}(\theta_1)$ is the third differential of $\psi(\theta)$ at point θ_1 .

Thus, we have

$$E_{\theta_1}[N_0^3] = \frac{1}{\mu_1^3}[ES_{N_0}^3 - 3\mu_1 E(N_0 S_{N_0}^2) + 3\mu_1^2\psi''(\theta_1)E(N_0^2 S_{N_0})$$

$$-3\psi''(\theta_1)E(N_0S_{N_0})+3\mu_1\psi''(\theta_1)EN_0^2-\psi^{(3)}(\theta_1)EN_0]$$

$$=\frac{1}{\mu_1^3}[(d+\rho_+)^3P(S_{N_0}>d)-3\mu_1(d+\rho_+)^2E(N_0|S_{N_0}>d)P(S_{N_0}>d)+3\mu_1\psi''(\theta_1)EN_0^2]$$
$$+O(\frac{(d+\rho_+)^2}{\mu_1^3})+O(\frac{1}{\mu_1^6}).$$

Combining with the results given in the proof of Lemma 4.7 and Lemma 4.8, the first result of Lemma 4.14 is obtained.

The second result of Lemma 4.14 follows by the following fact.

$$3(\frac{d+\rho_+}{\mu_1})^{-3}\psi''(\theta_1)\frac{d+\rho_+}{\mu_1^5}=\frac{1}{(d+\rho_+)^2\mu_1^2}\psi''(\theta_1)=o(\frac{1}{\mu_1(d+\rho)}).$$

To show (4.32), by Hölder inequality, we have

$$|E[(\frac{S_{N_0}}{E[N_0|S_{N_0}>d]}-\frac{E[S_{N_0}|S_{N_0}>d]}{E[N_0|S_{N_0}>d]})(\frac{N_0}{E[N_0|S_{N_0}>d]}-1)^2|S_{N_0}>d]|$$
$$\leq(E[|\frac{S_{N_0}}{E[N_0|S_{N_0}>d]}-\frac{E[S_{N_0}|S_{N_0}>d]}{E[N_0|S_{N_0}>d]}|^3|S_{N_0}>d])^{1/3}$$
$$(E[|\frac{N_0}{E[N_0|S_{N_0}>d]}-1|^3|S_{N_0}>d])^{2/3}$$
$$=\frac{1}{E[N_0|S_{N_0}>d]}(E|R_d-ER_d|^3)^{1/3}(E[|\frac{N_0}{E[N_0|S_{N_0}>d]}-1|^3|S_{N_0}>d])^{2/3}.$$

By the previous argument, we know that

$$E[|\frac{N_0}{E[N_0|S_{N_0}>d]}-1|^3|S_{N_0}>d]=O(1).$$

Thus, the left hand of (4.32) is at the order of $O(\frac{\theta}{d+\rho})$, which completes proof of (4.32).

The proof for (4.14) is technically similar to the proof of (4.9), but much more complicated in calculation. By the similar argument in proving (4.9), we know that, to show (4.14) is equivalent to show the following two equations.

$$E[(\frac{S_{N_M}}{E[N_M+\sigma_M|S_{N_M}>d]}-\frac{E[S_{N_M}|S_{N_M}>d]}{E[N_M+\sigma_M|S_{N_M}>d]})$$
$$(\frac{N_M+\sigma_M}{E[N_M+\sigma_M|S_{N_M}>d]}-1)^2|S_{N_M}>d]=o(\frac{1}{d+\rho}), \tag{4.37}$$

and

$$E[(\frac{N_M+\sigma_M}{E[N_M+\sigma_M|S_{N_M}>d]}-1)^3|S_{N_M}>d]=o(\frac{1}{\mu_1(d+\rho)}). \tag{4.38}$$

Similar to the results of Lemma 4.14, we have

Lemma 4.15: *Under condition (B),*

$$E[(N_M+\sigma_M-E[N_M+\sigma_M|S_{N_M}>d])^3|S_{N_M}>d]=\frac{3(d+\rho_+)}{\mu_1^5}\psi''(\theta_1)$$
$$+O(\frac{1}{\mu_1^6}). \tag{4.39}$$

Thus,

$$E[(\frac{N_M+\sigma_M}{E[N_M+\sigma_M|S_{N_M}>d]}-1)^3|S_{N_M}>d]=o(\frac{1}{\mu_1(d+\rho)}),$$
$$E[(\frac{S_{N_M}}{E[N_M+\sigma_M|S_{N_M}>d]}-\frac{E[S_{N_M}|S_{N_M}>d]}{E[N_M+\sigma_M|S_{N_M}>d]})$$
$$(\frac{N_M+\sigma_M}{E[N_M+\sigma_M|S_{N_M}>d]}-1)^2|S_{N_M}>d]=o(\frac{1}{d+\rho}).$$

Proof:

$$E[(N_M-E[N_M|S_{N_M}>d]+\sigma_M-E[\sigma_M|S_{N_M}>d])^3|S_{N_M}>d]$$

$$= E[(N_M - E[N_M|S_{N_M} > d])^3|S_{N_M} > d]$$

$$+3E[(N_M - E[N_M|S_{N_M} > d])^2(\sigma_M - E[\sigma_M|S_{N_M} > d])|S_{N_M} > d]$$

$$+3E[(N_M - E[N_M|S_{N_M} > d])(\sigma_M - E[\sigma_M|S_{N_M} > d])^2|S_{N_M} > d]$$

$$+E[(\sigma_M - E[\sigma_M|S_{N_M} > d])^3|S_{N_M} > d]. \tag{4.40}$$

Similar to the proof of Lemma 4.14, we can show that the first term on the right hand side of (4.40) is equal to

$$\frac{3(d+\rho)}{\mu_1^5} + O(\frac{1}{\Delta_0^6}).$$

For the second term on the right hand side of (4.40), we know that

$$E[(N_M - E[N_M|S_{N_M} > d])^2(\sigma_M - E[\sigma_M|S_{N_M} > d])|S_{N_M} > d]$$

$$= E[N_M^2\sigma_M|S_{N_M} > d] - E[N_M^2|S_{N_M} > d]E[\sigma_M|S_{N_M} > d]$$

$$-2E[N_M\sigma_M|S_{N_M} > d]E[N_M|S_{N_M} > d] + 2(E[N_M|S_{N_M} > d])^2E[\sigma_M|S_{N_M} > d] \tag{4.41}$$

In the following, we shall prove that (4.41) is at the order of $O(\frac{1}{\Delta_0^6})$.

To evaluate (4.41), using previous obtained results, we only need to approximate $E[N_M^2\sigma_M|S_{N_M} > d]$. In fact,

$$E[N_M^2\sigma_M, S_{N_M} > d] = E[N_M^2\sigma_M] - E[N_M^2\sigma_M, S_{N_M} \leq 0]$$

$$= E[N_M^2\sigma_M] + O(\frac{1}{\Delta_0^6}), \tag{4.42}$$

and

$$E[N_M^2\sigma_M] = E_{\theta_0}[\sigma_M E_{\theta_1}(N_M^2|M)]$$

$$= \frac{1}{\mu_1^2}[\psi''(\theta_1)E(N_M\sigma_M) + 2\mu_1 E[N_M\sigma_M(S_{N_M} - M)] - E[\sigma_M(S_{N_M} - M)^2]]. \quad (4.43)$$

From (4.20) and (4.22), we have

$$E(N_M\sigma_M) = \frac{d+\rho_+}{\mu_1}(-\frac{1}{\tilde{\mu}_0\Delta_0} + \frac{1}{\tilde{\mu}_0\Delta_1} + \frac{\Delta_0}{\tilde{\mu}_0\Delta_1(\Delta_1 - \Delta_0)})P_{\theta_1}(S_{N_M} > d) + O(\frac{1}{\Delta_0^4}).$$

Since

$$E[N_M\sigma_M(S_{N_M} - M)] = E[N_M\sigma_M S_{N_M}] - E[N_M\sigma_M M]$$

$$= (d+\rho_+)E[N_M\sigma_M, S_{N_M} > d] - \frac{1}{\mu_1}[E(\sigma_M M S_{N_M} - E(\sigma_M M^2)] + O(\frac{1}{\Delta_0^4})$$

$$= (d+\rho_+)E[N_M\sigma_M | S_{N_M} > d]P_{\theta_1}(S_{N_M} > d) - \frac{d+\rho_+}{\mu_1}E(\sigma_M M | S_{N_M} > d)$$

$$P_{\theta_1}(S_{N_M} > d) + O(\frac{1}{\Delta_0^4}),$$

and

$$E[\sigma_M(S_{N_M} - M)^2] = E[\sigma_M S_{N_M}^2] - 2E[\sigma_M S_{N_M} M] + E[\sigma_M M^2]$$

$$= (d+\rho_+)^2 E[\sigma_M | S_{N_M} > d)P_{\theta_1}(S_{N_M} > d) - 2(d+\rho_+)E[\sigma_M M | S_{N_M} > d)$$

$$P_{\theta_1}(S_{N_M} > d) + O(\frac{1}{\Delta_0^4}).$$

Combining above results, we have

$$E[N_M^2\sigma_m | S_{N_M} > d] - E[N_M^2 | S_{N_M} > d]E[\sigma_m | S_{N_M} > d] =$$

$$\frac{d+\rho_+}{\mu_1^2}(-\frac{4}{\tilde{\mu}_0\mu_1\Delta_0^2} + \frac{4}{\tilde{\mu}_0\mu_1\Delta_0\Delta_1} + \frac{4\Delta_0}{\tilde{\mu}_0\tilde{\mu}_1\Delta_1(\Delta_1 - \Delta_0)})$$

$$-\frac{d+\rho_+}{\mu_1^2}(\frac{2}{\Delta_0} - \frac{2}{\Delta_1} - \frac{2\Delta_0\mu_1}{\tilde{\mu}_1\Delta_1(\Delta_1 - \Delta_0)})(-\frac{1}{\tilde{\mu}_0\Delta_0} + \frac{1}{\tilde{\mu}_0\Delta_1} + \frac{\Delta_0}{\tilde{\mu}_0\Delta_1(\Delta_1 - \Delta_0)}) + O(\frac{1}{\Delta_0^4}). \quad (4.44)$$

On the other hand, by (4.23), Lemma 4.11 and 4.12, we have

$$E[N_M\sigma_M|S_{N_M} > d]E[N_M|S_{N_M} > d] - (E[N_M|S_{N_M} > d])^2E[\sigma_M|S_{N_M} > d]$$

$$= \frac{d+\rho_+}{\mu_1^2}\left(-\frac{2}{\tilde{\mu}_0\mu_1\Delta_0^2} + \frac{2}{\tilde{\mu}_0\mu_1\Delta_0\Delta_1} + \frac{2\Delta_0}{\tilde{\mu}_0\tilde{\mu}_1\Delta_1(\Delta_1 - \Delta_0)}\right)$$
$$-\frac{d+\rho_+}{\mu_1^2}\left(\frac{1}{\Delta_0} - \frac{1}{\Delta_1} - \frac{\Delta_0\mu_1}{\tilde{\mu}_1\Delta_1(\Delta_1-\Delta_0)}\right)\left(-\frac{1}{\tilde{\mu}_0\Delta_0} + \frac{1}{\tilde{\mu}_0\Delta_1} + \frac{\Delta_0}{\tilde{\mu}_0\Delta_1(\Delta_1-\Delta_0)}\right) + O\left(\frac{1}{\Delta_0^4}\right). \tag{4.45}$$

Insert (4.44) and (4.45) into (4.41), we have

$$E[(N_M - E[N_M|S_{N_M} > d])^2(\sigma_M - E[\sigma_M|S_{N_M} > d])|S_{N_M} > d] = O\left(\frac{1}{\Delta_0^6}\right).$$

Similarly, we can show that

$$E[(N_M - E[N_M|S_{N_M} > d])(\sigma_M - E[\sigma_M|S_{N_M} > d])^2|S_{N_M} > d] = O\left(\frac{1}{\Delta_0^6}\right).$$

Thus, (4.39) holds.

Similar to the proof of (4.32) and (4.33), we can prove (4.37) and (4.38).

This completes proof of Lemma 4.16.

REFERENCES

Cox, D. R. (1952). A note on the sequential estimation of mean. *Proc. Camb. Phil. Soc.*, **48**, 447-450.

Lorden, G. (1971). Procedures for reacting to a change in distribution. *Ann. Statist.*, **6**, 1897-1908.

Moustakides, G. V. (1986). Optimal stopping times for detecting changes in distribution. *Ann. Statist.* , **14**, 1379-1387.

Page, E. S. (1954). Continuous inspection schemes.*Biometrika*, **41**, 100-114.

Pollak, M. and Siegmund, D. (1985). A diffusion process and its applications to detecting a change in the drift of Brownian motion. *Biometrika*, **72**, 267-280.

Roberts, S. M. (1966). A comparison of some control chart procedures. *Technometrics*, **8**, 411-430.

Siegmund, D. (1978). Estimation following sequential test. *Biometrika*, **65**, 341-349.

Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals.* Springer. New York.

Srivastava, M. S. and Wu, Y. (1993). Comparison of EWMA, CUSUM and Shiryayev-Roberts procedures for detecting a shift in the mean. *Ann. Statist.*, **21**, 645-670.

Whitehead, J. (1986). On the bias of maximum likelihood estimation following a sequential test. *Biometrika*, **73**, 573-81.

Woodroofe, M. (1992). Estimation after sequential testing: A simple approach for a truncated sequential probability ratio test. *Biometrika*, **79**, 347-353.

Chapter 5

Confidence Intervals for Change Magnitude

This chapter addresses the question of constructing confidence intervals for the change magnitude after CUSUM test. To this end, we first obtain the asymptotic distribution function of change magnitude estimate. Then we derive confidence intervals based on this distribution function.

This chapter is organized as follows: Section 5.1 gives the general description of the problem and the method used to solve the problem. Section 5.2 presents the main results. Comparisons of the theoretical results with simulated results are given in Section 5.3. Finally, the rigorous proofs are presented in Section 5.4.

5.1 Introduction

In this chapter, attention is focused on monitoring the change in the mean of a normal process, a special case of the one parameter exponential family defined in Section 1.3. More specifically, it is assumed that $\{X_k\}$ follow $N(\theta_0, 1)$ for $k \le \nu$ and $N(\theta_1, 1)$ for $k > \nu$, where $\theta_0 < 0 < \theta_1$ and ν is the change point. The basic assumption is the Condition (B) given in Chapter 4.

For quick detection of the change point, the CUSUM procedure stops the process at the first time

$$N = \inf\{n > 0 : \ T_n > d\}, \tag{5.1}$$

where T_n is the CUSUM process

$$T_n = \max(0, T_{n-1} + X_n), \quad with \quad T_0 = 0, \tag{5.2}$$

and d is the control limit. Details of how to change the classical CUSUM procedure to the current version of CUSUM procedure have been given in Section 3.5. After the change has been detected, apparently, we want to estimate the change point and change magnitude based on the information provided by the control charts. As given in Chapter 4, we use

$$\hat{\nu} = \max\{n < N : \ T_n = 0\} \tag{5.3}$$

to estimate the change point, and use

$$\hat{\theta}_1 = T_N/(N - \hat{\nu}) \tag{5.4}$$

to estimate the change magnitude.

In practice, we might be more interested in constructing confidence intervals of change magnitude, so the distribution function of $\hat{\theta}_1$ is required. By a simple application of Anscombe's (1952) Theorem, as $d \to \infty$, $\sqrt{N - \hat{\nu}}(\hat{\theta}_1 - \theta_1)$ is asymptotically standard normal, which means the first order approximation to its distribution is not affected by its sequential nature. But, as we have already seen in Chapter 4, $\hat{\theta}_1$ is seriously biased in practical case (i.e. in the case that d is not extremely large), and thus simply using standard normal to approximate the distribution function of $\sqrt{N - \hat{\nu}}(\hat{\theta}_1 - \theta_1)$ yields very poor results; besides, from Figure 5.1, we can see that the distribution function of $\hat{\theta}_1$ is also seriously skewed and has a non-normal tendency. These are common for estimation after sequential test (Cox (1952), Woodroofe (1986, 1992), and Todd, Whitehead and Facey (1996)). To obtain more accurate approximations, Woodroofe (1986) introduced the so-called *very weak expansion* method and is found performing very well (Woodroofe (1992), and Todd, Whitehead and Tacey (1996)). By adopting this method to our case, $Z^* = \frac{\hat{\theta}_1 - E_{\theta_1}\hat{\theta}_1}{\sqrt{Var_{\theta_1}(\hat{\theta}_1)}}$ has an asymptotic distribution of standard normal under P_{θ_1} as $d \to \infty$.

The goal of this chapter is to obtain approximations for the distribution function of $\hat{\theta}_1$. From the results of Chapter 4, we only need to find the $Var_{\theta_1}(\hat{\theta}_1)$, which will be developed in the next section.

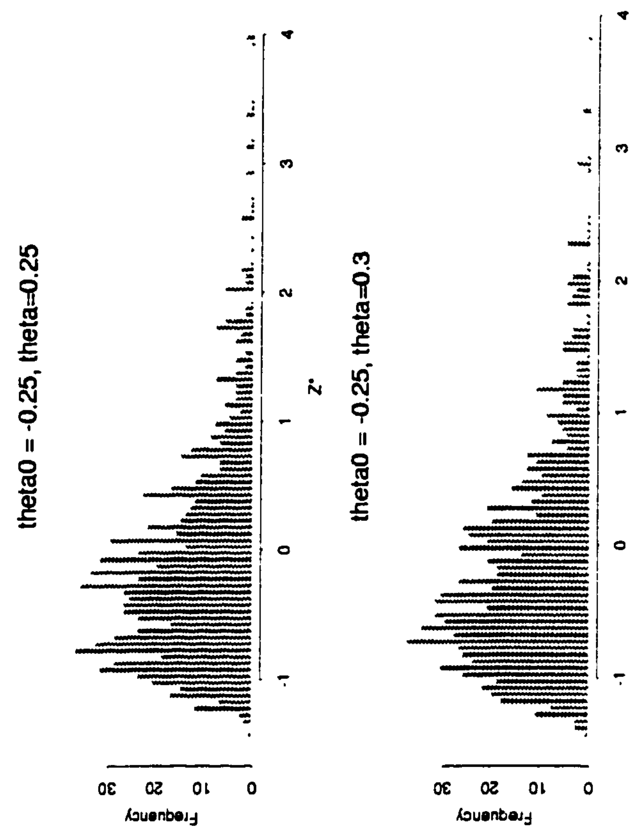

Figure 5.1: Histogram of Z^* with 1000 replications.

5.2 Asymptotic Distribution of $\hat{\theta}_1$

Before approximating the $Var_{\theta_1}(\hat{\theta}_1)$, a preliminary result is given for later use.

Lemma 5.1: *For a random variable X with $EX^2 < \infty$, and a given event A, let $m_A = E[X|A]$ and $m_{A^c} = E[X|A^c]$. Then*

$$Var(X) = Var(X|A)P(A) + Var(X|A^c)P(A^c) + (m_A - m_{A^c})^2 P(A)P(A^c).$$

Proof:

$$Var(X) = EX^2 - (EX)^2$$

$$= E(X^2|A)P(A) + E(X^2|A^c)P(A^c) - (m_A P(A) + m_{A^c} P(A^c))^2$$

$$= Var(X|A)P(A)+Var(A|A^c)P(A^c)+m_A^2 P(A)+m_{A^c}^2 P(A^c)-(m_A P(A)+m_{A^c} P(A^c))^2$$

$$= Var(X|A)P(A) + Var(X|A^c)P(A^c) + (m_A - m_{A^c})^2 P(A)P(A^c).$$

The proof is completed.

Next, note the fact that $[\frac{T_N}{N-\hat{\nu}}; \nu \leq \hat{\nu}|N > \nu]$ is stochastically equivalent to $[\frac{S_{N_0}}{N_0}, \nu \leq \hat{\nu}|S_{N_0} > d]$, and $[\frac{T_N}{N-\hat{\nu}}; \nu > \hat{\nu}|N > \nu]$ is asymptotically equivalent to $[\frac{S_{N_M}}{N_M+\sigma_M}, \nu > \hat{\nu}|S_{N_M} > d]$. Let $X = \frac{T_N}{N-\hat{\nu}}$ and $A = (\nu \leq \hat{\nu}|N > \nu)$, by Lemma 5.1, we have

$$Var(\hat{\theta}_1) = Var(\frac{S_{N_0}}{N_0}|S_{N_0} > d)P(\nu \leq \hat{\nu}|N > \nu)+$$

$$Var(\frac{S_{N_M}}{N_M + \sigma_M}|S_{N_M} > d)P(\nu > \hat{\nu}|N > \nu)+(m_1-m_2)^2 P(\nu \leq \hat{\nu}|N > \nu)P(\nu > \hat{\nu}|N > \nu).$$

where $m_1 = E_{\theta_1}[\frac{S_{N_0}}{N_0}|S_{N_0} > d]$ and $m_2 = E_{\theta_1}[\frac{S_{N_M}}{N_M+\sigma_M}|S_{N_M} > d]$.

By Lemma 3.1, it becomes

$$Var(\hat{\theta}_1) = Var(\frac{S_{N_0}}{N_0}|S_{N_0} > d)P_{\theta_1}(\tau_{-M} < \infty)+Var(\frac{S_{N_M}}{N_M + \sigma_M}|S_{N_M} > d)P_{\theta_1}(\tau_{-M} = \infty)$$

$$+(m_1 - m_2)^2 P_{\theta_1}(\tau_{-M} = \infty)P_{\theta_1}(\tau_{-M} < \infty) + O(\theta_0^4) + O(e^{2\theta_0 d}\theta_0^2). \quad (5.5)$$

So the remain problem is to find $Var(\frac{S_{N_0}}{N_0}|S_{N_0} > d)$ and $Var(\frac{S_{N_M}}{N_M+\sigma_M}|S_{N_M} > d)$. By borrowing some results from Chapter 4, we only need to approximate

$E_{\theta_1}[\frac{S^2_{N_M}}{(N_M+\sigma_M)^2}|S_{N_M} > d)$ and $E_{\theta_1}[\frac{S^2_{N_0}}{N_0^2}|S_{N_0} > d]$.

To evaluate these two quantities, we adopt the method used in Section 4.3. The following Taylor series expansion, similar to (4.8), will be used.

$$f(x,y) = \frac{x_0}{y_0} + \frac{1}{y_0}(x-x_0) - \frac{x_0}{y_0^2}(y-y_0) + \frac{x_0}{y_0^3}(y-y_0)^2 - \frac{1}{y_0^2}(x-x_0)(y-y_0)$$
$$-\frac{x_0}{{y_0}^4}(y-y_0)^3 + \frac{1}{{y_0}^3}(y-y_0)^2(x-x_0) + \frac{x_0}{{y_0}^5}(y-y_0)^3 - \frac{1}{{y_0}^4}(y-y_0)^3(x-x_0)$$
$$-\frac{x^*}{y^{*6}}(y-y_0)^5 + \frac{1}{y^{*5}}(y-y_0)^4(x-x_0), \tag{5.6}$$

where $|x^* - x_0| \le |x - x_0|$ and $|y^* - y_0| \le |y - y_0|$.

By letting $X = \frac{(S^2_{N_0}|S_{N_0}>d)}{E[N_0^2|S_{N_0}>d]}$, $Y = \frac{(N_0^2|S_{N_0}>d)}{E[N_0^2|S_{N_0}>d]}$, $x_0 = E[X]$ and $y_0 = E[Y]$, and denoting $A_1 = E[N_0^2|S_{N_0} > d]$, we get

$$E_\theta[\frac{S^2_{N_0}}{N_0^2}|S_{N_0} > d] = \frac{E[S^2_{N_0}|S_{N_0} > d]}{A_1} + \frac{E[S^2_{N_0}|S_{N_0} > d]}{A_1^3}Var(N_0^2|S_{N_0} > d)$$
$$-\frac{1}{A_1^2}Cov(N_0^2, S^2_{N_0}|S_{N_0} > d) - \frac{E[S^2_{N_0}|S_{N_0} > d]}{A_1^4}E[(N_0^2 - A_1])^3|S_{N_0} > d]$$
$$+\frac{1}{A_1^3}Cov((N_0^2 - A_1)^2, S^2_{N_0}|S_{N_0} > d) + \frac{E[S^2_{N_0}|S_{N_0} > d]}{A_1^5}E[(N_0^2 - A_1)^4|S_{N_0} > d]$$
$$-\frac{1}{A_1^4}Cov((N_0^2 - A_1)^3, S^2_{N_0}|S_{N_0} > d) + o(\frac{1}{(d+\rho_+)^2}), \tag{5.7}$$

where $\rho_+ \approx 0.583$. It will be shown at the end of Section 5.3 that the error of this approximation is at the order of $o(\frac{1}{(d+\rho_+)^2})$.

From Lemma 5.4-5.11 of Section 5.4, we get the following result.

Lemma 5.2: *Under condition (B), we have*

$$Var(\frac{S^2_{N_0}}{N_0^2}|S_{N_0} > d) = \frac{\theta_1}{d+\rho_+} + \frac{4}{(d+\rho_+)^2} + o(\frac{1}{(d+\rho_+)^2}).$$

On the other hand, by the Taylor series expansion (5.6), with $X = \frac{(S_{N_M}^2 | S_{N_M} > d)}{E[(N_M+\sigma_M)^2 | S_{N_M} > d]}$, $Y = \frac{((N_M+\sigma_M)^2 | S_{N_0} > d)}{E[(N_M+\sigma_M)^2 | S_{N_M} > d]}$, $x_0 = E[X]$ and $y_0 = E[Y]$, and denoting $A_2 = E[(N_M + \sigma_M)^2 | S_{N_M} > d]$, we get

$$E[\frac{S_{N_M}^2}{(N_M+\sigma_M)^2} | S_{N_M} > d] = \frac{E[S_{N_M}^2 | S_{N_M} > d]}{A_2}$$

$$+\frac{E[S_{N_M}^2 | S_{N_M} > d]}{A_2^3} Var((N_M+\sigma_M)^2 | S_{N_M} > d) - \frac{1}{A_2^2} Cov((N_M+\sigma_M)^2, S_{N_M}^2 | S_{N_M} > d)$$

$$-\frac{E[S_{N_M}^2 | S_{N_M} > d]}{A_2^4} E[((N_M+\sigma_M)^2 - A_2)^3 | S_{N_M} > d]$$

$$+\frac{1}{A_2^3} Cov(((N_M+\sigma_M)^2 - A_2)^2, S_{N_M}^2 | S_{N_M} > d)$$

$$+\frac{E[S_{N_M}^2 | S_{N_M} > d]}{A_2^5} E[((N_M+\sigma_M)^2 - A_2)^4 | S_{N_M} > d]$$

$$-\frac{1}{A_2^4} Cov(((N_M+\sigma_M)^2 - A_2)^3, S_{N_M}^2 | S_{N_M} > d) + o(\frac{1}{(d+\rho_+)^2}). \tag{5.8}$$

Again, using the related results in Lemma 5.4-5.11 of Section 5.4, we obtain

Lemma 5.3: *Under condition (B), we have*

$$Var(\frac{S_{N_M}^2}{(N_M+\sigma_M)^2} | S_{N_M} > d] = \frac{\theta_1}{d+\rho_+} + \frac{1}{(d+\rho_+)^2}(4 - \frac{2\theta_1^2}{\theta_0^2} + \frac{3\theta_1^4}{4\theta_0^4}) + o(\frac{1}{(d+\rho_+)^2}).$$

Combining Lemma 5.1-5.3, we obtain

Theorem 5.1: *Under Condition (B), we have*

$$Var(\hat{\theta}_1) = v + o(\frac{1}{d^2}).$$

where

$$v = \frac{\theta_1}{d} + \frac{1}{d^2}(4 - \frac{\theta_1^3}{\theta_0^2(\theta_1 - \theta_0)}(2 - \frac{\theta_1^2}{\theta_0^2} + \frac{\theta_1^3}{4\theta_0^2(\theta_1 - \theta_0)})). \tag{5.9}$$

Then, it follows that $\hat{\theta}_1$ *has an asymptotic normal distribution with mean* $\theta_1 + \frac{1}{d}(2 - \frac{\theta_1^3}{2\theta_0^2(\theta_1 - \theta_0)})$ *and variance* v.

5.3 Simulation Results

Simulation conducted here is to check the accuracy of coverage probabilities of confidence intervals derived from Theorem 5.1. In Table 5.1, 5.2 and 5.3, for a given confidence level, the actual coverage probabilities based on 10000 replications are presented for a fixed θ_0 and 3 different values of θ_1. From these tables, we can see that the actual coverage probabilities are slightly higher than the theoretical values for 90% and 95% confidence levels, while the actual coverage probabilities are very precise for 98% confidence level. In practice, the confidence levels are usually large (larger than 90%), our result gives a very satisfactory approximation.

The difference between the actual coverage probabilities and the theoretical values may be due to the following reasons.

First, under Condition (B), the expansions used in approximating mean and variance of $\hat{\theta}_1$ are very weak.

Second, and most importantly, the distribution of $\hat{\theta}_1$ is seriously skewed and has a non-normal tendency. This property has also been reported by Todd, Whitehead and Facey (1996) and Whitehead (1997) for estimation after other kind of sequential tests.

Third, from the proofs given in Section 5.2, we can see that $\hat{\theta}_1$ could be better approximated by a mixture of two independent normal distributions. In fact, in Section 5.2, we showed that, asymptotically, $\hat{\theta}_1$ comes from a normal population having mean $\theta_1 + \frac{2}{d}$ and variance $\frac{\theta_1}{d} + \frac{4}{d^2}$ with probability $\tilde{p}_0$, or comes from a normal population having mean $\theta_1 + \frac{2-\theta_1^2/(2\theta_0^2)}{d}$ and variance $\frac{\theta_1}{d} + \frac{4-2\theta_1^2/\theta_0^2+3\theta_1^4/(4\theta_0^4)}{d^2}$ with probability $1-\tilde{p}_0$, where $\tilde{p}_0 = -\frac{\theta_0}{\theta_1-\theta_0}e^{-\theta_1(\theta_1-\theta_0)}$. Under Condition (B), the term θ_1 in both means and the term $\frac{\theta_1}{d}$ in both variances are the dominant term, we therefore give the asymptotic distribution presented in Theorem 5.1 for the sake of easy application. But, in practice, the domination of these terms in either the means or the variances is very weak, this gives another source of inaccuracy.

Table 5.1: Coverage Probabilities of 90% Confidence Intervals

$\theta_1 \backslash \theta_0$	-0.2	-0.25	-0.3
0.2	94.46%		
0.25	94.21%	94.47%	
0.3	95.08%	95.21%	95.31%
0.35		95.48%	95.56%
0.4			95.98%

Table 5.2: Coverage Probabilities of 95% Confidence Intervals

$\theta_1 \backslash \theta_0$	-0.2	-0.25	-0.3
0.2	96.26%		
0.25	95.96%	96.38%	
0.3	96.68%	96.86%	96.78%
0.35		97.22%	97.51%
0.4			97.38%

Table 5.3: Coverage Probabilities of 98% Confidence Intervals

$\theta_1\backslash\theta_0$	-0.2	-0.25	-0.3
0.2	97.86%		
0.25	97.47%	96.46%	
0.3	98.14%	97.04%	97.87%
0.35		97.36%	98.36%
0.4			98.62%

5.4 Moments of N_0 and $N_M + \sigma_M$

This section provides theoretical foundation for the approximations in Section 5.2, i.e. it gives related moments of N_0 conditioning on $\{S_{N_0} > d\}$, and moments of $N_M + \sigma_M$ conditioning on $\{S_{N_M} > d\}$. These quantities will be given in a series of lemmas.

Lemma 5.4 to 5.8 are special case of results of Chapter 3 and 4, they are presented here for an easy reference.

Lemma 5.4: *Under Condition (B),*

$$E_{\theta_1}[S_{N_0}^2 | S_{N_0} > d] = (d + \rho_+)^2 + O(1),$$

and

$$E_{\theta_1}[S_{N_M}^2 | S_{N_M} > d] = (d + \rho_+)^2 + O(1).$$

Lemma 5.5: *Under Condition (B),*

$$E_{\theta_1}[N_0|S_{N_0} > d] = \frac{d+\rho_+}{\theta_1} - \frac{1}{\theta_1^2} + O(\frac{1}{\theta_1}),$$

and

$$E_{\theta_1}[N_M + \sigma_M|S_{N_0} > d] = \frac{d+\rho_+}{\theta_1} + \frac{1}{\theta_1^2}(-1 + \frac{\theta_1^2}{2\theta_0^2}) + O(\frac{1}{\theta_1}).$$

Lemma 5.6: *Under Condition (B),*

$$E_{\theta_1}[N_0^2|S_{N_0} > d] = \frac{(d+\rho_+)^2}{\theta_1^2} - \frac{d+\rho_+}{\theta_1^3} - \frac{1}{\theta_1^4} + O(\frac{1}{\theta_1^3}).$$

and

$$Var_{\theta_1}[N_0|S_{N_0} > d] = \frac{d+\rho_+}{\theta_1^3} - \frac{2}{\theta_1^4} + O(\frac{1}{\theta_1^3}).$$

On the other hand,

$$E_{\theta_1}[(N_M + \sigma_M)^2|S_{N_M} > d] = \frac{(d+\rho_+)^2}{\theta_1^2} + \frac{d+\rho_+}{\theta_1^3}(-1 + \frac{\theta_1^2}{\theta_0^2})$$
$$+\frac{1}{\theta_1^4}(-1 - \frac{\theta_1^2}{\theta_0^2} + \frac{\theta_1^4}{\theta_0^4}) + O(\frac{1}{\theta_1^3}),$$

and

$$Var_{\theta_1}[N_M + \sigma_M|S_{N_M} > d] = \frac{d+\rho_+}{\theta_1^3} + \frac{1}{\theta_1^4}(-2 + \frac{3\theta_1^4}{4\theta_0^4}) + O(\frac{1}{\theta_1^3}).$$

Lemma 5.7: *Under Condition (B),*

$$E_{\theta_1}[N_0^3|S_{N_0} > d] = \frac{(d+\rho_+)^3}{\theta_1^3} - 3\frac{d+\rho_+}{\theta_1^5} - 3\frac{1}{\theta_1^6} + O(\frac{1}{\theta_1^5}),$$

and

$$E_{\theta_1}[(N_0 - E_{\theta_1}(N_0|S_{N_0} > D))^3|S_{N_0} > D] = 3\frac{d+\rho_+}{\theta_1^5} + O(\frac{1}{\theta_1^6}).$$

While

$$E_{\theta_1}[(N_M + \sigma_M)^3|S_{N_M} > d] = \frac{(d+\rho_+)^3}{\theta_1^3} + \frac{3(d+\rho_+)^2}{2\theta_1^2\theta_0^2} +$$
$$\frac{d+\rho_+}{\theta_1^5}(-3 + \frac{3\theta_1^4}{\theta_0^4} - \frac{3\theta_1^2}{2\theta_0^2}) + O(\frac{1}{\theta_1^6}),$$

and

$$E_{\theta_1}[(N_M + \sigma_M - E_{\theta_1}(N_M + \sigma_M|S_{N_M} > d))^3|S_{N_M} > d] = 3\frac{d+\rho_+}{\theta_1^5} + O(\frac{1}{\theta_1^6}).$$

In the following, we will further give expansions for higher order moments of N_0 conditioning on $\{S_{N_0} > d\}$, and those of $N_M + \sigma_M$ conditioning on $\{S_{N_M} > d\}$. Detailed calculations are quite complicated with lots of technical difficulties. To save space we only give main steps.

Similar to Equation (4.35), we know that

$$\int e^{\theta S_n - \frac{1}{2}n\theta^2} dF_0(x_1)...dF_0(x_n) = 1, \tag{5.10}$$

and both sides of the equation are infinite times differentiable w.r.t θ.

Lemma 5.8: *Under Condition (B),*

$$E_{\theta_1}[N_0^4|S_{N_0} > d] = \frac{(d+\rho_+)^4}{\theta_1^4} + 2\frac{(d+\rho_+)^3}{\theta_1^5} - 3\frac{(d+\rho_+)^2}{\theta_1^6} + O(\frac{d+\rho_+}{\theta_1^7}), \tag{5.11}$$

$$E_{\theta_1}[(N_0 - E_{\theta_1}(N_0|S_{N_0} > D))^4|S_{N_0} > D] = 3\frac{(d+\rho_+)^2}{\theta_1^6} + O(\frac{d+\rho_+}{\theta_1^7}), \quad (5.12)$$

and

$$Var_{\theta_1}(N_0^2|S_{N_0} > d) = 4\frac{(d+\rho_+)^3}{\theta_1^5} - 2\frac{(d+\rho_+)^2}{\theta_1^6} + O(\frac{d+\rho_+}{\theta_1^7}). \quad (5.13)$$

While

$$E_{\theta_1}[(N_M + \sigma_M)^4|S_{N_0} > d] = \frac{(d+\rho_+)^4}{\theta_1^4} + \frac{(d+\rho_+)^3}{\theta_1^5}(2 + \frac{2\theta_1^2}{\theta_0^2})$$
$$+\frac{(d+\rho_+)^2}{\theta_1^6}(-3 + \frac{6\theta_1^4}{\theta_0^4}) + O(\frac{d+\rho_+}{\theta_1^7}), \quad (5.14)$$

$$E[(N_M + \sigma_M - E(N_M + \sigma_M|S_{N_M} > d))^4|S_{N_M} > d] = 3\frac{(d+\rho_+)^2}{\theta_1^6} + O(\frac{d+\rho_+}{\theta_1^7}). \quad (5.15)$$

and

$$Var[(N_M + \sigma_M)^2|S_{N_M} > d] = 4\frac{(d+\rho)^3}{\theta_1^5} + \frac{(d+\rho_+)^2}{\theta_1^6}(-2 + \frac{4\theta_1^2}{\theta_0^2} + \frac{3\theta_1^4}{\theta_0^4}) + O(\frac{d+\rho_+}{\theta_1^7}).$$
$$(5.16)$$

Proof: We first prove results of N_0. Differentiating both side of (5.10) four times w.r.t θ, for $S_0 = 0$,

$$\{(S_n - \theta_1 n)^4 - 6n(S_n - \theta_1 n)^2 + 3n^2, \Im_n\} \quad (5.17)$$

is a martingale under P_{θ_1}. Thus, we have

$$E_{\theta_1}[N_0^4] = \frac{1}{\theta_1^4}[-E_{\theta_1}S_{N_0}^4 + 4\theta_1 E_{\theta_1}(N_0 S_{N_0}^3) - 6\theta_1^2 E_{\theta_1}(N_0^2 S_{N_0}^2) + 4\theta_1^3 E_{\theta_1}(N_0^3 S_{N_0})$$

$$+6E_{\theta_1}(N_0 S_{N_0}^2) - 12\theta_1 E_{\theta_1}(N_0^2 S_{N_0}) + 6\theta_1^2 E_{\theta_1}(N_0^3) - 3E_{\theta_1}(N_0^2)$$

$$= \frac{1}{\theta_1^4}[(d+\rho_+)^4 + 2\frac{(d+\rho_+)^3}{\theta_1} - 3\frac{(d+\rho_+)^2}{\theta_1^2} + O(\frac{d+\rho_+}{\theta_1^3})]P_{\theta_1}(S_{N_0} > d).$$

Since

$$E_{\theta_1}[N_0^4,, S_{N_0} > d] = E_{\theta_1}[N_0^4] - E_{\theta_1}[N_0^4, S_{N_0} \le 0]$$

$$= E_{\theta_1}[N_0^4] + O(\frac{1}{\theta_1^7}).$$

Thus, we have

$$E_{\theta_1}[N_0^4 | S_{N_0} > d] = \frac{(d+\rho_+)^4}{\theta_1^4} + 2\frac{(d+\rho_+)^3}{\theta_1^5} - 3\frac{(d+\rho_+)^2}{\theta_1^6} + O(\frac{d+\rho_+}{\theta_1^7}).$$

This proves (5.11).

(5.12) and (5.13) are obtained by applying Lemma 5.5, 5.6 and some calculations.

To prove (5.14), (5.15) and (5.16), from Lemma 5.5 and 5.6, we only need to show (5.14).

Similar to (5.17), differentiating both sides of equation (5.10) four times w.r.t θ, for $S_0 = x$ and $0 \le x < d$, we know that

$$\{(S_n - x - \theta_1 n)^4 - 6n(S_n - x - \theta_1 n)^2 + 3n^2, \Im_n\} \tag{5.18}$$

is a martingale under P_{θ_1}. So

$$E_{\theta_1}[N_M^4] = \frac{1}{\theta_1^4}[-E_{\theta_1}(S_{N_M} - M)^4 + 4\theta_1 E_{\theta_1}(N_M(S_{N_M} - M)^3) - 6\theta_1^2 E_{\theta_1}(N_M^2(S_{N_M} - M)^2) +$$

$$4\theta_1^3 E_{\theta_1}(N_M^3(S_{N_M} - M)) + 6E_{\theta_1}(N_M(S_{N_M} - M)^2) - 12\theta_1 E_{\theta_1}(N_M^2(S_{N_M} - M))$$

$$+ 6\theta_1^2 E_{\theta_1}(N_M^3) - 3E_{\theta_1}(N_M^2)]$$

$$= [\frac{(d+\rho_+)^4}{\theta_1^4} + \frac{(d+\rho_+)^3}{\theta_1^5}(4 + \frac{2\theta_1}{\theta_0} + \frac{2\theta_0}{\theta_1 - \theta_0}) + \frac{(d+\rho_+)^2}{\theta_1^6}(6 + \frac{6\theta_1}{\theta_0} + \frac{3\theta_1^2}{\theta_0^2} + \frac{9\theta_0}{\theta_1 - \theta_0} + \frac{3\theta_0\theta_1}{(\theta_1 - \theta_0)^2})$$

$$+O(\frac{d+\rho_+}{\theta_1^7})]P_{\theta_1}(S_{N_M} > d).$$

Similarly, we can show that

$$E[N_M^3\sigma_m] = [\frac{(d+\rho_+)^3}{\theta_1^5}(\frac{\theta_1^2}{2\theta_0^2} - \frac{\theta_1}{2\theta_0} - \frac{\theta_1}{2(\theta_1-\theta_0)})$$

$$+\frac{(d+\rho_+)^2}{\theta_1^6}(-\frac{3\theta_1}{2\theta_0} + \frac{3\theta_1^3}{2\theta_0^3} - \frac{3\theta_1^2}{2(\theta_1-\theta_0)^2} - \frac{3\theta_1}{2(\theta_1-\theta_0)}) + O(\frac{d+\rho_+}{\theta_1^7})]P_{\theta_1}(S_{N_M} > d).$$

and

$$E[N_M^2\sigma_m^2] = [\frac{(d+\rho_+)^2}{\theta_1^6}(\frac{\theta_1^4}{\theta_0^4} - \frac{\theta_1^3}{\theta_0^3} + \frac{\theta_1^3}{2\theta_0(\theta_1-\theta_0)^2} - \frac{\theta_1^3}{2\theta_0^2(\theta_1-\theta_0)}) + O(\frac{d+\rho_+}{\theta_1^7})]P_{\theta_1}(S_{N_M} > d).$$

Combining above results, we have

$$E[(N_M+\sigma_m)^4|S_{N_M} > d] = \frac{(d+\rho_+)^4}{\theta_1^4} + \frac{(d+\rho_+)^3}{\theta_1^5}(2 + \frac{2\theta_1^2}{\theta_0^2})$$

$$+\frac{(d+\rho_+)^2}{\theta_1^6}(-3 + \frac{6\theta_1^4}{\theta_0^4}) + O(\frac{d+\rho_+}{\theta_1^7}).$$

This ends the proof of Lemma 5.8.

Lemma 5.9: *Under Condition (B),*

$$E_{\theta_1}[N_0^5|S_{N_0} > d] = \frac{(d+\rho_+)^5}{\theta_1^5} + 5\frac{(d+\rho_+)^4}{\theta_1^6} + 5\frac{(d+\rho_+)^3}{\theta_1^7} + O(\frac{(d+\rho_+)^2}{\theta_1^8}),$$

and

$$E_{\theta_1}[N_M^5|S_{N_M} > d] = \frac{(d+\rho_+)^5}{\theta_1^5} + \frac{(d+\rho_+)^4}{\theta_1^6}(\frac{15}{2} + \frac{5\theta_1}{2\theta_0} + \frac{5\theta_0}{2(\theta_1-\theta_0)}) +$$

$$\frac{(d+\rho_+)^3}{\theta_1^7}(25 + \frac{15\theta_1}{\theta_0} + \frac{5\theta_1^2}{\theta_0^2} + \frac{20\theta_0}{\theta_1-\theta_0} + \frac{5\theta_1\theta_0}{(\theta_1-\theta_0)^2}) + O(\frac{(d+\rho_+)^2}{\theta_1^8}).$$

Proof: The idea is similar to that used for proving Lemma 5.8, and thus only main steps are given. Differentiating both sides of Equation (5.10) five times w.r.t θ, if $S_0 = 0$, we know that

$$\{(S_n - \theta_1 n)^5 - 10n(S_n - \theta_1 n)^3 + 15n(S_n - \theta_1 n), \Im_n\} \tag{5.19}$$

is a martingale under P_{θ_1}. Thus, we have

$$E_{\theta_1}[N_0^5] = \frac{1}{\theta_1^5}[E_{\theta_1} S_{N_0}^5 - 5\theta_1 E_{\theta_1}(N_0 S_{N_0}^4) + 10\theta_1^2 E_{\theta_1}(N_0^2 S_{N_0}^3) - 10\theta_1^3 E_{\theta_1}(N_0^3 S_{N_0}^2) +$$

$$5\theta_1^4 E_{\theta_1}(N_0^4 S_{N_0}) - 10 E_{\theta_1}(N_0 S_{N_0}^3) + 30\theta_1 E_{\theta_1}(N_0^2 S_{N_0}^2) - 30\theta_1^2 E_{\theta_1}(N_0^3 S_{N_0}) + 10\theta_1^3 E_{\theta_1} N_0^4$$

$$+15 E_{\theta_1}(N_0^2 S_{N_0}) - 15\theta_1 E_{\theta_1} N_0^3]$$

$$= \frac{1}{\theta_1^5}[(d+\rho_+)^5 + 5\frac{(d+\rho_+)^4}{\theta_1} + 5\frac{(d+\rho_+)^3}{\theta_1^2} + O(\frac{(d+\rho_+)^2}{\theta_1^3})]P_{\theta_1}(S_{N_0} > d).$$

Since

$$E_{\theta_1}[N_0^5, S_{N_0} > d] = E_{\theta_1}[N_0^5] - E_{\theta_1}[N_0^5, S_{N_0} \le 0]$$

$$= E_{\theta_1}[N_0^5] + O(\frac{1}{\theta_1^9}),$$

we have

$$E_{\theta_1}[N_0^5 | S_{N_0} > d] = \frac{(d+\rho_+)^5}{\theta_1^5} + 5\frac{(d+\rho_+)^4}{\theta_1^6} + 5\frac{(d+\rho_+)^3}{\theta_1^7} + O(\frac{(d+\rho_+)^2}{\theta_1^8}),$$

which is the first result of Lemma 5.9.

To prove the second result, similar to (5.19), differentiating both sides of equation (5.10) five times w.r.t θ, for $S_0 = x$ and $0 \le x < d$, we know that

$$\{(S_n - x - \theta_1 n)^5 - 10n(S_n - x - \theta_1 n)^3 + 15n(S_n - x - \theta_1 n), \Im_n\} \tag{5.20}$$

is a martingale under P_{θ_1}. Thus

$$E_{\theta_1}[N_M^5] = \frac{1}{\theta_1^5}[E_{\theta_1}(S_{N_M} - M)^5 - 5\theta_1 E_{\theta_1}(S_{N_M} - M)^4 N_{N_M} + 10\theta_1^2 E_{\theta_1}(S_{N_M} - M)^3 N_{N_M}^2$$

$$-10\theta_1^3 E_{\theta_1}(S_{N_M} - M)^2 N_{N_M}^3 + 5\theta_1^4 E_{\theta_1}(S_{N_M} - M) N_{N_M}^4 - 10 E_{\theta_1}(S_{N_M} - M)^3$$

$$+30\theta_1 E_{\theta_1}(S_{N_M} - M)^2 N_{N_M}^2 - 30\theta_1^2 E_{\theta_1}(S_{N_M} - M) N_{N_M}^3 + 10\theta_1^3 E_{\theta_1} N_{N_M}^4$$

$$+15 E_{\theta_1}(S_{N_M} - M) N_{N_M}^2 - 15\theta_1 E_{\theta_1} N_{N_M}^3]$$

$$= [\frac{(d+\rho_+)^5}{\theta_1^5} + \frac{(d+\rho_+)^4}{\theta_1^6}(\frac{15}{2} + \frac{5\theta_1}{2\theta_0} + \frac{5\theta_0}{2(\theta_1 - \theta_0)}) +$$

$$\frac{(d+\rho_+)^3}{\theta_1^7}(25 + \frac{15\theta_1}{\theta_0} + \frac{5\theta_1^2}{\theta_0^2} + \frac{20\theta_0}{\theta_1 - \theta_0} + \frac{5\theta_0\theta_1}{(\theta_1 - \theta_0)^2}) + O(\frac{d+\rho_+}{\theta_1^7})]P_{\theta_1}(S_{N_M} > d).$$

Since

$$E[N_M^5, S_{N_M} > d] = E[N_M^5] - E[N_M^5, S_{N_M} \le 0]$$

$$= E[N_M^5] + O(\frac{1}{\theta_1^{10}}).$$

Then

$$E[N_M^5 | S_{N_M} > d] = \frac{(d+\rho_+)^5}{\theta_1^5} + \frac{(d+\rho_+)^4}{\theta_1^6}(\frac{15}{2} + \frac{5\theta_1}{2\theta_0} + \frac{5\theta_0}{2(\theta_1 - \theta_0)}) +$$

$$\frac{(d+\rho_+)^3}{\theta_1^7}(25 + \frac{15\theta_1}{\theta_0} + \frac{5\theta_1^2}{\theta_0^2} + \frac{20\theta_0}{\theta_1 - \theta_0} + \frac{5\theta_0\theta_1}{(\theta_1 - \theta_0)^2}) + O(\frac{d+\rho_+}{\theta_1^7}).$$

This completes the proof of Lemma 5.9.

Lemma 5.10: *Under Condition (B),*

$$E_{\theta_1}[N_0^6 | S_{N_0} > d] = \frac{(d+\rho_+)^6}{\theta_1^6} + 9\frac{(d+\rho_+)^5}{\theta_1^7} + 30\frac{(d+\rho_+)^4}{\theta_1^8} + O(\frac{(d+\rho_+)^3}{\theta_1^9}). \quad (5.21)$$

and

$$E_{\theta_1}[(N_0^2 - E_{\theta_1}(N_0^2|S_{N_0} > d))^3|S_{N_0} > d] = 48\frac{(d+\rho_+)^4}{\theta_1^8} + O(\frac{(d+\rho_+)^3}{\theta_1^9}). \quad (5.22)$$

On the other hand,

$$E_{\theta_1}[(N_M + \sigma_M)^6|S_{N_0} > d] = \frac{(d+\rho_+)^6}{\theta_1^6} + \frac{(d+\rho_+)^5}{\theta_1^7}(9 + \frac{3\theta_1^2}{\theta_0^2})$$

$$+\frac{(d+\rho_+)^4}{\theta_1^8}(30 + \frac{15\theta_1^2}{\theta_0^2} + \frac{15\theta_1^4}{\theta_0^4}) + O(\frac{(d+\rho_+)^3}{\theta_1^9}), \quad (5.23)$$

and

$$E_{\theta_1}[((N_M + \sigma_M)^2 - E_{\theta_1}((N_M + \sigma_M)^2|S_{N_M} > d))^3|S_{N_M} > d]$$

$$= 48\frac{(d+\rho_+)^4}{\theta_1^8} + O(\frac{(d+\rho_+)^3}{\theta_1^9}). \quad (5.24)$$

Proof: First, differentiating both sides of Equation (5.10) six times w.r.t θ. for $S_0 = 0$, we know that

$$\{(S_n - \theta_1 n)^6 - 15n(S_n - \theta_1 n)^4 + 45n^2(S_n - \theta_1 n)^2 - 15n^3, \Im_n\} \quad (5.25)$$

is a martingale under P_{θ_1}. Thus

$$E_{\theta_1}[N_0^6] = \frac{1}{\theta_1^6}[-E_{\theta_1}S_{N_0}^6 6\theta_1 E_{\theta_1}(N_0 S_{N_0}^5) - 15\theta_1^2 E_{\theta_1}(N_0^2 S_{N_0}^4) + 20\theta_1^3 E_{\theta_1}(N_0^3 S_{N_0}^3)$$

$$-15\theta_1^4 E_{\theta_1}(N_0^4 S_{N_0}^2) + 6\theta_1^5 E_{\theta_1}(N_0^5 S_{N_0}) + 15E_{\theta_1}(N_0 S_{N_0}^4) - 60\theta_1 E_{\theta_1}(N_0^2 S_{N_0}^3)$$

$$+90\theta_1^2 E_{\theta_1}(N_0^3 S_{N_0}^2) - 60\theta_1^3 E_{\theta_1}(N_0^4 S_{N_0}) + 15\theta_1^4 E_{\theta_1}(N_0^5) - 45E_{\theta_1}(N_0^2 S_{N_0}^2)$$

$$+90\theta_1 E_{\theta_1}(N_0^3 S_{N_0}) - 45\theta_1^2 E_{\theta_1}(N_0^4) + 15E_{\theta_1}(N_0^3)]$$

$$= [\frac{(d+\rho_+)^6}{\theta_1^6} + 9\frac{(d+\rho_+)^5}{\theta_1^7} + 30\frac{(d+\rho_+)^4}{\theta_1^8} + O(\frac{(d+\rho_+)^3}{\theta_1^9})]P_{\theta_1}(S_{N_0} > d).$$

Since

$$E_{\theta_1}[N_0^6, S_{N_0} > d] = E_{\theta_1}[N_0^6] - E_{\theta_1}[N_0^6, S_{N_0} \leq 0]$$

$$= E_{\theta_1}[N_0^6] + O(\frac{1}{\theta_1^{11}}).$$

Then, we obtain that

$$E_{\theta_1}[N_0^6 | S_{N_0} > d] = \frac{(d+\rho_+)^6}{\theta_1^6} + 9\frac{(d+\rho_+)^5}{\theta_1^7} + 30\frac{(d+\rho_+)^4}{\theta_1^8} + O(\frac{(d+\rho_+)^3}{\theta_1^9}).$$

(5.21) follows after some calculations.

As for approximating $E[(N_M + \sigma_M)^6 | S_{N_M} > d]$, first note that, for $S_0 = x$ and $0 \leq x < d$,

$$\{(S_n - x - \theta_1 n)^6 - 15n(S_n - x - \theta_1 n)^4 + 45n^2(S_n - x - \theta_1 n)^2 - 15n^3, \Im_n\} \quad (5.26)$$

is a martingale under P_{θ_1}. Thus

$$E_{\theta_1}[N_M^6] = \frac{1}{\theta_1^6}[-E(S_{N_M} - M)^6 + 6\theta_1 E(N_M(S_{N_M} - M)^5) - 15\theta_1^2 E(N_M^2(S_{N_M} - M)^4)$$

$$+20\theta_1^3 E(N_M^3(S_{N_M} - M)^3) - 15\theta_1^4 E(N_M^4(S_{N_M} - M)^2) + 6\theta_1^5 E(N_M^5(S_{N_M} - M))$$

$$+15E(N_M(S_{N_M} - M)^4) - 60\theta_1 E(N_M^2(S_{N_M} - M)^3) + 90\theta_1^2 E(N_M^3(S_{N_M} - M)^2)$$

$$-60\theta_1^3 E(N_M^4(S_{N_M} - M)) + 15\theta_1^4 E(N_M^5) - 45E(N_M^2(S_{N_M} - M)^2) + 90\theta_1 E(N_M^3(S_{N_M} - M))$$

$$-45\theta_1^2 E(N_M^4) + 15E(N_M^3)].$$

After some complicated calculation and omitting the terms with order lower than $\frac{(d+\rho_+)^4}{\theta_1^8}$, we obtain that

$$E_{\theta_1}[N_M^6] = [\frac{(d+\rho_+)^6}{\theta_1^6} + \frac{(d+\rho_+)^5}{\theta_1^7}(12 + \frac{3\theta_1}{\theta_0} + \frac{3\theta_0}{\theta_1 - \theta_0}) +$$

$$\frac{(d+\rho_+)^4}{\theta_1^8}(\frac{135}{2} + \frac{30\theta_1}{\theta_0} + \frac{75\theta_0}{\theta_1 - \theta_0} + \frac{15\theta_1^2}{2\theta_0^2} + \frac{15\theta_0\theta_1}{2(\theta_1 - \theta_0)^2}) + O(\frac{(d+\rho_+)^3}{\theta_1^9})]P_{\theta_1}(S_{N_M} > d).$$

Therefore, we have

$$E_{\theta_1}[N_M^6 | S_{N_M} > d] = \frac{(d+\rho_+)^6}{\theta_1^6} + \frac{(d+\rho_+)^5}{\theta_1^7}(12 + \frac{3\theta_1}{\theta_0} + \frac{3\theta_0}{\theta_1 - \theta_0})$$

$$+ \frac{(d+\rho_+)^4}{\theta_1^8}(\frac{135}{2} + \frac{30\theta_1}{\theta_0} + \frac{75\theta_0}{\theta_1 - \theta_0} + \frac{15\theta_1^2}{2\theta_0^2} + \frac{15\theta_0\theta_1}{2(\theta_1 - \theta_0)^2}) + O(\frac{(d+\rho_+)^3}{\theta_1^9}).$$

Similarly, we have

$$E[N_M^5 \sigma_M | S_{N_M} > d] = \frac{(d+\rho_+)^5}{\theta_1^7}(\frac{\theta_1^2}{2\theta_0^2} - \frac{\theta_1}{2\theta_0} - \frac{\theta_1}{2(\theta_1 - \theta_0)})$$

$$+ \frac{(d+\rho_+)^4}{\theta_1^8}(\frac{5\theta_1}{\theta_0} + \frac{5\theta_1^2}{2\theta_0^2} + \frac{5\theta_1^3}{2\theta_0^3} - \frac{5\theta_1}{\theta_1 - \theta_0} - \frac{5\theta_1^2}{2(\theta_1 - \theta_0)^2}) + O(\frac{(d+\rho_+)^3}{\theta_1^9}).$$

and

$$E[N_M^4 \sigma_M^2 | S_{N_M} > d] = \frac{(d+\rho_+)^4}{\theta_1^8}(\frac{\theta_1^4}{\theta_0^4} - \frac{\theta_1^3}{\theta_0^3} + \frac{\theta_1^3}{2\theta_0(\theta_1 - \theta_0)^2} - \frac{\theta_1^3}{2\theta_0^2(\theta_1 - \theta_0)}) + O(\frac{(d+\rho_+)^3}{\theta_1^9}).$$

(5.23) is obtained by applying Lemma 5.6. The proof is completed.

Lemma 5.11: *Under Condition (B),*

$$E_{\theta_1}[(N_0^2 - E_{\theta_1}(N_0^2 | S_{N_0} > d))^4 | S_{N_0} > d] = 48\frac{(d+\rho_+)^6}{\theta_1^10} + O(\frac{(d+\rho_+)^5}{\theta_1^{11}}). \quad (5.27)$$

and

$$E_{\theta_1}[((N_M+\sigma_M)^2-E_{\theta_1}((N_M+\sigma_M)^2|S_{N_M}>d))^4|S_{N_M}>d] = 48\frac{(d+\rho_+)^6}{\theta_1^10}+O(\frac{(d+\rho_+)^5}{\theta_1^{11}}). \quad (5.28)$$

Proof: To prove this Lemma, we note that, for $S_0 = 0$,

$$\{(S_n-\theta_1 n)^8-28n(S_n-\theta_1 n)^6+210n^2(S_n-\theta_1 n)^4-42(S_n-\theta_1 n)^2+105n^4, \Im_n\} \quad (5.29)$$

is a martingale under P_{θ_1}, and for $S_0 = x$ with $0 \le x < d$,

$$\{(S_n-x-\theta_1 n)^8-28n(S_n-x-\theta_1 n)^6+210n^2(S_n-x-\theta_1 n)^4-42n^3(S_n-x-\theta_1 n)^2$$

$$+105n^4, \Im_n\} \quad (5.30)$$

is a martingale under P_{θ_1}. Similar to the proof of Lemma 5.10, we obtain that

$$E_{\theta_1}[N_0^8|S_{N_0}>d] = \frac{(d+\rho_+)^8}{\theta_1^8}+20\frac{(d+\rho_+)^7}{\theta_1^9}+182\frac{(d+\rho_+)^6}{\theta_1^{10}}+O(\frac{(d+\rho_+)^5}{\theta_1^{11}}). \quad (5.31)$$

and

$$E_{\theta_1}[(N_M+\sigma_M)^8|S_{N_M}>d] = \frac{(d+\rho_+)^8}{\theta_1^8}+\frac{(d+\rho_+)^7}{\theta_1^9}(20+4\frac{\theta_1^2}{\theta_0^2})$$
$$+\frac{(d+\rho_+)^6}{\theta_1^{10}}(182+56\frac{\theta_1^2}{\theta_0^2}+28\frac{\theta_1^4}{\theta_0^4})+O(\frac{(d+\rho_+)^5}{\theta_1^{11}}). \quad (5.32)$$

Lemma 5.11 follows after some tedious calculations.

With above available results, the remain problem is the error checking for equations (5.7) and (5.8). We now prove that the error terms are indeed at the order of $o(\frac{1}{d^2})$.

Error Checking for (5.7) and (5.8):

Because of the similarity of the proofs, we only give the proof of (5.7).

First we note that conditioning on the event $\{S_{N_0} > d\}$,

$$\frac{S_{N_0}^2}{E_{\theta_1}[N_0^2|S_{N_0} > d]} - \frac{E_{\theta_1}[S_{N_0}^2|S_{N_0} > d]}{E_{\theta_1}[N_0^2|S_{N_0} > d]} \to 0 \ \ in \ \ probability \ \ as \ \ d \to \infty;$$

and

$$\frac{N_0^2}{E_{\theta_1}[N_0^2|S_{N_0} > d]} \to 1 \ \ in \ \ probability \ \ as \ \ d \to \infty.$$

Thus, we know that for the fifth order terms of Taylor expansion, the coefficient of $E_{\theta_1}[(\frac{S_{N_0}^2}{E[N_0^2|S_{N_0}>d]} - \frac{E_{\theta_1}[S_{N_0}^2|S_{N_0}>d]}{E_{\theta_1}[N_0^2|S_{N_0}>d]})(\frac{N_0^2}{E_{\theta_1}[N_0^2|S_{N_0}>d]} - 1)^4|S_{N_0} > d]$ converges to 1 in probability; while the coefficient of $E_{\theta_1}[(\frac{N_0^2}{E_{\theta_1}[N_0^2|S_{N_0}>d]} - 1)^5|S_{N_0} > d]$ is at the order of $\frac{E_{\theta_1}[S_{N_0}^2|S_{N_0}>d]}{E_{\theta_1}[N_0^2|S_{N_0}>d]} = O(\theta_1^2)$ in probability. Thus, we only need to show that

$$E_{\theta_1}[(\frac{S_{N_0}^2}{E[N_0^2|S_{N_0} > d]} - \frac{E_{\theta_1}[S_{N_0}^2|S_{N_0} > d]}{E_{\theta_1}[N_0^2|S_{N_0} > d]})(\frac{N_0^2}{E_{\theta_1}[N_0^2|S_{N_0} > d]} - 1)^4|S_{N_0} > d] = o(\frac{1}{(d+\rho)^2}), \tag{5.33}$$

and

$$E_{\theta_1}[(\frac{N_0^2}{E_{\theta_1}[N_0^2|S_{N_0} > d]} - 1)^5|S_{N_0} > d] = o(\frac{1}{\theta_1^2(d+\rho)^2}). \tag{5.34}$$

To prove (5.33) and (5.34), by the similar techniques used in Section 4.5, we have to evaluate $E_{\theta_1}[(N_0^2 - A_1)^5|S_{N_0} > d]$.

First, by noting that, for $S_0 = 0$,

$$\{(S_n - \theta_1 n)^9 - 36n(S_n - \theta_1 n)^7 + 378n^2(S_n - \theta_1 n)^5 - 1260n^3(S_n - \theta_1 n)^3$$

$$+94n^4(S_n - \theta_1 n), \Im_n\} \tag{5.35}$$

is a martingale under P_{θ_1}.

By the same method as used in proving Lemma 5.10 and after same simplification, we obtain that

$$E_{\theta_1}[N_0^9 | S_{N_0} > d] = \frac{(d+\rho_+)^9}{\theta_1^9} + 27\frac{(d+\rho_+)^8}{\theta_1^{10}} + 342\frac{(d+\rho_+)^7}{\theta_1^{11}} + O(\frac{(d+\rho_+)^6}{\theta_1^{12}}). \tag{5.36}$$

Similarly, by noting that, for $S_0 = 0$,

$$\{(S_n - \theta_1 n)^{10} - 45n(S_n - \theta_1 n)^8 + 630n^2(S_n - \theta_1 n)^6 - 3150n^3(S_n - \theta_1 n)^4$$

$$+4725n^4(S_n - \theta_1 n)^2 - 945n^5, \Im_n\} \tag{5.37}$$

is a martingale under P_{θ_1}. Thus, we have

$$E_{\theta_1}[N_0^{10} | S_{N_0} > d] = \frac{(d+\rho_+)^{10}}{\theta_1^{10}} + 35\frac{(d+\rho_+)^9}{\theta_1^{11}} + 585\frac{(d+\rho_+)^8}{\theta_1^{12}} + O(\frac{(d+\rho_+)^7}{\theta_1^{13}}). \tag{5.38}$$

Therefore,

$$E_{\theta_1}[(N_0^2 - E_{\theta_1}(N_0^2 | S_{N_0} > d))^5 | S_{N_0} > d] = O(\frac{(d+\rho_+)^7}{\theta_1^{13}}). \tag{5.39}$$

Thus

$$\begin{aligned} E_{\theta_1}[(\frac{N_0^2}{E_{\theta_1}[N_0^2 | S_{N_0} > d]} - 1)^5 | S_{N_0} > d] &= \frac{E_{\theta_1}[(N_0^2 - E_{\theta_1}(N_0^2 | S_{N_0} > d))^5 | S_{N_0} > d]}{(E_{\theta_1}[N_0^2 | S_{N_0} > d])^5} \\ &= O(\frac{(d+\rho_+)^7}{\theta_1^{13}}) / (\frac{(d+\rho_+)^{10}}{\theta_1^{10}}) \\ &= O(\frac{1}{(d+\rho_+)^3 \theta_1^3}) = o(\frac{1}{(d+\rho_+)^2 \theta_1^2}), \end{aligned}$$

which completes the proof of (5.34).

To show (5.33), by Hölder inequality, we have

$$E_{\theta_1}[(\frac{S_{N_0}^2}{E[N_0^2|S_{N_0} > d]} - \frac{E_{\theta_1}[S_{N_0}^2|S_{N_0} > d]}{E_{\theta_1}[N_0^2|S_{N_0} > d]})(\frac{N_0^2}{E_{\theta_1}[N_0^2|S_{N_0} > d]} - 1)^4|S_{N_0} > d] \leq$$

$$(E_{\theta_1}[(|\frac{S_{N_0}^2}{E[N_0^2|S_{N_0} > d]} - \frac{E_{\theta_1}[S_{N_0}^2|S_{N_0} > d]}{E_{\theta_1}[N_0^2|S_{N_0} > d]}|)^5])^{1/5}(E[|\frac{N_0^2}{E[N_0^2|S_{N_0} > d]} - 1|^5|S_{N_0} > d])^{4/5}$$

$$= O(\frac{\theta_1^2}{(d+\rho_+)}) * O(1) = o(\frac{1}{(d+\rho)^2}),$$

This completes the proof of (5.7).

Similarly, we can proof (5.8).

REFERENCES

Anscombe, F. (1952). Large sample theory of sequential estimation. *Proc. Camb. Phil. Soc.*, **48**, 600-607,

Cox, D. R. (1952). A note on the sequential estimation of mean. *Proc. Camb. Phil. Soc.*, **48**, 447-450.

Todd, S., Whitehead, J. and Facey, K. M. (1996). Point and interval estimation following a sequential clinical trial. *Biometrika*, **83**, 453-461.

Whitehead, J. (1997). *The design and analysis of sequential clinical trials.* 3rd ed., John Wiley & Sons, New York.

Woodroofe, M. (1986). Very weak expansion for sequential confidence levels. *Ann. Statist.*, **14**, 1049-1067.

Woodroofe, M. (1992). Estimation after sequential testing: A simple approach for a truncated sequential probability ratio test. *Biometrika*, **79**, 347-353.

Chapter 6

Summary and Topics for Future Research

This thesis develops some theory on inference problems after CUSUM test. The CUSUM test is extensively used in some industrial settings. For example, in Lucas (1985), James Lucas stated that "There are now more than 10000 CUSUM control schemes in use daily in DU Pond". With such significant use in applications, one would imagine that any improvement in our understanding of the properties of CUSUM procedure could be very useful. The main goal of this thesis is to provide some guidance to application.

In Chapter 2, properties of unbalanced two-sided random walk are investigated. The results obtained not only form the theoretical foundation for the coming chapters, in term of random walk, but also have their own independent interests as well. Chapter 3 discusses how to construct confidence intervals for the change point based on the information provided by the control charts. A practical method for constructing the confidence intervals is proposed for any given significance level by tracing back some zero points of the control charts. At the same time, the confidence intervals can also be built by tracing back certain number of items produced before the stopping time. Chapter 4 turns to studying the properties of estimates of change point and change magnitude. It is common that estimation after sequential test gives certain bias. Thus, derivations of biases of estimates of change point and change magnitude constitute the main part of Chapter 4. It is found that biases of change point and change magnitude are both not negligible, and bias of the estimate of change magnitude is especially substantial. Thus bias correction is definitely required. For this purpose, a practical method for bias correction is proposed therein. In Chapter 5, properties of estimate of change magnitude are further investigated . The asymptotic distribution function of the estimate of change magnitude is derived by Woodroofe's *very weak expansion* method. Based on the asymptotic distribution function, confidence intervals for change magnitude are obtained.

Results of this thesis are mainly developed under the frame of the standard one parameter exponential family. This frame allows our results to be applied to many special cases. Of particular important examples, we can apply the results to (i) the normal family, (ii) the non-negative exponential family, (iii) $Gamma(\alpha, \beta)$ family with α being fixed, and etc. These distribution families are among the most commonly used distribution functions in practice, and this makes our results very meaningful. Our results have been described in the context of quality control, but they are more widely applicable.

The ultimate goal is to apply the theory to practice. For this purpose, there are still some problems remained.

First, in reality, when changes happen, we rarely know the change magnitude, so we have to find a method to estimate it. To solve this problem, using results on the bias of change magnitude obtained in Chapter 4, in the normal case, the following iterative scheme is proposed:

$$\hat{\theta}_1(i) = \frac{\hat{\theta}_1}{1 + est.bias(\hat{\theta}_1(i-1))/\hat{\theta}_1(i-1)}, \quad for \quad i = 1, 2, \ldots,$$

with $\hat{\theta}_1(0) = -\theta_0$.

After change magnitude being estimated, we can further use it to estimate the bias of change point. With estimates for change magnitude and change point are obtained, the rest of questions become routine. But how well this method works is still under investigation.

Second, in this thesis, it is assumed that the in-control variance σ_0^2 is known. However, in practice, contrary to the problem of detecting change in the mean, there typically is no target value for the in-control variance, which should be as small as the process permits (Pollak and Siegmund (1991)); besides, in the literature, even in the case when the in-control variance is known, the change magnitude being detected is always assumed to be the relative ratio with respect to the in-control standard deviation. This reminds us to construct control charts which are invariant under scale transformation.

To illustrate this problem, the following simple case is given. As in previous chapters, we assume that the process being monitored is normally distributed with mean μ and variance σ_0^2. At each sampling point, random samples of size 3 are taken regularly. Let s_k^2 denote the sample variance of the kth sample. If the process standard deviation changes from σ_0 to $(1+\epsilon)\sigma_0$ at some unknown point ν, then $s_1^2, s_2^2, \ldots, s_\nu^2$ are independent and identically distributed as non-negative exponential random variable with mean σ_0^2; while $s_{\nu+1}^2, \ldots, s_n^2$ are independent and identically distributed as non-negative exponential random variable with mean $(1+\epsilon)^2\sigma_0^2$.

We also assume that the process is in control before sample point n_0, for some known value n_0, $0 < n_0 \leq \nu$. These data are used as the training samples which provide information for an initial estimate of σ_0^2.

Let

$$V_k = \sum_{i=1}^{k} s_i^2 \quad k = 1, 2, \ldots \tag{6.1}$$

$$Y_k = \frac{s_{k+1}^2}{V_{k+1}} \quad k = 1, 2, \ldots \tag{6.2}$$

$$Z_k = (1 - Y_k)^k \quad k = 1, 2, \ldots \tag{6.3}$$

$$\xi_k = -k \ln(1 - Y_k) \quad k = 1, 2, ... \tag{6.4}$$

It is well known that for any positive integer n, when the process variance has not changed, $(Y_1, Y_2, ..., Y_{n-1})$ is the maximum invariant statistics based on $(s_1^2, s_2^2, ..., s_n^2)$ under the scale transformation. and $Y_1, Y_2, ..., Y_{n-1}$ are also mutually independent (Phatarford 1971). Some simple calculations show that $Z_1, Z_2, ..., Z_{n-1}$ are iid random variables with uniform distribution on $(0, 1)$. Therefore, $\xi_1, \xi_2, ..., \xi_{n-1}$ are iid r.v.s having a non-negative exponential distribution function with mean 1.

When the process goes out of control, say, the standard deviation changes from σ_0 to $(1 + \epsilon)\sigma_0$ at some unknown point ν, then $\xi_{\nu+i}$ for $i \ll \nu$ are approximately independent and distributed as non-negative exponential random variable with mean $(1 + \epsilon)^2$. In fact, we have

$$\begin{aligned} \xi_{\nu+i} &= -(\nu + i) \ln(1 - Y_{\nu+i}) \\ &= (\nu + i) \ln(1 + \frac{s_{\nu+i+1}^2}{V_{\nu+i}}) \end{aligned}$$

$$\approx \frac{s_{\nu+i+1}^2}{\frac{1}{\nu+i} V_{\nu+i}}, \textit{ when } \nu \textit{ is large and } i \ll \nu. \tag{6.5}$$

With the distribution functions before and after change are known, CUSUM procedure based on the maximum invariant statistic is proposed as follows:

Stop the process at the time

$$N = \inf\{n > 0 : T_n > d\},$$

where

$$T_n = \max(0, T_{n-1} + (\xi_n - r)), \quad with \quad r = \frac{(1+\epsilon_0)^2 \ln(1+\epsilon_0)^2}{(1+\epsilon_0)^2 - 1}, \quad and \quad T_0 = 0,$$

ϵ_0 *is the reference value for* ϵ*, and* d *is the control limit.*

To end this question, we have to get rid of the restriction on the sample size. For a given sample size m, what kind of transformation shall we perform on $Y_k, k = 1, 2, \ldots$, and what are the distribution functions before and after change after the transformation? These are remained questions under study.

Finally, we want to extend our study to the multivariate case. This raises another series of similar problems. Details of these questions will be carried on in the future researches.

REFERENCES

Lucas, J. (1985): Cumulative sum (cusum) control scheme. *Commun. Statist. Theor. Meth.*, **14**, 2689-2704.

Phatarford, R.M. (1971): A sequential test for Gamma distributions. *J. Amer. Statist. Assoc.*, **66**, 876-878.

Pollak, M. and Siegmund, D. (1991): Sequential detection of a change in a normal mean when the initial value is unknown. *Ann. Statist.*, **19**, 394-416.