CHANGE-POINT ANALYSIS: SINGLE CHANGE-POINT IN A SEQUENCE OF INDEPENDENT GAUSSIAN AND EXPONENTIAL RANDOM VARIABLES

By

ELENA A KHAPALOVA

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

DOCTOR OF PHILOSOPHY

WASHINGTON STATE UNIVERSITY
College of Business

MAY 2009

To the Faculty of Washington State University:

The members of the Committee appointed to examine the dissertation of ELENA A KHAPALOVA find it satisfactory and recommend that it be accepted.

____

Stergios B. Fotopoulos, Ph.D., Chair

____

Venkata K. Jandhyala, Ph.D.

____

Sung K. Ahn, Ph.D.

____

Min-Chiang Wang, Ph.D.

____

K.A. Ariyawansa, Ph.D.

ACKNOWLEDGMENTS

I am very grateful to my advisor, Dr. Stergios B. Fotopoulos. I have been very fortunate to have his support and encouragement throughout my graduate studies and while writing this dissertation. I am thankful for his help with the theoretical and technical aspects of this research, in particular, the first two sections of chapters two and three are based on his work and are in large part his contribution. I would like to thank Dr. Jandhyala for the time he spent working with me on our research, especially this past summer, and also for his service as a member of my committee. I would also like to thank Dr. Ariyawansa for taking the time from his busy schedule to serve as a member of my committee. Moreover, I would like to thank Dr. Wang and Dr. Ahn for agreeing to serve on my committee. Dr. Wang has always been there to listen and offer advice, I very much appreciate the support and the time he gave me. Dr. Ahn has offered encouragement and time both for my research and my studies, I am incredibly grateful for his support and would like to say that he, as the department chair, takes very good care of us graduate students. Finally, I would like to offer my utmost gratitude to Janet Reid and Barbara Chatburn who were always there to listen and offer moral support.

I would like to thank everyone again for their patience, time, and encouragement, I will always look back at my time at Washington State University with appreciation and much fondness.

CHANGE-POINT ANALYSIS: SINGLE CHANGE-POINT IN A SEQUENCE OF INDEPENDENT GAUSSIAN AND EXPONENTIAL RANDOM VARIABLES

Abstract

by Elena A Khapalova, Ph.D.
Washington State University
May 2009

Chair: Stergios B. Fotopoulos

This thesis considers the problem of change-point analysis, more specifically the maximum likelihood estimates of a single abrupt change-point in sequence of independent, time-ordered exponential and normal random variables and the asymptotic distributions of those estimates. Exact computable expressions for the asymptotic distributions are calculated and the distributions are used to calculate confidence intervals for a change detected in the multivariate gaussian time-series of annual mean precipitation, a univariate gaussian time-series of temperature anomalies, and a time-ordered sequence of exponentially distributed time intervals between earthquakes. The accuracy of the asymptotic distribution and, in the gaussian case, robustness to departures from normality are investigated via simulations.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS iii

ABSTRACT iv

LIST OF TABLES vii

LISTOF FIGURES x

CHAPTER

1. INTRODUCTION TO CHANGE-POINT ANALYSIS 1

2. ASYMPTOTIC DISTRIBUTION OF THE MAXIMUM LIKELIHOOD ESTIMATE OF THE CHANGE-POINT WHEN A CHANGE OCCURS IN THE RATE PARAMETER OF AN EXPONENTIAL PROCESS 8

General results for the distribution of the maximum likelihood estimate of the change-point τ 10

The distribution of the maximum likelihood estimate of the change-point τ in the exponential case 21

Exact computable expressions for the asymptotic distribution of ξ_∞ in the exponential case 41

Change-point analysis of earthquake frequency data from the Sumatra, West Indonesia area 62

Simulations 73

3. ASYMPTOTIC DISTRIBUTION OF THE MAXIMUM LIKELIHOOD ESTIMATE OF THE CHANGE-POINT WHEN A CHANGE OCCURS IN THE MEAN PARAMETER OF A GAUSSIAN PROCESS 94

General results for the distribution of the maximum likelihood estimate of the change-point τ 96

Asymptotic distribution of ξ_∞ when the underlying process is Gaussian 100

Change-point analysis of multivariate global mean precipitation data 109

Change-point analysis of the mean annual temperature anomalies data 123

Simulations 130

APPENDIX

A.1 J-fold convolution densities 148

A.2 Determination of $u_k(dx)$ for $P(\xi_\infty = -k), k > 0$ 151

A.3 Calculations of $P(\xi_\infty = n),\ n < 0$ 162

A.4 The likelihood approach to hypothesis testing 179

BIBLIOGRAPHY 185

LIST OF TABLES

Table 2.1. The probability distribution function of $P(\xi_\infty = k)$ for $v_1/v_2 = 1.5, 2, 3, 4, 5,$ $and\ 6.$ 59

Table 2.2. The probability distribution function of $P(\xi_\infty = k)$ for $v_1/v_2 = 2, 6, 10, 14,$ $18, 22.$ 60

Table 2.3. The probability distribution function of $P(\xi_\infty = k)$ comparing the exact and approximate values for $k = -1, -2,$ ... -6. 61

Table 2.4. Time series of the 151 time intervals (in hours) between earthquakes. Date and Time indicate the time of detected earthquake. Hours indicates the time between the preceding earthquake and the current one. 65

Table 2.5. Probabilities and corresponding dates and times for $v_2/v_1 = 2.592$. 72

Table 2.6. Comparison of theoretical and simulated distribution of the mle, when the true change-point (τ) location changes for $v_1/v_2 = 2$. 77

Table 2.7 Comparison of theoretical and simulated distribution of the mle, when the true change-point (τ) location changes for $v_1/v_2 = 6$. 78

Table 2.8. Comparison of theoretical and simulated distribution of the mle, when the true change-point (τ) location changes for $v_1/v_2 = 14$. 79

Table 2.9. Comparison of theoretical and simulated distribution of the mle, when the true change-point (τ) location changes for $v_1/v_2 = 2$. 80

Table 2.10. Comparison of theoretical and simulated distribution of the mle, when the true change-point (τ) location changes for $v_1/v_2 = 6$. 81

Table 2.11. Comparison of theoretical and simulated distribution of the mle, when the true change-point (τ) location changes for $v_1/v_2 = 14$. 82

Table 2.12. Bias and MSE for v_1/v_2 = 2, 3, 6, 10, 14, 18, and 22 for varying sample sizes n and change-point τ locations. 83

Table 2.13. Comparison of theoretical and simulated (conditional and unconditional) distribution of the mle for v_1/v_2 = 2, 6, 10, 14, 18 and 22, when n = 40 and τ = 30. (Error rate for c mle 10^{-1} to 10^{-5}). 84

Table 2.14. Comparison of theoretical and simulated (conditional and unconditional) distribution of the mle for ν_1/ν_2 = 2, 6, 10, 14, 18 and 22, when n = 40 and τ = 20. (Error rate for c mle 10^{-1} to 10^{-5}). 85

Table 2.15. Comparison of theoretical and simulated (conditional and unconditional) distribution of the mle for ν_1/ν_2 = 2, 6, 10, 14, 18 and 22, when n = 40 and τ = 10. (Error rate for c mle 10^{-1} to 10^{-5}). 86

Table 2.16. Comparison of theoretical and simulated (conditional and unconditional) distribution of the mle for ν_1/ν_2 = 2, 6, 10, 14, 18 and 22, when n = 200 and τ = 100. (Error rate for c mle 10^{-4} to 10^{-5}). 87

Table 3.1. The probability distribution of $P(\xi_\infty = k)$, k = 0, ±1, ±2, … ±20 and η = 0.2, 0.6, 1.0, 1.5, 2.0, 2.5, 3.0. 108

Table 3.2. Time-series data. Annual precipitation in mm per day for the years 1901-2000 for the three bands: NL, Low, and SL. 111

Table 3.3. Datasets and corresponding test statistic U_n and w values. 114

Table 3.4. Datasets and corresponding p-values and change-point estimates $\hat{\tau}$. 115

Table 3.5. Datasets and corresponding test statistic U_n^* and w values. 117

Table 3.6. Datasets and corresponding p-values and change-point estimates $\hat{\tau}$. 117

Table 3.7. The unconditional probability distribution of $P(\xi_\infty = k)$ for the annual mean precipitation data. η = 2.448, 2.40, 1.39, 2.72, 2.11, 1.33. 121

Table 3.8. The conditional probability distribution, Cobb (1978), of the change-point mle for the annual mean precipitation data, with error rate around 10^{-5}. 122

Table 3.9. Annual mean temperature anomalies in .01° Celsius for the years 1919-1978. Anomaly values indicate the deviations from the corresponding 1951-1980 means. 126

Table 3.10. The unconditional probability distribution of $P(\xi_\infty = n)$ for the annual mean temperature anomalies data, η = 1.299. 130

Table 3.11. Bias and MSE for the distributions of mle and cmle of the change-point under normal, t_v, and χ_v^2 distributions. 135

Table 3.12. Theoretical distribution of the mle and the simulated distributions of the change-point mle and cmle under normal, t_v, and χ_v^2 distributions when n = 60 and τ = 20 and for η = 1.0, k = -20, …, 20. 136

Table 3.13. Theoretical distribution of the mle and the simulated distributions of the change-point mle and cmle under normal, t_v, and χ_v^2 distributions when n = 60 and τ=30 and for η = 1.0, k = -20, …, 20. 137

Table 3.14. Theoretical distribution of the mle and the simulated distributions of the change-point mle and cmle under normal, t_v, and χ_v^2 distributions when n = 60 and τ=20 and for η = 2.5, k = -20, …, 20. 138

Table 3.15. Theoretical distribution of the mle and the simulated distributions of the change-point mle and cmle under normal, t_v, and χ_v^2 distributions when n = 60 and τ=30 and for η = 2.5, k = -20, …, 20. 139

Table 3.16. Theoretical distribution of the mle and the simulated distributions of the change-point mle and cmle under normal, t_v, and χ_v^2 distributions when n = 100 and τ = 50 and for η = 2.5, k = -20, …, 20. 140

Table 3.17. Theoretical distribution of the mle and the simulated distributions of the change-point mle and cmle under normal, t_v, and χ_v^2 distributions when n = 100 and τ = 50 and for η = 1.0, k = -20, …, 20. 141

LIST OF FIGURES

Figure 2.1. The probability distribution function of $P(\xi_\infty = k)$ for $v_1/v_2 = 1.5, 2, 2.5, 3, \ldots 7$. 58

Figure 2.2. The probability distribution function of $P(\xi_\infty = k)$ for $v_1/v_2 = 2, 6, 10, 14, 18, 22$. 58

Figure 2.3. Time series of the 151 time intervals (in hours) between earthquakes. 64

Figure 2.4. Topographic map of Sumatra, West Indonesia. Our region of study is defined by latitude of 6° N to 6° S and longitude of 94° E to 107° E. 66

Figure 2.5. Plot of $-2\log\Lambda_t$ of the earthquake dataset. 68

Figure 2.6. Autocorrelation for the standardized dataset. 70

Figure 2.7. Partial Autocorrelation for the standardized dataset. 70

Figure 2.8. Exponential probability plot for data before and after change. Anderson-Darling statistics are 0.355 and 0.646 respectively. The p-values are 0.726 and 0.320 respectively. 71

Figure 2.9 (a-f). Comparison of theoretical and simulated (conditional and unconditional) distribution of the mle for ν_1/ν_2 = 2, 6, 10, 14,18, 22, n = 200 and τ = 100.. 89

Figure 2.10 (a-f). Comparison of theoretical and simulated (conditional and unconditional) distribution of the mle for ν_1/ν_2 = 2, 6, 10, 14, 18, 22, n = 40 and τ = 30. ... 91

Figure 3.1. The probability distribution of $P(\xi_\infty = k)$, k = 0, ±1, ±2, … ±20 and η = 0.2, 0.6, 1.0, 1.5, 2.0, 2.5, 3.0. 109

Figure 3.2. Time-series plots. Annual precipitation in mm per day for the years 1901-2000 for the three bands: (a) NL, (b) Low, (c) SL. 112

Figure 3.3. Annual mean temperature anomalies in .01° Celsius for the years 1919-1978. Anomaly values indicate the deviations from the corresponding 1951-1980 means. 127

Figure 3.4. Plot of $-2\log\Lambda_t$ of the annual mean temperature anomalies data. 127

Figure 3.5. Autocorrelation for the standardized observations. 128

Figure 3.6. Partial Autocorrelation for the standardized observations. 128

Figure 3.7. Annual mean temperature anomalies data check for normality. Both p-values (0.493 and 0.326) are large and indicate that we do not have sufficient evidence to reject the null hypothesis that the data are normal. 129

Figure 3.8 (a-e). Theoretical distribution of the mle (under normal distribution) and the simulated distributions of the change-point mle and cmle under normal, t_v, and χ_v^2 distributions when n = 100 and τ = 50 and for η=2.5. 142

Figure 3.9 (a-e). Theoretical distribution of the mle (under normal distribution) and the simulated distributions of the change-point mle and cmle under normal, t_v, and χ_v^2 distributions when n = 100 and τ = 50 and for η=1.0. 145

Figure A.4.1. Time series plot of intervals between failures of equipment. 181

Figure A.4.2. Graph of $-2\log\Lambda_{\mathrm{k}}$, you can see the highest peak at 4. 182

Figure A.4.3. Test for exponentiality. Anderson-Darling statistic = 0.257, p-value = 0.879. 182

This dissertation is dedicated to my parents and grandfather.

CHAPTER 1

INTRODUCTION TO CHANGE-POINT ANALYSIS

Page introduced the classical change-point problem in 1955 where he put forward the following question: if we have an independent time-ordered sequence of values, where the first *m* observations may have come from one distribution $F(x|\theta)$ and the remaining observation from another distribution $F(x|\theta')$, with $\theta \neq \theta'$, how would we detect and test for such a hypothesized change when *m* is unknown. Page (1955) suggested using cumulative sums to detect change, and gave a one sided test for change in the mean. In the subsequent years many approaches have been developed to try and solve this problem, and much has been written on the topic of change-point analysis.

Change-point analysis consists of two main parts: first we wish to detect and estimate the change-point in a sequence of random variables or in a regression model; second we wish to determine the distribution of the change-point estimate and compute a confidence region for the true-change point. The underlying distributions for a sequence of random variables generally considered include, most frequently the normal distribution, and somewhat less frequently the exponential, gamma, binomial, and Poisson; while the methods commonly used in analysis are the likelihood ratio, Bayesian, nonparametric, and information criteria. The bulk of literature seems to focus on estimation of change-points and hypothesis tests for change in univariate and multivariate sequences of random variables, and in a few cases, testing for multiple changes in a sequence of random variables. For a comprehensive overview of change-point detection and estimation see Csörgő & Horváth (1997) and Chen & Gupta (2000), in addition, some literature examples are given below.

Nonparameteric procedures for estimation and testing

Bhattacharya & Frierson (1981) came up with a nonparametric procedure to detect a change in distribution based on asymptotic behavior of cumulative sums. Ferger (1995) considered nonparametric tests for a change in distribution. Ritov (1990) took a sequence of random variables with unknown underlying distributions and found an asymptotic efficient estimation of the change-point. Carlstein's (1988) nonparametric estimate of the change-point was based on maximizing the distance between the empirical distribution functions of two time series sets of observations. Müller & Wang (1990) gave us a nonparametric estimator of a change-point in a hazard rate.

Bayesian procedures for estimation and testing

Raftery & Akman (1986) considered the Bayesian approach to estimate and test for a change in a Poisson process. Smith (1975) also took the Bayesian approach, in his case to estimate and test for change in the binomial and normal distributions. Barry & Hartigan (1993) took the Bayesian approach to detect sharp short-lived changes in parameters, while Jain (2001) used the Bayesian approach to estimate change-points in Erlang distributions. Chernoff & Zacks

(1964) found the Bayesian estimator for a change in the mean in a sequence of normal random variables.

Non-Bayesian procedures for estimation and testing

Matthews & Farewell (1982) applied the likelihood ratio test for a single change-point in the hazard rate and the log gamma family. Loader (1991) also used the likelihood ratio test to test for a change in hazard rate and in addition gave the approximate confidence regions and joint confidence regions for the location and size of change. Karasoy & Kadilar (2007) found an efficient estimate for a change in the rate of a hazard function. Akman & Raftery (1986) considered a non-Bayesian approach to testing for a change in the rate of a Poisson process in the special case where the change-point to sample size ratio stays constant as both approach infinity. Worsley (1986) considered a maximum likelihood ratio test for a change in independent exponential family of random variables, most notably the exponential distribution, and derived the null and alternative distributions of the test statistic, while Ramanayake & Gupta (2003) also computed a likelihood ratio test statistic for a change in the mean in a sequence of independent exponential random variables. Siegmund (1988) computed joint confidence sets for the change-point and exponential family parameters, whereas Bhattacharya (1987) found estimates for the change-points in multi-parameter exponential families. Lombard (1987) tested for one or more change-points in a series of independent random variables and derived test statistics and their asymptotic null distributions. Hsu (1979) worked on detecting changes in scale parameters in independent time ordered gamma random variables and obtained the distribution of the tests statistic. Gombay & Horvath (1990) also found the asymptotic distribution of their maximum likelihood test statistic when testing for change in the mean. Jandhyala et al. (2002) tested for change in the variance of individual observations and derived the null distribution of the likelihood ratio statistic. In addition, Kim & Siegmund (1989) found a likelihood ratio test to detect changes in linear regression.

A lot less work has been done do derive the asymptotic distributions of the change-point estimates. Hinkley (1970, 1971, 1972) was the first one to focus attention on the likelihood ratio, non-Bayesian, approach to estimating the unknown change-point and then obtaining the asymptotic distribution of the maximum likelihood estimate. In his work he considered the mle of a change-point when an abrupt change occurs in the distribution of a sequence of independent random variables. Unfortunately the form he gives for the asymptotic distribution is computationally intractable. Subsequently, multiple authors provided some sort of approximations or bounds for the distribution of the change point mle. For example, Jandhyala & Fotopoulos (1999) computed bounds and suggested two types of easily computed approximations for the distribution. Later Fotopoulos & Jandhyala (2001) found a form of the asymptotic distribution of the change-point mle, but again that expression was not in a computable form. Hu & Rukhin (1995) provided a lower bound for the distribution of the mle over or under estimating the true change point, while Borovkov (1999) provided both upper and lower bounds for the distribution of the mle. Müller and Wang (1990) gave the limiting distribution of nonparametric estimators of change-points in the hazard rate. Barry & Hartigan

(1993) also considered distributions of the change-points and parameters, while Worsley (1986) gave exact and approximate confidence regions for change-points. Cobb (1978) derived the conditional distribution of the maximum likelihood estimate of a change-point, his method could be applied to changes in the mean or variance for various distributions. Fotopoulos et al. (2009) looked at deriving the asymptotic distribution of the mle of a change point under contiguity.

Further, the change-point problems can be split into several groups, first we consider the abrupt versus smooth change problems. In the case of the abrupt change, the change (e.g., in the mean) is assumed to happen as a sudden shift from one distribution to another or one parameter value to another, it does not depend on sample size, this is the case that the majority of literature focuses on, e.g., Hinkley (1970), Worsley (1986), and Kim & Siegmund (1989). In a smooth change case the change-point is related to the sample size. For example, Lombard (1987) defines the smooth change model as follows

$$\theta_i = \begin{cases} \xi_1, & i \geq \tau_1 \\ \xi_1 + (i - \tau_1)(\xi_2 - \xi_1)/(\tau_2 - \tau_1), & \tau_1 < i \leq \tau_2 \\ \xi_2, & i > \tau_2 \end{cases}$$

and the abrupt change as

$$\theta_i = \begin{cases} \xi_1, & 1 \leq i \leq \tau \\ \xi_2, & \tau < i \leq n \end{cases}$$

where τ is the unknown change-point. While Fotopoulos et al. (2009) define a smooth change problem as one where the amount of change is a function of the sample size, such that as the sample size goes to infinity the amount of change goes to zero in a "smooth fashion".

In addition, there are two types of data we may consider for the change-point problem, online (or sequential) and fixed sample size off-line (or non-sequential) data. In a sequential change-point problem we are trying to identify a change (perhaps in mean or variance) and make a stopping rule. Sequential change-point literature includes Siegmund (1975) and Siegmund & Venkatraman (1995) who consider sequential detection of a change-point using likelihood ratio tests. Applications include statistical process control, see Siegmund (1975, 1995), and medicine, for example detection of change in estrogen levels, see Carter & Blight (1981).

When analyzing online (or sequential) data for possible changes we of course must also consider control charts which are commonly used in statistical process control to detect nonrandom variation in a process. Shewhart control charts and the cumulative sums (CUSUM) charts are examples of charts that are frequently used to detect changes in a parameter of a process. One concern with control charts is that the parameter values used in the formulas must be known, when they are not known we must estimate the parameters under the assumption that the process, upon which we base our estimates, was stable. In the case where the assumption of stability does not hold sequential change-point analysis could be used for better results.

In general, change-point analysis works well with offline data and larger historical datasets; it does not require that we know the true parameter values of the underlying distributions thus it can be a good complement to control charting. Control charts, on the other hand, are constantly updated and will detect abnormal observations and large changes fast, but they may miss small changes or take a while to detect them. Change-point analysis, applied to previously collected data, would, in some cases, be able to detect the smaller changes that were missed. In addition, it would provide more information about the location and number of the change-points than a control chart.

Finally, there are many application for change-point analysis in areas of transportation, science, medicine, agriculture, climatology, traffic control, inventory and production processes, telecommunications, queueing and statistical quality control. One example of application of (multiple) change-point methodology is in DNA sequencing, see Brown & Muller (1998). Change-point methodology can also be used for studying cancer therapies and their effectiveness, for example to determine if there was a change in the patients' relapse rates. Detecting changes in the interarrival times of customers would be quite useful for planning services or making staffing decisions, see Jain (2001).

CHAPTER 2

ASYMPTOTIC DISTRIBUTION OF THE MAXIMUM LIKELIHOOD ESTIMATE OF THE CHANGE-POINT WHEN A CHANGE OCCURS IN THE RATE PARAMETER OF AN EXPONENTIAL PROCESS

In this chapter we investigate an abrupt change-point problem with the change-point occurring in the rate of an exponential process. An exponential model is quite appropriate for use in many sciences, such as management, engineering, biology, climatology, and seismology. For example, assuming independence and exponentiality, we may wish to analyze intervals between successive failures of equipment (e.g., to see for if the times shorten towards life end), disease outbreaks (e.g., to see if preventative measures worked), weather events (e.g., for signs of climate change), or earthquakes in a region (e.g., to find evidence of dynamic triggering). We may also examine interarrival times of patients in a hospital, or interarrival times of customers in a calling center, as well as service times in that calling center, and determine if there is a change in the underlying exponential process as well as calculate a confidence interval for the unknown true change-point if a change is detected. Being able to conclude that there was a change in a data sequence and to specify a confidence interval for that change-point would greatly enhance the study of the above mentioned phenomena.

Thus we would like to make inferences about changes that occur in exponential time-ordered datasets and make available more statistical tools for the sciences and other areas. In order to do that we need to first test for change and then calculate the confidence interval for the change-point based on the asymptotic distribution of the change-point mle. Much of literature has focused on testing for change, for example Csörgö & Horváth (1997) and Chen & Gupta (2000) among many others provide robust parametric methodology for change-point tests of hypotheses based on maximum likelihood. But comparatively little literature has focused on exponential distribution and on the asymptotic distribution of the mle. Indeed, no published work yet gives the exact computable expression for such a distribution.

Multiple authors have provided some sort of approximations or bounds for the distribution of the change point mle. For example, Jandhyala & Fotopoulos (1999) computed bounds and suggested two types of easily computed approximations for the distribution. Later Fotopoulos & Jandhyala (2001) found a form of the asymptotic distribution of the change-point mle, but that expression was not in a computable form. Hu & Rukhin (1995) provided a lower bound for the distribution of the mle over or under estimating the true change point, while Borovkov (1999) provided both upper and lower bounds for the distribution of the mle.

Many other authors have studied various aspects of the change-point problem in an exponential or Poisson model. For example, Worsley (1986) used the maximum likelihood ratio method to test for change in an independent exponential sequence, and to compute an estimate of the unknown change-point and the confidence regions for the true change-point. He also considered the distributions of the test statistics used to test for change. Haccou, Meelis, & Geer (1988) also considered the likelihood ratio test and the asymptotic null-distribution of the test statistic. Haccou & Meelis (1988) tested for multiple change-points; while Ramanayake & Gupta (2003) looked at testing for epidemic change in an exponential model utilizing the likelihood ratio approach. Akman & Raftery (1986) investigated procedures to test for change in a Poisson process and to determine point and interval estimates of an unknown change-point also in a Poisson process. Raftery & Akman (1986) took the Bayesian approach to change-point inference for a Poisson process.

The rest of the chapter is organized as follows: in section 2.1 we will look at general results for the asymptotic distribution of the mle of the unknown change-point, then in section 2.2 we will focus specifically on the exponential distribution and derive the needed expressions from section 2.1 in order to later present, in section 2.3, the exact computable expressions for the asymptotic distribution of the mle of the unknown change-point when a change occurs abruptly in the rate of a sequence of independent exponential random variables. Then in section 2.4 we will present an application of the distribution theory previously derived by analyzing earthquake data from the Sumatra, West Indonesia region. Section 2.5 will conclude with a simulation study that considers the accuracy of the asymptotic distribution previously derived.

2.1 General results for the distribution of the maximum likelihood estimate of the change-point τ.

First, let us consider a time-ordered sequence of independent, real-valued random variables, $Y_1, Y_2, \dots, Y_n,\ n \geq 1$, defined on $(\Omega, \mathcal{F}, P)$. We assume that there exists a change-point $\tau_n \in \{1, 2, \dots, n-1\}$ such that the sequence $Y_1, Y_2, \dots, Y_{\tau_n}$, has a common probability density function f_1 and the sequence $Y_{\tau_n+1}, Y_{\tau_n+2}, \dots, Y_n$, has a common probability density function f_2, where $f_1 \neq f_2$. The true change-point τ_n is unknown and we wish to estimate it.

2.1.1 Expressions for the maximum likelihood estimate

We use the method from Hinkley (1970) to obtain following maximum likelihood estimate (mle) $\hat{\tau}_n$, based on a sample of n observations.

The likelihood function, if $\tau_n = t$, then is as follows

$$L(t) = \prod_{i=1}^{t} f_1(Y_i) \prod_{i=t+1}^{n} f_2(Y_i) = \prod_{i=1}^{t} \frac{f_1}{f_2}(Y_i) \prod_{i=1}^{n} f_2(Y_i) \qquad (2.1.1)$$

And the log likelihood function then is expressed as

$$\mathcal{L}(t) = \log L(t) = \sum_{i=1}^{t} \log f_1(Y_i) + \sum_{i=t+1}^{n} \log f_2(Y_i) \qquad (2.1.2)$$

Let $w_i = \log f_1(Y_i) - \log f_2(Y_i)$, then by adding and subtracting $\sum_{i=1}^{t} \log f_2(Y_i)$ the log likelihood function can be rewritten as

$$\mathcal{L}(t) = \sum_{i=1}^{t} w_i + \sum_{i=1}^{n} \log f_2(Y_i) \qquad (2.1.3)$$

where $t = 1, 2, \dots, n-1$.

The mle $\hat{\tau}_n$ is then equal to t that will maximize equation (2.1.3), and is expressed as

$$\hat{\tau}_n = arg \max_{1 \le t \le n-1} \sum_{i=1}^{t} \log \frac{f_1}{f_2}(Y_i) \tag{2.1.4}$$

To determine the distribution of the change-point estimate $\hat{\tau}_n$ it is more convenient to work with a centered mle $\hat{\tau}_n - \tau_n = k$. Thus we assume that the mle $\hat{\tau}_n$ occurs at $\hat{\tau}_n = \tau_n + k$, where k is located above or below the true change-point τ_n. Note that the mle can then be rewritten to reflect this as

$$\hat{\tau}_n - \tau_n = arg \max_{1-\tau_n \le k \le n-1-\tau_n} \sum_{i=1}^{\tau_n + k} \log \frac{f_1}{f_2}(Y_i) \tag{2.1.5}$$

2.1.2 Finite sample distribution for ξ_n

Now let us consider a time-ordered sequence of independent, real-valued random variables, $Y_1, Y_2, \dots, Y_n,\ n \ge 1,$ defined on $(\Omega, \mathcal{F}, P)$. We assume that there exists a change-point $\tau_n \in \{2, 3, \dots, n\}$ such that the sequence $Y_1, Y_2, \dots, Y_{\tau_n - 1}$, has a common probability density function f_1 and the sequence $Y_{\tau_n}, Y_{\tau_n+1}, \dots, Y_n$, has a common probability density function f_2, where $f_1 \ne f_2$ are densities of $F_1 \ne F_2$ with respect to some dominating measure $\mu(F_1, F_2 \ll \mu)$.

The true change-point τ_n is unknown and we wish to estimate it. Note that we have redefined what we call the true change-point τ_n, it is now the index of the first random variable with density function f_2 rather than the index of the last random variable with density f_1, therefore equation (2.1.4) will, in this case, estimate $\tau_n - 1$. And (2.1.1) is now $L(t) = \prod_{i=1}^{t-1} f_1(Y_i) \prod_{i=t}^{n} f_2(Y_i)$.

Let $t \in \{2, 3, \ldots, n\}$ and $\hat{\tau}_n - \tau_n = k \in \{2 - \tau_n, 3 - \tau_n, \ldots, n - \tau_n\}$. Then we define the following

$$R(t) \coloneqq \prod_{i=1}^{t-1} \frac{f_1}{f_2}(Y_i) \qquad (2.1.6)$$

$$Z(t) \coloneqq \log R(t) \qquad (2.1.7)$$

Let us assume that $k > 0$, and $\hat{\tau}_n < n$ is an "overestimate" of the true-change point τ_n. Then we can view the event $\hat{\tau}_n - \tau_n = k$ as follows

$$\{\hat{\tau}_n - \tau_n = k\colon\ k = 1, \ldots, n - \tau_n\}$$

$$= \Big\{R(\tau_n + k) > R(j), for\ j$$

$$= 1, \ldots, \tau_n + k - 1 \bigwedge R(\tau_n + k) \geq R(j), for\ j = \tau_n + k, \ldots, n\Big\} \quad (2.1.8)$$

$$= \Big\{Z(\tau_n + k) > Z(j), for\ j$$

$$= 1, \ldots, \tau_n + k - 1 \bigwedge Z(\tau_n + k) \geq Z(j), for\ j = \tau_n + k, \ldots, n\Big\} \quad (2.1.9)$$

Let

$$W_i \coloneqq \log \frac{f_1}{f_2}(Y_i) \qquad (2.1.10)$$

and define random variables X^* and X as follows

$$X_i^* \coloneqq W_{\tau_n + i}\ \ if\ \ i = 1, 2, \ldots, -\tau_n + n \qquad (2.1.11)$$

$$X_i \coloneqq -W_i \ \ if \ \ i = 1, 2, \dots, \tau_n - 1 \tag{2.1.12}$$

Note that X^* and X are real valued random variables with negative expected values.

$$E(X) = -\int \log \frac{f_1}{f_2}(x)\, f_1(x)\mu(dx) = -K(f_1, f_2) < 0 \tag{2.1.13}$$

$$E(X^*) = \int \log \frac{f_1}{f_2}(x)\, f_2(x)\mu(dx) = -K(f_2, f_1) < 0 \tag{2.1.14}$$

where K is the Kullback-Leibler information number, recall that $F_1 \neq F_2$.

Then the partial sums of the random variables X^* and X defined as

$$S_j = \sum_{i=1}^{j} X_i \quad and \quad S_j^* = \sum_{i=1}^{j} X_i^*, \quad S_0 = S_0^* = 0 \tag{2.1.15}$$

are two, independent of each other, random walks.

Now let us define the strict ascending and descending ladder epochs for the two random walks, borrowing from Asmussen (1987).

For the random walk S

$$T_1^- = min\{j \geq 1 : S_j < 0\}, \qquad T_1^+ = min\{j \geq 1 : S_j > 0\} \tag{2.1.16}$$

where T_1^- is the first strict descending ladder epoch, or the first time the random walk S enters $(-\infty, 0)$, and T_1^+ is the first strict ascending ladder epoch, or the first time the random walk S enters $(0, \infty)$. And similarly for the random walk S^*

$$T_1^{*-} = min\{j \geq 1: S_j^* < 0\}, \qquad T_1^{*+} = min\{j \geq 1\mathbb{N}: S_j^* > 0\} \tag{2.1.17}$$

In addition, let us define the finite and infinite horizon maxima, M_j is the maximum of the first j partial sums, and M_∞ is the total maximum, where $M_j \uparrow M_\infty \; as \; j \to \infty$.

$$M_j = \max_{0 \leq i \leq j} S_i \;\; and \;\; M_j^* = \max_{0 \leq i \leq j} S_i^* \tag{2.1.18}$$

$$M_\infty = \max_{0 \leq i < \infty} S_i \;\; and \;\; M_\infty^* = \max_{0 \leq i < \infty} S_i^* \tag{2.1.19}$$

We know both M_∞ and M_∞^* exist due to (2.1.13) and (2.1.14). Then for k > 0, and defining the random variable $\xi_n = \hat{\tau}_n - \tau_n$, we get the following probability using the above mentioned definitions for $T_1^{*\pm}$, $S_j^*$and M_j, and equation (2.1.9)

$$P(\xi_n = k) = P\left(Z(\tau_n + k) > \bigvee_{j=1}^{\tau_n + k - 1} Z(j) \, , Z(\tau_n + k) \geq \bigvee_{j=\tau_n + k}^{n} Z(j) \right)$$

$$= P\left(S_k^* > \bigvee_{j=1}^{\tau_n - 1} S_j \, , \bigwedge_{j=1}^{k-1} S_j^* > 0 \right) P\left(\bigvee_{j=1}^{n - \tau_n - k} S_j^* \leq 0 \right)$$

$$= P\left(T_1^{*-} > k, S_k^* > M_{\tau_n - 1}\right) P(T_1^{*+} \geq n - \tau_n - k) \tag{2.1.20}$$

Note that

$$P(\xi_n = k) \xrightarrow{(n-\tau_n) \; and \; \tau_n \uparrow \infty} P(T_1^{*-} > k, S_k^* > M_\infty) \, P(T_1^{*+} = \infty) \equiv P(\xi_\infty = k) \tag{2.1.21}$$

And similarly to (2.1.20) but for $k < 0$, i.e., $(-k \in \{2 - \tau_n, \ldots, -1\} \; and \; \tau_n > 2)$ we have

$$P(\xi_n = -k) = P\left(Z(\tau_n - k - 1) > \bigvee_{j=1}^{\tau_n - k - 1} Z(j)\,, Z(\tau_n - k - 1) \geq \bigvee_{j=1}^{\tau_n - k - 1} Z(j) \right)$$

$$= P\left(S_k > \bigvee_{j=1}^{n-\tau_n} S_j^*\,, \bigwedge_{j=1}^{k-1} S_j > 0 \right) P\left(\bigvee_{j=1}^{\tau_n - k - 1} S_j \leq 0 \right)$$

$$= P\big(T_1^- > k, S_k > M_{n-\tau_n}^*\big) P(T_1^+ \geq \tau_n - k) \qquad (2.1.22)$$

Note that

$$P(\xi_n = -k) \xrightarrow{(n-\tau_n)\ and\ \tau_n\ \to \infty} P(T_1^- > k, S_k > M_\infty^*) P(T_1^+ = \infty) \equiv P(\xi_\infty = -k) \quad (2.1.23)$$

And when $k = 0$, we have the standard expression

$$P(\xi_n = 0) = P\big(M_{\tau_n - 1} = 0\big) P\big(M_{n-\tau_n}^* = 0\big) \qquad (2.1.24)$$

Note that

$$P(\xi_n = 0) \xrightarrow{(n-\tau_n)\ and\ \tau_n\ \to \infty} P(M_\infty = 0) P(M_\infty^* = 0) \equiv P(\xi_\infty = 0) \qquad (2.1.25)$$

Then we can write the finite sample distribution of ξ_n as follows

$$P(\xi_n = k) = \begin{cases} P\big(T_1^{*-} > k, S_k^* > M_{\tau_n - 1}\big) P(T_1^{*+} \geq n - \tau_n - k), & k = 1, 2, \ldots \\ P\big(M_{\tau_n - 1} = 0\big) P\big(M_{n-\tau_n}^* = 0\big), & k = 0 \\ P\big(T_1^- > k, S_k \geq M_{n-\tau_n}^*\big) P(T_1^+ \geq \tau_n - k), & k = -1, -2, \ldots \end{cases} \qquad (2.1.26)$$

We can see that the finite sample distribution of ξ_n , (2.1.26) above, requires the unknown change-point τ_n, thus making is impossible to compute. Therefore we may choose to work with the asymptotic distribution of ξ_n by letting both $(n - \tau_n)\ and\ \tau_n\ \to \infty$, see (2.1.21), (2.1.23), and (2.1.21). Therefore the probability distribution of ξ_∞ , the limiting random variable, can be written as follows

$$P(\xi_\infty = k) = \begin{cases} P(T_1^{*-} > k, S_k^* > M_\infty)P(T_1^{*+} = \infty), & k = 1,2,\dots \\ P(M_\infty = 0)P(M_\infty^* = 0), & k = 0 \\ P\left(T_1^- > |k|, S_{|k|} > M_\infty^*\right)P(T_1^+ = \infty), & k = -1,-2,\dots \end{cases} \quad (2.1.27)$$

Note that ξ_∞ is a proper random variable since both M_∞ and M_∞^* exist and in view of (2.1.27) and (2.1.39), see Section 2.1.3.

2.1.3 Distribution of M_∞ and M_∞^*, general case

In Asmussen (1987) G_+ is defined as the strict ascending ladder height distribution which may be defective.

$$G_+(x) = P\left(T_1^+ < \infty, S_{T_1^+} \le x\right) \quad (2.1.28)$$

$$\|G_+\| = P(T_1^+ < \infty) < 1 \quad (2.1.29)$$

Note that $\|G_+\|$, in our case, is always < 1. Then it follows that

$$P(T_1^+ = \infty) = 1 - P(T_1^+ < \infty)$$

$$= 1 - \|G_+\| = P(M_\infty = 0) \quad (2.1.30)$$

Similarly

$$\|G_+^*\| = P(T_1^{*+} < \infty) < 1 \quad (2.1.31)$$

and

$$P(T_1^{*+} = \infty) = 1 - P(T_1^{*+} < \infty\ )$$

$$= 1 - \ \|G_+^*\| = P(M_\infty^* = 0) \qquad (2.1.32)$$

Note that $E(T_1^-) = (1 - \|G_+\|)^{-1}$ and $E(T_1^{*-}) = (1 - \|G_+^*\|)^{-1}$.

Recall that $M_\infty = \max_{0 \le j < \infty} S_j$ and T_1^- *and* T_1^+ are the first ascending and descending ladder epochs, see (2.1.16) for definitions. The distribution of M_∞ can then be defined as follows

$$P(M_\infty \le x) = P(T_1^+ = \infty) \sum_{j=0}^{\infty} P\big(T_1^- > j, S_j \le x\big)$$

$$= (1 - \|G_+\|) \sum_{j=0}^{\infty} P\big(T_1^- > j, S_j \in [0, x]\big)$$

$$= (1 - \|G_+\|) \left(1 + \sum_{j=1}^{\infty} P\big(T_1^- > j, S_j \in [0, x]\big) \right)$$

$$= (1 - \|G_+\|) \sum_{j=0}^{\infty} P\left(T_j^+ < \infty, S_{T_j^+} \in [0, x]\right)$$

$$= (1 - \|G_+\|) \left(1 + \sum_{j=1}^{\infty} P\left(T_j^+ < \infty, S_{T_j^+} \le x\right) \right) \qquad (2.1.33)$$

where

$$T_j^+ = min\{j > T_{j-1}^+ : S_j > 0\} \qquad (2.1.34)$$

Similar results apply for $M_\infty^* = \max_{0 \le j < \infty} S_j^*$:

$$P(M_\infty^* \le x) = P(T_1^{*+} = \infty) \sum_{j=0}^{\infty} P\big(T_1^{*-} > j, S_j^* \le x\big)$$

$$= (1 - \ \|G_+^*\|) \sum_{j=0}^{\infty} P\big(T_1^{*-} > j, S_j^* \in [0, x]\big)$$

$$= (1 - \ \|G_+^*\|) \left(1 + \sum_{j=1}^{\infty} P\big(T_1^{*-} > j, S_j^* \in [0, x]\big) \right)$$

$$= (1 - \ \|G_+^*\|) \sum_{j=0}^{\infty} P\left(T_j^{*+} < \infty, S_{T_j^{*+}}^* \in [0, x]\right)$$

$$= (1 - \ \|G_+^*\|) \left(1 + \sum_{j=1}^{\infty} P\left(T_j^{*+} < \infty, S_{T_j^{*+}}^* \le x\right) \right) \qquad (2.1.35)$$

where

$$T_j^{*+} = min\{j > T_{j-1}^{*+} : S_j^* > 0\} \qquad (2.1.36)$$

The following show that ξ_∞ is a proper random variable, see (2.1.27)

$$\sum_{k=1}^{\infty} P(\xi_\infty = k) = \sum_{k=1}^{\infty} P(T_1^{*-} > k, S_k^* > M_\infty) P(T_1^{*+} = \infty)$$

$$= (1 - \|G_+^*\|) \sum_{j=k}^{\infty} P(T_1^{*-} > k, S_k^* > M_\infty)$$

$$= P(M_\infty^* > M_\infty, \ M_\infty^* > 0\ ) \qquad (2.1.37)$$

And similarly,

$$\sum_{k=1}^{\infty} P(\xi_{\infty} = -k) = \sum_{k=1}^{\infty} P(T_1^- > k, S_k > M_{\infty}^*) P(T_1^+ = \infty)$$

$$= (1 - \|G_+\|) \sum_{k=1}^{\infty} P(T_1^- > k, S_k > M_{\infty}^*)$$

$$= P(M_{\infty} > M_{\infty}^*,\ M_{\infty} > 0\ ) \qquad (2.1.38)$$

Thus,

$$\sum_{k=-\infty}^{\infty} P(\xi_{\infty} = k) = 1 \qquad (2.1.39)$$

$$= P(M_{\infty} = M_{\infty}^* = 0) + P(M_{\infty}^* > M_{\infty},\ M_{\infty}^* > 0) + P(M_{\infty} > M_{\infty}^*,\ M_{\infty} > 0\ \ ) = 1$$

From literature we have

$$P(\xi_{\infty} = 0) = P(M_{\infty} = M_{\infty}^* = 0)$$

$$= P(M_{\infty} = 0) P(M_{\infty}^* = 0)$$

$$= (1 - \|G_+\|)(1 - \|G_+^*\|) \qquad (2.1.40)$$

Define the following for $k > 0$

$$q_k = P(T_1^- > k),\ \ q_0 = 1 \qquad (2.1.41)$$

$$q_k^* = P(T_1^{*-} > k),\ \ q_0^* = 1 \qquad (2.1.42)$$

Then the probability distribution of ξ_{∞} from (2.1.27) can be rewritten as follows

$$P(\xi_\infty = k) = \begin{cases} (1 - \|G_+^*\|)\left(q_k^* - \int_{0^+}^{\infty} P(M_\infty > x)P(S_k^* \in dx, T_1^{*-} > k)\right), k > 0 \\ (1 - \|G_+^*\|)(1 - \|G_+\|), \qquad k = 0 \\ (1 - \|G_+\|)\left(q_{|k|} - \int_{0^+}^{\infty} P(M_\infty^* > x)P\left(S_{|k|} \in dx, T_1^- > |k|\right)\right), k < 0 \end{cases} \tag{2.1.43}$$

and used to develop procedures that would allow us to compute the exact probabilities for ξ_∞ under different underlying densities $f_1 \neq f_2$. See Fotopoulos (2007) and Jandhyala & Fotopoulos (1999).

2.2 The distribution of the maximum likelihood estimate of the change-point τ in the exponential case

Consider a time-ordered sequence of independent real valued random variables, $Y_1, Y_2, \ldots, Y_n$, $n \geq 1$, defined on $(\Omega, \mathcal{F}, P)$. We assume that there exists a change-point $\tau_n \in \{2, 3, \ldots, n\}$ such that the sequence $Y_1, Y_2, \ldots, Y_{\tau_n - 1}$, has a common probability density function f_1 and the sequence $Y_{\tau_n}, Y_{\tau_n + 1}, \ldots, Y_n$, has a common probability density function f_2, where $f_1 \neq f_2$. Let, without loss of generality, $v_1 > v_2$, and

$$f(y_i, v) = \begin{cases} f_1(y_i, v_1) = Exp(v_1) = v_1 e^{-v_1 y}, & i = 1, \ldots, \tau - 1; \; y_i \geq 0 \\ f_2(y_i, v_2) = Exp(v_2) = v_2 e^{-v_2 y}, & i = \tau, \ldots, n; \; y_i \geq 0 \end{cases} \tag{2.2.1}$$

To calculate the distribution of ξ_∞ , see (2.1.43) previous section, in the exponential case we will need to determine the exact expressions for: $\|G_+^*\|, \|G_+\|$, q_k^*, M_∞, M_∞^*, X, X^*, $P(S_k^* \in dx, T_1^{*-} > k) = u_k^*(dx),$ and $P(S_k \in dx, T_1^- > k) = u_k(dx)$.

2.2.1 Expressions for the X in the exponential case

Let, $v_1 > v_2$ and consider the random variable

$$X = -log\frac{f_1}{f_2}(Y) \qquad (2.2.2)$$

Then applying the exponential probability distribution functions $(2.2.1)$ we get

$$X = -log\left(\frac{v_1 e^{-v_1 Y}}{v_2 e^{-v_2 Y}}\right) = log\left(\frac{v_2 e^{-v_2 Y}}{v_1 e^{-v_1 Y}}\right) = log\left(\frac{v_2}{v_1}\right) + log(e^{-v_2 Y + v_1 Y})$$

$$= log\left(\frac{v_2}{v_1}\right) + (v_1 - v_2)Y \qquad (2.2.3a)$$

where $Y \sim Exp(v_1)$. Also note that we can rewrite the above equation $(2.2.3a)$ as

$$X =_{\mathcal{D}} log\left(\frac{v_2}{v_1}\right) + \frac{v_1 - v_2}{v_1} E \qquad (2.2.3b)$$

where $E \sim Exp(1)$.

Then for $x > 0$ we have the following distribution of X

$$P(X > x) = P\left(log\frac{v_2}{v_1} + (v_1 - v_2)Y > x\right)$$

$$= P\left(Y > \frac{x - log\frac{v_2}{v_1}}{(v_1 - v_2)}\right) = e^{-v_1\left[\frac{x - log\frac{v_2}{v_1}}{v_1 - v_2}\right]} = e^{\frac{-v_1}{v_1 - v_2}\left[x - log\frac{v_2}{v_1}\right]}$$

$$= e^{\frac{-v_1 x}{v_1 - v_2} + log\left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1 - v_2}}} = \left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1 - v_2}} e^{\frac{-v_1 x}{v_1 - v_2}}, \qquad x > log\frac{v_2}{v_1} \qquad (2.2.4)$$

and

$$F_X(x) = P(X \le x) = 1 - \left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1 - v_2}} e^{\frac{-v_1 x}{v_1 - v_2}}, \qquad x > log\frac{v_2}{v_1} \qquad (2.2.5)$$

Since $\left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1-v_2}} \in (0,1)$ due to $v_1 > v_2$ the pdf of X is defined by

$$f_X(x) = \frac{v_1}{v_1 - v_2}\left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1 - v_2}} e^{\frac{-v_1 x}{v_1 - v_2}} I\left(x > log\frac{v_2}{v_1}\right) \qquad (2.2.6)$$

Note that $Y \sim Exp(v_1)$ and $y \ge 0$, then from $P\left(Y > \frac{x - log\frac{v_2}{v_1}}{(v_1 - v_2)}\right)$ we have $\frac{x - log\frac{v_2}{v_1}}{(v_1 - v_2)} > 0$ and therefore we must have $x > log\frac{v_2}{v_1}$.

2.2.2 Determination of the distribution of the maximum M_∞ in the exponential case

Now we need to derive an expression for M_∞, see equation (2.1.33) and notice the following

$$P(M_\infty \le x) = (1 - \|G_+\|)\left(1 + \sum_{j=1}^{\infty} P\left(T_j^+ < \infty, S_{T_j^+} \le x\right)\right)$$

$$= (1 - \|G_+\|)\left(1 + \sum_{j=1}^{\infty} P\left(T_1^+ < \infty, S_{T_1^+} \le x\right)^{j*}\right) \qquad (2.2.7)$$

Where $f^{j*}(x)$ is the j-fold convolution of the function $f(x)$. Recall that $T_j^+ = min\{j > T_{j-1}^+ : S_j > 0\}$ and we have noticed in (2.1.33) and (2.2.7) that $P(T_1^- > j, S_j \le x) = P\left(T_j^+ < \infty, S_{T_j^+} \le x\right) = P\left(T_1^+ < \infty, S_{T_1^+} \le x\right)^{j*}$. Therefore in order to determine $P(M_\infty \le x)$ we must first derive an expression for $P\left(S_{T_1^+} \leq x, T_1^+ < \infty\right)$ and then the renewal density $U(dx)$. We start with

$$P\left(S_{T_1^+} > x,\ \ T_1^+ = n\right) = P\left(\bigvee_{j=0}^{n-1} S_j \leq 0\,, S_n > x\right)$$

$$= E\left[I\left(\bigvee_{j=0}^{n-1} S_j \leq 0\right)P(X_n > x - S_{n-1}|\mathcal{F}_{n-1})\right] =$$

Note that $P(X_n > x - S_{n-1}) = P(X_n + S_{n-1} > x) = P(S_n > x)$. From (2.2.5) we recall that $P(X > x) = \left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1 - v_2}} e^{\frac{-v_1 x}{v_1 - v_2}}$, hence $P(X_n > x - S_{n-1}|\mathcal{F}_{n-1}) = \left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1-v_2}} e^{\frac{-v_1(x - S_{n-1})}{v_1 - v_2}}$, and

$$= E\left[I\left(\bigvee_{j=0}^{n-1} S_j \leq 0\right)\left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1-v_2}} e^{\frac{-v_1(x-S_{n-1})}{v_1-v_2}}\right]$$

$$= e^{\frac{-v_1 x}{v_1 - v_2}} E\left[I\left(\bigvee_{j=0}^{n-1} S_j \leq 0\right)\left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1-v_2}} e^{\frac{v_1 S_{n-1}}{v_1-v_2}}\right] =$$

Note that $P(X_n > -S_{n-1}|\mathcal{F}_{n-1}) = \left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1-v_2}} e^{\frac{-v_1(-S_{n-1})}{v_1-v_2}}$, since $S_{n-1} \leq 0$, and

$$= e^{\frac{-v_1 x}{v_1 - v_2}} E\left[I\left(\bigvee_{j=0}^{n-1} S_j \leq 0\right)P(X_n > -S_{n-1}|\mathcal{F}_{n-1})\right]$$

$$= e^{\frac{-v_1 x}{v_1 - v_2}} P(T_1^+ = n)$$

So in summary we have, from the above calculations,

$$P\left(S_{T_1^+} > x,\ T_1^+ = n\right) = e^{\frac{-v_1 x}{v_1 - v_2}} P(T_1^+ = n) \tag{2.2.8}$$

Summing the equation $(2.2.8)$ above over $n \geq 1$, and applying equation $(2.1.29)$ we obtain

$$\sum_{n=1}^{\infty} P\left(S_{T_1^+} > x,\ T_1^+ = n\right) = P\left(S_{T_1^+} > x,\ T_1^+ < \infty\right)$$

$$= e^{\frac{-v_1 x}{v_1 - v_2}} P(T_1^+ < \infty) = e^{\frac{-v_1 x}{v_1 - v_2}} \|G_+\|$$

Then

$$P\left(S_{T_1^+} \leq x, T_1^+ < \infty\right) = 1 - e^{\frac{-v_1 x}{v_1 - v_2}} \|G_+\| \tag{2.2.9}$$

And

$$P\left(S_{T_1^+} \in dx, T_1^+ < \infty\right) = d\left(1 - e^{\frac{-v_1 x}{v_1 - v_2}} \|G_+\|\right)$$

$$= dx \frac{v_1}{v_1 - v_2} e^{\frac{-v_1 x}{v_1 - v_2}} \|G_+\| \tag{2.2.10}$$

The distribution of $P\left(S_{T_j^+} \leq x, T_j^+ < \infty\right)$ is the j-fold convolution of density $P\left(S_{T_1^+} \in dx, T_1^+ < \infty\right)$. See Appendix [A.1] for details.

$$P\left(S_{T_1^+} \in dx, T_1^+ < \infty\right)^{j*} = dx\left(\frac{v_1}{v_1 - v_2} e^{\frac{-v_1 x}{v_1 - v_2}} \|G_+\|\right)^{j*}$$

$$= dx \|G_+\|^j \left(\frac{v_1}{v_1 - v_2} e^{\frac{-v_1 x}{v_1 - v_2}}\right)^{j*} =$$

$$= dx\|G_+\|^j \frac{\frac{v_1}{v_1 - v_2} e^{\frac{-v_1 x}{v_1 - v_2}} \left(\frac{v_1 x}{v_1 - v_2}\right)^{j-1}}{(j-1)!}, \qquad x > 0 \qquad (2.2.11)$$

and therefore

$$P\left(S_{T_1^+} > x, T_1^+ < \infty\right)^{j*} = \int_x^\infty \|G_+\|^j \frac{\frac{v_1}{v_1 - v_2} e^{\frac{-v_1 t}{v_1 - v_2}} \left(\frac{v_1 t}{v_1 - v_2}\right)^{j-1}}{(j-1)!} dt$$

$$= \|G_+\|^j \int_x^\infty \frac{\left(\frac{v_1}{v_1 - v_2}\right)^j t^{j-1} e^{\frac{-v_1}{v_1 - v_2} t}}{(j-1)!} dt$$

$$= \|G_+\|^j \frac{\Gamma\left(j, \frac{v_1}{v_1 - v_2} x\right)}{\Gamma(j)} \qquad (2.2.12)$$

We sum the convolution in equation (2.2.11) over $j \geq 1$ to obtain the density of the renewal function

$$U(dx) = \sum_{j=1}^\infty P\left(S_{T_1^+} \in dx, T_1^+ < \infty\right)^{j*}$$

$$= dx \sum_{j=1}^\infty \|G_+\|^j \frac{\frac{v_1}{v_1 - v_2} e^{\frac{-v_1 x}{v_1 - v_2}} \left(\frac{v_1 x}{v_1 - v_2}\right)^{j-1}}{(j-1)!}$$

$$= dx\|G_+\| \frac{v_1}{v_1 - v_2} e^{\frac{-v_1 x}{v_1 - v_2}(1 - \|G_+\|)}, \qquad x > 0 \qquad (2.2.13)$$

Note that $U(\{0\}) = 1$, and from (2.2.13) we obtain

$$U\big((0,x]\big) = \sum_{j=1}^{\infty} P\big(S_{T_1^+} \in (0,x], T_1^+ < \infty\big)^{j*} =$$

Note that we can now rewrite (2.2.7) as $P(M_\infty \le x) = (1 - \|G_+\|)\big(1 + U((0,x])\big)$

$$= \int_0^x dt \|G_+\| \frac{v_1}{v_1 - v_2} e^{\frac{-v_1 t}{v_1 - v_2}(1-\|G_+\|)}$$

$$= \frac{\|G_+\|}{(1-\|G_+\|)} \int_0^x (1 - \|G_+\|) \frac{v_1}{v_1 - v_2} e^{\frac{-v_1}{v_1 - v_2}(1-\|G_+\|)t} dt$$

$$= \frac{\|G_+\|}{(1-\|G_+\|)} \left(1 - e^{\frac{-v_1}{v_1 - v_2}(1-\|G_+\|)x}\right) \qquad (2.2.14)$$

where

$$\|G_+\| = P\big(S_{T_1^+} < \infty, T_1^+ < \infty\big) \; = P(T_1^+ < \infty)$$

And now to determine the expression for $\|G_+\|$. Note that for $x \le 0$

$$P\big(S_{T_1^-} \le x, T_1^- < \infty\big) = \int_0^\infty F_X(x-y)U(dy) \qquad (2.2.15)$$

With $E(X) < 0$, and from Asmussen (1987) we have

$$1 = \; P\big(S_{T_1^-} \le 0, T_1^- < \infty\big) = P(T_1^- < \infty)$$

$$= \int_0^\infty F_X(0-y)U(dy) = F_X(0) + \int_{0^+}^\infty F_X(-y)U(dy) =$$

$$= 1 - \left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1 - v_2}} + \|G_+\| \int_{log\frac{v_2}{v_1}}^{0} dx \frac{v_1}{v_1 - v_2} e^{\frac{v_1 x}{v_1 - v_2}(1 - \|G_+\|)} \left\{1 - \left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1 - v_2}} e^{\frac{v_1 x}{v_1 - v_2}}\right\}$$

$$= 1 - \left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1 - v_2}} + \frac{\|G_+\|}{(1 - \|G_+\|)} \int_{\frac{v_1(1-\|G_+\|)}{v_1 - v_2} log\frac{v_2}{v_1}}^{0} e^x dx - \left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1 - v_2}} \int_{\frac{v_1(1-\|G_+\|)}{v_1 - v_2} log\frac{v_2}{v_1}}^{0} e^x dx$$

$$= 1 - \left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1 - v_2}} + \frac{\|G_+\|}{(1 - \|G_+\|)} \left(1 - \left(\frac{v_2}{v_1}\right)^{\frac{v_1(1-\|G_+\|)}{v_1 - v_2}}\right) + \left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1 - v_2}} \left(1 - \left(\frac{v_2}{v_1}\right)^{\frac{-v_1\|G_+\|}{v_1 - v_2}}\right)$$

$$= 1 - \left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1 - v_2}} + \frac{\|G_+\|}{(1 - \|G_+\|)} - \frac{\|G_+\|}{(1 - \|G_+\|)} \left(\frac{v_2}{v_1}\right)^{\frac{v_1(1-\|G_+\|)}{v_1 - v_2}} + \left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1 - v_2}} - \left(\frac{v_2}{v_1}\right)^{\frac{v_1(1-\|G_+\|)}{v_1 - v_2}}$$

$$= 1 + \frac{\|G_+\|}{(1 - \|G_+\|)} - \left(1 + \frac{\|G_+\|}{(1 - \|G_+\|)}\right) \left(\frac{v_2}{v_1}\right)^{\frac{v_1(1-\|G_+\|)}{v_1 - v_2}} \qquad (2.2.16)$$

i.e.,

$$0 = \frac{\|G_+\|}{(1 - \|G_+\|)} - \left(1 + \frac{\|G_+\|}{(1 - \|G_+\|)}\right) \left(\frac{v_2}{v_1}\right)^{\frac{v_1(1-\|G_+\|)}{v_1 - v_2}}$$

or

$$0 = \|G_+\| - \left(\frac{v_2}{v_1}\right)^{\frac{v_1(1-\|G_+\|)}{v_1 - v_2}}$$

And since $0 < \|G_+\| < 1$, we have

$$\|G_+\| = \frac{v_2}{v_1} \qquad (2.2.17)$$

Now by taking $(2.2.7)$ and applying $(2.2.14)$ and $(2.2.17)$ we have

$$P(M_\infty \leq x) = (1 - \|G_+\|)\big(1 + U((0,x])\big)$$

$$= \left(1 - \frac{v_2}{v_1}\right)\left(1 + \frac{\|G_+\|}{(1 - \|G_+\|)}\left(1 - e^{\frac{-v_1}{v_1 - v_2}(1-\|G_+\|)x}\right)\right)$$

$$= \frac{v_1 - v_2}{v_1}\left(1 + \left(\frac{v_2}{v_1}\right)\left(\frac{v_1}{v_1 - v_2}\right)\left(1 - e^{\frac{-v_1}{v_1 - v_2}\left(\frac{v_1 - v_2}{v_1}\right)x}\right)\right)$$

$$= \frac{v_1 - v_2}{v_1} + \frac{v_2}{v_1} - \frac{v_2}{v_1}e^{-x}$$

Therefore

$$P(M_\infty \leq x) = 1 - \frac{v_2}{v_1}e^{-x}, \quad x \geq 0 \qquad (2.2.18)$$

and

$$P(M_\infty > x) = 1 - P(M_\infty \leq x)$$

$$= \|G_+\| e^{\frac{-v_1 x}{v_1 - v_2}(1-\|G_+\|)}$$

And thus we have

$$P(M_\infty > x) = \frac{v_2}{v_1}e^{-x}, \qquad x \geq 0 \qquad (2.2.19)$$

2.2.3 Expressions for the X^* in the exponential case

Let, $v_1 > v_2$ and consider the random variable

$$X^* = log\frac{f_1}{f_2}(Y) \qquad (2.2.20)$$

Then filling in the exponential probability distribution functions (2.2.1) we get

$$X^* = log\left(\frac{v_1e^{-v_1Y}}{v_2e^{-v_2Y}}\right) = log\left(\frac{v_1}{v_2}\right) + log(e^{-v_1Y+v_2Y})$$

$$= -\log\frac{v_2}{v_1} - (v_1 - v_2)Y \qquad (2.2.21a)$$

where $Y \sim Exp(v_2)$. Also note that we can rewrite the above equation $(2.2.21a)$ as

$$X^* =_{\mathcal{D}} log\left(\frac{v_1}{v_2}\right) + \frac{v_1 - v_2}{v_2}E^* \qquad (2.2.21b)$$

where $E^* \sim Exp(1)$.

Then

$$P(X^* > x) = P\left(-\log\frac{v_2}{v_1} - (v_1 - v_2)Y > x\right)$$

$$= P\left(\log\frac{v_2}{v_1} + (v_1 - v_2)Y \le -x\right) = P\left(Y \le \frac{-\left(x + \log\frac{v_2}{v_1}\right)}{(v_1 - v_2)}\right)$$

$$= 1 - e^{\frac{v_2\left(x+\log\frac{v_2}{v_1}\right)}{v_1-v_2}} = 1 - e^{\left(\log\frac{v_2}{v_1}\right)^{\frac{v_2}{v_1-v_2}}} e^{\frac{v_2x}{v_1-v_2}}$$

$$= 1 - \left(\frac{v_2}{v_1}\right)^{\frac{v_2}{v_1-v_2}} e^{\frac{v_2x}{v_1-v_2}}, \quad x \le -\log\frac{v_2}{v_1} \qquad (2.2.22)$$

Then

$$F_{X^*}(x) = P(X^* \le x) = \left(\frac{v_2}{v_1}\right)^{\frac{v_2}{v_1-v_2}} e^{\frac{v_2 x}{v_1-v_2}}, \quad x \le -\log\frac{v_2}{v_1} \qquad (2.2.23)$$

And the density function of X^*

$$f_{X^*}(x) = \frac{v_2}{v_1-v_2}\left(\frac{v_2}{v_1}\right)^{\frac{v_2}{v_1-v_2}} e^{\frac{v_2 x}{v_1-v_2}} I\left(x \le -\log\frac{v_2}{v_1}\right) \qquad (2.2.24)$$

Note that the expected value of X^* is negative as shown below

$$E(X^*) = \int_{-\infty}^{\infty} x f_{X^*}(x)dx = \int_{-\infty}^{-\log\frac{v_2}{v_1}} x\frac{v_2}{v_1-v_2}\left(\frac{v_2}{v_1}\right)^{\frac{v_2}{v_1-v_2}} e^{\frac{v_2 x}{v_1-v_2}} dx$$

$$= \left(\frac{v_2}{v_1}\right)^{\frac{v_2}{v_1-v_2}} \int_{-\infty}^{-\log\frac{v_2}{v_1}} x\frac{v_2}{v_1-v_2} e^{\frac{v_2 x}{v_1-v_2}} dx =$$

Let $u = x\frac{v_2}{v_1-v_2}$ $\quad du = \frac{v_2}{v_1-v_2}dx \quad dx = \frac{v_1-v_2}{v_2}du$

$$= \left(\frac{v_2}{v_1}\right)^{\frac{v_2}{v_1-v_2}} \frac{v_1-v_2}{v_2} \int_{-\infty}^{-\log\frac{v_2}{v_1}\left(\frac{v_2}{v_1-v_2}\right)} ue^u du$$

$$= \left(\frac{v_2}{v_1}\right)^{\frac{v_2}{v_1-v_2}} \frac{v_1-v_2}{v_2} (ue^u - e^u)\Big|_{-\infty}^{-\log\frac{v_2}{v_1}\left(\frac{v_2}{v_1-v_2}\right)}$$

$$= \left(\frac{v_2}{v_1}\right)^{\frac{v_2}{v_1-v_2}} \frac{v_1-v_2}{v_2}\left\{-\log\frac{v_2}{v_1}\left(\frac{v_2}{v_1-v_2}\right)\left(\frac{v_2}{v_1}\right)^{\frac{v_2}{v_1-v_2}} - \left(\frac{v_2}{v_1}\right)^{\frac{v_2}{v_1-v_2}}\right\}$$

Thus we have

$$E(X^*) = \log\frac{v_2}{v_1} - \left(\frac{v_1}{v_2} - 1\right) < 0 \qquad (2.2.25)$$

It is easier to work with a reflection of X^* to derive the remaining expressions and then revert those expressions back to X^* at the end. Let $X^{(r)} = -X^*$, and as before let $v_1 > v_2$. Then note that the expected value of the reflection will be positive

$$E\left(X^{(r)}\right) = -E(X^*) = -\left(\log\frac{v_2}{v_1} - \left(\frac{v_1}{v_2} - 1\right)\right) = -\log\frac{v_2}{v_1} + \left(\frac{v_1}{v_2} - 1\right) > 0 \quad (2.2.26)$$

The distribution function of $X^{(r)}$ is given below

$$F_{X^{(r)}}(x) = P\left(X^{(r)} \le x\right) = 1 - \left(\frac{v_2}{v_1}\right)^{\frac{v_2}{v_1 - v_2}} e^{\frac{v_2 x}{v_1 - v_2}}, \quad x > \log\frac{v_2}{v_1} \qquad (2.2.27)$$

It then follows that

$$P\left(X^{(r)} > x\right) = \left(\frac{v_2}{v_1}\right)^{\frac{v_2}{v_1 - v_2}} e^{\frac{-v_2 x}{v_1 - v_2}}, \quad x > \log\frac{v_2}{v_1} \qquad (2.2.28)$$

We define $S_j^{(r)}$ in a similar manner to $(2.1.15)$, as a partial sum of X

$$S_j^{(r)} = \sum_{i=1}^{j} X_i, \quad S_0^{(r)} = 0 \qquad (2.2.29)$$

Then the following probabilities can be determined by applying similar arguments to $S^{(r)}$ as were applied previously to S, $n \ge 1$, see $(2.2.8)$

$$P\left(S^{(r)}_{T_1^{(r)+}} > x, T_1^{(r)+} = n\right) = e^{\frac{-v_2 x}{v_1 - v_2}} P\left(T_1^{(r)+} = n\right) \qquad (2.2.30)$$

Then we sum over $n \geq 1$ as before and get the following probability

$$\sum_{n=1}^{\infty} P\left(S_{T_1^{(r)+}}^{(r)} > x, T_1^{(r)+} = n\right) = P\left(S_{T_1^{(r)+}}^{(r)} > x, T_1^{(r)+} < \infty\right)$$

$$= e^{\frac{-v_2 x}{v_1 - v_2}} P\left(T_1^{(r)+} < \infty\right)$$

Since $E\left(X^{(r)}\right) > 0$, $S^{(r)}$ drifts to ∞, the random variable $T_1^{(r)+}$is proper, in that $P\left(T_1^{(r)+} < \infty\right) = 1$. Therefore the probability above reduces to just

$$= e^{\frac{-v_2 x}{v_1 - v_2}} \qquad (2.2.31)$$

It follows that

$$P\left(S_{T_1^{(r)+}}^{(r)} \in dx, T_1^{(r)+} < \infty\right) = d\left(1 - e^{\frac{-v_2 x}{v_1 - v_2}}\right)$$

$$= dx \frac{v_2}{v_1 - v_2} e^{\frac{-v_2 x}{v_1 - v_2}} \qquad (2.2.32)$$

We also note that

$$U^{(r)}((x, \infty)) = \sum_{j=1}^{\infty} P\left(S_{T_1^{(r)+}}^{(r)} > x, T_1^{(r)+} < \infty\right)^{j*} \qquad (2.2.33)$$

The j-fold convolution of density in (2.2.32) is as follows

$$P\left(S_{T_1^{(r)+}}^{(r)} \in dx, T_1^{(r)+} < \infty\right)^{j*} = dx\left(\frac{v_2}{v_1 - v_2} e^{\frac{-v_2 x}{v_1 - v_2}}\right)^{j*} =$$

$$= dx \frac{\frac{v_2}{v_1 - v_2} e^{\frac{-v_2 x}{v_1 - v_2}} \left(\frac{v_2 x}{v_1 - v_2}\right)^{j-1}}{(j-1)!}, \quad x > 0 \qquad (2.2.34)$$

Then the renewal density based on (2.2.34) above is given by

$$U^{(r)}(dx) \sum_{j=1}^{\infty} P\left(S^{(r)}_{T_1^{(r)+}} \in dx, T_1^{(r)+} < \infty\right)^{j*} = dx \sum_{j=1}^{\infty} \frac{\frac{v_2}{v_1 - v_2} e^{\frac{-v_2 x}{v_1 - v_2}} \left(\frac{v_2 x}{v_1 - v_2}\right)^{j-1}}{(j-1)!}$$

$$= dx \frac{v_2}{v_1 - v_2}, \quad x > 0 \qquad (2.2.35)$$

It follows that

$$U^{(r)}((0, x]) = \frac{v_2 x}{v_1 - v_2}, \quad x > 0 \qquad (2.2.36)$$

Note that $U^{(r)}(\{0\}) = 1$.

Now we need to find a computable expression for $\left\| G_+^{(r)} \right\|$ for $x \in (-\infty, 0]$. See Asmussen (1987), and recall that $E(X) < 0$. Then

$$G_-^{(r)} = U_+^{(r)} * F^{(r)}(x)$$

$$= P\left(S^{(r)}_{T_1^{(r)-}} \leq x, T_1^{(r)-} < \infty\right) \qquad (2.2.37)$$

Let $x = 0$

$$P\left(S^{(r)}_{T_1^{(r)-}} \leq 0, T_1^{(r)-} < \infty\right) = F^{(r)}(0) + \frac{v_2}{v_1 - v_2} \int_{-\infty}^{0} F^{(r)}(y) dy =$$

$$= 1 - \left(\frac{v_2}{v_1}\right)^{\frac{v_2}{v_1 - v_2}} + \frac{v_2}{v_1 - v_2} \int_{-\infty}^{0} dy \int_{-\infty}^{y} F^{(r)}(du)$$

$$= 1 - \left(\frac{v_2}{v_1}\right)^{\frac{v_2}{v_1 - v_2}} + \frac{v_2}{v_1 - v_2} \int_{-\infty}^{0} \int_{u}^{y} dy F^{(r)}(du)$$

$$= 1 - \left(\frac{v_2}{v_1}\right)^{\frac{v_2}{v_1 - v_2}} + \frac{v_2}{v_1 - v_2} E\left(X^{(r)-}\right) \qquad (2.2.38)$$

Note that from (2.2.26) we know the following

$$E\left(X^{(r)}\right) = E\left(X^{(r)+}\right) - E\left(X^{(r)-}\right) = -\log\frac{v_2}{v_1} + \left(\frac{v_1}{v_2} - 1\right) > 0$$

And

$$E\left(X^{(r)+}\right) = \int_{0}^{\infty} \left(\frac{v_2}{v_1}\right)^{\frac{v_2}{v_1 - v_2}} e^{\frac{-v_2 x}{v_1 - v_2}} dx$$

$$= -\frac{v_1 - v_2}{v_2} \left(\frac{v_2}{v_1}\right)^{\frac{v_2}{v_1 - v_2}} \qquad (2.2.39)$$

While

$$E\left(X^{(r)-}\right) = -\frac{v_1 - v_2}{v_2} \left(\frac{v_2}{v_1}\right)^{\frac{v_2}{v_1 - v_2}} - \left(-\log\frac{v_2}{v_1} + \left(\frac{v_1}{v_2} - 1\right)\right)$$

$$= -\frac{v_1 - v_2}{v_2} \left(\frac{v_2}{v_1}\right)^{\frac{v_2}{v_1 - v_2}} + \log\frac{v_2}{v_1} - \frac{v_1 - v_2}{v_2} \qquad (2.2.40)$$

Thus, repeating the process that resulted in equation (2.2.38) and applying the equation (2.2.40) above we have

$$P\left(S^{(r)}_{T_1^{(r)-}} \le 0, T_1^{(r)-} < \infty\right) = P\left(T_1^{(r)-} < \infty\right) = \int_0^\infty F_{X^{(r)}}(-y)U^{(r)}(dy)$$

$$= F_{X^{(r)}}(0) + \frac{v_2}{v_1 - v_2}\int_{0^+}^\infty F_{X^{(r)}}(-y)dy$$

$$= 1 - \left(\frac{v_2}{v_1}\right)^{\frac{v_2}{v_1 - v_2}} + \frac{v_2}{v_1 - v_2}E\left(X^{(r)-}\right)$$

$$= 1 - \left(\frac{v_2}{v_1}\right)^{\frac{v_2}{v_1 - v_2}} + \frac{v_2}{v_1 - v_2}\left\{-\frac{v_1 - v_2}{v_2}\left(\frac{v_2}{v_1}\right)^{\frac{v_2}{v_1 - v_2}} + \log\frac{v_2}{v_1} - \frac{v_1 - v_2}{v_2}\right\}$$

$$= -\frac{v_2}{v_1 - v_2}\log\frac{v_2}{v_1} \in (0,1) \qquad (2.2.41)$$

Now we switch back to working with X^*, and utilize the expressions we derived above. For $x > 0$,

$$P\left(S^{(r)}_{T_1^{(r)-}} \le -x, T_1^{(r)-} < \infty\right) = \sum_{n=1}^\infty P\left(\bigwedge_{j<n}(-S_j^*) > 0\,, -S_n^* \le x\right)$$

$$= \sum_{n=1}^\infty P\left(-\bigwedge_{j<n}(-S_j^*) \le 0\,, S_n^* \ge x\right) =$$

$$= \sum_{n=1}^{\infty} P\left(\bigvee_{j<n} S_j^* \leq 0\,, S_n^* \geq x\right)$$

$$= \sum_{n=1}^{\infty} P\left(S_{T_1^{*+}}^* \geq x, T_1^{*+} = n\right)$$

$$= P\left(S_{T_1^{*+}}^* \geq x, T_1^{*+} < \infty\right) \qquad (2.2.42)$$

Hence, letting $x = 0$ and applying (2.2.41),

$$\|G_+^*\| = P(T_1^{*+} < \infty) = P\left(S_{T_1^{*+}}^* > 0, T_1^{*+} < \infty\right)$$

$$= P\left(S_{T_1^{(r)-}}^{(r)} \leq 0, T_1^{(r)-} < \infty\right) = P\left(T_1^{(r)-} < \infty\right)$$

Thus

$$\|G_+^*\| = -\frac{v_2}{v_1 - v_2} \log \frac{v_2}{v_1} \in (0,1) \qquad (2.2.43)$$

Letting $x > 0$, then applying (2.2.22) $P(X^* > x) = 1 - \left(\frac{v_2}{v_1}\right)^{\frac{v_2}{v_1 - v_2}} e^{\frac{v_2 x}{v_1 - v_2}}$, $x \leq -\log\frac{v_2}{v_1}$, and Asmussen (1987) it follows that

$$P\left(S_{T_1^{*+}}^* > x, T_1^{*+} < \infty\right) = 1 - \left(\frac{v_2}{v_1}\right)^{\frac{v_2}{v_1 - v_2}} e^{\frac{v_2 x}{v_1 - v_2}} + \frac{v_2}{v_1 - v_2} \int_x^{-\log\frac{v_2}{v_1}} \left(1 - \left(\frac{v_2}{v_1}\right)^{\frac{v_2}{v_1 - v_2}} e^{\frac{v_2 x}{v_1 - v_2}}\right) dy$$

$$= 1 - \left(\frac{v_2}{v_1}\right)^{\frac{v_2}{v_1 - v_2}} e^{\frac{v_2 x}{v_1 - v_2}} + \frac{v_2}{v_1 - v_2}\left(-\log\frac{v_2}{v_1} - x\right) - \left(1 - \left(\frac{v_2}{v_1}\right)^{\frac{v_2}{v_1 - v_2}} e^{\frac{v_2 x}{v_1 - v_2}}\right) =$$

$$= -\frac{v_2}{v_1 - v_2}\log\frac{v_2}{v_1} - \frac{v_2 x}{v_1 - v_2} \tag{2.2.44}$$

Hence, from (2.2.43) and (2.2.44) we obtain

$$P\left(S^*_{T_1^{*+}} \in (0, x], T_1^{*+} < \infty\right) = P(T_1^{*+} < \infty) - P\left(S^*_{T_1^{*+}} > x, T_1^{*+} < \infty\right)$$

$$= \|G^*_+\| - P\left(S^*_{T_1^{*+}} > x, T_1^{*+} < \infty\right)$$

$$= -\frac{v_2}{v_1 - v_2}\log\frac{v_2}{v_1} - \left(-\frac{v_2}{v_1 - v_2}\log\frac{v_2}{v_1} - \frac{v_2 x}{v_1 - v_2}\right)$$

$$= \frac{v_2 x}{v_1 - v_2}, \qquad x \in \left(0, -\log\frac{v_2}{v_1}\right] \tag{2.2.45}$$

Deriving the expression for the j-fold convolution of $P\left(S^*_{T_1^{*+}} \in (0, x], T_1^{*+} < \infty\right)$, which is the cumulative probability of a random variable that follows a defective uniform distribution, is the next step. The j-fold convolution of a Uniform distribution on interval $(0, b]$ is given by

$$F_{S_j}(x) = \begin{cases} 0, & x < 0 \\ \dfrac{1}{j!\, b^j}\displaystyle\sum_{k=0}^{\tilde{n}(j,x)} (-1)^k \binom{j}{k} (x - kb)^j, & 0 < x < jb \\ 1, & x \geq jb \end{cases} \tag{2.2.46}$$

Where $\tilde{n}(n, x) := \left[\frac{x}{a}\right]$ =largest integer less than $\frac{x}{a}$. See Appendix [A.1] for details.

Let $y = \frac{x}{-\log\frac{v_2}{v_1}}$ and $j \geq 1$ then

$$P\left(S^*_{T_1^{*+}} \in (0, x], T_1^{*+} < \infty\right)^{j*} = \left(\frac{v_2}{v_1 - v_2}\right)^j \frac{1}{j!}\sum_{k=0}^{\tilde{n}(j,x)} (-1)^k \binom{j}{k}\left(x - k\left(-\log\frac{v_2}{v_1}\right)\right)^j =$$

$$= \left(-\frac{v_2}{v_1 - v_2}\log\frac{v_2}{v_1}\right)^j \frac{1}{j!}\sum_{k=0}^{\tilde{n}(j,x)} (-1)^k \binom{j}{k} (x-k)_+^j$$

$$= \begin{cases} \left(-\frac{v_2}{v_1 - v_2}\log\frac{v_2}{v_1}\right)^j \frac{1}{j!}\sum_{k=0}^{[y]} (-1)^k \binom{j}{k} (y-k)_+^j, \ y \in [0, j] \\ \left(-\frac{v_2}{v_1 - v_2}\log\frac{v_2}{v_1}\right)^j, y > j \end{cases} \tag{2.2.47}$$

Using the above, we can now derive $U^*\big((0,x]\big)$, the renewal function, by summing over $j \geq 1$

$$U^*\big((0,x]\big) = \sum_{j=1}^{\infty} P\left(S^*_{T_1^{*+}} \in (0,x], T_1^{*+} < \infty\right)^{j*}$$

$$= \sum_{j=1}^{\infty} \left(-\frac{v_2}{v_1 - v_2}\log\frac{v_2}{v_1}\right)^j \frac{1}{j!}\sum_{k=0}^{[y]} (-1)^k \binom{j}{k} (y-k)_+^j$$

$$= \sum_{k=0}^{\infty} (-1)^k \left(-\frac{v_2(y-k)}{v_1 - v_2}\log\frac{v_2}{v_1}\right)^k e^{\left(\frac{-v_2(y-k)}{v_1 - v_2}\log\frac{v_2}{v_1}\right)} \frac{I(y>k)}{k!} - 1 \tag{2.2.48}$$

where $y = \frac{x}{-\log\frac{v_2}{v_1}}$.

Let $x = \infty$ in equation (2.2.48) above, and in view of equation (2.2.43), we get

$$U^*\big((0,\infty]\big) = \frac{\|G_+^*\|}{1 - \|G_+^*\|} = \frac{\frac{-v_2}{v_1 - v_2}\log\frac{v_2}{v_1}}{1 + \frac{v_2}{v_1 - v_2}\log\frac{v_2}{v_1}} \tag{2.2.49}$$

In addition, we have

$$U^*((x,\infty)) = U^*\big((0,\infty)\big) - U^*\big((0,x]\big) =$$

$$= \frac{\frac{-v_2}{v_1 - v_2}\log\frac{v_2}{v_1}}{1 + \frac{v_2}{v_1 - v_2}\log\frac{v_2}{v_1}} - U^*\big((0,x]\big), \quad x > 0 \qquad (2.2.50)$$

And finally the let us define and derive the renewal density function

$$U^*(dx) = u^*(y)dy \qquad (2.2.51)$$

$$U^*(dx)$$

$$= \left(-\log\frac{v_2}{v_1}\right) dy \left\{ \left(\frac{-v_2}{v_1 - v_2}\log\frac{v_2}{v_1}\right)\left(U^*\big((0,x]\big) + 1\right) \right.$$

$$\left. - \left(\frac{-v_2}{v_1 - v_2}\log\frac{v_2}{v_1}\right)\sum_{i=0}^{\infty}(-1)^i \left(-\frac{v_2\big(y-(i+1)\big)}{v_1 - v_2}\log\frac{v_2}{v_1}\right)^i e^{\left(\frac{-v_2(y-(i+1))}{v_1 - v_2}\log\frac{v_2}{v_1}\right)} \frac{I\big(y > (i+1)\big)}{i!} \right\}$$

where $y = \frac{x}{-\log\frac{v_2}{v_1}}$.

2.2.4 Determination of the distribution of the maximum M_∞^* in the exponential case

We now need to determine $P(M_\infty^* \le x)$, based on (2.2.48) we have

$$1 + U^*\big((0,x]\big) = \sum_{i=0}^{\infty}(-1)^i \left(\frac{-v_2(y-i)}{v_1 - v_2}\log\frac{v_2}{v_1}\right)^i e^{\left(\frac{-v_2(y-i)}{v_1 - v_2}\log\frac{v_2}{v_1}\right)} \frac{I(y > i)}{i!} \qquad (2.2.52)$$

where $y = \frac{x}{-\log\frac{v_2}{v_1}}$.

Thus, recalling (2.1.35) and $\sum_{j=1}^{\infty} P\left(T_j^{*+} < \infty, S_{T_j^{*+}}^* \le x\right) = \sum_{j=1}^{\infty} P\left(S_{T_1^{*+}}^* \in (0,x], T_1^{*+} < \infty\right)^{j*}$,

and applying (2.2.43) and (2.2.52) we have

$$P(M_\infty^* \le x) = (1 - \|G_+^*\|)\left(1 + \sum_{j=1}^{\infty} P\left(T_j^{*+} < \infty, S_{T_j^{*+}}^* \le x\right)\right)$$

$$= \left(1 + \frac{v_2}{v_1 - v_2}\log\frac{v_2}{v_1}\right)\{1 + U^*((0,x])\}$$

$$= \left(1 + \frac{v_2}{v_1 - v_2}\log\frac{v_2}{v_1}\right)\left\{\sum_{i=0}^{\infty}(-1)^i\left(-\frac{v_2(y-i)}{v_1 - v_2}\log\frac{v_2}{v_1}\right)^i e^{\left(\frac{-v_2(y-i)}{v_1-v_2}\log\frac{v_2}{v_1}\right)}\frac{I(y > i)}{i!}\right\} \quad (2.2.53)$$

where $y = \frac{x}{-\log\frac{v_2}{v_1}}$.

2.3 Exact computable expressions for the asymptotic distribution of ξ_∞ in the exponential case

We will now derive the exact computable expressions for the asymptotic distribution of ξ_∞ in the exponential case, where we have the following underlying distribution, with $\tau \in (2, 3 \dots, n)$ and $v_1 > v_2$

$$f(y_i, v) = \begin{cases} f_1(y_i, v_1) = Exp(v_1) = v_1 e^{-v_1 y}, & i = 1, \dots, \tau - 1;\ \ y_i \ge 0 \\ f_2(y_i, v_2) = Exp(v_2) = v_2 e^{-v_2 y}, & i = \tau, \dots n;\ \ y_i \ge 0 \end{cases}$$

Recall (2.1.43)

$$P(\xi_\infty = k) = \begin{cases} (1 - \|G_+^*\|)\left(q_k^* - \int_{0^+}^{\infty} P(M_\infty > x)P(S_k^* \in dx, T_1^{*-} > k)\right), k > 0 \\ (1 - \|G_+^*\|)(1 - \|G_+\|), \quad k = 0 \\ (1 - \|G_+\|)\left(q_{|k|} - \int_{0^+}^{\infty} P(M_\infty^* > x)P\left(S_{|k|} \in dx, T_1^- > |k|\right)\right), k < 0 \end{cases}$$

2.3.1 Determination of $P(\xi_\infty = 0)$

By applying $(2.2.17)\ \|G_+\| = \frac{\nu_2}{\nu_1}$ and $(2.2.43)\ \|G_+^*\| = -\frac{\nu_2}{\nu_1-\nu_2}\ln\frac{\nu_2}{\nu_1}$ to the above equation $(2.1.43)$ when $k = 0$ we obtain

$$P(\xi_\infty = 0) = (1 - \|G_+^*\|)(1 - \|G_+\|)$$

$$= \left(1 + \frac{\nu_2}{\nu_1 - \nu_2}\ln\frac{\nu_2}{\nu_1}\right)\left(1 - \frac{\nu_2}{\nu_1}\right), \qquad k = 0 \qquad (2.3.1)$$

Section 2.3.2 Determination of $P(\xi_\infty = k), k > 0$

Jandhyala & Fotopoulos (1999) have worked out the computational procedure for this case through bounds and approximations. As it turns out their suggested method gives us approximations for $P(\xi_\infty = k)$ that are exact due to identical bounds (upper and lower) for the maximum of the random walk S. That paper though does not provide the exact distribution of the maximum, which was found earlier, recall equations $(2.2.18)\ and\ (2.2.19)$.

$$P(M_\infty \le x) = 1 - \frac{\nu_2}{\nu_1}e^{-x} \quad x \ge 0$$

$$P(M_\infty > x) = \frac{\nu_2}{\nu_1}e^{-x} \qquad x \ge 0$$

We now derive the density of ξ_∞ from $(2.1.43)$, when $k > 0$ (applying the existing computational procedures).

$$P(\xi_\infty = k) = (1 - \|G_+^*\|)\left(q_k^* - \int_{0^+}^{\infty} P(M_\infty > x)P(S_k^* \in dx, T_1^{*-} > k)\right), k > 0$$

From (2.2.43) we know that $(1-\|G_+^*\|)=\left(1+\frac{v_2}{v_1-v_2}\log\frac{v_2}{v_1}\right)$. Next, let us derive a computable expression for $P(S_k^*\in dx, T_1^{*-}>k)$.

Let

$$u_k^*(dx)=P(S_k^*\in dx, T_1^{*-}>k) \qquad (2.3.2)$$

and rewrite the expression from (2.1.43) for $k>0$ as

$$P(\xi_\infty=k)=\left(1+\frac{v_2}{v_1-v_2}\log\frac{v_2}{v_1}\right)\left(q_k^*-\int_{0^+}^{\infty}\frac{v_2}{v_1}e^{-x}u_k^*(dx)\right), k>0 \qquad (2.3.3)$$

Then

$$\int_{0^+}^{\infty}e^{-\lambda x}u_k^*(dx)=\tilde{u}_k^*(\lambda) \qquad (2.3.4)$$

which is the Laplace transform of the sequence $\{u_k^*(dx): k=1,2,\dots\}$. Note that the Laplace transform at $\lambda=1$ is just

$$\tilde{u}_k^*(1)=\tilde{u}_k^* \qquad (2.3.5)$$

Then equation (2.3.3) is simplified to

$$P(\xi_\infty=k)=\left(1+\frac{v_2}{v_1-v_2}\log\frac{v_2}{v_1}\right)\left(q_k^*-\frac{v_2}{v_1}\tilde{u}_k^*\right), k>0 \qquad (2.3.6)$$

We use the following iterative (*Leibnitz*) procedure to calculate $\tilde{u}_k^*$ in (2.3.5)

$$n\tilde{u}_k^* = \sum_{j=0}^{k-1} \tilde{b}_{k-j}^* \tilde{u}_j^*, \quad k \geq 1 \qquad \tilde{u}_0^* = 1 \tag{2.3.7}$$

where,

$$\tilde{b}_k^* = \int_0^\infty e^{-x} P(S_k^* \in dx) \tag{2.3.8}$$

Note that

$$P(S_k^* < x) = P\left(-k \log \frac{v_2}{v_1} - \frac{v_1 - v_2}{v_2} \sum E_i^* < x\right)$$

$$= P\left(\sum E_i^* > \left(-x - k \log \frac{v_2}{v_1}\right) \frac{v_2}{v_1 - v_2}\right) \tag{2.3.9}$$

where $E_i^* \sim Exp(1)$

$$dP(S_k^* < x) = P(S_k^* \in dx)$$

$$= dx \frac{v_2}{v_1 - v_2} \left(\frac{v_2}{v_1 - v_2}\left(-x - k \log \frac{v_2}{v_1}\right)\right)^{k-1} e^{\left(x + k \log \frac{v_2}{v_1}\right)\frac{v_2}{v_1 - v_2}} I\left(-\infty, -k \log \frac{v_2}{v_1}\right)$$

$$= dx \left(\frac{v_2}{v_1 - v_2}\right)^k (-1)^{k-1} \left(x + k \log \frac{v_2}{v_1}\right)^{k-1} e^{\frac{x v_2}{v_1 - v_2}} \left(\frac{v_2}{v_1}\right)^{\frac{k v_2}{v_1 - v_2}} I\left(-\infty, -k \log \frac{v_2}{v_1}\right) \tag{2.3.10}$$

Thus from (2.3.8)

$$\tilde{b}_k^* = \int_0^\infty e^{-x} P(S_k^* \in dx) =$$

$$= (-1)^{n-1} \int_0^{-k\log\frac{v_2}{v_1}} \frac{1}{\Gamma(k)} e^{\frac{x(-v_1+2v_2)}{v_1-v_2}} \left(\frac{v_2}{v_1-v_2}\right)^k \left(\frac{v_2}{v_1}\right)^{\frac{kv_2}{v_1-v_2}} \left(x + k\log\frac{v_2}{v_1}\right)^{k-1} dx =$$

$let\ \delta = \frac{v_1}{v_2}$

$$= -\left(\frac{-1}{(\delta-1)\delta^{\frac{1}{\delta-1}}}\right)^k \int_0^{k\log\delta} \frac{1}{\Gamma(k)} e^{-x\left(\frac{\delta-2}{\delta-1}\right)} (x - k\log\delta)^{k-1} dx$$

$$= -\left(\frac{1}{(\delta-1)\delta^{\frac{1}{\delta-1}}}\right)^k \int_0^{k\log\delta} \frac{1}{\Gamma(k)} e^{-x\left(\frac{\delta-2}{\delta-1}\right)} (-x + k\log\delta)^{k-1} dx =$$

$let\ u = -x + k\log\delta$

$$= -\left(\frac{1}{(\delta-1)\delta^{\frac{1}{\delta-1}}}\right)^k \int_{k\log\delta}^{0} \frac{1}{\Gamma(k)} e^{-\left(\frac{\delta-2}{\delta-1}\right)(-u-k\log\delta)} (u)^{k-1} du$$

$$= \left(\frac{1}{\delta(\delta-1)}\right)^k \int_0^{n\log\delta} \frac{1}{\Gamma(k)} e^{\left(\frac{\delta-2}{\delta-1}\right)u} (u)^{k-1} du =$$

If $\delta \in (1,2)$, $\frac{\delta-2}{\delta-1} < 0$ then

$$= \left(\frac{1}{\delta(2-\delta)}\right)^k \int_0^{k\log\delta\left(\frac{2-\delta}{\delta-1}\right)} \frac{1}{\Gamma(k)} e^{-u} (u)^{k-1} du$$

$$= \left(\frac{1}{\delta(2-\delta)}\right)^k \frac{\gamma\left(k, \frac{n(2-\delta)\log\delta}{\delta-1}\right)}{\Gamma(k)}$$

If $\delta = 2$, $\frac{\delta-2}{\delta-1} = 0$ then

$$= \left(\frac{k\ln 2}{2}\right)^k \frac{1}{k!}$$

If $\delta > 2$, $\frac{\delta-2}{\delta-1} > 0$ then

$$= \left(\frac{1}{\delta(\delta-2)}\right)^k \left[\delta^{k\frac{\delta-2}{\delta-1}} \sum_{j=0}^{k-1} \frac{(-1)^j \left(k\frac{\delta-2}{\delta-1}\log\delta\right)^{k-1-j}}{\Gamma(k-j)} + (-1)^k\right]$$

Summing up the above we ultimately obtain the following expression

$$\tilde{b}_k^* = \begin{cases} \left(\frac{1}{\delta(2-\delta)}\right)^k \frac{\gamma\left(k, \frac{k(2-\delta)\log\delta}{\delta-1}\right)}{\Gamma(k)}, & \delta \in (1,2) \\ \left(\frac{k\log 2}{2}\right)^n \frac{1}{k!}, & \delta = 2 \\ \left(\frac{1}{\delta(\delta-2)}\right)^k \left[\delta^{k\frac{\delta-2}{\delta-1}} \sum_{j=0}^{k-1} \frac{(-1)^j \left(k\frac{\delta-2}{\delta-1}\log\delta\right)^{k-1-j}}{\Gamma(k-j)} + (-1)^k\right], & \delta > 2 \end{cases} \quad (2.3.11)$$

where $\delta = \frac{v_1}{v_2}$ and incomplete gamma function $\frac{\gamma(\alpha,x)}{\Gamma(a)} = \frac{1}{\Gamma(a)}\int_0^x t^{\alpha-1}e^{-t}dt$

The following iterative procedure is used to calculate $q_k^* = P(T_1^{*-} > k)$ in (2.1.42)

$$kq_k^* = \sum_{j=0}^{k-1} b_{k-j}^* q_j^*, k \geq 1 \qquad q_0^* = 1 \qquad (2.3.12)$$

where

$$b_k^* = P(S_k^* > 0) =$$

$$= P\left(-k\log\frac{v_2}{v_1} - \frac{v_1 - v_2}{v_2}\sum_{i=1}^{k} E_i^* > 0\right)$$

$$= P\left(\sum_{i=1}^{k} E_i^* < -k\log\frac{v_2}{v_1}\left(\frac{v_2}{v_1 - v_2}\right)\right)$$

$$= \int_0^{-k\log\frac{v_2}{v_1}\left(\frac{v_2}{v_1 - v_2}\right)} \frac{x^{k-1}}{\Gamma(k)} e^{-x} dx$$

Then

$$b_k^* = \frac{\gamma\left(k, -k\left(\frac{v_2}{v_1 - v_2}\right)\log\frac{v_2}{v_1}\right)}{\Gamma(k)} \qquad (2.3.13)$$

where $E_i^* \sim Exp(1)$ with the incomplete gamma function being

$$\frac{\gamma(\alpha, x)}{\Gamma(a)} = \frac{1}{\Gamma(a)}\int_0^x t^{\alpha-1}e^{-t}dt$$

and $\Gamma(\alpha)$ is the gamma function as shown below

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx = (\alpha - 1)!$$

2.3.3 Determination of $P(\xi_\infty = k), k < 0$

We now find a computable expression to determine the exact probabilities for ξ_∞ equaling negative values of k. From $(2.1.43)$ we have

$$P(\xi_\infty = k) = (1 - \|G_+\|)\left(q_{|k|} - \int_{0^+}^{\infty} P(M_\infty^* > x)P\left(S_{|k|} \in dx, T_1^- > |k|\right)\right), k < 0$$

We have already calculated $\|G_+\| = \frac{v_2}{v_1}$ see (2.2.17) and $P(M_\infty^* \leq x) = \left(1 + \frac{v_2}{v_1 - v_2}\log\frac{v_2}{v_1}\right)\{1 + U^*\left((0, x]\right)\}$, see (2.2.53). Then

$$P(M_\infty^* > x) = 1 - P(M_\infty^* \leq x)$$

$$= 1 - \left(1 + \frac{v_2}{v_1 - v_2}\log\frac{v_2}{v_1}\right)\{1 + U^*\left((0, x]\right)\}$$

$$= 1 - \left(1 + \frac{v_2}{v_1 - v_2}\log\frac{v_2}{v_1}\right) - \left(1 + \frac{v_2}{v_1 - v_2}\log\frac{v_2}{v_1}\right)U^*\left((0, x]\right)$$

$$= -\frac{v_2}{v_1 - v_2}\log\frac{v_2}{v_1} - \left(1 + \frac{v_2}{v_1 - v_2}\log\frac{v_2}{v_1}\right)U^*\left((0, x]\right) \qquad (2.3.14)$$

where by (2.2.52)

$$U^*\left((0, x]\right) = \sum_{i=0}^{\infty}(-1)^i\left(-\frac{v_2(y - i)}{v_1 - v_2}\log\frac{v_2}{v_1}\right)^i e^{\left(\frac{-v_2(y-i)}{v_1 - v_2}\log\frac{v_2}{v_1}\right)}\frac{I(y > i)}{i!} - 1$$

where $y = \frac{x}{-\log\frac{v_2}{v_1}}$.

Let

$$u_k(dx) = P(S_k \in dx, T_1^- > k) \qquad (2.3.15)$$

$$u_k(\infty) = \int_{0^+}^{\infty} u_k(dx) = P(T_1^- > k) = q_k \qquad (2.3.16)$$

Now, rewriting the section for $k < 0$ from equation (2.1.43), and applying (2.2.17) and (2.3.15) we have

$$P(\xi_\infty = -k) = \left(1 - \frac{v_2}{v_1}\right)\left(q_k - \int_{0^+}^{\infty} P(M_\infty^* > x)u_k(dx)\right), k > 0 \qquad (2.3.17)$$

We consider just the inner part of (2.3.17) for the moment and apply equations (2.3.13), (2.3.14) and (2.3.16)

$$q_k - \int_{0^+}^{\infty} P(M_\infty^* > x)u_k(dx)$$

$$= q_k - \int_{0^+}^{\infty} \left[1 - \left(1 + \frac{v_2}{v_1 - v_2}\log\frac{v_2}{v_1}\right)\{1 + U^*((0, x])\}\right] u_k(dx)$$

$$= q_k - \int_{0^+}^{\infty} u_k(dx) + \int_{0^+}^{\infty} \left[\left(1 + \frac{v_2}{v_1 - v_2}\log\frac{v_2}{v_1}\right)\{1 + U^*((0, x])\}\right] u_k(dx)$$

$$= q_k - q_k + \left(1 + \frac{v_2}{v_1 - v_2}\log\frac{v_2}{v_1}\right)\int_{0^+}^{\infty} [1 + U^*((0, x])]u_k(dx)$$

$$= \left(1 + \frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}\right)\int_{0^+}^{\infty}\left[\sum_{\mathrm{i}=0}^{\infty}(-1)^{\mathrm{i}}\left(-\frac{\mathrm{v}_2(\mathrm{y}-\mathrm{i})}{\mathrm{v}_1 - \mathrm{v}_2}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}\right)^{\mathrm{i}} \mathrm{e}^{\left(\frac{-\mathrm{v}_2(\mathrm{y}-\mathrm{i})}{\mathrm{v}_1-\mathrm{v}_2}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}\right)}\frac{\mathrm{I}(\mathrm{y} > i)}{\mathrm{i}!}\right]\mathrm{u}_\mathrm{k}(\mathrm{dx})$$

$$= \left(1 + \frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}\right)\sum_{\mathrm{i}=0}^{\infty}\frac{(-1)^{\mathrm{i}}}{\mathrm{i}!}\int_{0^+}^{\infty}\left[\left(\frac{-\mathrm{v}_2(\mathrm{y}-\mathrm{i})}{\mathrm{v}_1 - \mathrm{v}_2}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}\right)^{\mathrm{i}} \mathrm{e}^{\left(\frac{-\mathrm{v}_2(\mathrm{y}-\mathrm{i})}{\mathrm{v}_1-\mathrm{v}_2}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}\right)}\mathrm{I}(\mathrm{y} > i)\right]\mathrm{u}_\mathrm{k}(\mathrm{dx}) =$$

Note: $y = \frac{x}{-\log\frac{v_2}{v_1}}$, $x = -y\log\frac{v_2}{v_1}$ and $-(y-i)\log\frac{v_2}{v_1} = -\left(\frac{x}{-\log\frac{v_2}{v_1}} - i\right)\log\frac{v_2}{v_1} = x + i\log\frac{v_2}{v_1}$

$$= \left(1 + \frac{v_2}{v_1 - v_2}\log\frac{v_2}{v_1}\right)\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}\int_{0^+}^{\infty}\left[\left(\frac{v_2}{v_1 - v_2}\left(x + i\log\frac{v_2}{v_1}\right)\right)^i e^{\frac{v_2}{v_1-v_2}\left(x+i\log\frac{v_2}{v_1}\right)} I\left(x > -i\log\frac{v_2}{v_1}\right)\right] u_k(dx)$$

$$= \left(1 + \frac{v_2}{v_1 - v_2}\log\frac{v_2}{v_1}\right)\sum_{i=0}^{\infty}\frac{(-1)^i}{i!} \times$$

$$\times \int_{-i\log\frac{v_2}{v_1}}^{\infty}\left[\left(\frac{v_2}{v_1 - v_2}\left(x + i\log\frac{v_2}{v_1}\right)\right)^i e^{\frac{v_2}{v_1-v_2}\left(x+i\log\frac{v_2}{v_1}\right)}\right] u_k(dx) \qquad (2.3.18)$$

Applying the above (2.3.18) to (2.3.17), we obtain the following expression for the asymptotic distribution of ξ_∞ when $k < 0$.

$$P(\xi_\infty = -k) = \frac{v_1 - v_2}{v_1}\left(1 + \frac{v_2}{v_1 - v_2}\log\frac{v_2}{v_1}\right)\sum_{i=0}^{\infty}\frac{(-1)^i}{i!} \times$$

$$\times \int_{-i\log\frac{v_2}{v_1}}^{\infty}\left[\left(\frac{v_2}{v_1 - v_2}\left(x + i\log\frac{v_2}{v_1}\right)\right)^i e^{\frac{v_2}{v_1-v_2}\left(x+i\log\frac{v_2}{v_1}\right)}\right] u_k(dx), k > 0 \quad (2.3.19)$$

Note that since the q_k is canceled out and does not appear in the expression above, we do not need to compute its values here, but for completeness' sake the necessary expressions for calculating q_k are presented below

$$q_0 = 1, \qquad kq_k = \sum_{j=0}^{k-1} b_{k-j} q_j , \qquad k > 0$$

$$b_k = P\left(k \log \frac{\mathrm{v}_2}{\mathrm{v}_1} + \frac{v_1 - v_2}{v_1} \sum_{i=1}^{k} E_i > 0 \right)$$

$$= P\left(\sum_{i=1}^{k} E_i > -k \frac{v_1}{v_1 - v_2} \log \frac{\mathrm{v}_2}{\mathrm{v}_1} \right) = \frac{\Gamma\left(k, -k \frac{v_1}{v_1 - v_2} \log \frac{\mathrm{v}_2}{\mathrm{v}_1}\right)}{\Gamma(k)}$$

where $E_i \sim Exp(1)$ and $\Gamma(a,x)/\Gamma(a) = \int_x^\infty t^{a-1} e^{-t} dt / \Gamma(a)$ is the complement to the incomplete gamma function.

The only thing that is left is to compute $\mathrm{u_k(dx)}$. As it turns out, we cannot get one general computable expression for $\mathrm{u_k(dx)}$ in (2.3.19), and thus we must first determine $u_k(x)$ for each $k = 1, 2, \ldots$ individually, using induction, and then calculate the desired $u_k(dx)$.

Let

$$\bar{u}_k(x) = P(S_k > x, T_1^- > k) , \qquad x > 0 \qquad (2.3.20)$$

be the probability that the first time the random walk S_k enters $(-\infty, 0)$ is after the kth step, with ladder height at that kth step being greater than x, and $k \geq 1$. For $k \geq 2$, and $u_k(dx) = P(S_k \in dx, T_1^- > k)$ as defined in (2.3.15) we can obtain the following expression

$$u_k(dx) = \frac{v_1}{v_1 - v_2} dx \left[\bar{u}_k(x) - \bar{u}_{k-1}\left(x + \log \frac{v_2}{v_1}\right) \right] \qquad (2.3.21)$$

Thus once we have determined an expression for $\bar{u}_1(x)$, we just need to find the subsequent differences $\bar{u}_k(x) - \bar{u}_{k-1}\left(x + \log\frac{v_2}{v_1}\right)$ to obtain the needed expressions for $u_k(dx)$, $k \geq 2$. See Appendix [A.2] for all the expressions and the details on how they were derived, for now we just show the computations and results for $\bar{u}_1(x)$ and $\bar{u}_2(x)$ as examples.

Let $k = 1$, then applying $(2.2.4)$ we obtain

$$\bar{u}_1(x) = P(S_1 > x, T_1^- > 1) = P(S_1 > x) = P(X_1 > x)$$

$$= e^{\frac{-v_1}{v_1 - v_2}\left[x - log\left(\frac{v_2}{v_1}\right)\right]} \qquad (2.3.22)$$

and

$$u_1(dx) = dx\,\frac{v_1}{v_1 - v_2}\,e^{\frac{-v_1}{v_1 - v_2}\left[x - log\left(\frac{v_2}{v_1}\right)\right]} \qquad (2.3.23)$$

Let $k = 2$, then

$$\bar{u}_2(x) = e^{\frac{-v_1}{v_1 - v_2}\left[x - 2log\left(\frac{v_2}{v_1}\right)\right]}\left\{1 + \frac{v_1}{v_1 - v_2}\left(x - log\,\frac{v_2}{v_1}\right)\right\} \qquad (2.3.24)$$

and

$$u_2(dx) = dx\,\frac{v_1}{v_1 - v_2}\left\{\bar{u}_2(x) - \bar{u}_1\left(x - log\,\frac{v_2}{v_1}\right)\right\}$$

$$= dx\,\frac{v_1}{v_1 - v_2}\,e^{-\frac{v_1}{v_1 - v_2}\left(x - 2log\frac{v_2}{v_1}\right)}\left\{\frac{v_1}{v_1 - v_2}\left(x - log\,\frac{v_2}{v_1}\right)\right\} \qquad (2.3.25)$$

Now that we have obtained expressions for $u_k(dx)$ we can derive the exact probabilities for $P(\xi_\infty = -k),\ k > 0$ by applying the above method in $(2.3.21)$ and the expression we obtained in $(2.3.19)$. The full details of computations of the following probabilities can be found in Appendix [A.3], we do not show them here directly as the computations, while straightforward, are quite long. Note that we have computed the exact probabilities for $P(\xi_\infty = -k),\ k = 1, 2, \dots, 6$, and it is quite possible to continue the calculations beyond that, but for larger values of k the computations become very tedious and time consuming. Instead we note that Jandhyala & Fotopoulos (1999) have derived a method to approximate these probabilities, and it can be seen in Table 3 in Section 2.3.4 that those approximations are quite accurate and once the values for k get close to 7 the differences become truly minor.

THEOREM 2.1 *Let $\xi_n = \hat{\tau}_n - \tau_n$ be a random variable with the asymptotic distribution ξ_∞ as both $n - \tau_n$ and $\tau_n \to \infty$, when $k < 0$ and the underlying process is exponential, with $Y \sim Exp(v_1)$ before the change at τ_n and $Y \sim Exp(v_2)$ after the change, with $v_1 > v_2$. Then following are the expressions for $P(\xi_\infty = k)$ when $k = -1, -2, \dots, -6$,*

$$P(\xi_\infty = -1) = \frac{\left(1 + \frac{v_2}{v_1 - v_2}\ln\frac{v_2}{v_1}\right)\left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1 - v_2}}}{\left(1 + \frac{v_2}{v_1 - v_2}\left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1 - v_2}}\right)}$$

$$P(\xi_\infty = -2) = \frac{\left(1 + \frac{v_2}{v_1 - v_2}\ln\frac{v_2}{v_1}\right)\left(\frac{v_2}{v_1}\right)^{\frac{2v_1}{v_1 - v_2}}\left(1 - \ln\frac{v_2}{v_1}\right)\left(\frac{v_1}{v_1 - v_2}\right)}{\left(1 + \frac{v_2}{v_1 - v_2}\left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1 - v_2}}\right)^2}$$

$$P(\xi_\infty=-3)=\left(1+\frac{\nu_2}{\nu_1-\nu_2}\ln\frac{\nu_2}{\nu_1}\right)\left(\frac{\nu_2}{\nu_1}\right)^{\frac{3\nu_1}{\nu_1-\nu_2}}\frac{\left(\frac{\nu_1}{\nu_1-\nu_2}\right)^2}{2!}\left\{\frac{2\left(1-\ln\frac{\nu_2}{\nu_1}\right)^2}{\left(1+\frac{\nu_2}{\nu_1-\nu_2}\left(\frac{\nu_2}{\nu_1}\right)^{\frac{\nu_1}{\nu_1-\nu_2}}\right)^3}\right.$$

$$\left.+\frac{\left(-\ln\frac{\nu_2}{\nu_1}\right)^2}{\left(1+\frac{\nu_2}{\nu_1-\nu_2}\left(\frac{\nu_2}{\nu_1}\right)^{\frac{\nu_1}{\nu_1-\nu_2}}\right)^2}\right\}$$

$$P(\xi_\infty=-4)=\left(1+\frac{\nu_2}{\nu_1-\nu_2}\ln\frac{\nu_2}{\nu_1}\right)\left(\frac{\nu_2}{\nu_1}\right)^{\frac{4\nu_1}{\nu_1-\nu_2}}\frac{\left(\frac{\nu_1}{\nu_1-\nu_2}\right)^3}{3!}\left\{\frac{6\left(1-\ln\frac{\nu_2}{\nu_1}\right)^3}{\left(1+\frac{\nu_2}{\nu_1-\nu_2}\left(\frac{\nu_2}{\nu_1}\right)^{\frac{\nu_1}{\nu_1-\nu_2}}\right)^4}\right.$$

$$\left.+\frac{6\left(-\ln\frac{\nu_2}{\nu_1}\right)^2\left(1-\ln\frac{\nu_2}{\nu_1}\right)}{\left(1+\frac{\nu_2}{\nu_1-\nu_2}\left(\frac{\nu_2}{\nu_1}\right)^{\frac{\nu_1}{\nu_1-\nu_2}}\right)^3}+\frac{4\left(-\ln\frac{\nu_2}{\nu_1}\right)^3}{\left(1+\frac{\nu_2}{\nu_1-\nu_2}\left(\frac{\nu_2}{\nu_1}\right)^{\frac{\nu_1}{\nu_1-\nu_2}}\right)^2}\right\}$$

$$P(\xi_\infty=-5)=\left(1+\frac{\nu_2}{\nu_1-\nu_2}\ln\frac{\nu_2}{\nu_1}\right)\left(\frac{\nu_2}{\nu_1}\right)^{\frac{5\nu_1}{\nu_1-\nu_2}}\frac{\left(\frac{\nu_1}{\nu_1-\nu_2}\right)^4}{4!}\left\{\frac{24\left(1-\ln\frac{\nu_2}{\nu_1}\right)^4}{\left(1+\frac{\nu_2}{\nu_1-\nu_2}\left(\frac{\nu_2}{\nu_1}\right)^{\frac{\nu_1}{\nu_1-\nu_2}}\right)^5}\right.$$

$$+\frac{36\left(-\ln\frac{\nu_2}{\nu_1}\right)^2\left(1-\ln\frac{\nu_2}{\nu_1}\right)^2}{\left(1+\frac{\nu_2}{\nu_1-\nu_2}\left(\frac{\nu_2}{\nu_1}\right)^{\frac{\nu_1}{\nu_1-\nu_2}}\right)^4}+\frac{2\left(-\ln\frac{\nu_2}{\nu_1}\right)^3\left(16-19\ln\frac{\nu_2}{\nu_1}\right)}{\left(1+\frac{\nu_2}{\nu_1-\nu_2}\left(\frac{\nu_2}{\nu_1}\right)^{\frac{\nu_1}{\nu_1-\nu_2}}\right)^3}$$

$$\left.+\frac{27\left(-\ln\frac{\nu_2}{\nu_1}\right)^4}{\left(1+\frac{\nu_2}{\nu_1-\nu_2}\left(\frac{\nu_2}{\nu_1}\right)^{\frac{\nu_1}{\nu_1-\nu_2}}\right)^2}\right\}$$

$$P(\xi_\infty = -6) = \left(1 + \frac{\nu_2}{\nu_1 - \nu_2}\ln\frac{\nu_2}{\nu_1}\right)\left(\frac{\nu_2}{\nu_1}\right)^{\frac{6\nu_1}{\nu_1-\nu_2}} \frac{\left(\frac{\nu_1}{\nu_1-\nu_2}\right)^5}{5!} \left\{ \frac{120\left(1-\ln\frac{\nu_2}{\nu_1}\right)^5}{\left(1+\frac{\nu_2}{\nu_1-\nu_2}\left(\frac{\nu_2}{\nu_1}\right)^{\frac{\nu_1}{\nu_1-\nu_2}}\right)^6} \right.$$

$$+ \frac{240\left(-\ln\frac{\nu_2}{\nu_1}\right)^2\left(1-\ln\frac{\nu_2}{\nu_1}\right)^3}{\left(1+\frac{\nu_2}{\nu_1-\nu_2}\left(\frac{\nu_2}{\nu_1}\right)^{\frac{\nu_1}{\nu_1-\nu_2}}\right)^5} + \frac{30\left(-\ln\frac{\nu_2}{\nu_1}\right)^3\left(8-19\ln\frac{\nu_2}{\nu_1}+11\left(-\ln\frac{\nu_2}{\nu_1}\right)^2\right)}{\left(1+\frac{\nu_2}{\nu_1-\nu_2}\left(\frac{\nu_2}{\nu_1}\right)^{\frac{\nu_1}{\nu_1-\nu_2}}\right)^4}$$

$$\left. + \frac{54\left(-\ln\frac{\nu_2}{\nu_1}\right)^4\left(5-7\ln\frac{\nu_2}{\nu_1}\right)}{\left(1+\frac{\nu_2}{\nu_1-\nu_2}\left(\frac{\nu_2}{\nu_1}\right)^{\frac{\nu_1}{\nu_1-\nu_2}}\right)^3} + \frac{256\left(-\ln\frac{\nu_2}{\nu_1}\right)^5}{\left(1+\frac{\nu_2}{\nu_1-\nu_2}\left(\frac{\nu_2}{\nu_1}\right)^{\frac{\nu_1}{\nu_1-\nu_2}}\right)^2} \right\}$$

Beyond this point, for $k = -7, -8, \ldots$ we shall use the approximations from Jandhyala & Fotopoulos (1999), see *pp. 137 section 6.2 exponential distribution*. Below is a brief summary of the expressions used to compute the approximate probabilities.

$$P(\xi_\infty = -k) \approx e^{-B(1)}\{q_k - m^* \tilde{u}_k\} \qquad (2.3.26)$$

where

$$\tilde{u}_k(\lambda) = \int_0^\infty e^{-\lambda x} u_k(dx)$$

is the Laplace transform of the sequence $\{u_k(dx), k = 1, 2, \ldots\}$, same as in $(2.3.4)$. Setting $\lambda = 1$ we obtain $\tilde{u}_k(1) = \tilde{u}_k$.

To calculate $\tilde{u}_k$ we use the following iterations

$$\tilde{u}_0 = 1, \ \ n\tilde{u}_n = \sum_{j=0}^{n-1} \tilde{b}_{n-j}\tilde{u}_j\,, n > 0 \qquad (2.3.27)$$

where

$$\tilde{b}_n = \left(\frac{\delta^2}{\delta^2 - 1}\right)^n P(\chi^2_{2n} < 2n(\delta + 1)d) \qquad (2.3.28)$$

Then

$$m^* = d\sum_{n=1}^{\infty} P(\chi^2_{2n} < 2n(\delta - 1)d) - \frac{1}{\delta - 1}\sum_{n=1}^{\infty} P\left(\chi^2_{2(n+1)} < 2n(\delta - 1)d\right) \qquad (2.3.29)$$

where

$$\delta = \frac{v_2^{-1}}{v_2^{-1} - v_1^{-1}}, \quad d = \log\frac{\delta}{\delta - 1}$$

To calculate q_k we use the following iterations

$$q_0 = 1, \ \ nq_n = \sum_{j=0}^{n-1} b_{n-j}q_j\,, n > 0 \qquad (2.3.30)$$

where

$$b_n = P(\chi^2_{2n} < 2\delta nd) \qquad (2.3.31)$$

And

$$B(1) = \sum_{n=1}^{\infty} \frac{b_n}{n} \qquad (2.3.32)$$

2.3.4 Probability graphs and tables for ξ_∞

The following graphs and tables were calculated using the exact probabilities by applying the expressions derived in section 2.3 for $k > -6$ and using the approximations calculated applying the method in $(2.3.26)$ for $k < -6$. The total probabilities in Tables 1, 2 and 3 were calculated based on 60 or more values.

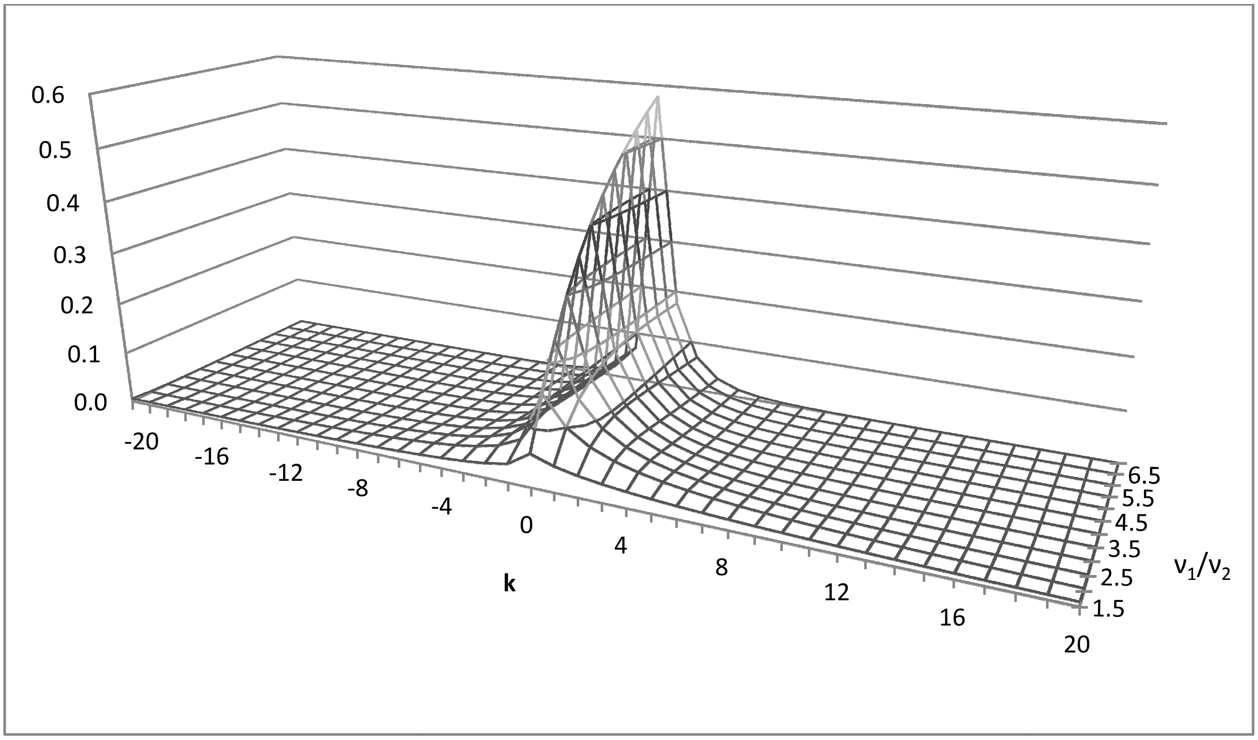


Figure 2.1. The probability distribution function of $P(\xi_\infty = k)$ for $v_1/v_2 = 1.5, 2, 2.5, 3, \dots 7$.

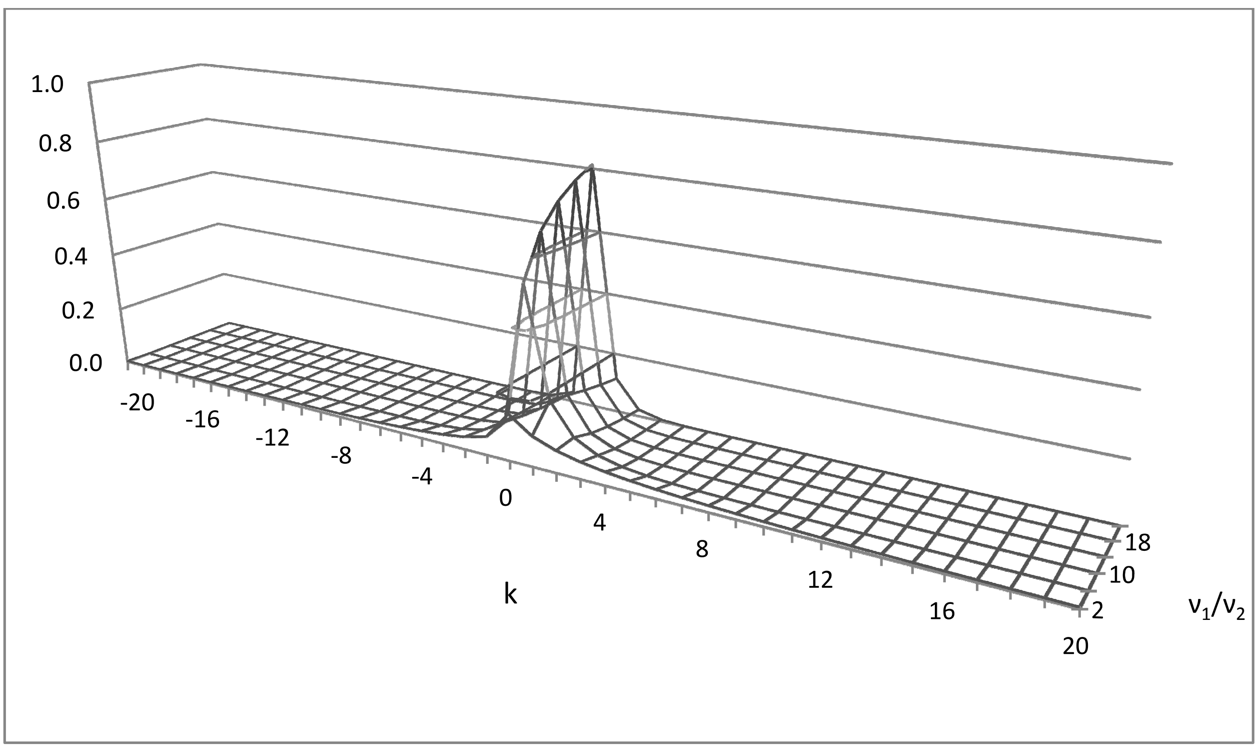


Figure 2.2. The probability distribution function of $P(\xi_\infty = k)$ for $v_1/v_2 = 2, 6, 10, 14, 18, 22$.

TABLE 2.1

n	$v_1/v_2 = 1.5$	$v_1/v_2 = 2$	$v_1/v_2 = 3$	$v_1/v_2 = 4$	$v_1/v_2 = 5$	$v_1/v_2 = 6$
-20	0.0054	0.0025	0.0003	0.0001	0.0000	0.0000
-15	0.0074	0.0044	0.0010	0.0002	0.0001	0.0000
-10	0.0108	0.0085	0.0032	0.0012	0.0005	0.0002
-9	0.0117	0.0099	0.0042	0.0017	0.0008	0.0004
-8	0.0129	0.0116	0.0055	0.0025	0.0012	0.0006
-7	0.0142	0.0136	0.0072	0.0036	0.0019	0.0011
-6	0.0156	0.0164	0.0098	0.0055	0.0032	0.0019
-5	0.0175	0.0198	0.0134	0.0083	0.0053	0.0035
-4	0.0199	0.0244	0.0189	0.0130	0.0091	0.0065
-3	0.0231	0.0311	0.0278	0.0215	0.0165	0.0129
-2	0.0276	0.0416	0.0437	0.0383	0.0327	0.0278
-1	0.0352	0.0614	0.0791	0.0805	0.0773	0.0730
0	0.0630	0.1534	0.3005	0.4034	0.4781	0.5347
1	0.0490	0.1003	0.1538	0.1735	0.1793	0.1790
2	0.0404	0.0726	0.0927	0.0913	0.0845	0.0769
3	0.0345	0.0556	0.0607	0.0531	0.0445	0.0372
4	0.0300	0.0440	0.0419	0.0328	0.0250	0.0193
5	0.0265	0.0358	0.0299	0.0211	0.0147	0.0105
6	0.0236	0.0296	0.0218	0.0139	0.0089	0.0059
7	0.0213	0.0248	0.0163	0.0094	0.0055	0.0034
8	0.0193	0.0210	0.0123	0.0065	0.0035	0.0020
9	0.0176	0.0179	0.0094	0.0045	0.0022	0.0012
10	0.0162	0.0154	0.0073	0.0032	0.0014	0.0007
15	0.0110	0.0079	0.0022	0.0006	0.0002	0.0001
20	0.0080	0.0044	0.0008	0.0001	0.0000	0.0000
Total:	0.99445	0.99996	0.99996	0.99998	0.99999	1.00000

Table 2.1. The probability distribution function of $P(\xi_\infty = k)$ for $v_1/v_2 = 1.5, 2, 3, 4, 5, and\ 6$.

TABLE 2.2

n	$v_1/v_2 = 2$	$v_1/v_2 = 6$	$v_1/v_2 = 10$	$v_1/v_2 = 14$	$v_1/v_2 = 18$	$v_1/v_2 = 22$
-20	0.0025	0.0000	0.0000	0.0000	0.0000	0.0000
-15	0.0044	0.0000	0.0000	0.0000	0.0000	0.0000
-10	0.0085	0.0002	0.0000	0.0000	0.0000	0.0000
-9	0.0099	0.0004	0.0000	0.0000	0.0000	0.0000
-8	0.0116	0.0006	0.0001	0.0000	0.0000	0.0000
-7	0.0136	0.0011	0.0002	0.0000	0.0000	0.0000
-6	0.0164	0.0019	0.0004	0.0001	0.0000	0.0000
-5	0.0198	0.0035	0.0009	0.0003	0.0001	0.0001
-4	0.0244	0.0065	0.0022	0.0010	0.0005	0.0003
-3	0.0311	0.0129	0.0056	0.0030	0.0018	0.0012
-2	0.0416	0.0278	0.0161	0.0105	0.0075	0.0056
-1	0.0614	0.0730	0.0571	0.0463	0.0388	0.0334
0	0.1534	0.5347	0.6697	0.7401	0.7839	0.8140
1	0.1003	0.1790	0.1617	0.1429	0.1275	0.1151
2	0.0726	0.0769	0.0525	0.0380	0.0290	0.0230
3	0.0556	0.0372	0.0195	0.0117	0.0076	0.0053
4	0.0440	0.0193	0.0078	0.0039	0.0022	0.0014
5	0.0358	0.0105	0.0033	0.0014	0.0007	0.0004
6	0.0296	0.0059	0.0015	0.0005	0.0002	0.0001
7	0.0248	0.0034	0.0007	0.0002	0.0001	0.0000
8	0.0210	0.0020	0.0003	0.0001	0.0000	0.0000
9	0.0179	0.0012	0.0001	0.0000	0.0000	0.0000
10	0.0154	0.0007	0.0001	0.0000	0.0000	0.0000
15	0.0079	0.0001	0.0000	0.0000	0.0000	0.0000
20	0.0044	0.0000	0.0000	0.0000	0.0000	0.0000
Total:	0.99996	1.00000	1.00000	1.00000	1.00000	1.00000

Table 2.2. The probability distribution function of $P(\xi_\infty = k)$ for $v_1/v_2 = 2, 6, 10, 14, 18, 22$.

TABLE 2.3.

n	$v_1/v_2 = 1.5$		$v_1/v_2 = 3$		$v_1/v_2 = 6$		$v_1/v_2 = 10$		$v_1/v_2 = 20$	
	exact	*approx*	*exact*	*approx*	*exact*	*approx*	*exact*	*approx*	*exact*	*approx*
-20		0.00540		0.00034		0.00000		0.00000		0.00000
-15		0.00741		0.00099		0.00002		0.00000		0.00000
-10		0.01078		0.00323		0.00022		0.00002		0.00000
-9		0.01174		0.00418		0.00037		0.00004		0.00000
-8		0.01286		0.00546		0.00063		0.00008		0.00000
-7		0.01417		0.00722		0.00109		0.00017		0.00001
-6	0.01565	0.01573	0.00980	0.00971	0.00194	0.00191	0.00040	0.00039	0.00003	0.00003
-5	0.01753	0.01764	0.01339	0.01332	0.00347	0.00344	0.00091	0.00090	0.00011	0.00010
-4	0.01992	0.02004	0.01890	0.01877	0.00650	0.00642	0.00220	0.00216	0.00038	0.00037
-3	0.02308	0.02319	0.02779	0.02753	0.01286	0.01267	0.00564	0.00554	0.00148	0.00145
-2	0.02759	0.02764	0.04373	0.04308	0.02785	0.02729	0.01609	0.01573	0.00643	0.00628
-1	0.03518	0.03485	0.07912	0.07676	0.07303	0.07057	0.05713	0.05517	0.03589	0.03474
0	0.06302		0.30046		0.53471		0.66974		0.80021	
1	0.04902		0.15383		0.17902		0.16171		0.12099	
2	0.04044		0.09267		0.07687		0.05250		0.02572	
3	0.03446		0.06072		0.03719		0.01953		0.00635	
4	0.02998		0.04186		0.01928		0.00785		0.00171	
5	0.02648		0.02986		0.01046		0.00332		0.00048	
10	0.01615		0.00728		0.00070		0.00007		0.00000	
15	0.01103		0.00222		0.00006		0.00000		0.00000	
20	0.00798		0.00076		0.00001		0.00000		0.00000	
Total:	0.99445		0.99996		1.00000		1.00000		1.00000	

Table 2.3. The probability distribution function of $P(\xi_\infty = k)$ comparing the exact and approximate values for $k = -1, -2, \ldots - 6$.

2.4 Change-point analysis of earthquake frequency data from the Sumatra, West Indonesia area

2.4.1 Introduction

One of the more promising approaches to understanding earthquake mechanics is the study of stress/strain changes and how they affect the frequency and distribution of earthquakes.

Seismic waves, for example from a manmade event or a distant earthquake, may stress an existing fault making it more unstable and hasten the occurrence of the next earthquake, in literature this is referred to as dynamic triggering. Iwata and Nakanishi (2004) tested a hypothesis to see if such dynamic triggering occurs by looking at intervals between the time of the passage of a seismic wave and the origin time of the earthquake immediately following it They compared those intervals to time intervals between random earthquakes, ones not affected by seismic waves. The data used in their hypothesis test was obtained from Matsushiro, Nagano Prefecture, central Japan. In the end, they found a statistically significant decrease in the time intervals observed after a seismic wave, and concluded that further studies of dynamic triggering should focus on analyzing such time intervals.

In addition, multiple researchers, as mentioned by Iwata and Nakanishi (2004), have claimed that stress changes accelerate the occurrence of the subsequent earthquakes. It has also been suggested in literature that after a large stress/strain change an increase in seismicity occurs within a few days and the larger the change the larger the degree of the dynamic triggering.

The data we chose to utilize in the following analysis, see Figure 2.3, Table 2.4, and Figure 2.4 below, was obtained from the International Seismological Centre On-line Bulletin and consists of 151 time intervals between earthquakes observed between January 1st 2004 and March 22nd 2005, measured in hours, between earthquake occurrences at 50 to 500 km depth around the Sumatra, West Indonesia region as defined by latitude of $-6°$ to $6°$ and longitude of $94°$ to $107°$. The great Sumatra-Andaman (Aceh-Andaman) earthquake, of magnitude 9.1, occurred on December 26th, 2004 (at 00:58 at $3.30°$ latitude and $95.78°$ longitude), thus we would like to see if, as claimed in previous research on dynamic triggering, we would observe shortened periods between subsequent earthquakes in the region due to elevated stress. In other words, is there a change in the distribution of the time intervals, and when did the change occur if there was one. Another possible approach to studying dynamic triggering is to analyze the data first and subsequently check for possible causes of a change in the length of time intervals between events. This type of approach may be useful to help identify seismic events that are hard to detect otherwise (for example if seismic monitors do not exist in or data are not available from a certain region).

Having statistical evidence that there is in fact a change in the time intervals between earthquakes following a major seismic event, and when that change occurred (utilizing confidence intervals), furthers the study of seismology and allows for better assessment of the expected seismic activity in the area of interest. Further studies may also consider the effects of magnitude and distance of the triggering event on the time and degree of change.

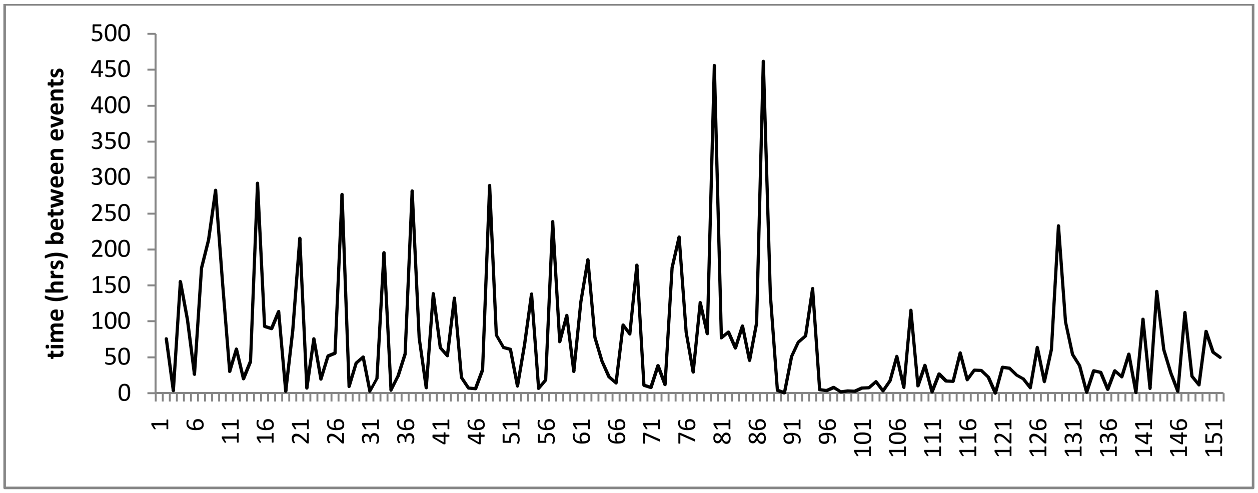


Figure 2.3. Time series of the 151 time intervals (in hours) between earthquakes.

TABLE 2.4

	Date	Time	Hours
	2-Jan-04	15:05:45	
1	5-Jan-04	18:59:17	75.89
2	5-Jan-04	22:24:17	3.42
3	12-Jan-04	9:46:38	155.37
4	16-Jan-04	16:22:55	102.60
5	17-Jan-04	18:44:27	26.36
6	25-Jan-04	1:00:08	174.26
7	2-Feb-04	22:12:53	213.21
8	14-Feb-04	16:40:54	282.47
9	21-Feb-04	0:36:10	151.92
10	22-Feb-04	6:46:27	30.17
11	24-Feb-04	20:20:15	61.56
12	25-Feb-04	16:29:40	20.16
13	27-Feb-04	12:47:47	44.30
14	10-Mar-04	17:03:34	292.26
15	14-Mar-04	14:06:19	93.05
16	18-Mar-04	8:10:53	90.08
17	23-Mar-04	1:43:20	113.54
18	23-Mar-04	4:29:46	2.77
19	26-Mar-04	19:50:41	87.35
20	4-Apr-04	19:09:21	215.31
21	5-Apr-04	2:17:09	7.13
22	8-Apr-04	6:08:46	75.86
23	9-Apr-04	1:55:48	19.78
24	11-Apr-04	5:34:44	51.65
25	13-Apr-04	13:22:00	55.79
26	25-Apr-04	1:57:22	276.59
27	25-Apr-04	11:24:04	9.45
28	27-Apr-04	5:06:42	41.71
29	29-Apr-04	7:22:06	50.26
30	29-Apr-04	9:50:35	2.47
31	30-Apr-04	6:08:10	20.29
32	8-May-04	9:26:35	195.31
33	8-May-04	13:25:30	3.98
34	9-May-04	13:05:04	23.66
35	11-May-04	19:34:22	54.49
36	23-May-04	13:09:42	281.59
37	26-May-04	17:53:42	76.73
38	27-May-04	1:25:48	7.54
39	1-Jun-04	19:49:12	138.39
40	4-Jun-04	10:54:22	63.09
41	6-Jun-04	14:50:28	51.94
42	12-Jun-04	3:03:18	132.21
43	13-Jun-04	0:42:02	21.65
44	13-Jun-04	7:59:45	7.30
45	13-Jun-04	14:05:10	6.09
46	14-Jun-04	22:26:49	32.36
47	26-Jun-04	23:12:51	288.77
48	30-Jun-04	7:38:05	80.42
49	2-Jul-04	23:31:36	63.89
50	5-Jul-04	12:26:59	60.92
51	5-Jul-04	22:15:09	9.80
52	8-Jul-04	17:41:53	67.45
53	14-Jul-04	11:49:44	138.13
54	14-Jul-04	18:25:05	6.59
55	15-Jul-04	12:39:09	18.23
56	25-Jul-04	11:30:35	238.86
57	28-Jul-04	11:09:37	71.65
58	1-Aug-04	23:10:44	108.02
59	3-Aug-04	5:27:19	30.28
60	8-Aug-04	12:17:49	126.84
61	16-Aug-04	6:03:50	185.77
62	19-Aug-04	11:32:08	77.47
63	21-Aug-04	8:00:29	44.47
64	22-Aug-04	6:43:08	22.71
65	22-Aug-04	20:53:12	14.17
66	26-Aug-04	19:44:51	94.86
67	30-Aug-04	5:55:18	82.17
68	6-Sep-04	16:01:15	178.10
69	7-Sep-04	3:23:01	11.36
70	7-Sep-04	11:12:33	7.83
71	9-Sep-04	1:41:25	38.48
72	9-Sep-04	13:34:18	11.88
73	16-Sep-04	19:56:37	174.37
74	25-Sep-04	21:21:03	217.41
75	29-Sep-04	9:51:53	84.51
76	30-Sep-04	15:25:19	29.56
77	5-Oct-04	21:29:43	126.07
78	9-Oct-04	8:17:06	82.79
79	28-Oct-04	8:21:52	456.08
80	31-Oct-04	13:26:27	77.08
81	4-Nov-04	2:40:24	85.23
82	6-Nov-04	17:33:48	62.89
83	10-Nov-04	14:56:57	93.39
84	12-Nov-04	12:19:01	45.37
85	16-Nov-04	13:58:29	97.66
86	5-Dec-04	19:41:43	461.72
87	11-Dec-04	12:15:57	136.57
88	11-Dec-04	16:28:09	4.20
89	11-Dec-04	16:45:27	0.29
90	13-Dec-04	18:52:04	50.11
91	16-Dec-04	18:18:14	71.44
92	20-Dec-04	2:05:52	79.79
93	26-Dec-04	3:51:13	145.76
94	26-Dec-04	8:47:45	4.94
95	26-Dec-04	12:18:52	3.52
96	26-Dec-04	20:14:49	7.93
97	26-Dec-04	22:14:03	1.99
98	27-Dec-04	1:22:23	3.14
99	27-Dec-04	4:06:14	2.73
100	27-Dec-04	11:19:48	7.23
101	27-Dec-04	19:03:21	7.73
102	28-Dec-04	11:17:44	16.24
103	28-Dec-04	14:30:42	3.22
104	29-Dec-04	7:52:52	17.37
105	31-Dec-04	10:58:25	51.09
106	31-Dec-04	19:03:39	8.09
107	5-Jan-05	14:34:33	115.52
108	6-Jan-05	0:56:29	10.37
109	7-Jan-05	15:35:02	38.64
110	7-Jan-05	17:33:25	1.97
111	8-Jan-05	20:12:59	26.66
112	9-Jan-05	13:15:53	17.05
113	10-Jan-05	5:40:15	16.41
114	12-Jan-05	13:58:19	56.30
115	13-Jan-05	8:52:44	18.91
116	14-Jan-05	17:08:35	32.26
117	16-Jan-05	0:54:25	31.76
118	16-Jan-05	22:57:40	22.05
119	16-Jan-05	22:58:23	0.01
120	18-Jan-05	10:51:31	35.89
121	19-Jan-05	21:24:27	34.55
122	20-Jan-05	22:47:22	25.38
123	21-Jan-05	18:55:56	20.14
124	22-Jan-05	2:24:14	7.47
125	24-Jan-05	17:59:22	63.59
126	25-Jan-05	9:54:23	15.92
127	27-Jan-05	23:05:03	61.18
128	6-Feb-05	15:50:47	232.76
129	10-Feb-05	19:17:55	99.45
130	13-Feb-05	1:22:08	54.07
131	14-Feb-05	15:46:22	38.40
132	14-Feb-05	17:06:52	1.34
133	16-Feb-05	0:25:22	31.31
134	17-Feb-05	5:31:27	29.10
135	17-Feb-05	10:57:58	5.44
136	18-Feb-05	18:07:04	31.15
137	19-Feb-05	17:02:15	22.92
138	21-Feb-05	23:19:37	54.29
139	22-Feb-05	0:10:10	0.84
140	26-Feb-05	6:52:52	102.71
141	26-Feb-05	13:27:14	6.57
142	4-Mar-05	11:07:42	141.67
143	6-Mar-05	23:36:10	60.47
144	8-Mar-05	3:51:49	28.26
145	8-Mar-05	6:22:36	2.51
146	12-Mar-05	22:33:12	112.18
147	13-Mar-05	22:12:44	23.66
148	14-Mar-05	9:37:06	11.41
149	17-Mar-05	23:20:47	85.73
150	20-Mar-05	8:08:41	56.80
151	22-Mar-05	10:03:58	49.92

Table 2.4. Time series of the 151 time intervals (in hours) between earthquakes. Date and Time indicate the time of detected earthquake. Hours indicates the time between the preceding earthquake and the current one.

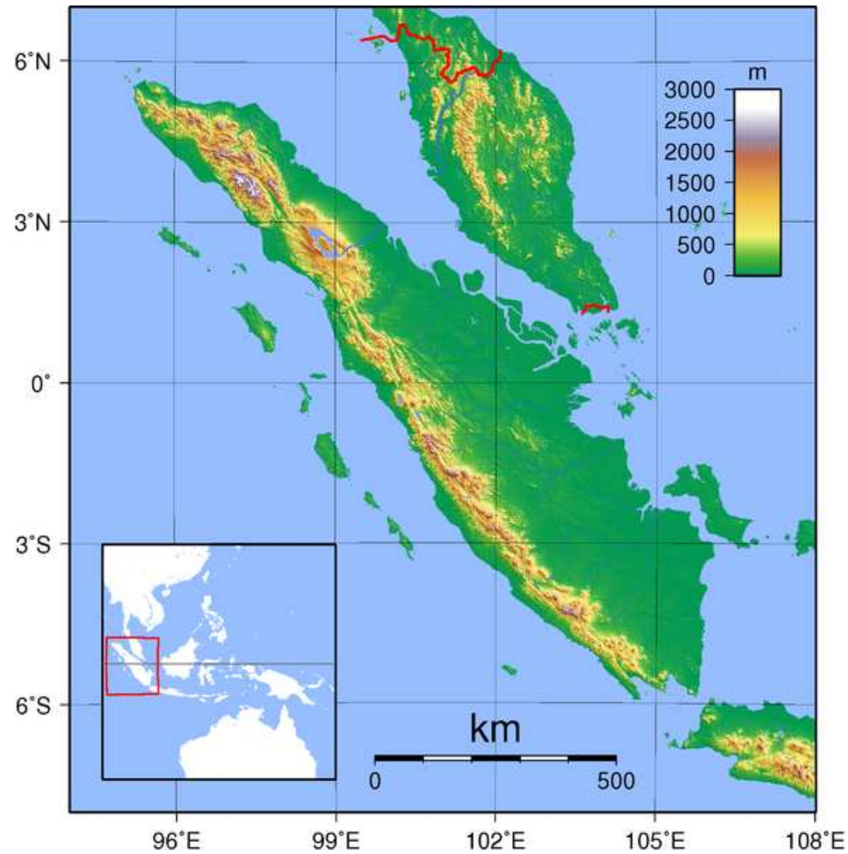


Figure 2.4. Topographic map of Sumatra, West Indonesia. Our region of study is defined by latitude of $6°$ N to $6°$ S and longitude of $94°$ E to $107°$ E.

2.4.2 Analysis of earthquake data

We will now conduct a change-point analysis of the earthquake interval data presented in Table 2.4 above. We begin with detection, in other words, do we have evidence that a change-point exists in the distribution of the data? If we conclude that a change-point exists we will apply estimation methodology to determine the asymptotic distribution of the change-point maximum likelihood estimate.

We start with 151 observations $Y_1, Y_2, \dots, Y_{151}$, representing the time interval in hours between the i^{th} and $(i-1)^{st}$, $i = 1, 2, \dots, 151$ observed earthquake, that we assume to be independent. Our null hypothesis is that there is no change in the distribution of the observations, while the alternative hypothesis is that a change in distribution has occurred at some unknown point in time, which we call τ_n, so that $Y_1, Y_2, \dots, Y_{\tau_n - 1} \sim f_1$ and $Y_{\tau_n}, Y_{\tau_n + 1}, \dots, Y_{151} \sim f_2$. We assume that the data is exponential with $Y_i \sim Exp(v_i)$, thus we can write out the hypotheses as follows

$$H_0\colon\ v_1 = \cdots = v_n \qquad vs. \qquad H_1\colon\ v_1 = \cdots = v_{\tau_n - 1} \neq v_{\tau_n} = \cdots = v_n$$

We then take the likelihood approach to hypothesis testing and calculate the likelihood ratio statistic U_n, its asymptotic theory is presented in Csörgö & Horváth (1997), see Appendix [A.4] for details. It should be noted that asymptotic theory of the likelihood ratio statistic U_n is quite robust to departures from the assumptions of both exponentiality and independence. Thus let

$$U_n = \max_{1 \le t < n} (-2 \log \Lambda_{\mathrm{t}})$$

where $-2 \log \Lambda_{\mathrm{t}} = 2\{\mathrm{k} \log \hat{v}_{\mathrm{t}} + (\mathrm{n} - \mathrm{k}) \log \hat{v}_{\mathrm{t}}^{*} - \mathrm{n} \log \hat{v}_{\mathrm{n}}\}$ and $\frac{1}{\hat{v}_{\mathrm{n}}} = \frac{1}{n} \sum_{i=1}^{n} Y_i$, $\frac{1}{\hat{v}_{\mathrm{t}}} = \frac{1}{k} \sum_{i=1}^{t} Y_i$, $and\, \frac{1}{\hat{v}_{\mathrm{k}}^{*}} = \frac{1}{n-t} \sum_{i=t+1}^{n} Y_i$. See Figure 2.5 for a graph of the $-2 \log \Lambda_{\mathrm{t}}$. Therefore if U_n is large we will reject the null hypothesis and conclude that the change-point occurs at $t+1$, i.e., $\hat{\tau} = t + 1$. From Csörgö and Horváth (1997) we know that the asymptotic distribution of the test statistic is as follows

$$P\left\{W \leq (2\log\log n\, U_n)^{1/2} - \left(2\log\log n + \frac{d}{2}\log\log\log n - \log\Gamma\left(\frac{d}{2}\right)\right) = w\right\} \sim e^{-2e^{-w}}$$

For the earthquake data, with d= 1 and $n = 151$, we obtain the following: $U_n = 29.22$, $w = 6.817$ and $\hat{\tau} = 94$. Then we calculate the p-value and obtain the following result

$$P\{W > 6.817\} + P\{W \leq -6.817\} = \left(1 - e^{-2e^{-6.817}}\right) + e^{-2e^{6.817}} = 0.00219$$

Now that we have strong evidence, due to the very small p-value, that there is a change in the parameter of the underlying exponential distribution at an unknown change-point τ_n which we estimate to be 94, we should test the underlying assumptions of independence and exponentiality before proceeding with determining the asymptotic distribution of the change-point mle.

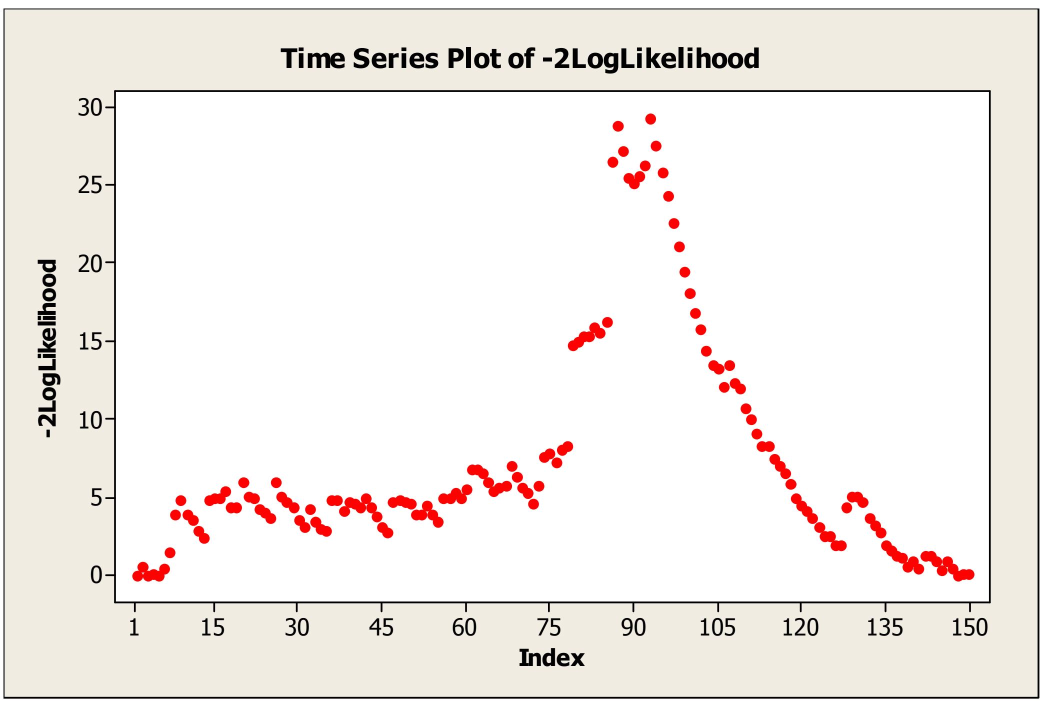


Figure 2.5. Plot of $-2\log\Lambda_t$ of the earthquake dataset.

We validate the assumption of independence by generating autocorrelation and partial autocorrelation plots of the standardized data, taking into account the different rates before and after the change and checking for significant lags. As you can see in Figures 2.6 and 2.7 below, none of the lags are significant, thus we can conclude that the assumption of independence is not violated. To check for exponentiality we utilize the Anderson-Darling goodness of fit test which we apply to the data before the change-point, and then to the data after the change-point. As you can see in Figure 2.8 the p-values are 0.726 and 0.320 signifying that the assumption of exponentiality was not violated.

We are able to assume that the two parameter estimates are true parameter values, $\hat{v}_1 = v_1 = 0.0108$ and $\hat{v}_2 = v_2 = 0.0280$ since Hinkley (1972) showed the asymptotic distribution of the change-point maximum likelihood estimate will not be affected if we use estimates rather than the unknown parameters in our calculations. We note that unlike the previous sections of the chapter where we assumed $v_1 > v_2$ in this case we have $v_1 < v_2$, so we need to reverse the time series so that $\tilde{v}_1 = 0.0280 > \tilde{v}_2 = 0.0108$ obtaining a ratio of 2.592 to use with the expressions we have derived previously, then we can compute the distribution of ξ_∞ based on sub chapter 2.3 and $(2.1.43)$. Once we have done the computations we can work out a 95% confidence interval for the true change-point τ_n, this turns out to cover points $\{80, 81, \dots, 104\}$ and corresponds to the interval from October 31st, 2004 13:26 through December 29th, 2004 7:53, with the mle estimate corresponding to December 26th, 2004 8:47. See Table 2.5 below for probabilities and corresponding dates and times.

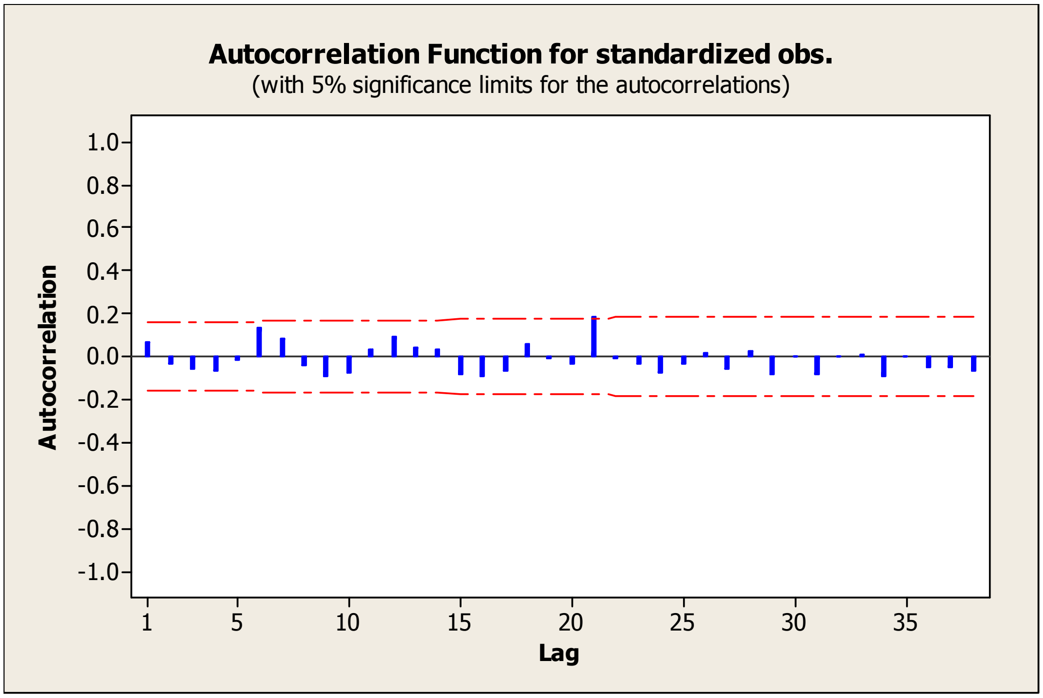


Figure 2.6. Autocorrelation for the standardized dataset.

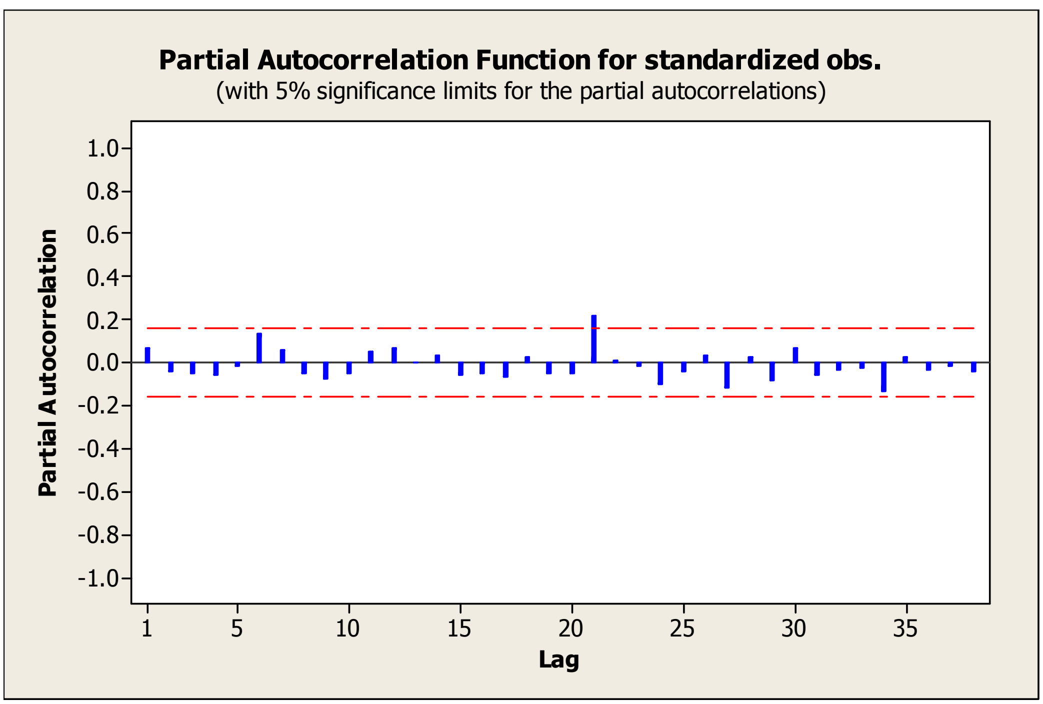


Figure 2.7. Partial Autocorrelation for the standardized dataset.

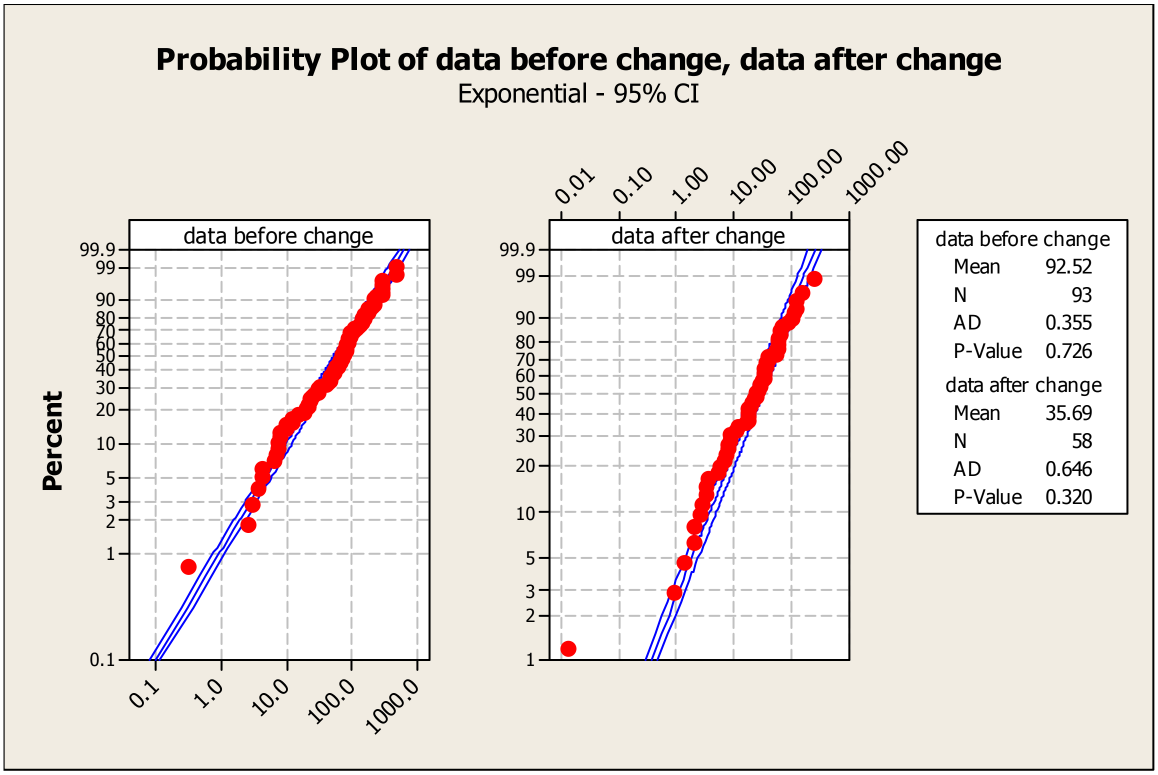


Figure 2.8. Exponential probability plot for data before and after change. Anderson-Darling statistics are 0.355 and 0.646 respectively. The p-values are 0.726 and 0.320 respectively.

TABLE 2.5.

95% confidence interval and $P(\xi_\infty = n)$ for $v_2/v_1 = 2.592$.

n	*prob.*	*index*	*day*	*time*
-14	0.0046	80	10/31/04	13:26
-13	0.0056	81	11/04/04	2:40
-12	0.0068	82	11/06/04	17:33
-11	0.0083	83	11/10/04	14:56
-10	0.0102	84	11/12/04	12:19
-9	0.0126	85	11/16/04	13:58
-8	0.0158	86	12/05/04	19:41
-7	0.0200	87	12/11/04	12:15
-6	0.0257	88	12/11/04	16:28
-5	0.0335	89	12/11/04	16:45
-4	0.0448	90	12/13/04	18:52
-3	0.0617	91	12/16/04	18:18
-2	0.0889	92	12/20/04	2:05
-1	0.1380	93	12/26/04	3:51
0	0.2468	94	12/26/04	8:47
1	0.0752	95	12/26/04	12:18
2	0.0447	96	12/26/04	20:14
3	0.0302	97	12/26/04	22:14
4	0.0217	98	12/27/04	1:22
5	0.0161	99	12/27/04	4:06
6	0.0124	100	12/27/04	11:19
7	0.0096	101	12/27/04	19:03
8	0.0076	102	12/28/04	11:17
9	0.0060	103	12/28/04	14:30
10	0.0049	104	12/29/04	7:52

Table 2.5. Probabilities and corresponding dates and times for $v_2/v_1 = 2.592$.

In conclusion, the Sumatra-Andaman earthquake, of magnitude 9.1, occurred on December 26th, 2004 at 00:58 and our point estimate for the change-point corresponds to December 26th, 2004 8:47, in addition the 95% confidence interval for the true change-point clearly includes the time and date of the occurrence of the great Sumatra-Andaman earthquake. Thus we can conclude that we do in fact have evidence that dynamic triggering has occurred due to seismic waves from the event increasing stress changes and shortening the periods between occurrences of the subsequent earthquakes as shown by the change in the estimated rate from 0.0280 to 0.0108.

2.5 Simulations

In this section we investigate, through simulations, the accuracy of the asymptotic distribution we have calculated earlier for varying sample sizes n, amounts of change v_1/v_2, and true change-point location τ. In addition we look at the conditional distribution of the change-point mle proposed by Cobb (1978) and compare it to the unconditional distribution we have derived in section 2.3.

We chose the following sample sizes and corresponding change-points for our simulations,

Sample sizes (n)	Change-points (τ)
500	250
200	100
100	10, 20, 30, 40, 50, 60, 70, 80, 90
60	10, 20, 30, 50
40	10, 20, 30

for each we consider $v_1/v_2 = 2, 3, 6, 10, 14, 18, \text{and } 22$. Tables 2.6 through 2.11 show a detailed comparison of the simulated and theoretical results for $v_1/v_2 = 2, 6, \text{and } 14$ for $k = -20, \dots, +20$. Table 2.12 shows the Bias and MSE values calculated for selected values of n and τ with varying amounts of change v_1/v_2. We find that the Bias is near zero in all cases when $v_1/v_2 > 3$ and also when we consider large sample sizes (n) and the change-point τ is located near the middle of the dataset regardless of the size of change v_1/v_2. It is interesting to note that in the cases of small changes (i.e., $v_1/v_2 < 6$) the Bias, when the change-point occurs close to the start of the dataset, is always larger than the Bias when τ is closer to the end of the dataset. For example, when $v_1/v_2 = 2$ and $n = 100$ the mle Bias for $\tau = 10$ is 1.328 while the Bias for $\tau = 90$ is -2.092. We also note that (when v_1/v_2 is small) Bias is positive when the change τ occurs early and negative when τ occurs late, this is most likely due to the asymmetry of the distribution. In general, as sample size decreases Bias becomes smaller when the change occurs early or late and larger when the change occurs towards the center of the data. Therefore, Bias is not a significant issue when v_1/v_2 is greater than 3, the sample size is large, or if the sample size is small the change-point is not near the center of the data series.

We also note that as long as n is large (i.e., $n > 100$) or v_1/v_2 is at least 6 the mle MSE values are very close to the theoretical MSE values. For small changes (i.e., $v_1/v_2 < 6$) the MSE becomes smaller, less than the theoretical value, as sample size decreases or as the change-point location moves further away from the center of the data series. It is of note that the MSE value is smaller if the change-point is at the upper end of the data series than if it is at the lower end. For example, when $v_1/v_2 = 2$ and $n = 100$ the mle MSE for $\tau = 10$ is 70.293 while the MSE for $\tau = 90$ is 50.839. As before, this is probably due to the asymmetry of the distribution of the mle.

Our derived distribution, based on Hinkley (1970, 1972), depends strictly on the parameters of the underlying distribution. In his paper Cobb (1978) suggests taking Hinkley's approach and then conditioning on a sufficient number of data values to either side of the change-point mle. His conditional solution is as follows

$$P(\tau - \hat{\tau} = k | X_{\delta}^{*}(\hat{\tau}) = y) = P(\tau - \hat{\tau} = k | Y_{\hat{\tau}-\delta+1}, \dots, Y_{\hat{\tau}+\delta})$$

$$\cong p_n(Y; \hat{\tau} + k) \Big/ \sum_{k=-\delta}^{\delta} p_n(Y; \hat{\tau} + k)$$

See Cobb (1978) for the method of choosing δ and details of the distribution, the error rate, and its calculation. Note that we must have $k \in \{-\delta, \dots, \delta\}$.

From Tables 2.6 – 2.11 we see that the simulated distribution of the mle clearly matches the theoretical distribution, regardless of the amount of change, sample size, or change-point location. The only thing of note is that the simulated mle probabilities tend to be a bit larger than the theoretical probabilities when the change-point is very far away from the center of the data series and the amount of change is small ($v_1/v_2 = 2$).

Tables 2.13 – 2.16 and Figures 2.9 (a)-(f) and 2.10 (a)-(f) compare the theoretical and simulated (conditional and unconditional) distribution of the mle for v_1/v_2 = 2, 6, 10, 14, 18 and 22, when n = 40, 200 and τ =10, 20, 30, 100. (In all our calculations of the conditional distribution of the mle we try to get an error rate that is as close to 10^{-5} as possible, but in many cases, especially when sample size is small, the smallest error rate we can obtain is 10^{-1}). As we can see, the theoretical distribution and the simulated unconditional distribution closely match and both are asymmetric, regardless of the amount of change, sample size, and change-point location. The only time the simulated distribution of the mle is not a near perfect match is when the sample size and amount of change are small and the change-point is close to the edge of the dataset, see Figure 2.10 (a). The simulated conditional distribution of the mle is more symmetric and flatter than the theoretical distribution. Thus we can conclude that the unconditional distribution of the mle outperforms the conditional distribution of the mle in all the cases we have considered.

TABLE 2.6 $v_1/v_2 = 2$

Comparison of theoretical and simulated results for changes in the true change-point (τ) location.

k	*Theor.*	n = 100								
		τ = 10	τ = 20	τ = 30	τ = 40	τ = 50	τ = 60	τ = 70	τ = 80	τ = 90
-20	0.0025	0.0000	0.0000	0.0029	0.0024	0.0026	0.0029	0.0026	0.0022	0.0023
-15	0.0044	0.0000	0.0051	0.0048	0.0040	0.0043	0.0044	0.0047	0.0043	0.0048
-10	0.0085	0.0000	0.0088	0.0086	0.0083	0.0088	0.0084	0.0086	0.0085	0.0089
-9	0.0099	0.0191	0.0100	0.0105	0.0097	0.0103	0.0104	0.0100	0.0112	0.0112
-8	0.0116	0.0182	0.0133	0.0122	0.0125	0.0119	0.0114	0.0111	0.0111	0.0122
-7	0.0136	0.0185	0.0148	0.0138	0.0128	0.0139	0.0139	0.0138	0.0138	0.0151
-6	0.0164	0.0194	0.0176	0.0170	0.0155	0.0172	0.0162	0.0164	0.0166	0.0175
-5	0.0198	0.0243	0.0206	0.0192	0.0215	0.0209	0.0199	0.0200	0.0191	0.0212
-4	0.0244	0.0281	0.0250	0.0245	0.0242	0.0243	0.0249	0.0253	0.0252	0.0255
-3	0.0311	0.0344	0.0320	0.0308	0.0308	0.0303	0.0302	0.0319	0.0302	0.0329
-2	0.0416	0.0440	0.0419	0.0423	0.0426	0.0410	0.0413	0.0431	0.0424	0.0460
-1	0.0614	0.0690	0.0598	0.0612	0.0638	0.0619	0.0626	0.0605	0.0637	0.0647
0	0.1534	0.1657	0.1549	0.1554	0.1533	0.1525	0.1543	0.1540	0.1561	0.1673
1	0.1003	0.1063	0.1030	0.1001	0.1010	0.1002	0.0995	0.1027	0.1004	0.1078
2	0.0726	0.0767	0.0760	0.0736	0.0738	0.0735	0.0741	0.0731	0.0740	0.0796
3	0.0556	0.0591	0.0591	0.0567	0.0558	0.0557	0.0562	0.0549	0.0570	0.0622
4	0.0440	0.0460	0.0442	0.0449	0.0446	0.0428	0.0425	0.0445	0.0467	0.0492
5	0.0358	0.0370	0.0364	0.0352	0.0369	0.0359	0.0363	0.0354	0.0362	0.0447
6	0.0296	0.0302	0.0300	0.0295	0.0296	0.0293	0.0301	0.0303	0.0307	0.0376
7	0.0248	0.0256	0.0248	0.0273	0.0238	0.0252	0.0245	0.0248	0.0263	0.0358
8	0.0210	0.0210	0.0204	0.0211	0.0207	0.0214	0.0198	0.0205	0.0219	0.0335
9	0.0179	0.0178	0.0176	0.0169	0.0179	0.0188	0.0173	0.0183	0.0185	0.0560
10	0.0154	0.0150	0.0157	0.0148	0.0161	0.0149	0.0158	0.0151	0.0172	0.0000
15	0.0079	0.0081	0.0081	0.0075	0.0074	0.0083	0.0075	0.0078	0.0097	0.0000
20	0.0044	0.0045	0.0040	0.0045	0.0041	0.0042	0.0044	0.0047	0.0000	0.0000

Table 2.6. Comparison of theoretical and simulated distribution of the mle, when the true change-point (τ) location changes for $v_1/v_2 = 2$.

TABLE 2.7 $v_1/v_2 = 6$

Comparison of theoretical and simulated results for changes in change-point (τ) location.

		n = 100								
k	*Theor.*	τ = 10	τ = 20	τ = 30	τ = 40	τ = 50	τ = 60	τ = 70	τ = 80	τ = 90
-20	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-15	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001
-10	0.0002	0.0000	0.0003	0.0003	0.0003	0.0002	0.0002	0.0002	0.0001	0.0003
-9	0.0004	0.0006	0.0002	0.0005	0.0004	0.0003	0.0003	0.0003	0.0003	0.0004
-8	0.0006	0.0007	0.0005	0.0007	0.0008	0.0006	0.0006	0.0007	0.0006	0.0006
-7	0.0011	0.0011	0.0011	0.0013	0.0011	0.0012	0.0011	0.0010	0.0011	0.0007
-6	0.0019	0.0019	0.0020	0.0021	0.0017	0.0016	0.0019	0.0015	0.0024	0.0021
-5	0.0035	0.0034	0.0040	0.0034	0.0038	0.0038	0.0028	0.0035	0.0032	0.0032
-4	0.0065	0.0068	0.0064	0.0060	0.0067	0.0067	0.0066	0.0066	0.0065	0.0064
-3	0.0129	0.0134	0.0132	0.0121	0.0132	0.0119	0.0126	0.0128	0.0127	0.0129
-2	0.0278	0.0280	0.0280	0.0282	0.0282	0.0281	0.0277	0.0287	0.0290	0.0278
-1	0.0730	0.0746	0.0711	0.0727	0.0730	0.0755	0.0730	0.0747	0.0730	0.0738
0	0.5347	0.5335	0.5353	0.5352	0.5340	0.5321	0.5362	0.5352	0.5362	0.5340
1	0.1790	0.1773	0.1794	0.1810	0.1782	0.1802	0.1789	0.1770	0.1771	0.1818
2	0.0769	0.0781	0.0764	0.0758	0.0771	0.0758	0.0733	0.0765	0.0778	0.0761
3	0.0372	0.0374	0.0369	0.0369	0.0369	0.0367	0.0397	0.0358	0.0364	0.0371
4	0.0193	0.0191	0.0201	0.0191	0.0197	0.0196	0.0204	0.0193	0.0191	0.0193
5	0.0105	0.0099	0.0113	0.0096	0.0104	0.0113	0.0104	0.0103	0.0101	0.0100
6	0.0059	0.0061	0.0062	0.0063	0.0056	0.0054	0.0053	0.0064	0.0060	0.0060
7	0.0034	0.0036	0.0028	0.0032	0.0038	0.0035	0.0035	0.0033	0.0032	0.0036
8	0.0020	0.0018	0.0017	0.0021	0.0021	0.0018	0.0021	0.0024	0.0018	0.0020
9	0.0012	0.0012	0.0010	0.0012	0.0013	0.0014	0.0012	0.0013	0.0013	0.0017
10	0.0007	0.0007	0.0007	0.0010	0.0006	0.0007	0.0008	0.0007	0.0006	0.0000
15	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0000	0.0001	0.0000	0.0000
20	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 2.7 Comparison of theoretical and simulated distribution of the mle, when the true change-point (τ) location changes for $v_1/v_2 = 6$.

TABLE 2.8 $v_1/v_2 = 14$

Comparison of theoretical and simulated results for changes in change-point (τ) location.

		n = 100								
k	*Theor.*	τ = 10	τ = 20	τ = 30	τ = 40	τ = 50	τ = 60	τ = 70	τ = 80	τ = 90
-20	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-15	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-8	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-7	0.0000	0.0001	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000	0.0001
-6	0.0001	0.0000	0.0001	0.0001	0.0002	0.0001	0.0002	0.0001	0.0001	0.0001
-5	0.0003	0.0002	0.0005	0.0005	0.0002	0.0005	0.0003	0.0003	0.0004	0.0006
-4	0.0010	0.0009	0.0009	0.0012	0.0011	0.0010	0.0010	0.0009	0.0010	0.0009
-3	0.0030	0.0029	0.0030	0.0031	0.0030	0.0027	0.0031	0.0031	0.0031	0.0028
-2	0.0105	0.0107	0.0110	0.0103	0.0106	0.0106	0.0111	0.0104	0.0110	0.0111
-1	0.0463	0.0450	0.0468	0.0442	0.0468	0.0469	0.0452	0.0464	0.0460	0.0461
0	0.7401	0.7398	0.7398	0.7422	0.7395	0.7391	0.7425	0.7379	0.7403	0.7415
1	0.1429	0.1444	0.1436	0.1434	0.1427	0.1439	0.1418	0.1438	0.1414	0.1430
2	0.0380	0.0383	0.0367	0.0385	0.0387	0.0374	0.0375	0.0394	0.0384	0.0365
3	0.0117	0.0115	0.0119	0.0112	0.0109	0.0121	0.0118	0.0118	0.0123	0.0112
4	0.0039	0.0039	0.0038	0.0033	0.0043	0.0036	0.0034	0.0039	0.0038	0.0038
5	0.0014	0.0014	0.0012	0.0015	0.0012	0.0013	0.0014	0.0011	0.0015	0.0015
6	0.0005	0.0006	0.0004	0.0004	0.0004	0.0004	0.0004	0.0005	0.0004	0.0004
7	0.0002	0.0002	0.0002	0.0001	0.0002	0.0001	0.0002	0.0001	0.0002	0.0001
8	0.0001	0.0001	0.0001	0.0001	0.0000	0.0001	0.0001	0.0001	0.0001	0.0001
9	0.0000	0.0001	0.0000	0.0000	0.0001	0.0001	0.0000	0.0000	0.0000	0.0001
10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
15	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
20	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 2.8. Comparison of theoretical and simulated distribution of the mle, when the true change-point (τ) location changes for $v_1/v_2 = 14$.

TABLE 2.9 $v_1/v_2 = 2$

Comparison of theoretical and simulated results for changes in change-point (τ) location.

		n = 60					n = 40			n = 500
k	*Theor.*	τ = 10	τ = 20	τ = 30	τ = 40	τ = 50	τ = 10	τ = 20	τ = 30	τ = 250
-20	0.0025	0.0000	0.0000	0.0025	0.0025	0.0026	0.0000	0.0000	0.0032	0.0023
-15	0.0044	0.0000	0.0050	0.0039	0.0045	0.0046	0.0000	0.0049	0.0050	0.0043
-10	0.0085	0.0000	0.0095	0.0089	0.0087	0.0098	0.0000	0.0090	0.0090	0.0082
-6	0.0164	0.0203	0.0162	0.0161	0.0171	0.0174	0.0201	0.0164	0.0172	0.0161
-5	0.0198	0.0237	0.0200	0.0207	0.0207	0.0211	0.0239	0.0205	0.0212	0.0201
-4	0.0244	0.0288	0.0260	0.0248	0.0242	0.0262	0.0288	0.0253	0.0260	0.0243
-3	0.0311	0.0335	0.0315	0.0315	0.0311	0.0328	0.0351	0.0307	0.0337	0.0316
-2	0.0416	0.0454	0.0438	0.0414	0.0422	0.0447	0.0450	0.0420	0.0444	0.0418
-1	0.0614	0.0660	0.0614	0.0653	0.0618	0.0659	0.0678	0.0641	0.0659	0.0618
0	0.1534	0.1606	0.1548	0.1569	0.1580	0.1638	0.1633	0.1591	0.1655	0.1537
1	0.1003	0.1066	0.1036	0.0995	0.1036	0.1097	0.1095	0.1039	0.1086	0.1008
2	0.0726	0.0763	0.0757	0.0726	0.0728	0.0793	0.0747	0.0766	0.0800	0.0700
3	0.0556	0.0590	0.0573	0.0563	0.0572	0.0624	0.0583	0.0563	0.0620	0.0549
4	0.0440	0.0469	0.0443	0.0445	0.0467	0.0515	0.0474	0.0461	0.0501	0.0446
5	0.0358	0.0387	0.0368	0.0366	0.0369	0.0437	0.0382	0.0397	0.0427	0.0365
6	0.0296	0.0321	0.0299	0.0305	0.0308	0.0382	0.0309	0.0300	0.0387	0.0296
10	0.0154	0.0158	0.0165	0.0158	0.0167	0.0000	0.0164	0.0161	0.0000	0.0152
15	0.0079	0.0084	0.0081	0.0083	0.0097	0.0000	0.0079	0.0093	0.0000	0.0080
20	0.0044	0.0048	0.0045	0.0045	0.0000	0.0000	0.0045	0.0000	0.0000	0.0043

Table 2.9. Comparison of theoretical and simulated distribution of the mle, when the true change-point (τ) location changes for $v_1/v_2 = 2$.

TABLE 2.10 $v_1/v_2 = 6$

Comparison of theoretical and simulated results for changes in change-point (τ) location.

		n = 60					n = 40			n = 500
k	*Theor.*	τ = 10	τ = 20	τ = 30	τ = 40	τ = 50	τ = 10	τ = 20	τ = 30	τ = 250
-20	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-15	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-10	0.0002	0.0000	0.0002	0.0002	0.0002	0.0001	0.0000	0.0002	0.0002	0.0003
-6	0.0019	0.0017	0.0016	0.0022	0.0021	0.0019	0.0016	0.0020	0.0021	0.0019
-5	0.0035	0.0036	0.0035	0.0036	0.0036	0.0030	0.0033	0.0035	0.0036	0.0036
-4	0.0065	0.0065	0.0062	0.0068	0.0066	0.0069	0.0062	0.0057	0.0064	0.0066
-3	0.0129	0.0127	0.0137	0.0131	0.0122	0.0128	0.0129	0.0120	0.0128	0.0124
-2	0.0278	0.0285	0.0268	0.0261	0.0301	0.0277	0.0279	0.0289	0.0273	0.0262
-1	0.0730	0.0739	0.0714	0.0752	0.0722	0.0738	0.0704	0.0716	0.0721	0.0743
0	0.5347	0.5378	0.5356	0.5289	0.5367	0.5337	0.5396	0.5384	0.5365	0.5352
1	0.1790	0.1773	0.1784	0.1847	0.1774	0.1806	0.1769	0.1782	0.1809	0.1780
2	0.0769	0.0765	0.0805	0.0776	0.0757	0.0781	0.0773	0.0748	0.0766	0.0788
3	0.0372	0.0364	0.0362	0.0353	0.0371	0.0364	0.0375	0.0385	0.0386	0.0375
4	0.0193	0.0188	0.0187	0.0195	0.0189	0.0190	0.0189	0.0186	0.0183	0.0183
5	0.0105	0.0103	0.0105	0.0108	0.0102	0.0104	0.0107	0.0111	0.0094	0.0102
6	0.0059	0.0062	0.0060	0.0058	0.0061	0.0058	0.0058	0.0056	0.0058	0.0058
10	0.0007	0.0007	0.0007	0.0008	0.0006	0.0000	0.0008	0.0006	0.0000	0.0009
15	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0000
20	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 2.10. Comparison of theoretical and simulated distribution of the mle, when the true change-point (τ) location changes for $v_1/v_2 = 6$.

TABLE 2.11 $v_1/v_2 = 14$

Comparison of theoretical and simulated results for changes in change-point (τ) location.

		n = 60					n = 40			n = 500
k	*Theor.*	τ = 10	τ = 20	τ = 30	τ = 40	τ = 50	τ = 10	τ = 20	τ = 30	τ = 250
-20	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-15	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-6	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0002	0.0001	0.0001
-5	0.0003	0.0004	0.0005	0.0003	0.0004	0.0003	0.0002	0.0003	0.0003	0.0004
-4	0.0010	0.0010	0.0010	0.0010	0.0012	0.0012	0.0010	0.0011	0.0008	0.0010
-3	0.0030	0.0033	0.0032	0.0031	0.0028	0.0030	0.0032	0.0028	0.0026	0.0029
-2	0.0105	0.0101	0.0107	0.0112	0.0103	0.0109	0.0106	0.0113	0.0107	0.0104
-1	0.0463	0.0463	0.0453	0.0460	0.0471	0.0471	0.0474	0.0466	0.0462	0.0469
0	0.7401	0.7390	0.7401	0.7386	0.7409	0.7397	0.7383	0.7393	0.7400	0.7400
1	0.1429	0.1448	0.1428	0.1419	0.1417	0.1416	0.1420	0.1434	0.1430	0.1427
2	0.0380	0.0374	0.0394	0.0389	0.0378	0.0379	0.0389	0.0368	0.0377	0.0385
3	0.0117	0.0118	0.0112	0.0127	0.0120	0.0118	0.0122	0.0120	0.0123	0.0127
4	0.0039	0.0038	0.0040	0.0040	0.0037	0.0042	0.0040	0.0040	0.0040	0.0027
5	0.0014	0.0012	0.0011	0.0015	0.0013	0.0014	0.0014	0.0014	0.0013	0.0011
6	0.0005	0.0006	0.0004	0.0004	0.0004	0.0005	0.0004	0.0005	0.0006	0.0004
10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000
15	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
20	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 2.11. Comparison of theoretical and simulated distribution of the mle, when the true change-point (τ) location changes for $v_1/v_2 = 14$.

TABLE 2.12 Bias and MSE

		$v_1/v_2 = 2$		$v_1/v_2 = 3$		$v_1/v_2 = 6$		$v_1/v_2 = 10$		$v_1/v_2 = 14$		$v_1/v_2 = 18$		$v_1/v_2 = 22$	
		Theor.	mle	Theor.	mle	Theor.	mle	Theor.	mle	Theor.	mle	Theor.	mle	Theor.	mle
n = 500	Bias	0.000	0.055	0.000	-0.004	0.000	-0.001	0.000	0.004	0.000	-0.004	0.000	-0.002	0.000	0.000
τ = 250	MSE	114.676	114.202	18.733	18.705	2.816	2.815	1.079	1.085	0.641	0.616	0.452	0.465	0.348	0.348
n = 200	Bias	0.000	-0.014	0.000	-0.005	0.000	-0.003	0.000	-0.006	0.000	0.000	0.000	0.000	0.000	-0.001
τ = 100	MSE	114.676	113.044	18.733	18.691	2.816	2.777	1.079	1.072	0.641	0.628	0.452	0.456	0.348	0.347
n = 100	Bias	0.000	1.328	0.000	0.218	0.000	-0.001	0.000	0.004	0.000	0.005	0.000	0.000	0.000	0.001
τ = 10	MSE	114.676	70.293	18.733	15.703	2.816	2.749	1.079	1.070	0.641	0.642	0.452	0.459	0.348	0.348
n = 100	Bias	0.000	0.613	0.000	0.062	0.000	0.002	0.000	0.002	0.000	-0.005	0.000	-0.002	0.000	0.002
τ = 20	MSE	114.676	83.356	18.733	17.463	2.816	2.820	1.079	1.106	0.641	0.635	0.452	0.446	0.348	0.348
n = 100	Bias	0.000	-0.051	0.000	-0.027	0.000	0.001	0.000	0.006	0.000	-0.002	0.000	0.001	0.000	0.005
τ = 50	MSE	114.676	99.672	18.733	18.518	2.816	2.854	1.079	1.119	0.641	0.638	0.452	0.454	0.348	0.349
n = 100	Bias	0.000	-0.932	0.000	-0.074	0.000	-0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002
τ = 80	MSE	114.676	74.305	18.733	17.402	2.816	2.768	1.079	1.060	0.641	0.647	0.452	0.447	0.348	0.351
n = 100	Bias	0.000	-2.092	0.000	-0.341	0.000	-0.012	0.000	-0.012	0.000	-0.005	0.000	-0.002	0.000	0.000
τ = 90	MSE	114.676	50.839	18.733	12.376	2.816	2.603	1.079	1.050	0.641	0.635	0.452	0.443	0.348	0.347
n = 60	Bias	0.000	1.276	0.000	0.179	0.000	-0.007	0.000	0.002	0.000	0.000	0.000	0.001	0.000	-0.002
τ = 10	MSE	114.676	62.824	18.733	15.049	2.816	2.675	1.079	1.065	0.641	0.635	0.452	0.446	0.348	0.346
n = 60	Bias	0.000	-0.117	0.000	0.004	0.000	-0.004	0.000	-0.008	0.000	0.002	0.000	0.002	0.000	0.004
τ = 30	MSE	114.676	70.487	18.733	17.798	2.816	2.796	1.079	1.074	0.641	0.654	0.452	0.461	0.348	0.351
n = 60	Bias	0.000	-2.015	0.000	-0.344	0.000	-0.011	0.000	0.004	0.000	-0.001	0.000	0.001	0.000	0.000
τ = 50	MSE	114.676	47.437	18.733	12.761	2.816	2.617	1.079	1.084	0.641	0.647	0.452	0.450	0.348	0.346
n = 40	Bias	0.000	0.907	0.000	0.167	0.000	0.016	0.000	-0.001	0.000	0.002	0.000	0.001	0.000	0.001
τ = 10	MSE	114.676	47.100	18.733	14.654	2.816	2.745	1.079	1.055	0.641	0.644	0.452	0.454	0.348	0.360
n = 40	Bias	0.000	-0.294	0.000	-0.030	0.000	0.002	0.000	0.001	0.000	-0.002	0.000	0.001	0.000	0.005
τ = 20	MSE	114.676	47.944	18.733	15.750	2.816	2.834	1.079	1.091	0.641	0.652	0.452	0.451	0.348	0.358
n = 40	Bias	0.000	-1.799	0.000	-0.330	0.000	-0.017	0.000	-0.005	0.000	0.003	0.000	-0.002	0.000	0.000
τ = 30	MSE	114.676	37.342	18.733	12.320	2.816	2.575	1.079	1.081	0.641	0.649	0.452	0.452	0.348	0.345

Table 2.12. Bias and MSE for v_1/v_2 = 2, 3, 6, 10, 14, 18, and 22 for varying sample sizes n and change-point τ locations.

TABLE 2.13

Comparison of theoretical and simulated (conditional and unconditional mle) results for v_1/v_2 = 2, 6, 10, 14, 18 and 22.

	$v_1/v_2 = 2$			$v_1/v_2 = 6$			$v_1/v_2 = 10$			$v_1/v_2 = 14$			$v_1/v_2 = 18$			$v_1/v_2 = 22$		
n	Theor	mle	c mle	Theor	mle	c mle	Theor	mle	c mle	Theor	mle	c mle	Theor	mle	c mle	Theor	mle	c mle
-20	0.002	0.003	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-15	0.004	0.005	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-10	0.009	0.009	0.015	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-9	0.010	0.011	0.018	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-8	0.012	0.012	0.020	0.001	0.001	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-7	0.014	0.015	0.023	0.001	0.001	0.004	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-6	0.016	0.017	0.026	0.002	0.002	0.006	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-5	0.020	0.021	0.030	0.003	0.004	0.011	0.001	0.001	0.003	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.000
-4	0.024	0.026	0.035	0.007	0.006	0.019	0.002	0.003	0.008	0.001	0.001	0.004	0.001	0.000	0.002	0.000	0.000	0.001
-3	0.031	0.034	0.041	0.013	0.013	0.036	0.006	0.006	0.019	0.003	0.003	0.012	0.002	0.002	0.008	0.001	0.001	0.005
-2	0.042	0.044	0.049	0.028	0.027	0.072	0.016	0.015	0.050	0.011	0.011	0.037	0.007	0.009	0.029	0.006	0.005	0.022
-1	0.061	0.066	0.059	0.073	0.072	0.155	0.057	0.057	0.149	0.046	0.046	0.134	0.039	0.039	0.122	0.033	0.033	0.110
0	0.153	0.165	0.072	0.535	0.537	0.385	0.670	0.673	0.537	0.740	0.740	0.621	0.784	0.782	0.676	0.814	0.817	0.721
1	0.100	0.109	0.058	0.179	0.181	0.156	0.162	0.161	0.149	0.143	0.143	0.135	0.127	0.129	0.124	0.115	0.113	0.109
2	0.073	0.080	0.047	0.077	0.077	0.071	0.053	0.051	0.049	0.038	0.038	0.037	0.029	0.029	0.028	0.023	0.023	0.023
3	0.056	0.062	0.038	0.037	0.039	0.036	0.020	0.020	0.019	0.012	0.012	0.012	0.008	0.007	0.007	0.005	0.006	0.006
4	0.044	0.050	0.032	0.019	0.018	0.018	0.008	0.008	0.008	0.004	0.004	0.004	0.002	0.002	0.002	0.001	0.001	0.001
5	0.036	0.043	0.026	0.010	0.009	0.010	0.003	0.003	0.003	0.001	0.001	0.001	0.001	0.001	0.001	0.000	0.000	0.000
6	0.030	0.039	0.022	0.006	0.006	0.006	0.001	0.001	0.001	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000
7	0.025	0.035	0.018	0.003	0.003	0.003	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.021	0.034	0.014	0.002	0.002	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9	0.018	0.059	0.011	0.001	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 2.13. Comparison of theoretical and simulated (conditional and unconditional) distribution of the mle for ν_1/ν_2 = 2, 6, 10, 14, 18 and 22, when n = 40 and τ = 30. (Error rate for c mle 10^{-1} to 10^{-5}).

TABLE 2.14

Comparison of theoretical and simulated (conditional and unconditional mle) results for v_1/v_2 = 2, 6, 10, 14, 18 and 22.

	$v_1/v_2 = 2$			$v_1/v_2 = 6$			$v_1/v_2 = 10$			$v_1/v_2 = 14$			$v_1/v_2 = 18$			$v_1/v_2 = 22$		
n	Theor	mle	c mle	Theor	mle	c mle	Theor	mle	c mle	Theor	mle	c mle	Theor	mle	c mle	Theor	mle	c mle
-19	0.003	0.005	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-15	0.004	0.005	0.008	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-10	0.009	0.009	0.016	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-6	0.016	0.016	0.027	0.002	0.002	0.006	0.000	0.001	0.002	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000
-5	0.020	0.020	0.032	0.003	0.003	0.010	0.001	0.001	0.003	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.000
-4	0.024	0.025	0.037	0.007	0.006	0.019	0.002	0.002	0.008	0.001	0.001	0.004	0.001	0.001	0.002	0.000	0.000	0.001
-3	0.031	0.031	0.044	0.013	0.012	0.036	0.006	0.005	0.019	0.003	0.003	0.011	0.002	0.002	0.008	0.001	0.001	0.005
-2	0.042	0.042	0.053	0.028	0.029	0.072	0.016	0.016	0.050	0.011	0.011	0.037	0.007	0.007	0.028	0.006	0.006	0.023
-1	0.061	0.064	0.065	0.073	0.072	0.155	0.057	0.056	0.148	0.046	0.047	0.135	0.039	0.039	0.122	0.033	0.033	0.110
0	0.153	0.159	0.079	0.535	0.538	0.387	0.670	0.672	0.535	0.740	0.739	0.621	0.784	0.786	0.678	0.814	0.811	0.717
1	0.100	0.104	0.064	0.179	0.178	0.155	0.162	0.159	0.148	0.143	0.143	0.135	0.127	0.125	0.121	0.115	0.117	0.112
2	0.073	0.077	0.053	0.077	0.075	0.071	0.053	0.054	0.052	0.038	0.037	0.036	0.029	0.029	0.028	0.023	0.024	0.023
3	0.056	0.056	0.045	0.037	0.038	0.036	0.020	0.020	0.019	0.012	0.012	0.012	0.008	0.008	0.008	0.005	0.006	0.006
4	0.044	0.046	0.038	0.019	0.019	0.019	0.008	0.008	0.008	0.004	0.004	0.004	0.002	0.002	0.002	0.001	0.001	0.001
5	0.036	0.040	0.032	0.010	0.011	0.011	0.003	0.003	0.003	0.001	0.001	0.001	0.001	0.001	0.001	0.000	0.001	0.001
6	0.030	0.030	0.027	0.006	0.006	0.006	0.001	0.002	0.002	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000
10	0.015	0.016	0.015	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
15	0.008	0.009	0.008	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
19	0.005	0.016	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 2.14. Comparison of theoretical and simulated (conditional and unconditional) distribution of the mle for ν_1/ν_2 = 2, 6, 10, 14, 18 and 22, when n = 40 and τ = 20. (Error rate for c mle 10^{-1} to 10^{-5}).

TABLE 2.15

Comparison of theoretical and simulated (conditional and unconditional mle) results for v_1/v_2 = 2, 4, 6, 10, and 20.

	$v_1/v_2 = 2$			$v_1/v_2 = 6$			$v_1/v_2 = 10$			$v_1/v_2 = 14$			$v_1/v_2 = 18$			$v_1/v_2 = 22$		
n	Theor	mle	c mle	Theor	mle	c mle	Theor	mle	c mle	Theor	mle	c mle	Theor	mle	c mle	Theor	mle	c mle
-9	0.010	0.019	0.013	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-8	0.012	0.017	0.015	0.001	0.001	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-7	0.014	0.018	0.018	0.001	0.001	0.003	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-6	0.016	0.020	0.022	0.002	0.002	0.005	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-5	0.020	0.024	0.026	0.003	0.003	0.010	0.001	0.001	0.003	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.000
-4	0.024	0.029	0.031	0.007	0.006	0.019	0.002	0.002	0.008	0.001	0.001	0.004	0.001	0.001	0.002	0.000	0.000	0.001
-3	0.031	0.035	0.038	0.013	0.013	0.035	0.006	0.005	0.019	0.003	0.003	0.012	0.002	0.002	0.007	0.001	0.001	0.005
-2	0.042	0.045	0.046	0.028	0.028	0.071	0.016	0.015	0.051	0.011	0.011	0.038	0.007	0.007	0.028	0.006	0.006	0.023
-1	0.061	0.068	0.056	0.073	0.070	0.154	0.057	0.057	0.149	0.046	0.047	0.135	0.039	0.040	0.122	0.033	0.034	0.111
0	0.153	0.163	0.070	0.535	0.540	0.385	0.670	0.671	0.534	0.740	0.738	0.620	0.784	0.784	0.679	0.814	0.811	0.717
1	0.100	0.109	0.058	0.179	0.177	0.153	0.162	0.161	0.149	0.143	0.142	0.135	0.127	0.126	0.121	0.115	0.117	0.112
2	0.073	0.075	0.049	0.077	0.077	0.072	0.053	0.052	0.050	0.038	0.039	0.037	0.029	0.030	0.029	0.023	0.023	0.022
3	0.056	0.058	0.042	0.037	0.037	0.036	0.020	0.020	0.020	0.012	0.012	0.012	0.008	0.008	0.008	0.005	0.005	0.005
4	0.044	0.047	0.036	0.019	0.019	0.019	0.008	0.008	0.008	0.004	0.004	0.004	0.002	0.002	0.002	0.001	0.002	0.002
5	0.036	0.038	0.031	0.010	0.011	0.010	0.003	0.003	0.003	0.001	0.001	0.001	0.001	0.001	0.001	0.000	0.000	0.000
6	0.030	0.031	0.027	0.006	0.006	0.006	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.025	0.026	0.023	0.003	0.003	0.003	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.021	0.022	0.020	0.002	0.002	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9	0.018	0.019	0.018	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
10	0.015	0.016	0.016	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
15	0.008	0.008	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20	0.004	0.004	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 2.15. Comparison of theoretical and simulated (conditional and unconditional) distribution of the mle for ν_1/ν_2 = 2, 6, 10, 14, 18 and 22, when n = 40 and τ = 10. (Error rate for c mle 10^{-1} to 10^{-5}).

TABLE 2.16

Comparison of theoretical and simulated (conditional and unconditional mle) results for v_1/v_2 = 2, 6, 10, 14, 18 and 20.

	$v_1/v_2 = 2$			$v_1/v_2 = 6$			$v_1/v_2 = 10$			$v_1/v_2 = 14$			$v_1/v_2 = 18$			$v_1/v_2 = 22$		
n	Theor	mle	c mle	Theor	mle	c mle	Theor	mle	c mle	Theor	mle	c mle	Theor	mle	c mle	Theor	mle	c mle
-20	0.002	0.002	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-15	0.004	0.005	0.008	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-10	0.009	0.009	0.016	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-6	0.016	0.017	0.027	0.002	0.002	0.006	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-5	0.020	0.021	0.031	0.003	0.003	0.010	0.001	0.001	0.003	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.000
-4	0.024	0.024	0.036	0.007	0.006	0.019	0.002	0.002	0.008	0.001	0.001	0.004	0.001	0.001	0.002	0.000	0.000	0.001
-3	0.031	0.031	0.043	0.013	0.013	0.036	0.006	0.006	0.019	0.003	0.003	0.011	0.002	0.002	0.008	0.001	0.001	0.005
-2	0.042	0.041	0.051	0.028	0.029	0.072	0.016	0.016	0.051	0.011	0.010	0.037	0.007	0.008	0.029	0.006	0.006	0.023
-1	0.061	0.061	0.062	0.073	0.073	0.156	0.057	0.059	0.150	0.046	0.046	0.136	0.039	0.039	0.122	0.033	0.033	0.110
0	0.153	0.156	0.076	0.535	0.538	0.386	0.670	0.673	0.537	0.740	0.741	0.622	0.784	0.780	0.675	0.814	0.815	0.720
1	0.100	0.102	0.062	0.179	0.179	0.155	0.162	0.159	0.147	0.143	0.143	0.134	0.127	0.130	0.124	0.115	0.114	0.110
2	0.073	0.073	0.051	0.077	0.075	0.071	0.053	0.051	0.050	0.038	0.038	0.037	0.029	0.029	0.028	0.023	0.023	0.022
3	0.056	0.055	0.043	0.037	0.036	0.035	0.020	0.019	0.019	0.012	0.012	0.012	0.008	0.007	0.007	0.005	0.005	0.005
4	0.044	0.043	0.036	0.019	0.020	0.019	0.008	0.008	0.008	0.004	0.004	0.004	0.002	0.002	0.002	0.001	0.002	0.001
5	0.036	0.036	0.031	0.010	0.010	0.010	0.003	0.003	0.004	0.001	0.002	0.001	0.001	0.001	0.001	0.000	0.000	0.000
6	0.030	0.030	0.027	0.006	0.006	0.006	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
10	0.015	0.015	0.015	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
15	0.008	0.008	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20	0.004	0.004	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 2.16. Comparison of theoretical and simulated (conditional and unconditional) distribution of the mle for ν_1/ν_2 = 2, 6, 10, 14, 18 and 22, when n = 200 and τ = 100. (Error rate for c mle 10^{-4} to 10^{-5}).

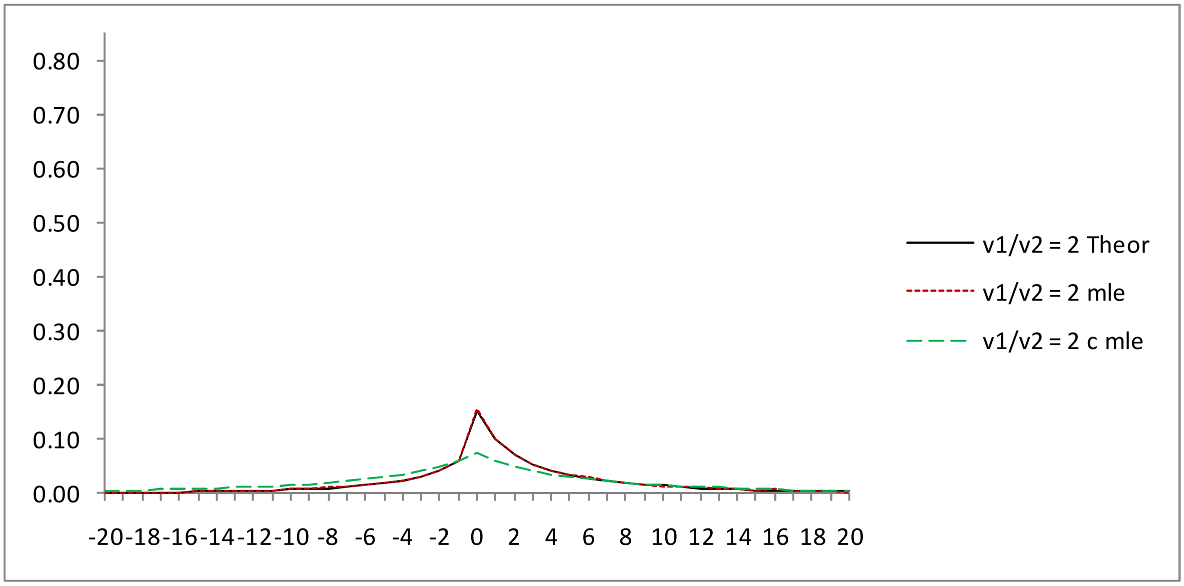


Figure 2.9 (a). Comparison of theoretical and simulated (conditional and unconditional) distribution of the mle for ν_1/ν_2 = 2, n = 200 and τ = 100.

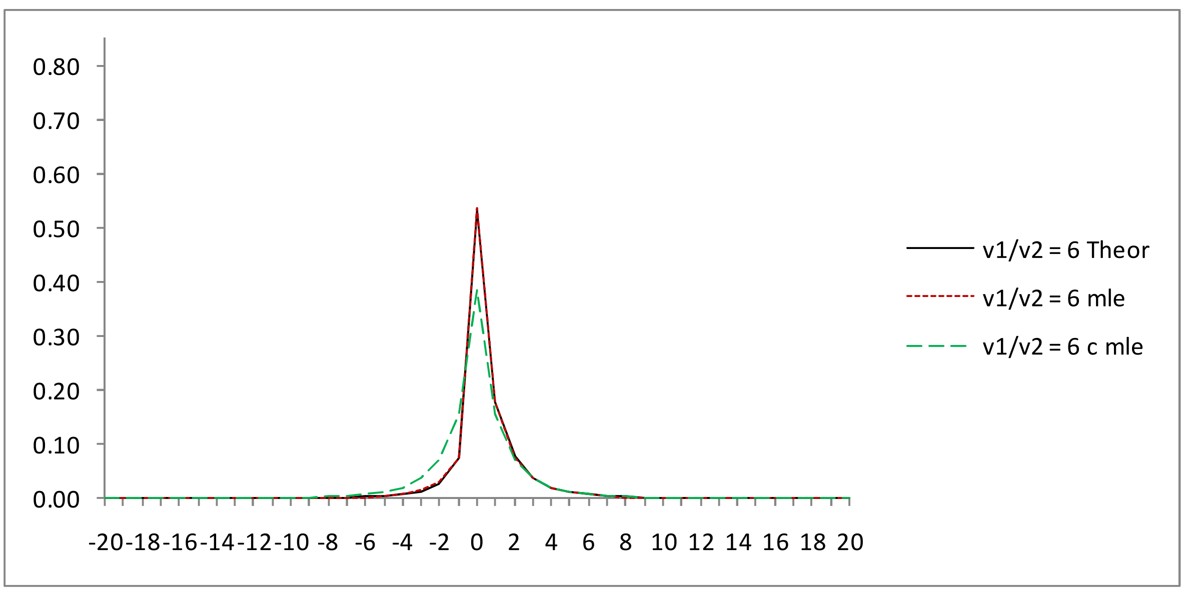


Figure 2.9 (a). Comparison of theoretical and simulated (conditional and unconditional) distribution of the mle for ν_1/ν_2 = 6, n = 200 and τ = 100.

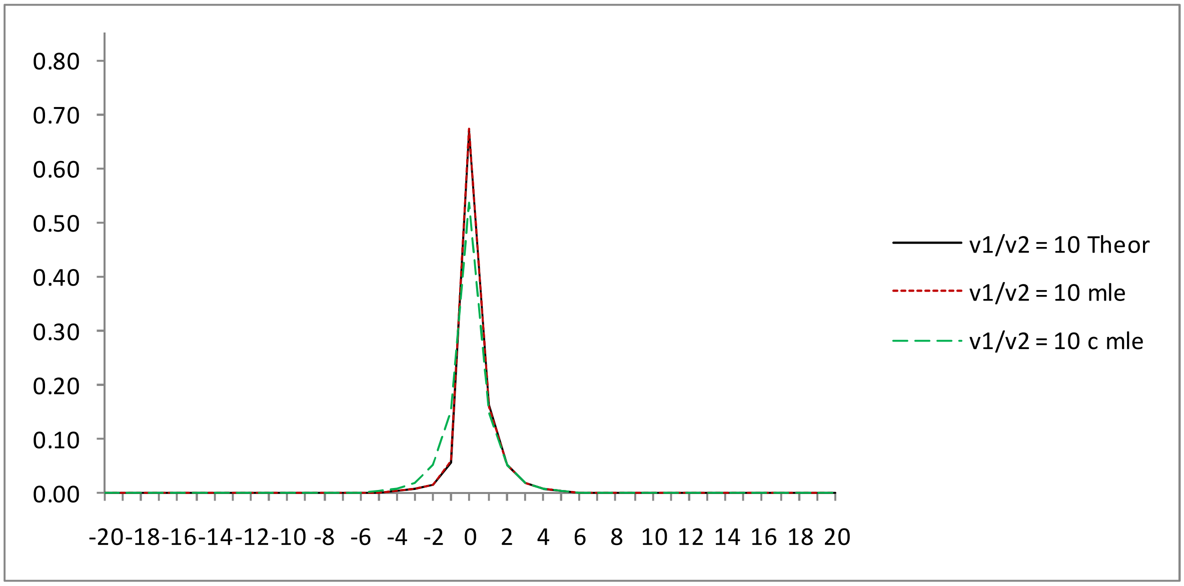


Figure 2.9 (c). Comparison of theoretical and simulated (conditional and unconditional) distribution of the mle for ν_1/ν_2 = 10, n = 200 and τ = 100.

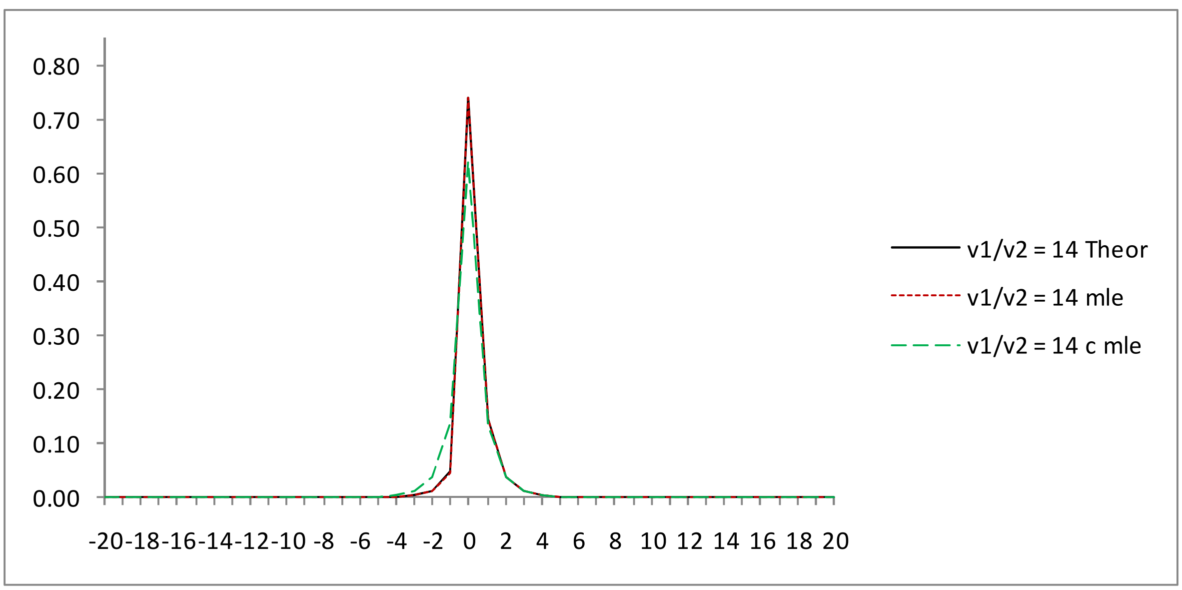


Figure 2.9 (d). Comparison of theoretical and simulated (conditional and unconditional) distribution of the mle for ν_1/ν_2 = 14, n = 200 and τ = 100.

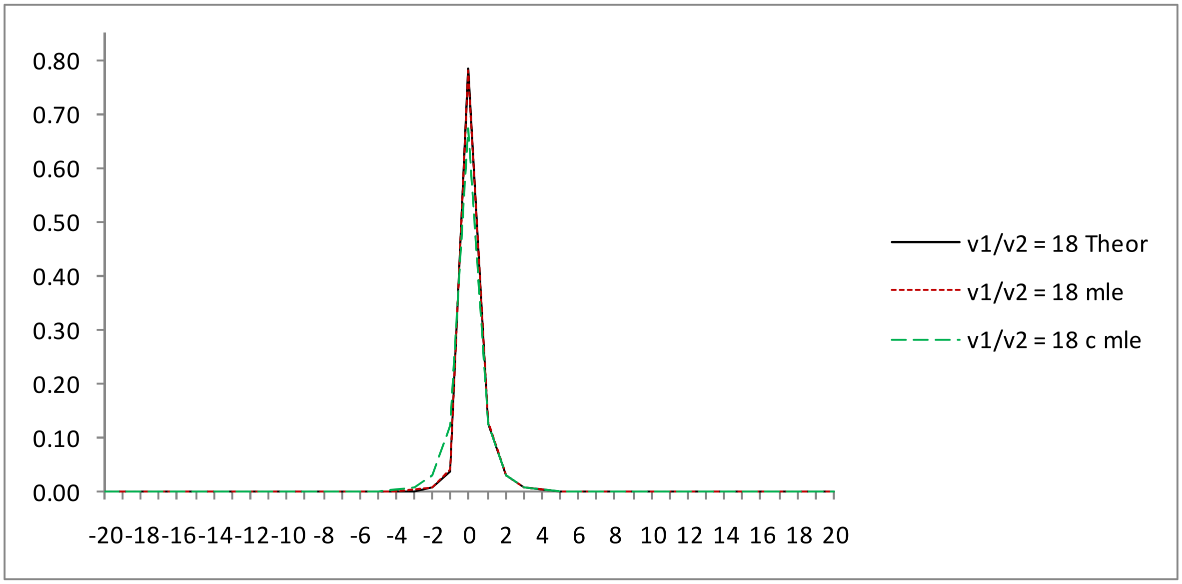


Figure 2.9 (e). Comparison of theoretical and simulated (conditional and unconditional) distribution of the mle for ν_1/ν_2 = 18, n = 200 and τ = 100.

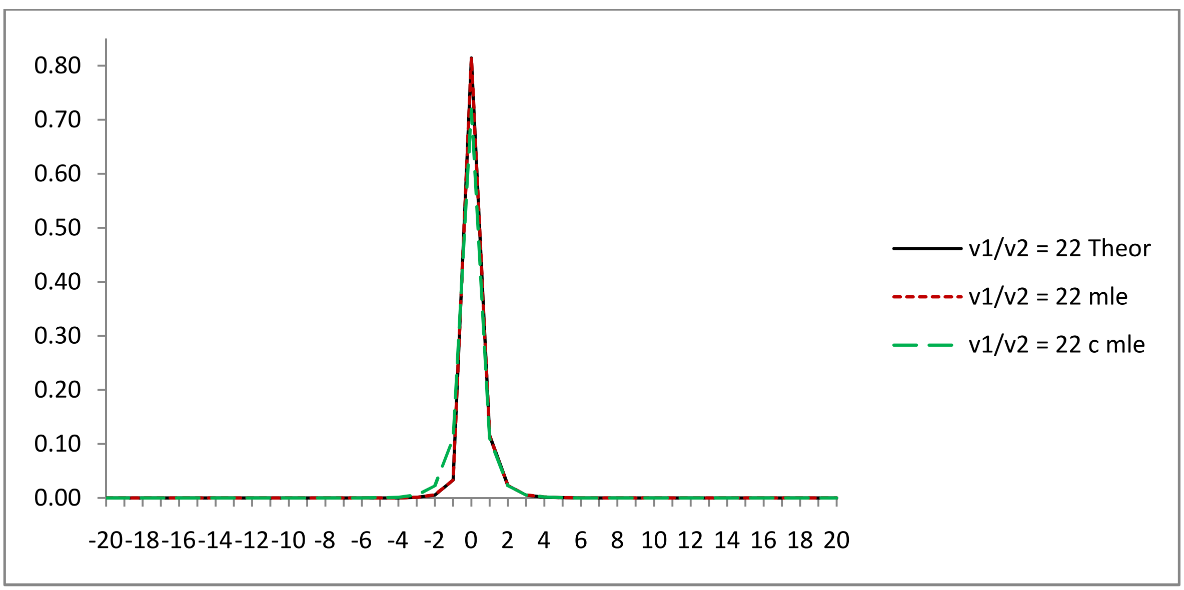


Figure 2.9 (f). Comparison of theoretical and simulated (conditional and unconditional) distribution of the mle for ν_1/ν_2 = 22, n = 200 and τ = 100.

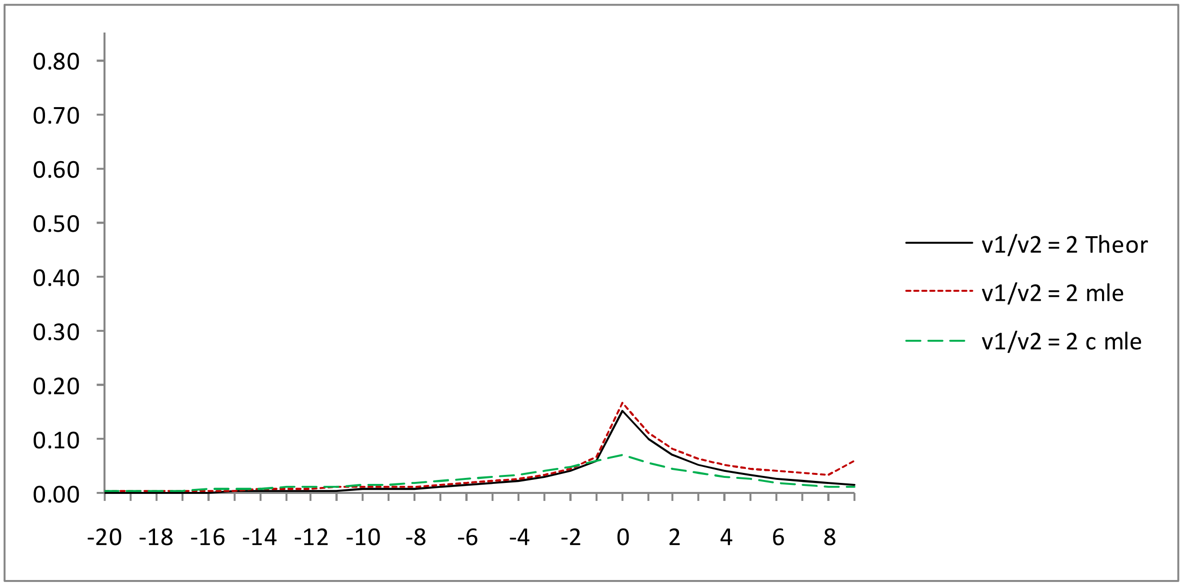


Figure 2.10 (a). Comparison of theoretical and simulated (conditional and unconditional) distribution of the mle for ν_1/ν_2 = 2, n = 40 and τ = 30.

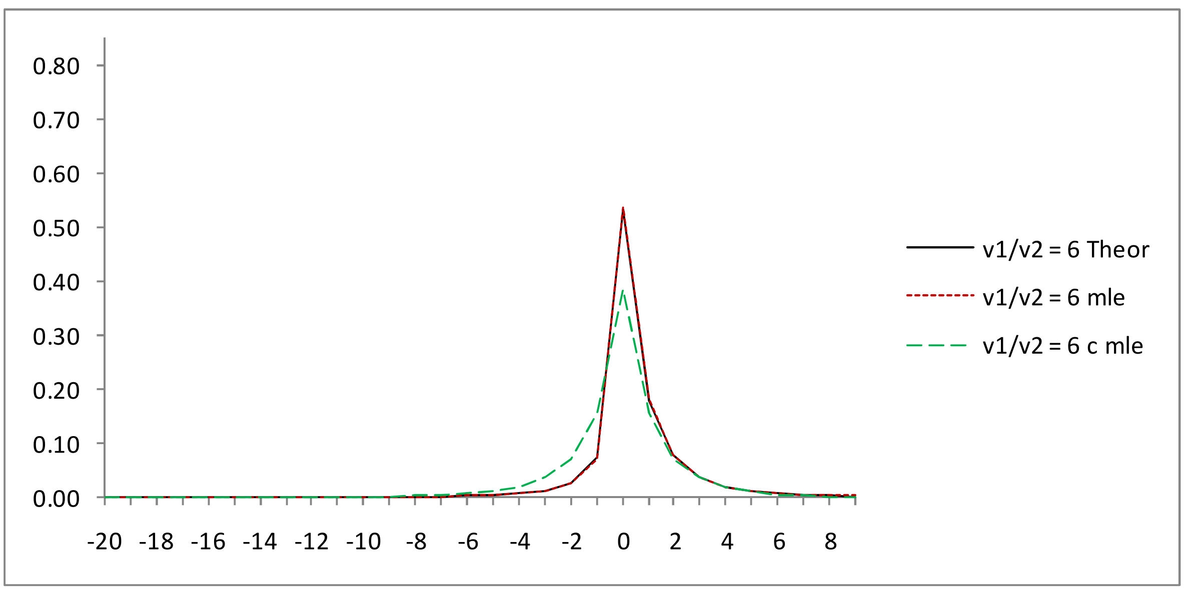


Figure 2.10 (b). Comparison of theoretical and simulated (conditional and unconditional) distribution of the mle for ν_1/ν_2 = 6, n = 40 and τ = 30.

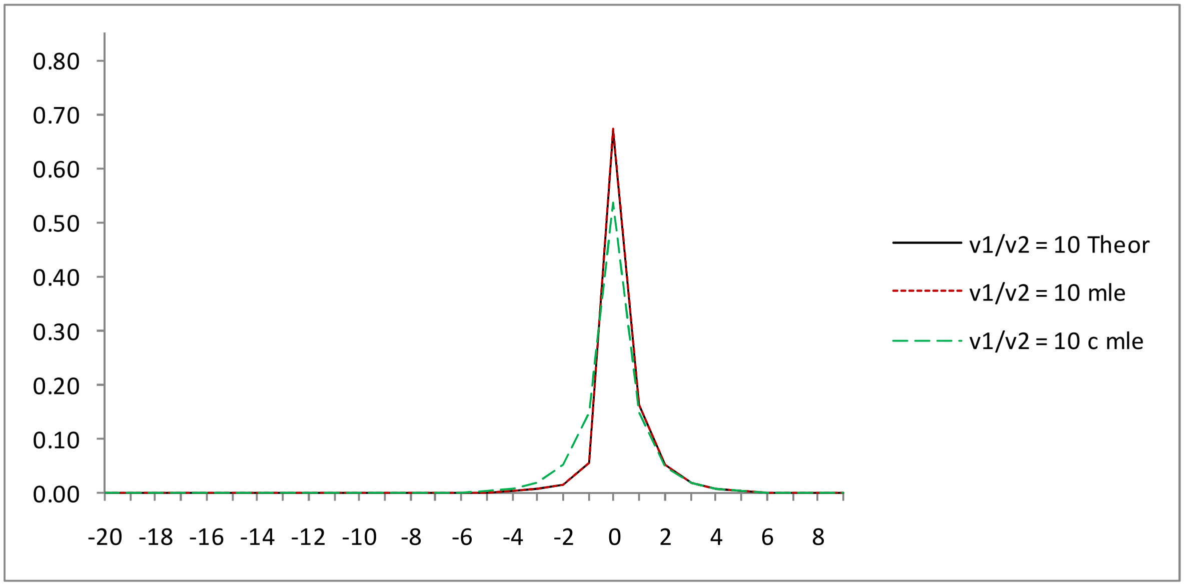


Figure 2.10 (c). Comparison of theoretical and simulated (conditional and unconditional) distribution of the mle for ν_1/ν_2 = 10, n = 40 and τ = 30.

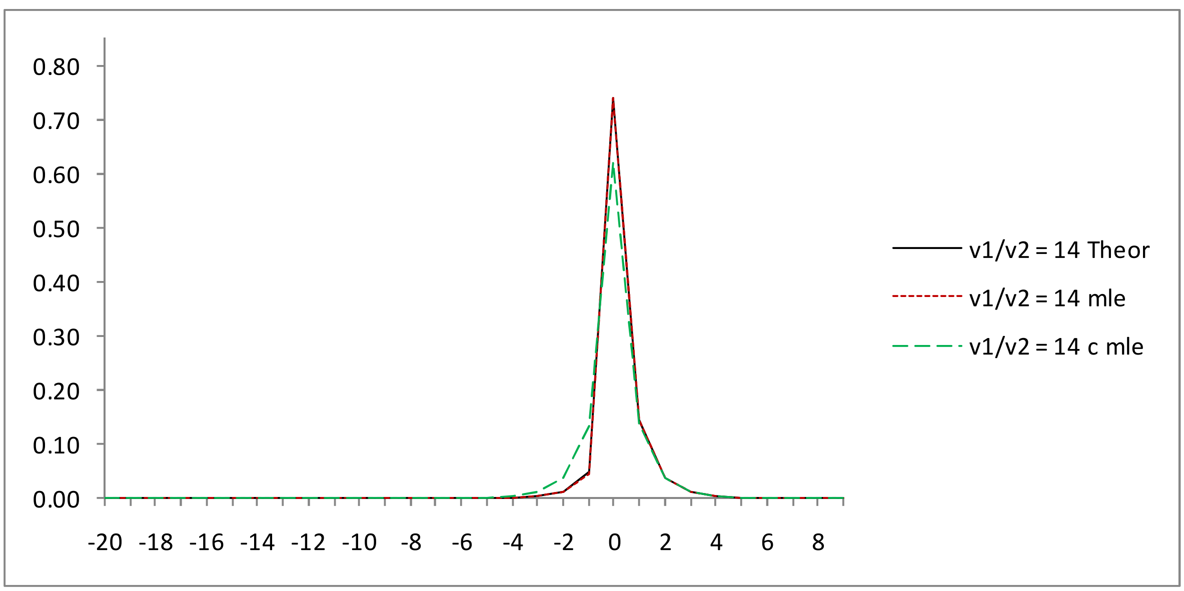


Figure 2.10 (d). Comparison of theoretical and simulated (conditional and unconditional) distribution of the mle for ν_1/ν_2 = 14, n = 40 and τ = 30.

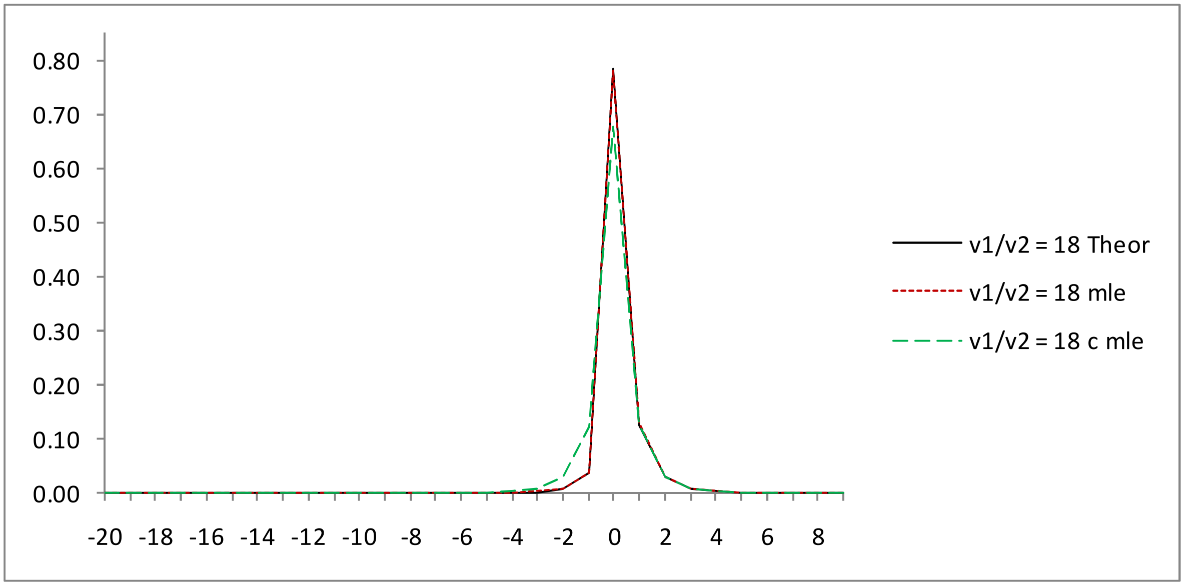


Figure 2.10 (e). Comparison of theoretical and simulated (conditional and unconditional) distribution of the mle for ν_1/ν_2 = 18, n = 40 and τ = 30.

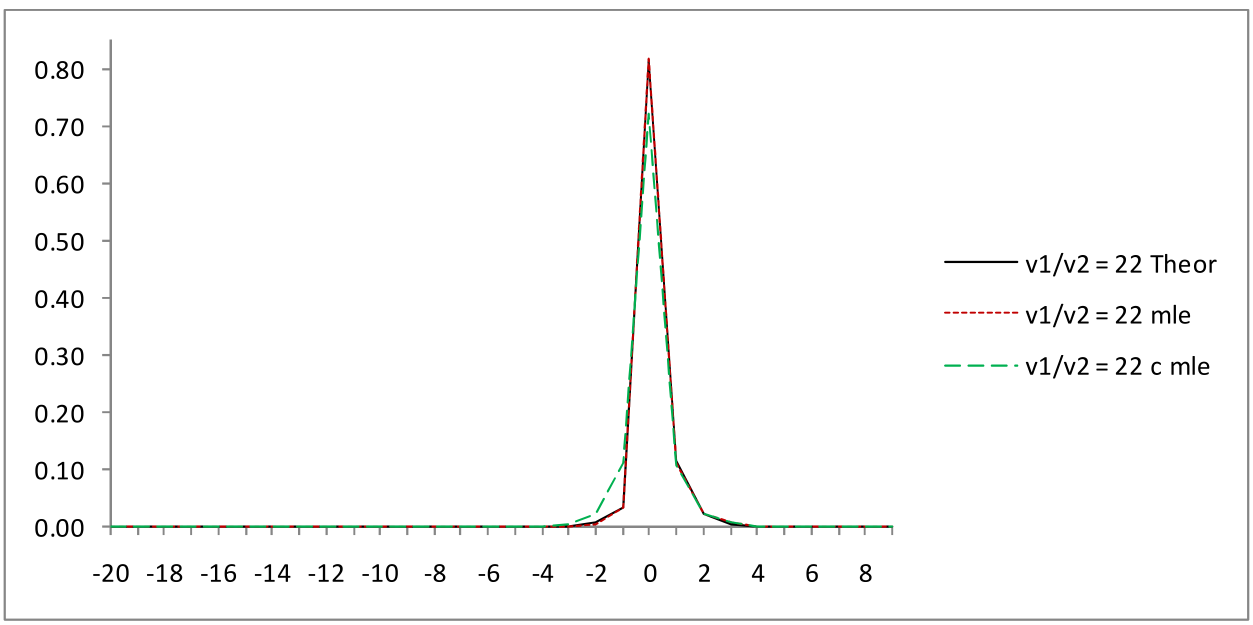


Figure 2.10 (f). Comparison of theoretical and simulated (conditional and unconditional) distribution of the mle for ν_1/ν_2 = 22, n = 40 and τ = 30.

CHAPTER 3

ASYMPTOTIC DISTRIBUTION OF THE MAXIMUM LIKELIHOOD ESTIMATE OF THE CHANGE-POINT WHEN A CHANGE OCCURS IN THE MEAN PARAMETER OF A GAUSSIAN PROCESS

In this chapter we investigate an abrupt change-point problem with the change-point occurring in the mean only of a Gaussian process. A Gaussian model is useful in many sciences, such as biology, climatology, and management. For example, assuming independence and normality, we may wish to analyze temperature or precipitation data in order to determine if there was a change in the climate and when it occurred. Many industrial data follow the normal distribution, thus we may utilize change-point analysis in quality control. If a change in the process is detected, the distribution of the maximum likelihood estimator will give a confidence interval that will allow managers to track down possible causes of the change. Agricultural data may also follow the normal distribution, and we may wish to know if there was a change between one year and another, perhaps due to weather, new irrigation or planting methods or new machinery employed.

Thus we would like to make inferences about changes that occur in Gaussian independent time-series data and make more statistical tools available for the sciences and other areas. Some work has already been done on this subject, Hinkley (1970) gives approximations for the distribution for the mle for a change in the mean only of a normal distribution. Many others provide upper and lower bounds for the distribution of the change-point mle. Borovkov (1999) provides upper and lower bounds, Hu & Rukhin (1995) provide a lower bound, Jandhyala & Fotopoulos (1999) and Fotopoulos & Jandhyala (2001) provide the upper and lower bounds for the asymptotic distribution of the mle, Jandhyala & Fotopoulos (1999) also provide two approximations of the change-point mle distribution. In addition, Cobb (1978) suggests a conditional approach to the distribution of the change-point mle, where Hinkley's approach would be labeled as unconditional, since it depends only on the parameters. And much work has been done on developing tests for change in a (univariate and multivariate) normal distribution, see Csörgö & Horváth (1997) and Chen & Gupta (2000) for example, who refer to many works on the subject, e.g., Worsley (1979), Sen & Srivastava (1975), and Srivastava & Worsley (1986).

The rest of the chapter is organized as follows: in section 3.1 we will look at general results for the asymptotic distribution of the mle of the unknown change-point, then in section 2.2 we will focus on the univariate and multivariate Gaussian distribution and derive the computable expressions for the asymptotic distribution of the mle of the unknown change-point when a change occurs abruptly in the mean only of a sequence of independent normal random variables. Then in sections 2.3 and 2.4 we will present applications of the distribution theory using global precipitation data and surface temperature anomalies data. Section 2.5 will conclude with a simulation study of the accuracy of the asymptotic distribution and its robustness to departures from normality.

3.1 General results for the distribution of the maximum likelihood estimate of the change-point τ.

As in section 2.1.1, consider a sequence of time-ordered, independent, real valued random variables, $Y_1, Y_2, \dots, Y_n, n \geq 1$, defined on a probability space $(\Omega, \mathcal{F}, P)$. We assume there exists a change-point $\tau_n \in \{1, 2, \dots, n-1\}$ such that $Y_1, Y_2, \dots, Y_{\tau_n}$ have a common density function f_1 and the remaining sequence $Y_{\tau_n+1}, Y_{\tau_n+2}, \dots, Y_n$ has a common probability density function f_2, with $f_1 \neq f_2$. Note that f_1 and f_2 are densities of F_1 and F_2 with respect to some dominating measure $\mu(F_1, F_2 \ll \mu)$. Following Hinkley (1970), the change-point τ_n is an unknown parameter that we wish to estimate. The likelihood function then is as follows $L(t) = \prod_{i=1}^{\tau_n} f_1(Y_i) \prod_{i=\tau_n+1}^{n} f_2(Y_i)$, and the mle of τ_n can be expressed as $\hat{\tau}_n = arg \max_{1 \leq t \leq n-1} \sum_{i=1}^{t} \log \frac{f_1}{f_2}(Y_i), i = 1, 2, \dots, n-1$. For the sake of convenience we work with the difference between the true change-point and the maximum likelihood estimate, $\hat{\tau}_n - \tau_n$. In this case the mle function can be rewritten as

$$\hat{\tau}_n - \tau_n = arg \max_{1-\tau_n \leq k \leq n-1-\tau_n} \sum_{i=1}^{\tau_n+k} \log \frac{f_1}{f_2}(Y_i), \qquad \hat{\tau}_n - \tau_n \in \{1-\tau_n, \dots, n-1-\tau_n\}$$

where the maximizer is a result of a two-sided random walk below

$$\Gamma_n(k, \tau_n) = \begin{cases} \sum_{i=1}^{k} X_i^* = S_k^*, & k = 1, \dots, n-1-\tau_n \\ 0, & k = 0 \\ \sum_{i=1}^{-k} X_i = S_{-k}, & k = -1, \dots, 1-\tau_n \end{cases} \qquad (3.1.1)$$

We define the following random variables

$$X_i^* = \log\frac{f_1}{f_2}(Y_i^*) \quad and \quad X_i = -\log\frac{f_1}{f_2}(Y_i), \qquad i = 1, \dots, n-1$$

and the partial sums

$$S_j = \sum_{i=1}^{j} X_i \qquad and \qquad S_j^* = \sum_{i=1}^{j} X_i^*, \qquad S_0 = S_0^* = 0$$

Note that that in the above X^* and X are real random variables defined on $\mathbf{R}$, where the sequences $\{Y, Y_i: i \geq 1\}$ and $\{Y^*, Y_i^*: i \geq 1\}$ are independent with identical and independent copies on $(\mathbf{R}, \mathcal{R})$, such that $Y \sim f_1$ and $Y^* \sim f_2$. Thus the partial sums S_j and S_j^* of the random variables X^* and X are two, independent of each other, random walks.

Since $F_1 \neq F_2$ and thus $f_1 \neq f_2$ the expected values of X^* and X are negative:

$$E(X) = -\int \log\frac{f_1}{f_2}(x)\, f_1(x)\mu(dx) = -K(f_1, f_2) = -E_{f_1}\left(\log\frac{f_1}{f_2}(Y)\right) < 0$$

$$E(X^*) = \int \log\frac{f_1}{f_2}(x)\, f_2(x)\mu(dx) = -K(f_2, f_1) = E_{f_2}\left(\log\frac{f_1}{f_2}(Y^*)\right) < 0$$

We show below that for Jandhyala & Fotopoulos' (1999, p. 132) Proposition 1 and Stoyan (1976, p.83) , $\theta = \theta^* = 1$ and the three assumptions for X_1 and X_1^* in the proposition are not needed as they hold whenever $F_1 \neq F_2$ and $P(X > 0) > 0$.

Assume that $P(X > 0) > 0$, and for $\theta > 0$ let $\phi(\theta) = E\{e^{\theta X}\}$ and $\psi(\theta) = E\{e^{\theta X^*}\}$, then we see that $\phi(\theta) = \psi(1-\theta)$ and

$$\phi(\theta) = E\{e^{\theta X}\} = \int f_1(x)\left\{\frac{f_1}{f_2}\right\}^{-\theta} \mu(dx) = \int f_1^{1-\theta}(x) f_2^{\theta}(x)\mu(dx)$$

$$\leq \left\{\int f_1(x)\mu(dx)\right\}^{1-\theta} \left\{\int f_2(x)\mu(dx)\right\}^{\theta} = 1 \qquad (3.1.2)$$

thus $\phi(\theta) \leq 1 \; \forall \; \theta \in [0,1]$. We know that the asymptotic tail of the total maximum

$$M = sup\{S_j : j \in \mathbf{N}\}$$

can be described by one of the following cases: (1) $\theta = 0$, sub-exponential case, (2) $\theta > 0$ and $\phi(\theta) < 1$, an intermediate case, and (3) $\theta > 0$ and $\phi(\theta) = 1$, Cramer's case. We see that since $\theta > 0$ to satisfy case (3) we must have $\theta = 1$, see equation (3.1.2), thus X satisfies Cramer's condition. It then follows that X^* satisfies Cramer's condition as well, since as we noted before $\phi(\theta) = \psi(1-\theta)$. Thus $\theta = \theta^* = 1$. It follows that $\phi(\theta) \leq 1 \; \forall \; \theta \in [0,1]$ and ϕ is strictly convex on $\theta \in [0,1]$ suggesting $\phi(\theta)$ reaches the minimum at a unique $\theta_0 \in [0,1]$, i.e., $\phi(\theta_0) = \inf_{\theta \in [0,1]} \phi(\theta) < 1$. Hence the assumptions for X_1 and X_1^* in Jandhyala & Fotopoulos (1999) are no longer needed.

Let $\xi_n = \hat{\tau}_n - \tau_n$. Since the finite sample distribution of ξ_n requires the unknown change-point parameter τ_n, we choose to work with the asymptotic distribution of ξ_n. As in chapter 2 we let $n \to \infty$ so that both $\tau_n \to \infty$ and $n - \tau_n \to \infty$, then the limiting random variable ξ_∞ will have the following probability distribution, see Fotopoulos (2007) and Jandhyala & Fotopoulos (1999)

THEOREM 3.1. *When $F_1 \neq F_2$ and $P(X > 0) > 0$, the probability distribution of ξ_∞ is as follows*

$P(\xi_\infty = k) =$

$$= \begin{cases} P(T_1^{*+} = \infty)\left\{P(T_1^{*-} > k) - \int_{0^+}^{\infty} P(M_\infty \geq x)P(S_k^* \in dx, T_1^{*-} > k)\right\}, k \geq 1 \\ P(T_1^{+} = \infty)P(T_1^{*+} = \infty), \qquad k = 0 \\ P(T_1^{+} = \infty)\left\{P(T_1^{-} > |k|) - \int_{0^+}^{\infty} P(M_\infty^* \geq x)P\left(S_{|k|} \in dx, T_1^{-} > |k|\right)\right\}, k \leq -1 \end{cases} \qquad (3.1.3)$$

Where

$$T_1^- := inf\{j \geq 1: S_j < 0\}, \qquad T_1^+ := inf\{j \geq 1: S_j > 0\}$$

$$T_1^{*-} := inf\{j \geq 1: S_j^* < 0\}, \qquad T_1^{*+} := inf\{j \geq 1: S_j^* > 0\}$$

are the first strict descending and ascending ladder epochs and

$$M_\infty := \max_{0 \leq i < \infty} S_i \quad and \quad M_\infty^* := \max_{0 \leq i < \infty} S_i^*$$

are the total maxima.

The expressions in the equation (3.1.3) above are not in a computable form, but we can see that the distribution of M_∞^* and M_∞ depends on the underlying distributions f_1 and f_2. By assuming the underlying distributions to be Gaussian we can derive a computable form for the asymptotic distribution of the change-point mle.

3.2 Asymptotic distribution of ξ_∞ when the underlying process is Gaussian

We first consider case where the change occurs in the mean only of the underlying univariate Gaussian distribution. Then we show how this can be extended to the general multivariate Gaussian case.

3.2.1 Univariate Gaussian case

Consider a time-ordered sequence of independent real valued random variables, $Y_1, Y_2, \dots, Y_n$, $n \geq 1$, defined on $(\Omega, \mathcal{F}, P)$. We assume that there exists a change-point $\tau_n \in \{1, 2, \dots, n-1\}$ such that the sequence $Y_1, Y_2, \dots, Y_{\tau_n}$, has a common probability density function f_1, i.e., $Y_i \sim N(\mu_1, \sigma^2)$ for $i = 1, \dots, \tau_n$, and the sequence $Y_{\tau_{n+1}}, Y_{\tau_n+2}, \dots, Y_n$, has a common probability density function f_2, i.e., $Y_i \sim N(\mu_2, \sigma^2)$ for $i = \tau_n + 1, \dots n$, where $f_1 \neq f_2$ and $\mu_1 \neq \mu_2$.

$$f(y_i, v) = \begin{cases} f_1(y_i; \mu_1, \sigma^2) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-(y_i-\mu_1)^2/2\sigma^2}, & i = 1, \dots, \tau_n;\ \ y_i \geq 0 \\ f_2(y_i; \mu_2, \sigma^2) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-(y_i-\mu_2)^2/2\sigma^2}, & i = \tau_n + 1, \dots, n;\ \ y_i \geq 0 \end{cases}$$

Note that a change occurs in the mean only, the variance σ^2 stays constant, and is assumed to be known. The random variables $X_i = -\log \frac{f_1}{f_2}(Y_i) = \log \frac{f_2}{f_1}(Y_i)$ and $X_i^* = \log \frac{f_1}{f_2}(Y_i^*)$, where $Y \sim N(\mu_1, \sigma^2)$ and $Y^* \sim N(\mu_2, \sigma^2)$, can then be expressed as

$$X = \log \left\{ \frac{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(Y-\mu_2)^2/2\sigma^2}}{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(Y-\mu_1)^2/2\sigma^2}} \right\} = -\frac{(Y-\mu_2)^2}{2\sigma^2} + \frac{(Y-\mu_1)^2}{2\sigma^2} =$$

$$= _{\mathcal{D}} \frac{-(\mu_1 + Z\sigma - \mu_2)^2 + (\mu_1 + Z\sigma - \mu_1)^2}{2\sigma^2}$$

$$= \frac{-(\mu_1 - \mu_2)^2 - 2(\mu_1 - \mu_2)Z\sigma - (Z\sigma)^2 + (Z\sigma)^2}{2\sigma^2}$$

$$= -\frac{(\mu_1 - \mu_2)^2}{2\sigma^2} - \frac{2(\mu_1 - \mu_2)Z\sigma}{2\sigma^2}$$

$$= -\frac{(\mu_1 - \mu_2)^2}{2\sigma^2} - \frac{(\mu_1 - \mu_2)}{\sigma}Z \qquad (3.2.1)$$

and

$$X^* = \log\left\{\frac{\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(Y^*-\mu_1)^2/2\sigma^2}}{\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(Y^*-\mu_2)^2/2\sigma^2}}\right\} = -\frac{(Y^* - \mu_1)^2}{2\sigma^2} + \frac{(Y^* - \mu_2)^2}{2\sigma^2}$$

$$= _{\mathcal{D}} \frac{-(\mu_2 + Z^*\sigma - \mu_1)^2 + (\mu_2 + Z^*\sigma - \mu_2)^2}{2\sigma^2}$$

$$= \frac{-(\mu_1 - \mu_2)^2 + 2(\mu_1 - \mu_2)Z^*\sigma - (Z^*\sigma)^2 + (Z^*\sigma)^2}{2\sigma^2}$$

$$= -\frac{(\mu_1 - \mu_2)^2}{2\sigma^2} + \frac{2(\mu_1 - \mu_2)Z^*\sigma}{2\sigma^2}$$

$$= -\frac{(\mu_1 - \mu_2)^2}{2\sigma^2} + \frac{(\mu_1 - \mu_2)}{\sigma}Z^* \qquad (3.2.2)$$

Where $Z \sim N(0,1)$ and $Z^* \sim N(0,1)$ are independent of each other. We note in the above that X and X^* are identically distributed with identical negative means $E(X) = E(X^*) = -\eta^2/2 < 0$

$$E(X) = E(X^*) = -\frac{(\mu_1 - \mu_2)^2}{2\sigma^2} = -\left(\frac{\mu_1 - \mu_2}{\sigma}\right)^2/2 < 0$$

where η is the standardized amount of change in the mean

$$\eta = \frac{|\mu_1 - \mu_2|}{\sigma} \qquad (3.2.3)$$

The variances are also identical

$$Var(X) = Var(X^*) = \left(\frac{\mu_1 - \mu_2}{\sigma}\right)^2 = \eta^2$$

Due to this we can limit out analysis to one side only of the random walk $\Gamma_n(\cdot)$ in equation (3.1.1) and the partial sum can now be written as

$$S_j = \sum_{i=1}^{j} X_i = \sum_{i=1}^{j}\left(-\frac{(\mu_1 - \mu_2)^2}{2\sigma^2} - \frac{(\mu_1 - \mu_2)}{\sigma}Z\right)$$

$$= \sum_{i=1}^{j}(-\eta^2/2 - \eta Z) = -j\eta^2/2 - \eta\sum_{i=1}^{j} Z$$

$$=_{\mathcal{D}} -\frac{j\eta^2}{2} - \eta Z\sqrt{j}, \qquad Z \sim N(0,1) \qquad (3.2.4)$$

In Asmussen (1987) G_+ is defined as the strict ascending ladder height distribution which is defective when $E(X) < 0$.

$$G_+(dx) = P\left(T_1^+ < \infty, S_{T_1^+} \in dx\right) \qquad (3.2.5)$$

Hence, as in chapter 2, $\|G_+\| = P(T_1^+ < \infty) < 1$ and $1/E(T_1^-) = 1 - \|G_+\| = P(T_1^+ = \infty) = P(M_\infty = 0)$.

THEOREM 3.2. *Given a time-ordered sequence* $Y_1, Y_2, \dots, Y_n, n \geq 1$ *where* $Y_i \sim N(\mu_1, \sigma^2), i = 1, \dots, \tau_n$ *and* $Y_i \sim N(\mu_2, \sigma^2), i = \tau_n + 1, \dots, n$*, and* $\mu_1 \neq \mu_2$ *the probability distribution of* ξ_∞*, see equation* (3.1.3)*, can be written as follows*

$$P(\xi_\infty = k) = \begin{cases} (1 - \|G_+\|)\left(q_{|k|} - \|G_+\|\tilde{q}_{|k|}\right), & k = \pm 1, \pm 2, \dots \\ (1 - \|G_+\|)^2, & k = 0 \end{cases} \qquad (3.2.6)$$

where

$$1 - \|G_+\| = P(T_1^+ = \infty) = P(M_\infty = 0) = e^{-\sum_{j=1}^{\infty} \frac{1}{j} \bar{\Phi}(\eta\sqrt{j}/2)} \qquad (3.2.7)$$

and

$$q_k = E[I(T_1^- > k)] \qquad \tilde{q}_k = E[e^{-S_k} I(T_1^- > k)], \qquad k = 1,2, \dots \quad and \quad q_0 = \tilde{q}_0 = 1 \qquad (3.2.8)$$

Proof follows by applying Lemmas 3.1, 3.2, and 3.3. From Shiryaev et al., (1993) we have the following:

LEMMA 3.1, for $x \geq 0$

$$P\left(\max_{m \leq n} S_n\right) = \Phi\left(\frac{x + n\eta^2/2}{\sigma\sqrt{n}}\right) - e^{-x}\Phi\left(\frac{-x + n\eta^2/2}{\sigma\sqrt{n}}\right) \to 1 - e^{-x} = P(M_\infty \leq x) \ as \ n \to \infty$$

LEMMA 3.2. for $x \geq 0$

$$P(S_n \in dx,\, T_1^- > n) = \eta^{-1} E\left\{(\,T_1^- > n - 1) \cap \varphi\left(\frac{x - S_{n-1} + \eta^2/2}{\eta}\right)\right\}, n \geq 1$$

Proof. For $x > 0$,

$$P(S_n \in (0,x],\ T_1^- > n) = P\left(\bigcap_{j=0}^{n-1}(S_j > 0) \cap S_n \in (0,x]\right) =$$

$$= E\left\{I\left(\bigcap_{j=0}^{n-1}(S_j > 0)\right) P(X_n \in (-S_{n-1}, x - S_{n-1}] | \mathcal{F}_{n-1})\right\}$$

$$= E\left\{I\left(\bigcap_{j=0}^{n-1}(S_j > 0)\right) P\left(Z_n \in \left(\frac{-S_{n-1} + \eta^2/2}{\eta}, \frac{x - S_{n-1} + \eta^2/2}{\eta}\right] | \mathcal{F}_{n-1}\right)\right\}$$

$$= E\left\{I(\ T_1^- > n-1) \cap \Phi\left(\frac{x - S_{n-1} + \eta^2/2}{\eta}\right) - \Phi\left(\frac{-S_{n-1} + \eta^2/2}{\eta}\right)\right\}, n \geq 1$$

The rest follows by differentiating w.r.t. x.

LEMMA 3.3

$$\int_{0^+}^{\infty} P(M_\infty \geq x) P\ (S_n \in dx,\ T_1^- > n) = \|G_+^*\| E\{e^{-S_n} I(\ T_1^- > n)\}, n \geq 1$$

Proof. Note that $P(M_\infty \leq x) = P(M_\infty \geq x | M_\infty > 0) P(M_\infty > 0) = \|G_+\| e^{-x}, x > 0,$ and applying Lemma 3.2

$$\int_{0^+}^{\infty} P(M_\infty \geq x) P\ (S_n \in dx,\ T_1^- > n) =$$

$$= \eta^{-1}\|G_+^*\| E\left\{ I(\,T_1^- > n-1) \int_{0^+}^{\infty} e^{-x}\varphi\left(\frac{x - S_{n-1} + \eta^2/2}{\eta}\right) dx \right\}$$

$$= \|G_+^*\| E\{I(\,T_1^- > n-1) e^{-S_n} I(\eta Z_n > -S_{n-1} + \eta^2/2)\}$$

$$= \|G_+^*\| E\{e^{-S_n} I(\,T_1^- > n)\}, n \geq 1$$

To compute the equations (3.2.8), we look to Feller (1971, pg 416) and Chover et al. (1973). The generating functions of the sequences satisfy the following

$$\sum_{k=1}^{\infty} s^k q_k = e^{\sum_{k=1}^{\infty} \frac{s^k b_k}{k}} \quad and \quad \sum_{k=1}^{\infty} s^k \tilde{q}_k = e^{\sum_{k=1}^{\infty} \frac{s^k \tilde{b}_k}{k}} \qquad (3.2.9)$$

where (in the Gaussian case)

$$b_k = P(S_k > 0) = \overline{\Phi}\big(\eta\sqrt{k}/2\big), \quad k \geq 1 \qquad (3.2.10)$$

and

$$\tilde{b}_k = E[e^{-S_k} I(S_k > 0)] = e^{k\eta^2}\overline{\Phi}\big(3\,\eta\sqrt{k}/2\big), \quad k \geq 1 \qquad (3.2.10)$$

Thus (applying the Leibnitz rule) we can use the following familiar iterative procedures to compute q_k and $\tilde{q}_k$

$$q_0 = \tilde{q}_0 = 1, \qquad k q_k = \sum_{j=0}^{k-1} b_{k-j} q_j\,, \qquad k\tilde{q}_k = \sum_{j=0}^{k-1} \tilde{b}_{k-j}\tilde{q}_j\,, \qquad k > 0 \quad (3.2.11)$$

In order to show that the probabilities in equation (3.2.6) sum to one we look to Hinkley (1970) and note that $P(M_\infty \geq x) = P(M_\infty \geq x | M_\infty > 0)P(M_\infty > 0) = \|G_+\|e^{-x},\ x > 0$. It then follows that

$$P(\xi_\infty > 0) = P(M_\infty^* > M_\infty, M_\infty^* > 0) = \int_{0^+}^{\infty} P(M_\infty < x)P(M_\infty^* \in dx)$$

$$= \int_{0^+}^{\infty} (1 - \|G_+\|e^{-x})\|G_+\|e^{-x}dx = [1 - (1 - \|G_+\|)^2]/2$$

Note that we know $P(\xi_\infty = 0) = (1 - \|G_+\|)^2$ and that ξ_∞ is symmetric, see (3.2.6). Hence we see that the probabilities must sum to one.

$$1 = (1 - \|G_+\|)^2 + 2\left(\frac{1 - (1 - \|G_+\|)^2}{2}\right)$$

3.2.1 Multivariate Gaussian case

Let us consider a sequence of independent time-ordered random vectors $\boldsymbol{Y}_i \in \boldsymbol{R}^d,\ i = 1, \dots, n$ with density function $f(y; \boldsymbol{\mu}, \boldsymbol{\Sigma})$. The change in parameter $\boldsymbol{\mu}$ from $\boldsymbol{\mu}_1$ to $\boldsymbol{\mu}_2$, $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$ occurs at some unknown change-point $\tau_n \in \{1,2, \dots, n-1\}$, with $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in \Theta$, with a common variance-covariance matrix $\boldsymbol{\Sigma}$, which we assume to be positive definite.

Let $\langle x, y \rangle$ denote the inner product defined and the extended seminorm is defined if there exists a covariance operator Σ such that $\|x\|_\Sigma^2 = \langle \Sigma \mathrm{x}, \mathrm{x} \rangle$. Then $Y =_{\mathcal{D}} \mu_1 + \Sigma^{1/2}\mathbf{Z}$ for data before the change-point τ_n. Note that $\mathbf{Z}$ is a *d*-variate standard normal vector. Therefore, for values before the change-point, we may express the random variable X as

$$X = -\log\frac{f(y;\mu_1,\Sigma)}{f(y;\mu_2,\Sigma)}$$

$$= -\frac{1}{2}\{\langle\Sigma^{-1}(\mathrm{Y}-\mu_2),(\mathrm{Y}-\mu_2)\rangle - \langle\Sigma^{-1}(\mathrm{Y}-\mu_1),(\mathrm{Y}-\mu_1)\rangle\}$$

$$=_{\mathcal{D}} -\frac{1}{2}\|\mu_1-\mu_2\|^2_{\Sigma^{-1}} - \|\mu_1-\mu_2\|_{\Sigma^{-1}}Z \qquad (3.2.12)$$

where Z is a standard normal variable with mean zero and variance one.

And in a similar fashion, for values after the change-point τ_n, $Y^* =_{\mathcal{D}} \mu_2 + \Sigma^{1/2}\mathbf{Z}^*$. Note that $\mathbf{Z}^*$ is a *d*-variate standard normal vector. Then we may express the random variable X^* as

$$X^* = \log\frac{f(y;\mu_1,\Sigma)}{f(y;\mu_2,\Sigma)} =_{\mathcal{D}} -\frac{1}{2}\|\mu_1-\mu_2\|^2_{\Sigma^{-1}} + \|\mu_1-\mu_2\|_{\Sigma^{-1}}Z^* \qquad (3.2.13)$$

where Z^* is a standard normal variable with mean zero and variance one, independent of Z. We let η be the extent of standardized change in the multivariate mean vectors

$$\eta^2 = \|\mu_1-\mu_2\|^2_{\Sigma^{-1}} = (\mu_1-\mu_2)^T\Sigma^{-1}(\mu_1-\mu_2) \qquad (3.2.14)$$

Thus we can see that the multivariate case can be transformed into a corresponding univariate case with $\eta = \|\mu_1-\mu_2\|_{\Sigma^{-1}}$.

Table 3.1 and Figure 3.1 below give the theoretical probability distribution of the change-point mle, the totals are based on 150 or more values.

k	η = 0.2	η = 0.6	η = 1.0	η = 1.5	η = 2.0	η = 2.5	η = 3.0
0	0.0178	0.1270	0.2802	0.4749	0.6409	0.7674	0.8568
1	0.0154	0.0757	0.1181	0.1313	0.1152	0.0876	0.0599
2	0.0136	0.0539	0.0689	0.0590	0.0385	0.0208	0.0097
3	0.0123	0.0416	0.0454	0.0310	0.0156	0.0061	0.0020
4	0.0113	0.0335	0.0318	0.0177	0.0069	0.0020	0.0005
5	0.0105	0.0277	0.0231	0.0106	0.0033	0.0007	0.0001
6	0.0099	0.0233	0.0173	0.0066	0.0016	0.0003	0.0000
7	0.0093	0.0199	0.0132	0.0042	0.0008	0.0001	0.0000
8	0.0088	0.0172	0.0102	0.0027	0.0004	0.0000	0.0000
9	0.0084	0.0150	0.0080	0.0018	0.0002	0.0000	0.0000
10	0.0080	0.0132	0.0064	0.0012	0.0001	0.0000	0.0000
11	0.0077	0.0116	0.0051	0.0008	0.0001	0.0000	0.0000
12	0.0074	0.0103	0.0041	0.0005	0.0000	0.0000	0.0000
13	0.0071	0.0092	0.0033	0.0004	0.0000	0.0000	0.0000
14	0.0068	0.0082	0.0027	0.0003	0.0000	0.0000	0.0000
15	0.0066	0.0074	0.0022	0.0002	0.0000	0.0000	0.0000
16	0.0063	0.0067	0.0018	0.0001	0.0000	0.0000	0.0000
17	0.0061	0.0060	0.0015	0.0001	0.0000	0.0000	0.0000
18	0.0059	0.0055	0.0012	0.0001	0.0000	0.0000	0.0000
19	0.0057	0.0050	0.0010	0.0000	0.0000	0.0000	0.0000
20	0.0056	0.0045	0.0008	0.0000	0.0000	0.0000	0.0000
Total:	0.9908	1.0227	1.0201	1.0120	1.0061	1.0025	1.0011

Table 3.1. The probability distribution of $P(\xi_\infty = k)$, k = 0, ±1, ±2, … ±20 and η = 0.2, 0.6, 1.0, 1.5, 2.0, 2.5, 3.0.

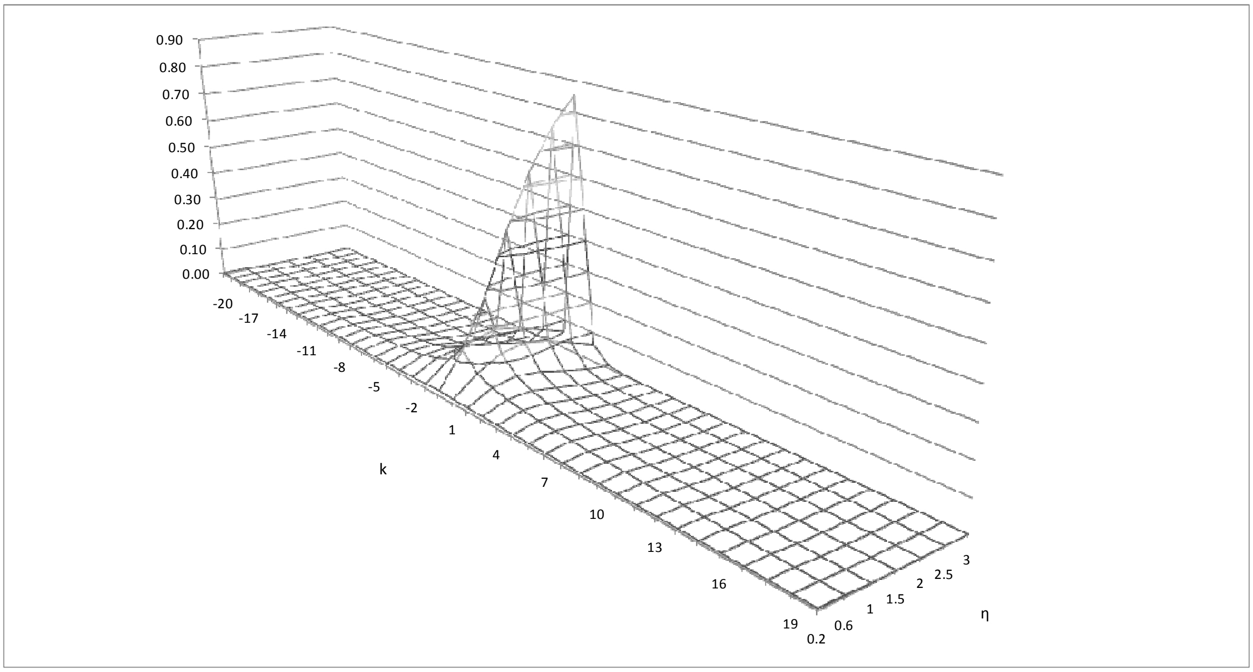


Figure 3.1. The probability distribution of $P(\xi_\infty = k)$, k = 0, ±1, ±2, … ±20 and η = 0.2, 0.6, 1.0, 1.5, 2.0, 2.5, 3.0.

3.3 Change-point analysis of multivariate global mean precipitation data

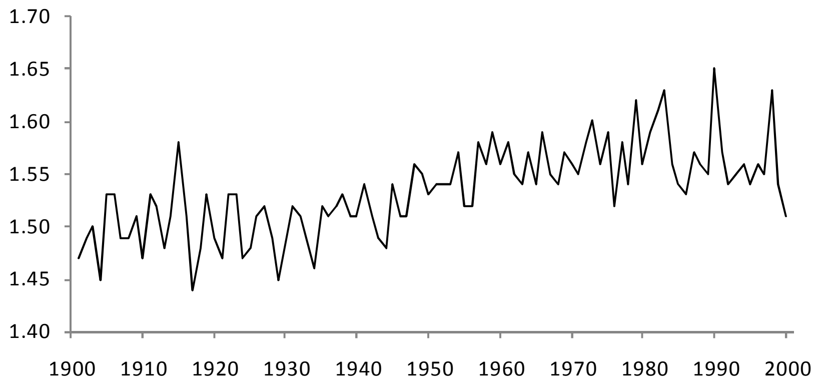
3.3.1 Introduction

The dataset we analyze covers the major climate zones: tropical, temperate, and arctic, with the temperate and arctic zones grouped together. The Northern Latitudes (NL) band includes the Northern Temperate Zone and the Arctic Circle and lies north of the Tropic of Cancer (24° to 90° N), it covers 30% of the global area and includes North America, Europe, and Asia (this includes Canada, the United States, China, Russia, Egypt, and all of the European countries). The Low Latitudes (Low) band includes the Tropical Zone and extends from the Tropic of Capricorn to the Tropic of Cancer (24° S to 24° N), it covers 40% of the global area and includes Africa, Asia, Australia, Central and South America (this includes Australia, India, Sri Lanka, Mexico, Cuba, Brazil, Ecuador, Peru, Nigeria, and Ethiopia). And finally, the Southern Latitudes (SL) band includes the Southern Temperate Zone and the Antarctic Circle, it lies South of the Tropic of Capricorn (24° S to 90° S), it covers 30% of the global area and includes South America, Africa, Antarctica, and Australia (this includes Argentina, South Africa, Australia, and New Zealand). The majority of Earth's land mass, and the world's population is located in the Northern and Southern Temperate Zones. While the temperate zones have four seasons: winter, spring, summer and fall, the tropical zone has two seasons: dry and wet.

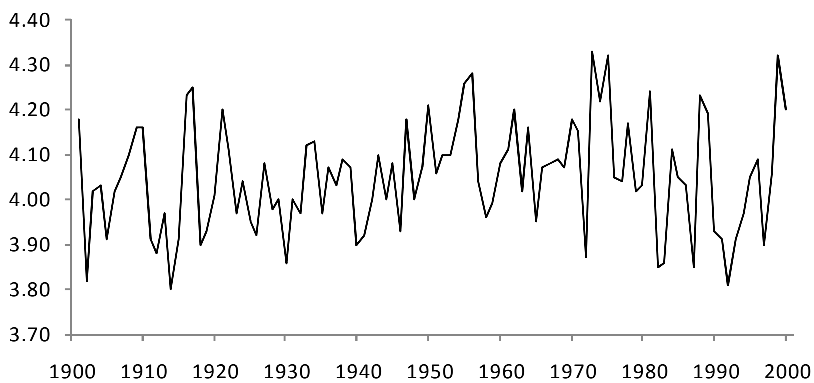
We will now conduct a change-point analysis of the annual precipitation data presented in Table 3.2 and Figure 3.2 below. We begin with detection, in other words, do we have evidence that a change-point exists in the distribution of the multivariate data? If we conclude that a change-point exists we will apply estimation methodology to determine the asymptotic distribution of the change-point maximum likelihood estimate.

Year	NL	Low	SL	Year	NL	Low	SL	Year	NL	Low	SL	Year	NL	Low	SL
1901	1.47	4.18	1.86	**1926**	1.51	3.92	1.91	**1951**	1.54	4.06	1.84	**1976**	1.52	4.05	2.14
1902	1.49	3.82	1.81	**1927**	1.52	4.08	1.85	**1952**	1.54	4.10	1.95	**1977**	1.58	4.04	2.04
1903	1.50	4.02	1.95	**1928**	1.49	3.98	1.90	**1953**	1.54	4.10	1.98	**1978**	1.54	4.17	2.09
1904	1.45	4.03	1.93	**1929**	1.45	4.00	1.86	**1954**	1.57	4.18	1.98	**1979**	1.62	4.02	1.97
1905	1.53	3.91	1.88	**1930**	1.49	3.86	1.93	**1955**	1.52	4.26	2.06	**1980**	1.56	4.03	1.96
1906	1.53	4.02	1.81	**1931**	1.52	4.00	1.95	**1956**	1.52	4.28	2.11	**1981**	1.59	4.24	2.03
1907	1.49	4.05	1.89	**1932**	1.51	3.97	1.96	**1957**	1.58	4.04	1.97	**1982**	1.61	3.85	1.92
1908	1.49	4.10	1.82	**1933**	1.49	4.12	1.82	**1958**	1.56	3.96	2.04	**1983**	1.63	3.86	2.07
1909	1.51	4.16	1.86	**1934**	1.46	4.13	2.01	**1959**	1.59	3.99	1.99	**1984**	1.56	4.11	2.12
1910	1.47	4.16	1.94	**1935**	1.52	3.97	1.90	**1960**	1.56	4.08	1.99	**1985**	1.54	4.05	1.97
1911	1.53	3.91	1.92	**1936**	1.51	4.07	1.92	**1961**	1.58	4.11	2.05	**1986**	1.53	4.03	2.05
1912	1.52	3.88	1.85	**1937**	1.52	4.03	1.81	**1962**	1.55	4.20	1.87	**1987**	1.57	3.85	2.01
1913	1.48	3.97	1.90	**1938**	1.53	4.09	1.90	**1963**	1.54	4.02	2.15	**1988**	1.56	4.23	1.91
1914	1.51	3.80	2.03	**1939**	1.51	4.07	2.08	**1964**	1.57	4.16	1.91	**1989**	1.55	4.19	1.97
1915	1.58	3.91	1.91	**1940**	1.51	3.90	1.92	**1965**	1.54	3.95	1.97	**1990**	1.65	3.93	2.08
1916	1.51	4.23	1.84	**1941**	1.54	3.92	2.03	**1966**	1.59	4.07	1.98	**1991**	1.57	3.91	1.96
1917	1.44	4.25	2.00	**1942**	1.51	4.00	2.03	**1967**	1.55	4.08	1.92	**1992**	1.54	3.81	2.10
1918	1.48	3.90	2.02	**1943**	1.49	4.10	1.86	**1968**	1.54	4.09	1.96	**1993**	1.55	3.91	1.99
1919	1.53	3.93	1.88	**1944**	1.48	4.00	1.79	**1969**	1.57	4.07	1.93	**1994**	1.56	3.97	1.90
1920	1.49	4.01	2.03	**1945**	1.54	4.08	1.93	**1970**	1.56	4.18	1.93	**1995**	1.54	4.05	2.01
1921	1.47	4.20	2.03	**1946**	1.51	3.93	2.05	**1971**	1.55	4.15	2.03	**1996**	1.56	4.09	2.04
1922	1.53	4.11	1.88	**1947**	1.51	4.18	2.05	**1972**	1.58	3.87	2.02	**1997**	1.55	3.90	2.09
1923	1.53	3.97	1.90	**1948**	1.56	4.00	1.90	**1973**	1.60	4.33	2.15	**1998**	1.63	4.06	2.04
1924	1.47	4.04	1.78	**1949**	1.55	4.07	1.93	**1974**	1.56	4.22	2.14	**1999**	1.54	4.32	2.01
1925	1.48	3.95	1.96	**1950**	1.53	4.21	2.06	**1975**	1.59	4.32	2.13	**2000**	1.51	4.20	2.12

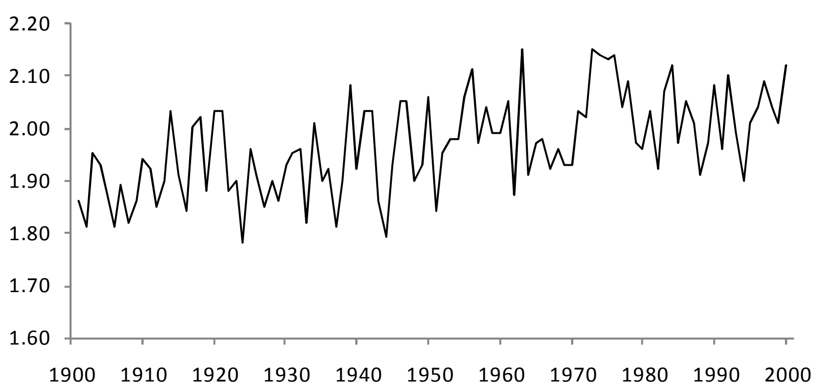
Table 3.2. Time-series data. Annual precipitation in mm per day for the years 1901-2000 for the three bands: NL, Low, and SL.



(a) NL



(b) Low



(c) SL

Figure 3.2. Time-series plots. Annual precipitation in mm per day for the years 1901-2000 for the three bands: (a) NL, (b) Low, (c) SL.

3.3.2 Analysis of precipitation data

We start with 100 observations $Y_1, Y_2, \dots, Y_{100}$, where Y_i represents the annual mean precipitation in mm per day for the years 1901-2000 for three latitude bands NL, Low, and SL, so in this case the dimension $d = 3$ and $n = 100$. Our null hypothesis is that there is no change in the mean vector in the data series, while the alternative hypothesis is that a change in the mean vector has occurred at some unknown point in time, which we call τ_n, so that $Y_1, Y_2, \dots, Y_{\tau_n - 1} \sim f_1(\mu_1, \Sigma)$ and $Y_{\tau_n}, Y_{\tau_n + 1}, \dots, Y_{100} \sim f_2(\mu_2, \Sigma)$. We assume that the data is independent and normal with $Y_i \sim MVNormal\left(\mu^{(i)}, \Sigma\right), i = 1, \dots, n$, and that no change occurs in the variance-covariance matrix, thus we can write out the hypotheses as follows

$$H_0\text{:}\ \ \mu^{(1)} = \cdots = \mu^{(n)} = \mu_1$$

$$H_1\text{:}\ \mu^{(1)} = \cdots = \mu^{(\tau_n - 1)} = \mu_1 \neq \mu^{(\tau_n)} = \cdots = \mu^{(n)} = \mu_2$$

We then take the likelihood approach to hypothesis testing and calculate the likelihood ratio statistic U_n, its asymptotic theory is presented in Csörgö & Horváth (1997), see Appendix [A.4] for details. It should be noted that asymptotic theory of the likelihood ratio statistic U_n is quite robust to departures from the assumptions of both exponentiality and independence. Thus let

$$U_n = \max_{1 \le t < n} (-2 \log \Lambda_{\mathrm{t}}) = \max_{1 \le k < n} n \log \left(\frac{\left|\widehat{\boldsymbol{\Sigma}}_{\mathrm{n}}\right|}{\left|\widehat{\boldsymbol{\Sigma}}_{\mathrm{k}}\right|} \right)$$

Where $\widehat{\boldsymbol{\Sigma}}_{\mathrm{k}} = \frac{1}{\mathrm{n}} \left\{ \sum_{i=1}^{k} (\boldsymbol{Y}_i - \widehat{\boldsymbol{\mu}}_{\mathrm{k}})(\boldsymbol{Y}_i - \widehat{\boldsymbol{\mu}}_{\mathrm{k}})^T + \sum_{i=k+1}^{n} (\boldsymbol{Y}_{\boldsymbol{i}} - \widehat{\boldsymbol{\mu}}_{\mathrm{k}}^*)(\boldsymbol{Y}_{\boldsymbol{i}} - \widehat{\boldsymbol{\mu}}_{\mathrm{k}}^*)^T \right\}$, $\widehat{\boldsymbol{\mu}}_{\mathrm{k}} = \frac{1}{k} \sum_{i=1}^{k} \boldsymbol{Y}_i$, and $\widehat{\boldsymbol{\mu}}_{\mathrm{k}}^* = \frac{1}{n-k} \sum_{i=k+1}^{n} \boldsymbol{Y}_i, k = 1, \dots, n$.

The asymptotic distribution of $W_n = (2\log\log n\, U_n)^{1/2} - \left(2\log\log n + \frac{d}{2}\log\log\log n - \log\Gamma\left(\frac{d}{2}\right)\right)$ has a double exponential form, $\lim_{n\to\infty} P(W_n \le w) = e^{-2e^{-w}}$, thus we obtain the following

$$P\left\{W_n \le (2\log\log n\, U_n)^{1/2} - \left(2\log\log n + \frac{d}{2}\log\log\log n - \log\Gamma\left(\frac{d}{2}\right)\right) = w\right\} \sim e^{-2e^{-w}}$$

Here d represents the number of parameters that change under H_α, hence $d = 3$ when we consider all three bands, $d = 2$ when we consider pairs of bands, and $d = 1$ when we consider each band individually. We obtain the following test statistic and w values:

Datasets	U_n	w
NL	74.92	12.43
SL	36.13	7.81
Low	6.6	1.79
NL SL	92.09	13.29
NL Low	89.07	13.02
SL Low	38.83	7.41
NL SL Low	104.12	14.02

Table 3.3. Datasets and corresponding test statistic U_n and w values.

Then from the values in Table 3.3 above we obtain the following p-values and change-point estimates

Test for change in mean		
Datasets	$\hat{\tau}$	p-value
NL	47	0.0000
SL	45	0.0008
Low	46	0.2829
NL SL	44	0.0000
NL Low	47	0.0000
SL Low	45	0.0012
NL SL Low	46	0.0000

Table 3.4. Datasets and corresponding p-values and change-point estimates $\hat{\tau}$.

Based on the results shown in Table 3.4 above we may conclude that there is strong evidence that a change in the mean annual precipitation in the three bands (NL, Low, SL) has occurred at $\hat{\tau} = 46$. In the bivariate cases we also see that changes occurred at $\hat{\tau} = 44, 47, 45$. In the univariate case we detect changes in the NL and SL data series at $\hat{\tau} = 47, 45$ but not in the Low data series.

We must now check our three assumptions: independence, normality, and constant variance-covariance. To test for independence, we computed the standardized residuals for each of the series and generated autocorrelation and partial autocorrelation plots. We note no significant lags. Each pair was then checked for cross-correlations, none were significant. Hence we conclude that the assumption of independence was not violated. To check for normality Mardia's (1970) skewness and kurtosis tests, and Henze-Zirkler's (1990) t-test were utilized to check for multivariate normality of the standardized residuals. There were no indications of violations of the normality assumption.

Normality Test

Equation	Test Statistic	Value	Prob
NL	Shapiro-Wilk W	0.97	0.0617
Low	Shapiro-Wilk W	0.97	0.2604
SL	Shapiro-Wilk W	0.96	0.0475
System	Mardia Skewness	17.54	0.0632
	Mardia Kurtosis	-0.39	0.6997
	Henze-Zirkler T	0.98	0.3263

And finally, to check for constant variance-covariance structure we calculated deviations from the (estimated) means for each vector. We then tested for a change in the variance-covariance matrix $\mathbf{\Sigma}$, with the null hypothesis being that the covariance matrix did not change over time, versus the alternative hypothesis that the covariance matrix did change at some unknown point of time k. The generalized log-likelihood ratio test statistic is given by

$$U_n^* = \max_{1\le k<n} \log\left(\frac{|\widehat{\mathbf{\Sigma}}_{\mathrm{n}}|^n}{|\widehat{\mathbf{\Sigma}}_{\mathrm{k}}|^k |\widehat{\mathbf{\Sigma}}_{\mathrm{k}}^*|^{n-k}}\right)$$

where the estimators of the variance-covariance matrix are

$$\widehat{\mathbf{\Sigma}}_{\mathrm{n}} = \frac{1}{\mathrm{n}}\left\{\sum_{i=1}^{n}(\boldsymbol{X}_i - \widehat{\boldsymbol{\mu}})(\boldsymbol{X}_i - \widehat{\boldsymbol{\mu}})^T\right\} \qquad \widehat{\mathbf{\Sigma}}_{\mathrm{k}} = \frac{1}{\mathrm{k}}\left\{\sum_{i=1}^{k}(\boldsymbol{X}_i - \widehat{\boldsymbol{\mu}})(\boldsymbol{X}_i - \widehat{\boldsymbol{\mu}})^T\right\}$$

$$\widehat{\mathbf{\Sigma}}_{\mathrm{k}}^* = \frac{1}{\mathrm{n}-\mathrm{k}}\left\{\sum_{i=k+1}^{n}(\boldsymbol{X}_i - \widehat{\boldsymbol{\mu}})(\boldsymbol{X}_i - \widehat{\boldsymbol{\mu}})^T\right\} \qquad \widehat{\boldsymbol{\mu}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i$$

Note that if we know a change has occurred in the mean at $\tau_n = t$ then $\boldsymbol{X}_i$ are the deviations from the mean

$$\boldsymbol{X}_i = \begin{cases} (\boldsymbol{Y}_i - \widehat{\boldsymbol{\mu}}_t), & i = 1, \dots, t \\ (\boldsymbol{Y}_i - \widehat{\boldsymbol{\mu}}_t^*), & i = t+1, \dots, n \end{cases}$$

where $\widehat{\boldsymbol{\mu}}_t$ and $\widehat{\boldsymbol{\mu}}_t^*$ are the estimators of the mean based on the first t and the last $n-t$ values respectively. If no change in the mean vector has occurred then $\boldsymbol{X}_i = (\boldsymbol{Y}_i - \widehat{\boldsymbol{\mu}}_n)$. We obtain the following results

Datasets	U_n^*	w
NL	5.20	1.29
SL	1.73	-0.39
Low	8.45	2.39
NL SL	5.93	0.44
NL Low	14.85	2.92
SL Low	8.94	1.42
NL SL Low	17.25	3.63

Table 3.5. Datasets and corresponding test statistic U_n^* and w values.

Test for change in variance		
Datasets	$\hat{\tau}$	p-value
NL	78	0.4240
SL	13	0.7930
Low	71	0.1681
NL SL	78	0.7671
NL Low	71	0.1019
SL Low	71	0.3850
NL SL Low	71	0.0518

Table 3.6. Datasets and corresponding p-values and change-point estimates $\hat{\tau}$.

$$\widehat{\boldsymbol{\Sigma}}_{\hat{\tau}=46} = \begin{bmatrix} 0.00083 & -0.00104 & -0.00010 \\ -0.00104 & 0.01397 & 0.00051 \\ -0.00010 & 0.00051 & 0.00580 \end{bmatrix}$$

$$\widehat{\boldsymbol{\mu}}_{\hat{\tau}=46} = [1.5019 \quad 4.0159 \quad 1.9150]$$

$$\widehat{\boldsymbol{\mu}}^*_{\hat{\tau}=46} = [1.5611 \quad 4.0778 \quad 2.0107]$$

The mean vectors and the variance-covariance matrix for the three bands [NL, Low, SL] and $\hat{\tau} = 46$ are shown above.

All the univariate and bivariate tests, shown in Tables 3.5 and 3.6, indicate that the variance and covariance did not change for the duration of the time series. In the case of all three bands, the p-value is somewhat small, but still greater than 0.05, thus we conclude that the assumption of constant variance-covariance was not violated.

We now compute the exact limiting distribution of the change-point mle, using the equations derived in section 3.2, and also, for comparison purposes, the Cobb (1978) conditional distribution of the mle (choosing delta that gives us an error rate closest to 10^{-5}). See Tables 3.7 and 3.8 below. We note that the limiting distribution of the unconditional mle has a high peak at zero and tapers off very quickly for the larger η values, and it is still quite peaky for the smaller, $\eta < 2$ values (Low SL, SL). On the other hand, the conditional distribution of the mle tends to asymmetry (see especially the NL SL bivariate case) and is not as tight around the peaks at zero, which are also much lower than the ones in the unconditional case.

Thus we have detected a change in mean precipitation, and can with reasonable confidence suggest that the true change-point lies between the years 1944 and 1948. More specifically, we have detected a change when considering the data for all three bands simultaneously, any combination of the two bands, and Northern and Southern Latitude bands individually. We did not detect a change in the Low Latitudes band, which covers the tropics. We also note that when considering the entire data series (NL SL Low) the estimated change-point $\hat{\tau}$=46 gives us a 99% confidence interval of (44, 48) with the standardized amount of change $\eta = 2.72$ (see Table 3.7).

From the available climate literature we found support for change in precipitation at around that interval of time. Lau and Wu (2006) mention that the studies have shown an increase in precipitation since the 1950s in high latitudes (which are covered by the NL and SL bands). New et al. (2001) also state that global land precipitation has gone up over the past century, with a peak noted in 1950s-1960s. They also mention a large increase in precipitation in the Northern Hemisphere that may be due to an improved measuring gauge introduced in the 1950s. New et al. (2001) also state that in the 40-60°S (included in the SL band) below average precipitation was observed before the 1930s, which was then followed by above average precipitation in the 1930s-1960s. Plummer et al. (1999) assert the same regarding precipitation in New Zealand (part of the SL band) where more serious droughts were observed in the period from 1920-1951 than in the subsequently following 30 year period. Qian and Zhu (2001) report that China experienced high temperatures and periods of below normal precipitation in the 1920s-1940s.

Groisman et al. (1999) investigated climate change in Canada, United States, Mexico, Russia, China, Australia, Norway, and Poland and state that changes in the (monthly) mean precipitation are associated with much larger changes in extreme precipitation. Thus an increase in the mean precipitation is of particular concern as it may indicate an increase in extreme or heavy precipitation that may lead to floods and other natural disasters, this in turn can lead to loss of human life and have a negative impact on the economy.

There may be several reasons why there is a lack of change in the mean tropical precipitation. First, a large part of the tropics is water (ocean), data from before 1970s, when satellites started being used to monitor precipitation over the large water areas, is based on land precipitation data collected by land stations . This data is not necessarily accurate, it is subject to errors, and since land stations in the tropics are few any errors generated are not minimized when data is combined with that from other stations. Second, according to various studies (e.g., Allen & Ingram (2002); New et al. (2001)) the Low band (24° S to 24° N) covers both areas where precipitation has increased and areas where precipitation has decreased. For example, there are some indications that in the 20-40° S band the precipitation has increased, but in the 10° N/S to 20° N/S bands precipitations has decreased. It may be that those changes cancel each other out and thus no overall change is detected.

In conclusion, being able to identify changes in the precipitation allows for a better understanding of Earth’s climate system. It allows us to identify changes in the climate that may be due to human development (e.g., air and water pollution, deforestation) and are not part of the general climate variation. This will allow scientist to build better climate models, and thus give better predictions for various climate scenarios.

k	NL SL η =2.48	NL Low η = 2.40	SL Low η = 1.39	NL SL Low η = 2.72	NL η = 2.11	SL η = 1.33
-20			0.0001			0.0001
-15			0.0003			0.0005
-14			0.0005			0.0006
-13			0.0007			0.0009
-12			0.0009			0.0012
-11			0.0013			0.0016
-10			0.0018		0.0001	0.0023
-9			0.0026		0.0001	0.0032
-8		0.0001	0.0038		0.0003	0.0046
-7	0.0001	0.0002	0.0057		0.0005	0.0066
-6	0.0003	0.0004	0.0085	0.0001	0.0011	0.0097
-5	0.0008	0.0010	0.0131	0.0003	0.0024	0.0146
-4	0.0021	0.0026	0.0209	0.0011	0.0054	0.0226
-3	0.0064	0.0075	0.0348	0.0038	0.0129	0.0368
-2	0.0214	0.0239	0.0628	0.0151	0.0341	0.0646
-1	0.0887	0.0934	0.1316	0.0750	0.1096	0.1311
0	0.7631	0.7453	0.4338	0.8110	0.6721	0.4108
1	0.0887	0.0934	0.1316	0.0750	0.1096	0.1311
2	0.0214	0.0239	0.0628	0.0151	0.0341	0.0646
3	0.0064	0.0075	0.0348	0.0038	0.0129	0.0368
4	0.0021	0.0026	0.0209	0.0011	0.0054	0.0226
5	0.0008	0.0010	0.0131	0.0003	0.0024	0.0146
6	0.0003	0.0004	0.0085	0.0001	0.0011	0.0097
7	0.0001	0.0002	0.0057		0.0005	0.0066
8		0.0001	0.0038		0.0003	0.0046
9			0.0026		0.0001	0.0032
10			0.0018		0.0001	0.0023
11			0.0013			0.0016
12			0.0009			0.0012
13			0.0007			0.0009
14			0.0005			0.0006
20			0.0001			0.0001

Table 3.7. The unconditional probability distribution of $P(\xi_\infty = k)$ for the annual mean precipitation data. η = 2.448, 2.40, 1.39, 2.72, 2.11, 1.33.

	NL SL	NL Low	SL Low	NL SL Low	NL	SL
delta	12	7	15	8	10	18
epsilon	2.4E-05	3.3E-06	7.9E-05	3.0E-06	6.6E-05	1.2E-05
d						
-14						
-13						
-12			0.0003			0.0004
-11			0.0001			0.0002
-10			0.0004			0.0005
-9			0.0007			0.0011
-8			0.0105			0.0166
-7			0.0265			0.0493
-6			0.0031			0.0061
-5			0.0106			0.0127
-4		0.0003	0.0050		0.0009	0.0038
-3		0.0376	0.0018	0.0001	0.0433	0.0011
-2		0.0137	0.0087	0.1634	0.0240	0.0069
-1	0.0003	0.3024	0.1994	0.0983	0.1250	0.1434
0	0.2733	0.5492	0.3087	0.4697	0.6514	0.2513
1	0.2425	0.0784	0.0989	0.1735	0.0811	0.0528
2	0.1981	0.0133	0.0137	0.0704	0.0213	0.0111
3	0.1617	0.0031	0.0465	0.0198	0.0249	0.0330
4	0.0591	0.0014	0.0745	0.0008	0.0138	0.0578
5	0.0254	0.0004	0.0078	0.0028	0.0077	0.0102
6	0.0040	0.0001	0.0616	0.0010	0.0042	0.0876
7	0.0194		0.0631	0.0002	0.0003	0.1077
8	0.0118		0.0384	0.0000	0.0006	0.0780
9	0.0041		0.0178		0.0015	0.0565
10	0.0002		0.0016			0.0099
11	0.0001		0.0001			0.0007
12						0.0006
13						0.0002
14						0.0001
15						0.0001
16						0.0000
17						0.0001
18						

Table 3.8. The conditional probability distribution, Cobb (1978), of the change-point mle for the annual mean precipitation data, with error rate around 10^{-5}.

3.4 Change-point analysis of the mean annual temperature anomalies data

The univariate dataset we analyze now covers the upper part of the northern hemisphere, 64° - 90° N, we have 60 observations $Y_1, Y_2, \dots, Y_{60}$, $d = 1$ and $n = 60$, where Y_i represents the annual mean temperature anomalies, relative to the base period 1951-1980, in .01° Celsius. To be more precise, the anomaly values making up the dataset are the differences from the corresponding 1951-1980 means. See Figure 3.3 and Table 3.9. While absolute temperatures may vary a lot over short distances, annual temperature anomalies are representative of much larger areas, thus working with temperature anomalies is much more valuable. Our purpose in analyzing this data is to improve existing global climate models and to see if we can detect a change in the temperature and thus possible effects of global climate change due to man-made interference. Hansen & Lebedeff (1987) noted several trends in global surface air temperature: the 1880-1940 warming period, the 1940-1965 cooling period, and the 1965-1985 warming period. The dataset we have covers the strong cooling period in the upper northern latitudes (including Alaska, Canada, and Greenland) that the authors mention. Koenker & Schorfheide (1994) reviewed the cooling period (1940-1965), found by Hansen & Lebedeff (1987), and obtained statistical evidence that there is a break in the warming trend that is observed in that interval of time. By applying change-point analysis we wish to determine a better estimate and confidence interval of when that change took place.

Our null hypothesis is that there is no change in the mean in the temperature anomalies dataset, while the alternative hypothesis is that a change in the mean only has occurred at some unknown point in time, which we call τ_n, so that $Y_1, Y_2, \dots, Y_{\tau_n - 1} \sim f_1(\mu_1, \sigma^2)$ and $Y_{\tau_n}, Y_{\tau_n + 1}, \dots, Y_{60} \sim f_2(\mu_2, \sigma^2)$. We assume that the data is independent and normal with $Y_i \sim Normal(\mu^{(i)}, \sigma^2)$, $i = 1, \dots, n$, and that no change occurs in the variance, thus we can write out the hypotheses as follows

$$H_0: \ \mu^{(1)} = \cdots = \mu^{(n)} = \ \mu_1$$

$$H_1: \mu^{(1)} = \cdots = \mu^{(\tau_n - 1)} = \ \mu_1 \neq \mu^{(\tau_n)} = \cdots = \mu^{(n)} = \ \mu_2$$

We then take the likelihood approach to hypothesis testing and calculate the likelihood ratio statistic U_n, its asymptotic theory is presented in Csörgö & Horváth (1997), see Appendix [A.4] and the previous section 3.3 for more details. We obtain the following results

$$U_n = \max_{1 \le t < n} (-2 \log \Lambda_t) = 20.3953$$

See Figure 3.4 for the graph of $-2 \log \Lambda_t$. The limiting distribution of $W_n = (2 \log \log n \, U_n)^{1/2} - \left(2 \log \log n + \frac{d}{2} \log \log \log n - \log \Gamma\left(\frac{d}{2}\right)\right)$ has a double exponential form, $\lim_{n \to \infty} P(W_n \le w) = e^{-2e^{-w}}$, thus we have $w = 5.1643$ and obtain a p-value of 0.01137. Therefore we reject the null hypothesis and conclude that there is strong evidence that a change in the mean annual temperature anomalies in the northern hemisphere, 64° - 90° N, occurred at $\hat{\tau} = 36$ which corresponds to the year 1954, this matches the time interval from Koenker & Schorfheide (1994)

We now check our three assumptions: independence, normality, and constant variance. To test for independence, we computed standardized observation and generated autocorrelation and partial autocorrelation plots, see Figures 3.5 and 3.6. We note no significant lags. To check for normality we applied the Anderson-Darling goodness of fit test to the data before and after change, see Figure 3.7 for results. There were no indications of violations of the normality assumption. And finally, we tested for a change in the variance, with the null hypothesis being that the variance did not change over time, versus the alternative hypothesis that the variance did change at some unknown point of time. See section 3.3 analysis for details, note that $d = 1$ in this instance. We then obtained the following results: $U_n^* = 10.25701$ and $w = 2.90349$, hence the p-value equals 0.10386 and we can conclude that the constant variance assumption was not violated.

We then calculate $\eta = \frac{|\mu_1 - \mu_2|}{\sigma} = \frac{|47.83+7.46|}{42.57} = 1.299$ and compute the exact limiting distribution of the change-point mle, using the equations derived in section 3.2. See Table 3.10 below. Thus we can conclude that the change in the mean of the annual mean temperature anomalies has occurred between the years 1949 and 1959 with a reasonably high (almost 95%) level of confidence. We note that this interval falls within the time interval considered by Koenker & Schorfheide (1994), hence our results agree with their findings and in addition give us a confidence interval for the time of change.

Year	Temp.	Year	Temp.
1919	-75	1949	34
1920	26	1950	23
1921	6	1951	13
1922	-15	1952	24
1923	32	1953	102
1924	40	1954	82
1925	-3	1955	-29
1926	50	1956	-6
1927	2	1957	13
1928	72	1958	-18
1929	19	1959	55
1930	66	1960	37
1931	68	1961	-15
1932	46	1962	48
1933	-8	1963	-14
1934	97	1964	-67
1935	39	1965	-30
1936	40	1966	-85
1937	131	1967	38
1938	151	1968	-32
1939	72	1969	-4
1940	100	1970	-21
1941	-6	1971	-13
1942	48	1972	-38
1943	132	1973	13
1944	115	1974	-20
1945	61	1975	17
1946	4	1976	-7
1947	111	1977	18
1948	23	1978	-19

Table 3.9. Annual mean temperature anomalies in .01° Celsius for the years 1919-1978. Anomaly values indicate the deviations from the corresponding 1951-1980 means.

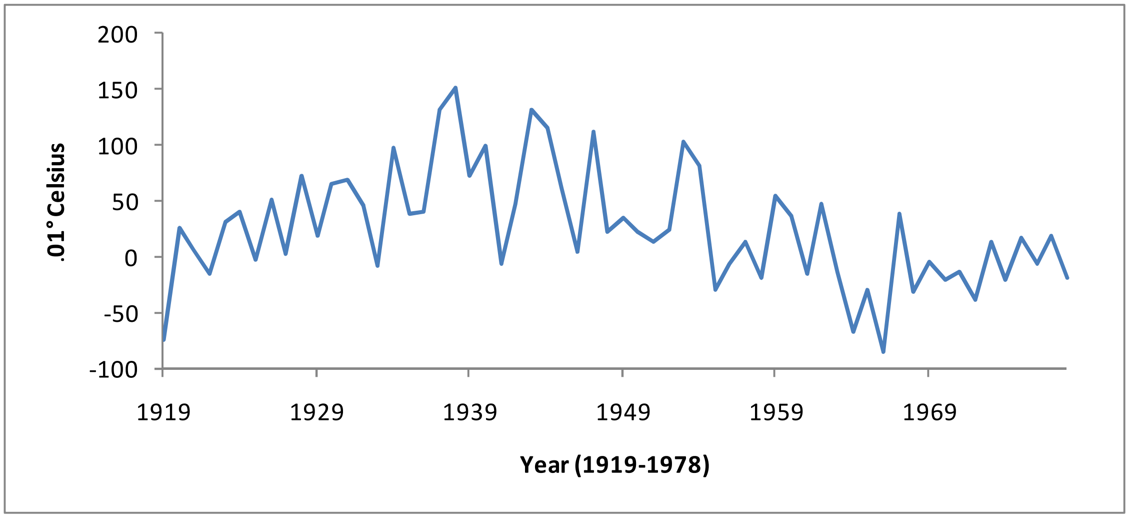


Figure 3.3. Annual mean temperature anomalies in .01° Celsius for the years 1919-1978. Anomaly values indicate the deviations from the corresponding 1951-1980 means.

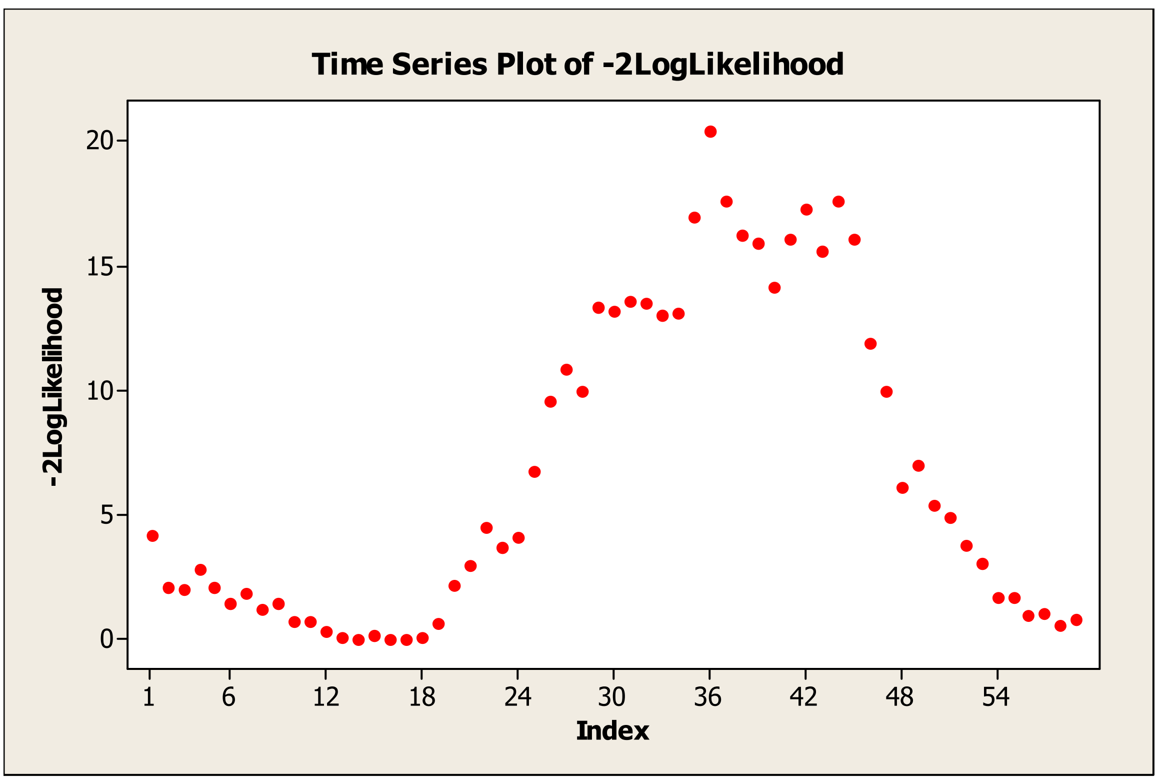


Figure 3.4. Plot of $-2\log\Lambda_t$ of the annual mean temperature anomalies data.

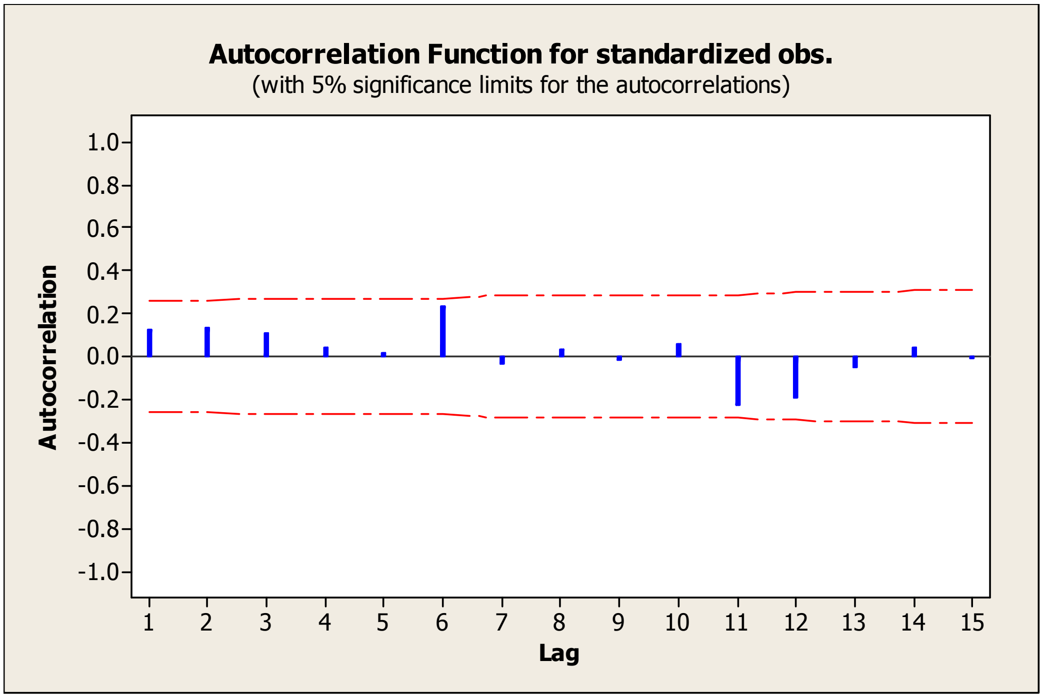


Figure 3.5. Autocorrelation for the standardized observations.

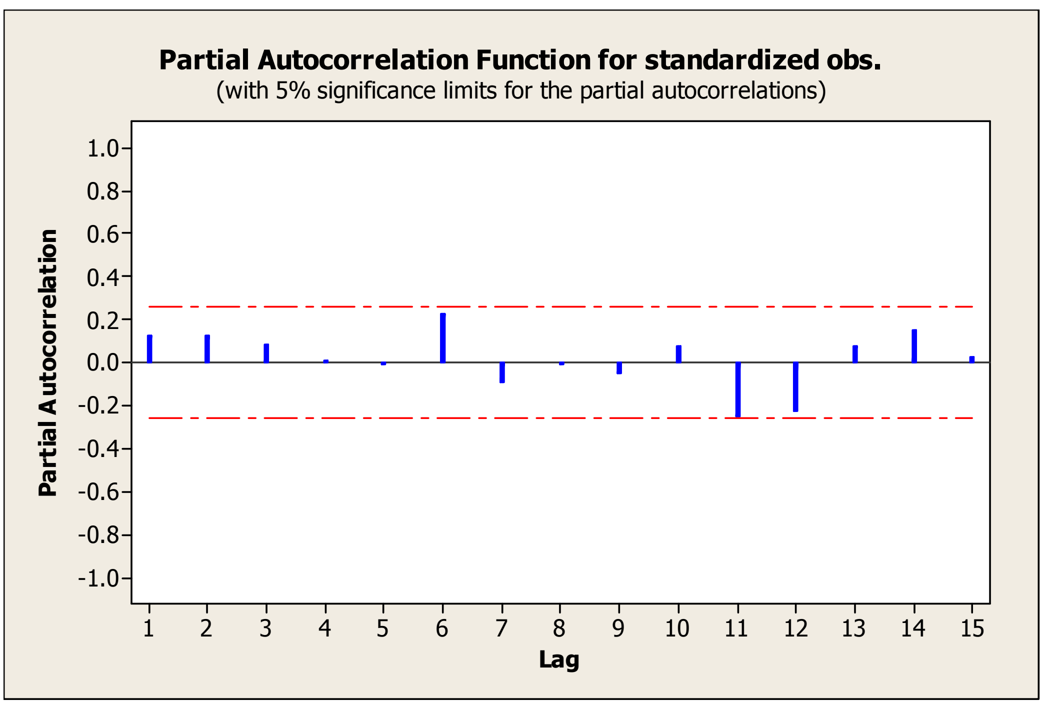


Figure 3.6. Partial Autocorrelation for the standardized observations.

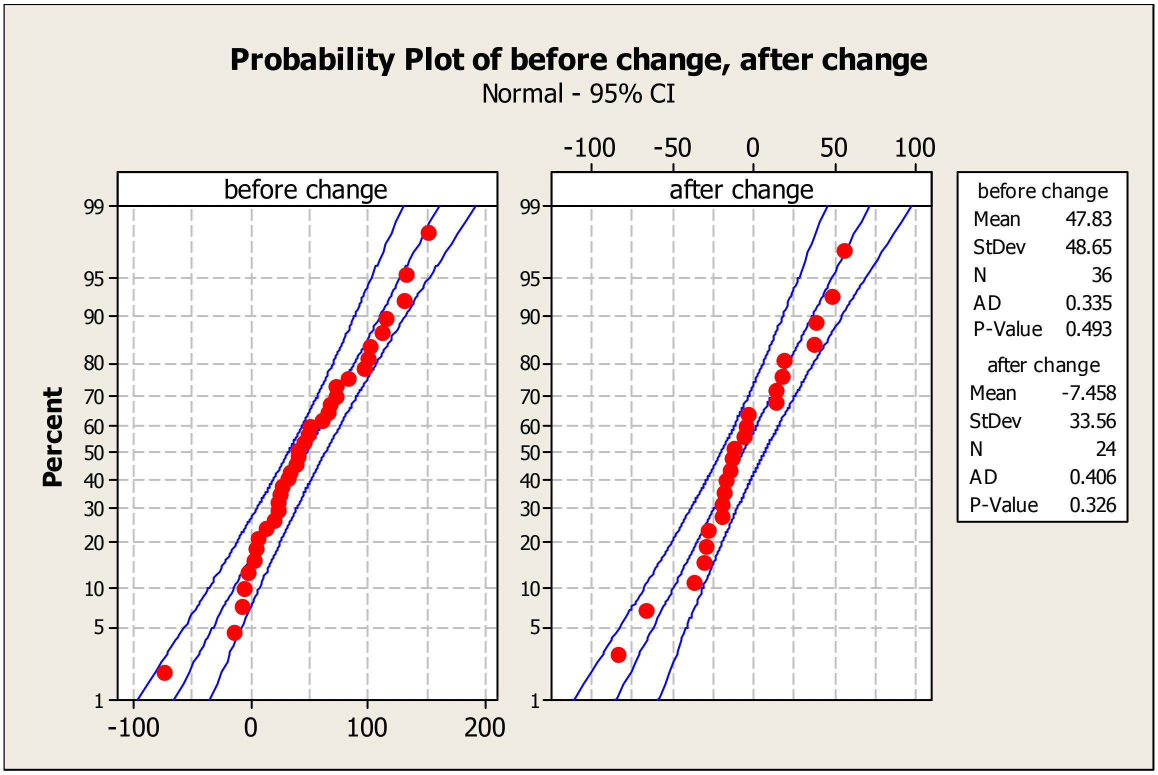


Figure 3.7. Annual mean temperature anomalies data check for normality. Both p-values (0.493 and 0.326) are large and indicate that we do not have sufficient evidence to reject the null hypothesis that the data are normal.

n	η=1.299
-10	0.0025
-9	0.0035
-8	0.0050
-7	0.0071
-6	0.0103
-5	0.0154
-4	0.0236
-3	0.0378
-2	0.0654
-1	0.1306
0	0.3988
1	0.1306
2	0.0654
3	0.0378
4	0.0236
5	0.0154
6	0.0103
7	0.0071
8	0.0050
9	0.0035
10	0.0025

Table 3.10. The unconditional probability distribution of $P(\xi_\infty = n)$ for the annual mean temperature anomalies data, η = 1.299, given for $n = -10, \ldots + 10$.

3.5 Simulations

In this section we will use simulations to investigate the accuracy of the exact limiting distribution, derived in section 3.2, with respect to various sample sizes, amounts of standardized change in the mean η, and locations of the true change-point τ. Furthermore, we would like to test the robustness of the exact limiting distribution to departures from normality (i.e., if the family of distributions is not Gaussian), by utilizing the symmetric standardized t_v-distribution and the asymmetric standardized χ_v^2-distribution.

In addition to the above, we would like to compare our distribution of the change-point mle with the distribution of the change-point mle proposed by Cobb (1978). Our derived distribution, based on Hinkley (1970, 1972), depends strictly on the parameters of the underlying distribution, thus we shall call it the unconditional distribution of the mle. Cobb, on the other hand, proposed conditioning on a sufficient number of data values to either side of the change-point in order to derive the distribution of the change-point mle. With δ representing the number of observations around $\hat{\tau}_n$, the conditional distribution of the change-point mle would be as follows

$$P\left(\hat{\tau}_n - \tau_n = d | Y_{\hat{\tau}_{n-\delta+1}}, \dots, Y_{\hat{\tau}_{n+\delta}}\right) \cong \frac{p_n(Y; \hat{\tau}_n + d)}{\sum_{d=-\delta}^{\delta} p_n(Y; \hat{\tau}_n + d)}$$

Note that we must have $d \in \{-\delta, \dots, \delta\}$, here we chose delta such that the error rate is as close as possible to 10^{-5} (for $\eta = 1.0$ the best error rate we actually obtained was near 10^{-3}). See Cobb (1978) for details on choosing δ, the error rate and detailed probability calculations.

We chose the following sample sizes and corresponding change-points for our simulations,

Sample sizes n	Change-points τ
100	20, 30, 40, 50
60	20, 30
40	20

for each we consider $\eta = 1.0, 2.5$, and the following values of the degrees of freedom v for the t_v-distribution ($d.f. = v = 5, 10, 20$) and the χ_v^2-distribution ($d.f. = v = 1, 5, 20$).

We performed 50,000 simulations in each case and the results are presented in the following Tables 3.11 – 3.17. Table 3.11 shows that the Bias for the distribution of the unconditional mle is very small, in fact it is close to zero for both the normal distribution and the t_v-distribution in all cases except when $\eta = 1.0$ and the change-point is far away from the center of the dataset, even then though, the Bias is negligible. In the case of the conditional mle, Cobb (1978), the Bias is somewhat larger overall, but is still quite small and is not a concern. Looking at the χ_v^2-distribution, we see that the Bias is more significant for both the conditional and unconditional mle, especially when $v = 1$. This is not too surprising considering the asymmetry of the distribution in this case.

The MSE values are fairly close to the theoretical MSE values in almost all cases, they change only marginally as the change-point moves further away from the center and as the sample size gets smaller. We do note however that for $\eta = 2.5$ and the χ_1^2-distribution the MSE values are significantly different. Comparing the mle and cmle distributions, we see that the cmle distribution generally has somewhat larger MSE values.

Figures 3.8 (a) and 3.9 (a) show that when the underlying distribution is normal, the simulated distribution of the unconditional mle clearly matches the theoretical distribution, regardless of the amount of change. On the other hand, the simulated distribution of the conditional mle matches the theoretical distribution when the change is large ($\eta = 2.5$), but is flatter and more spread out when the change is small ($\eta = 1.0$).

Figures 3.8 (b, c) and 3.9 (b, c) show that when the underlying (simulated) distribution is central t_v the unconditional distribution is still in fairly close agreement with the theoretical distribution regardless of the amount of change (η) or the degrees of freedom ($v = 5{,}20$). The simulated distribution of the conditional mle is a fairly close match to the theoretical distribution when $\eta = 2.5$, but when the change is small, $\eta = 1.0$ it again, as in the case of the underlying normal distribution, is flatter and more spread out than the theoretical distribution.

Figures 3.8 (d, e) and 3.9 (d, e) show that when the simulated distribution is χ_v^2 the conditional distribution of the mle is symmetric and closer to the theoretical, and thus performs better, when the degrees of freedom are small $v = 1$ (and the χ_v^2 distribution is highly asymmetric) and the change is large, $\eta = 2.5$. In this instance the distribution of the unconditional mle is somewhat skewed to the right. When the degrees of freedom increase to 20 (and the χ_v^2 distribution becomes more bell-shaped and symmetric) with $\eta = 2.5$, both the conditional and unconditional distributions perform quite well. When the change is small, $\eta = 1.0$ the unconditional distribution of the mle is quite close to the theoretical, and the conditional distribution of the mle is once more flatter and more spread out.

Tables 3.16 and 3.17 give the numerical values for the theoretical and simulated distributions of the mle under departures from normality, for n = 100 and τ = 50 with η = 1.0, 2.5, same as the figures just discussed. In addition, Tables 3.12 – 3.15 give the same values only a for smaller sample size n = 60 and two different change-points τ = 20, 30.

In conclusion, we see that the change-point distribution we derived in section 3.2 is quite robust to departures from normality as long as the true underlying distribution is bell-shaped and symmetric, or if the amount of change is large, $\eta \geq 2.5$. And in all cases, except when the underlying distribution is very skewed and the change is small, the unconditional distribution of the mle does better than the conditional distribution. This is especially noticeable when the amount of change is small.

				normal		central t						chi-square					
						5 *df*		10 *df*		20 *df*		1 *df*		5 *df*		20 *df*	
			Theor.	mle	c mle	mle	c mle	mle	c mle	mle	c mle	mle	c mle	mle	c mle	mle	c mle
n = 100, τ = 20	η = 1.0	Bias	0.000	0.069	0.397	0.081	0.410	0.112	0.440	0.118	0.441	1.070	1.121	0.459	0.659	0.290	0.555
		MSE	26.586	23.206	34.452	22.136	34.278	22.981	34.913	22.988	34.545	28.357	45.845	24.584	38.061	24.176	36.804
	η = 2.5	Bias	0.000	-0.002	0.014	-0.001	0.016	-0.008	0.012	-0.004	0.011	0.274	0.218	0.191	0.137	0.096	0.082
		MSE	0.592	0.602	0.903	0.673	0.977	0.599	0.927	0.597	0.914	0.837	1.181	0.591	0.968	0.593	0.928
n = 100, τ = 30	η = 1.0	Bias	0.000	0.000	0.164	0.079	0.241	-0.016	0.146	0.033	0.201	1.020	0.874	0.453	0.523	0.276	0.364
		MSE	26.586	24.138	37.272	24.743	37.705	24.629	37.476	24.574	37.108	27.010	43.309	25.608	40.118	25.861	39.468
	η = 2.5	Bias	0.000	0.001	0.010	0.005	0.012	-0.007	0.005	-0.001	0.005	0.271	0.205	0.193	0.131	0.092	0.071
		MSE	0.592	0.574	0.878	0.674	0.988	0.591	0.903	0.589	0.887	0.840	1.151	0.601	0.939	0.574	0.889
n = 100, τ = 40	η = 1.0	Bias	0.000	0.012	0.081	0.011	0.069	-0.022	0.064	-0.019	0.056	1.030	0.767	0.460	0.413	0.230	0.249
		MSE	26.586	26.134	38.478	26.000	39.461	25.235	38.254	25.060	37.847	26.793	41.736	26.035	39.571	25.587	38.942
	η = 2.5	Bias	0.000	0.001	0.004	0.007	0.008	0.003	0.008	0.001	0.003	0.273	0.201	0.189	0.123	0.086	0.063
		MSE	0.592	0.590	0.888	0.671	0.972	0.604	0.910	0.591	0.894	0.839	1.113	0.602	0.943	0.591	0.895
n = 100, τ = 50	η = 1.0	Bias	0.000	0.018	0.016	-0.037	-0.040	0.015	-0.001	0.016	-0.001	1.009	0.673	0.418	0.305	0.219	0.160
		MSE	26.586	25.607	38.381	25.143	38.823	25.622	38.638	26.129	39.020	25.996	39.195	26.426	39.215	26.240	39.242
	η = 2.5	Bias	0.000	0.000	-0.001	0.001	-0.003	-0.003	-0.003	0.002	0.003	0.271	0.200	0.197	0.125	0.103	0.069
		MSE	0.592	0.577	0.887	0.671	0.983	0.601	0.909	0.570	0.875	0.852	1.125	0.612	0.936	0.599	0.900
n = 60, τ = 20	η = 1.0	Bias	0.000	0.048	0.272	0.103	0.331	0.101	0.311	0.070	0.293	0.977	0.896	0.429	0.538	0.284	0.425
		MSE	26.586	23.361	31.990	21.862	31.105	22.888	31.838	22.841	31.559	24.933	35.109	23.877	33.530	23.173	32.538
	η = 2.5	Bias	0.000	0.003	0.010	0.005	0.011	-0.003	0.006	0.002	0.011	0.273	0.207	0.189	0.130	0.088	0.074
		MSE	0.592	0.566	0.880	0.682	1.003	0.601	0.918	0.584	0.899	0.833	1.184	0.601	0.959	0.581	0.919
n = 60, τ = 30	η = 1.0	Bias	0.000	0.041	0.027	0.003	-0.006	0.010	-0.001	0.011	0.004	0.914	0.569	0.373	0.241	0.181	0.115
		MSE	26.586	23.402	32.923	23.792	32.647	22.979	32.893	24.130	33.522	21.791	30.758	23.797	32.510	23.938	33.000
	η = 2.5	Bias	0.000	-0.008	-0.007	-0.001	0.000	0.274	-0.001	-0.002	-0.003	0.276	0.203	0.191	0.122	0.094	0.066
		MSE	0.592	0.581	0.890	0.633	0.948	0.837	0.907	0.583	0.881	0.869	1.170	0.596	0.932	0.589	0.901
n = 40, τ = 20	η = 1.0	Bias	0.000	0.044	0.044	0.021	0.009	0.058	0.031	-0.005	-0.001	0.745	0.431	0.317	0.170	0.147	0.073
		MSE	26.586	20.091	24.829	19.193	23.906	19.907	24.671	19.738	24.568	17.515	21.582	19.813	24.297	20.192	24.638
	η = 2.5	Bias	0.000	0.000	0.000	-0.005	-0.003	-0.002	-0.002	0.004	0.006	0.277	0.203	0.191	0.127	0.088	0.060
		MSE	0.592	0.581	0.885	0.678	0.997	0.623	0.934	0.593	0.894	0.881	1.212	0.624	0.965	0.563	0.875

Table 3.11. Bias and MSE for the distributions of mle and cmle of the change-point under normal, t_v, and χ_v^2 distributions .

		normal		central t						chi-square					
				5 *df*		10 *df*		20 *df*		1 *df*		5 *df*		20 *df*	
	Theor.	mle	c mle	mle	c mle	mle	c mle	mle	c mle	mle	c mle	mle	c mle	mle	c mle
-20	0.001														
-19	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.000	0.001	0.001	0.001	0.001	0.001
-18	0.001	0.002	0.001	0.002	0.001	0.001	0.001	0.002	0.001	0.000	0.001	0.001	0.001	0.001	0.001
-17	0.002	0.002	0.002	0.001	0.002	0.001	0.002	0.002	0.002	0.000	0.001	0.001	0.001	0.001	0.001
-16	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.000	0.001	0.001	0.002	0.002	0.002
-15	0.002	0.002	0.003	0.002	0.003	0.002	0.003	0.002	0.003	0.000	0.002	0.001	0.002	0.002	0.002
-14	0.003	0.003	0.004	0.003	0.003	0.003	0.004	0.003	0.004	0.001	0.002	0.002	0.003	0.002	0.003
-13	0.003	0.004	0.005	0.003	0.004	0.003	0.004	0.004	0.004	0.001	0.003	0.002	0.004	0.003	0.004
-12	0.004	0.005	0.006	0.004	0.005	0.004	0.006	0.004	0.006	0.001	0.003	0.003	0.005	0.004	0.005
-11	0.005	0.005	0.007	0.004	0.006	0.005	0.007	0.005	0.007	0.002	0.005	0.004	0.006	0.005	0.007
-10	0.006	0.006	0.009	0.006	0.008	0.006	0.009	0.006	0.009	0.002	0.006	0.005	0.008	0.006	0.009
-9	0.008	0.008	0.011	0.008	0.011	0.008	0.011	0.008	0.011	0.003	0.008	0.006	0.010	0.007	0.011
-8	0.010	0.010	0.014	0.009	0.013	0.010	0.014	0.010	0.014	0.005	0.010	0.008	0.013	0.011	0.014
-7	0.013	0.014	0.018	0.011	0.017	0.013	0.018	0.013	0.018	0.007	0.014	0.012	0.017	0.011	0.017
-6	0.017	0.018	0.023	0.017	0.022	0.016	0.022	0.017	0.023	0.011	0.019	0.016	0.022	0.016	0.022
-5	0.023	0.022	0.029	0.021	0.028	0.022	0.029	0.024	0.029	0.016	0.025	0.022	0.029	0.022	0.029
-4	0.032	0.031	0.038	0.029	0.038	0.031	0.038	0.032	0.039	0.028	0.035	0.031	0.038	0.030	0.038
-3	0.045	0.043	0.052	0.041	0.051	0.042	0.051	0.044	0.052	0.041	0.050	0.044	0.052	0.044	0.052
-2	0.069	0.066	0.073	0.064	0.073	0.067	0.073	0.067	0.073	0.071	0.074	0.069	0.073	0.067	0.073
-1	0.118	0.115	0.108	0.111	0.110	0.114	0.109	0.113	0.108	0.135	0.115	0.122	0.109	0.120	0.109
0	0.280	0.281	0.175	0.317	0.184	0.292	0.177	0.285	0.176	0.308	0.201	0.283	0.179	0.277	0.175
1	0.118	0.115	0.105	0.109	0.106	0.112	0.105	0.112	0.105	0.105	0.106	0.110	0.105	0.116	0.106
2	0.069	0.068	0.070	0.063	0.070	0.065	0.069	0.067	0.069	0.060	0.067	0.064	0.069	0.065	0.069
3	0.045	0.043	0.049	0.042	0.049	0.043	0.049	0.043	0.049	0.043	0.047	0.043	0.049	0.045	0.050
4	0.032	0.031	0.036	0.028	0.036	0.031	0.036	0.031	0.036	0.030	0.035	0.031	0.036	0.031	0.037
5	0.023	0.022	0.027	0.022	0.027	0.022	0.027	0.023	0.027	0.023	0.027	0.023	0.028	0.023	0.028
6	0.017	0.016	0.021	0.016	0.021	0.017	0.021	0.017	0.021	0.019	0.021	0.017	0.022	0.018	0.022
7	0.013	0.012	0.016	0.013	0.017	0.013	0.017	0.012	0.017	0.014	0.017	0.013	0.017	0.014	0.017
8	0.010	0.011	0.014	0.009	0.013	0.010	0.013	0.010	0.013	0.011	0.014	0.011	0.014	0.011	0.014
9	0.008	0.008	0.011	0.007	0.010	0.008	0.011	0.008	0.011	0.009	0.011	0.009	0.012	0.008	0.011
10	0.006	0.007	0.009	0.006	0.008	0.007	0.009	0.006	0.009	0.008	0.009	0.008	0.010	0.006	0.009
11	0.005	0.005	0.007	0.005	0.007	0.005	0.007	0.006	0.007	0.007	0.008	0.006	0.008	0.006	0.008
12	0.004	0.004	0.006	0.004	0.006	0.004	0.006	0.004	0.006	0.005	0.007	0.005	0.007	0.004	0.006
13	0.003	0.003	0.005	0.003	0.005	0.003	0.005	0.003	0.005	0.005	0.006	0.004	0.006	0.003	0.005
14	0.003	0.002	0.004	0.003	0.004	0.003	0.004	0.003	0.004	0.004	0.005	0.003	0.005	0.003	0.005
15	0.002	0.002	0.003	0.002	0.003	0.002	0.004	0.003	0.004	0.003	0.004	0.003	0.004	0.002	0.004
16	0.002	0.002	0.003	0.002	0.003	0.002	0.003	0.002	0.003	0.003	0.004	0.002	0.003	0.002	0.003
17	0.002	0.001	0.002	0.002	0.003	0.001	0.002	0.002	0.003	0.002	0.003	0.002	0.003	0.002	0.003
18	0.001	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.002	0.002	0.003	0.002	0.002	0.001	0.002
19	0.001	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.002	0.002	0.002	0.002	0.002	0.001	0.002
20	0.001	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.002

Table 3.12. Theoretical distribution of the mle and the simulated distributions of the change-point mle and cmle under normal, t_v, and χ_v^2 distributions when n = 60 and τ = 20 and for η = 1.0, k = -20, …, 20.

		normal		central t						chi-square					
				5 *df*		10 *df*		20 *df*		1 *df*		5 *df*		20 *df*	
	Theor.	mle	c mle	mle	c mle	mle	c mle	mle	c mle	mle	c mle	mle	c mle	mle	c mle
-20	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.000	0.001	0.001	0.001	0.001	0.001
-19	0.001	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.002	0.000	0.001	0.001	0.001	0.001	0.001
-18	0.001	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.002	0.000	0.001	0.001	0.001	0.001	0.002
-17	0.002	0.001	0.002	0.002	0.002	0.001	0.002	0.001	0.002	0.000	0.001	0.001	0.002	0.001	0.002
-16	0.002	0.002	0.003	0.002	0.003	0.002	0.003	0.002	0.003	0.000	0.002	0.001	0.002	0.001	0.002
-15	0.002	0.002	0.003	0.002	0.003	0.002	0.003	0.002	0.003	0.000	0.002	0.001	0.003	0.002	0.003
-14	0.003	0.003	0.004	0.002	0.004	0.003	0.004	0.003	0.004	0.001	0.002	0.002	0.003	0.003	0.004
-13	0.003	0.003	0.005	0.003	0.005	0.003	0.005	0.003	0.005	0.001	0.003	0.002	0.004	0.003	0.005
-12	0.004	0.003	0.006	0.004	0.006	0.004	0.006	0.004	0.006	0.001	0.004	0.003	0.005	0.004	0.006
-11	0.005	0.005	0.007	0.004	0.007	0.005	0.007	0.005	0.007	0.002	0.005	0.004	0.007	0.004	0.007
-10	0.006	0.006	0.009	0.006	0.008	0.006	0.009	0.006	0.009	0.002	0.006	0.005	0.008	0.006	0.009
-9	0.008	0.008	0.011	0.007	0.011	0.008	0.011	0.009	0.011	0.003	0.008	0.006	0.010	0.007	0.011
-8	0.010	0.010	0.014	0.010	0.013	0.010	0.014	0.010	0.014	0.005	0.010	0.008	0.013	0.010	0.014
-7	0.013	0.012	0.017	0.013	0.017	0.012	0.017	0.012	0.018	0.007	0.014	0.012	0.017	0.011	0.017
-6	0.017	0.017	0.022	0.015	0.021	0.017	0.022	0.017	0.022	0.010	0.019	0.015	0.022	0.016	0.022
-5	0.023	0.021	0.028	0.021	0.027	0.021	0.028	0.023	0.029	0.016	0.025	0.021	0.028	0.022	0.028
-4	0.032	0.031	0.038	0.028	0.036	0.031	0.038	0.030	0.038	0.025	0.034	0.030	0.037	0.031	0.038
-3	0.045	0.044	0.051	0.040	0.050	0.044	0.051	0.043	0.051	0.041	0.049	0.047	0.051	0.043	0.051
-2	0.069	0.067	0.072	0.063	0.071	0.065	0.072	0.067	0.072	0.071	0.073	0.070	0.073	0.071	0.073
-1	0.118	0.112	0.106	0.110	0.109	0.114	0.108	0.113	0.107	0.138	0.115	0.121	0.109	0.117	0.108
0	0.280	0.281	0.175	0.315	0.184	0.294	0.177	0.284	0.174	0.307	0.203	0.282	0.180	0.282	0.176
1	0.118	0.114	0.107	0.112	0.109	0.114	0.107	0.114	0.106	0.107	0.109	0.109	0.106	0.110	0.106
2	0.069	0.070	0.073	0.063	0.071	0.064	0.071	0.066	0.072	0.063	0.069	0.064	0.070	0.066	0.072
3	0.045	0.045	0.052	0.042	0.051	0.042	0.051	0.043	0.051	0.042	0.048	0.044	0.050	0.044	0.050
4	0.032	0.031	0.038	0.030	0.037	0.031	0.037	0.030	0.037	0.030	0.035	0.031	0.037	0.031	0.037
5	0.023	0.023	0.029	0.022	0.028	0.023	0.028	0.021	0.028	0.023	0.027	0.024	0.028	0.023	0.028
6	0.017	0.017	0.022	0.015	0.021	0.016	0.022	0.018	0.022	0.018	0.021	0.019	0.022	0.017	0.022
7	0.013	0.012	0.017	0.012	0.017	0.012	0.017	0.013	0.018	0.015	0.017	0.013	0.018	0.013	0.017
8	0.010	0.010	0.014	0.010	0.013	0.010	0.014	0.010	0.014	0.012	0.014	0.011	0.014	0.010	0.014
9	0.008	0.007	0.011	0.008	0.011	0.008	0.011	0.008	0.011	0.009	0.011	0.009	0.012	0.008	0.011
10	0.006	0.007	0.009	0.006	0.009	0.006	0.009	0.006	0.009	0.008	0.010	0.007	0.009	0.007	0.009
11	0.005	0.005	0.007	0.004	0.007	0.005	0.007	0.005	0.007	0.007	0.008	0.005	0.008	0.005	0.008
12	0.004	0.004	0.006	0.004	0.006	0.005	0.006	0.004	0.006	0.005	0.007	0.004	0.006	0.005	0.006
13	0.003	0.003	0.005	0.003	0.005	0.003	0.005	0.004	0.005	0.004	0.005	0.004	0.005	0.004	0.005
14	0.003	0.002	0.004	0.002	0.004	0.003	0.004	0.003	0.004	0.004	0.005	0.003	0.004	0.003	0.004
15	0.002	0.002	0.003	0.002	0.003	0.002	0.003	0.002	0.003	0.003	0.004	0.003	0.004	0.003	0.004
16	0.002	0.002	0.003	0.002	0.003	0.001	0.003	0.002	0.003	0.004	0.004	0.002	0.003	0.002	0.003
17	0.002	0.001	0.002	0.001	0.002	0.002	0.002	0.002	0.002	0.003	0.003	0.002	0.003	0.002	0.003
18	0.001	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.002	0.002	0.003	0.002	0.002	0.002	0.002
19	0.001	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.002	0.002	0.002	0.002	0.002	0.001	0.002
20	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.002	0.002	0.001	0.002	0.001	0.001

Table 3.13. Theoretical distribution of the mle and the simulated distributions of the change-point mle and cmle under normal, t_v, and χ_v^2 distributions when n = 60 and τ=30 and for η = 1.0, k = -20, …, 20.

		normal		central t						chi-square					
				5 *df*		10 *df*		20 *df*		1 *df*		5 *df*		20 *df*	
	Theor.	mle	c mle	mle	c mle	mle	c mle	mle	c mle	mle	c mle	mle	c mle	mle	c mle
-20	0.000														
-19	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-18	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-17	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-16	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-15	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-14	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-13	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-12	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-11	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-9	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-6	0.000	0.000	0.000	0.001	0.001	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-5	0.001	0.001	0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.001
-4	0.002	0.002	0.003	0.002	0.004	0.003	0.003	0.002	0.003	0.000	0.001	0.000	0.001	0.001	0.002
-3	0.006	0.006	0.009	0.007	0.010	0.006	0.010	0.006	0.009	0.000	0.003	0.000	0.004	0.003	0.007
-2	0.021	0.020	0.030	0.017	0.027	0.020	0.030	0.020	0.029	0.000	0.013	0.003	0.019	0.013	0.025
-1	0.088	0.086	0.115	0.069	0.103	0.080	0.112	0.086	0.116	0.000	0.070	0.033	0.099	0.071	0.110
0	0.767	0.769	0.679	0.800	0.703	0.780	0.686	0.772	0.681	0.855	0.739	0.821	0.699	0.779	0.682
1	0.088	0.087	0.115	0.072	0.105	0.079	0.110	0.083	0.113	0.083	0.102	0.095	0.115	0.093	0.117
2	0.021	0.021	0.031	0.019	0.028	0.020	0.030	0.021	0.031	0.032	0.036	0.028	0.035	0.026	0.035
3	0.006	0.006	0.010	0.007	0.010	0.006	0.010	0.006	0.010	0.015	0.016	0.011	0.014	0.008	0.012
4	0.002	0.002	0.004	0.003	0.004	0.002	0.004	0.002	0.004	0.007	0.008	0.005	0.006	0.004	0.005
5	0.001	0.001	0.001	0.001	0.002	0.001	0.002	0.001	0.002	0.004	0.004	0.002	0.003	0.001	0.002
6	0.000	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.002	0.002	0.001	0.001	0.001	0.001
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.001	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.000	0.000	0.000
9	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.000	0.000	0.000
10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
11	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
12	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
13	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
14	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
15	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
16	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
17	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
18	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
19	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 3.14. Theoretical distribution of the mle and the simulated distributions of the change-point mle and cmle under normal, t_v, and χ_v^2 distributions when n = 60 and τ=20 and for η = 2.5, k = -20, …, 20.

		normal		central t						chi-square					
				5 *df*		10 *df*		20 *df*		1 *df*		5 *df*		20 *df*	
	Theor.	mle	c mle	mle	c mle	mle	c mle	mle	c mle	mle	c mle	mle	c mle	mle	c mle
-20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-19	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-18	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-17	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-16	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-15	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-14	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-13	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-12	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-11	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-9	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-6	0.000	0.000	0.000	0.001	0.001	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-5	0.001	0.001	0.001	0.001	0.002	0.000	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.001
-4	0.002	0.002	0.003	0.002	0.004	0.000	0.003	0.002	0.003	0.000	0.001	0.000	0.001	0.001	0.002
-3	0.006	0.007	0.010	0.006	0.010	0.000	0.010	0.006	0.010	0.000	0.003	0.000	0.005	0.003	0.007
-2	0.021	0.021	0.031	0.020	0.028	0.000	0.030	0.020	0.031	0.000	0.013	0.002	0.019	0.013	0.026
-1	0.088	0.086	0.116	0.072	0.105	0.000	0.114	0.083	0.114	0.000	0.069	0.033	0.099	0.070	0.109
0	0.767	0.769	0.677	0.798	0.701	0.856	0.684	0.775	0.682	0.855	0.742	0.821	0.701	0.780	0.683
1	0.088	0.085	0.115	0.070	0.104	0.080	0.112	0.082	0.113	0.083	0.102	0.094	0.114	0.092	0.117
2	0.021	0.021	0.030	0.019	0.029	0.032	0.030	0.021	0.030	0.032	0.036	0.029	0.035	0.025	0.034
3	0.006	0.006	0.009	0.007	0.010	0.015	0.010	0.006	0.010	0.013	0.015	0.012	0.014	0.010	0.013
4	0.002	0.002	0.003	0.003	0.004	0.007	0.003	0.002	0.003	0.008	0.008	0.004	0.006	0.004	0.005
5	0.001	0.001	0.001	0.001	0.002	0.004	0.001	0.001	0.001	0.004	0.004	0.002	0.003	0.001	0.002
6	0.000	0.000	0.000	0.000	0.001	0.002	0.001	0.000	0.000	0.002	0.002	0.001	0.001	0.001	0.001
7	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.001	0.000	0.001	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.001	0.000	0.000	0.000	0.000
9	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
11	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
12	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
13	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
14	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
15	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
16	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
17	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
18	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
19	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 3.15. Theoretical distribution of the mle and the simulated distributions of the change-point mle and cmle under normal, t_v, and χ_v^2 distributions when n = 60 and τ=30 and for η = 2.5, k = -20, …, 20.

		chi-square						central t						normal	
		1 *df*		5 *df*		20 *df*		5 *df*		10 *df*		20 *df*			
	Theor.	mle	c mle	mle	c mle	mle	c mle	mle	c mle	mle	c mle	mle	c mle	mle	c mle
-20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-19	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-18	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-17	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-16	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-15	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-14	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-13	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-12	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-11	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-9	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.001	0.000	0.000	0.000	0.001
-5	0.001	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
-4	0.002	0.000	0.001	0.000	0.001	0.001	0.002	0.003	0.004	0.003	0.004	0.002	0.003	0.002	0.004
-3	0.006	0.000	0.003	0.000	0.005	0.003	0.007	0.006	0.010	0.006	0.010	0.006	0.009	0.006	0.010
-2	0.021	0.000	0.012	0.003	0.020	0.011	0.025	0.019	0.029	0.020	0.030	0.020	0.030	0.021	0.031
-1	0.088	0.000	0.074	0.033	0.101	0.070	0.111	0.071	0.105	0.078	0.112	0.083	0.114	0.086	0.116
0	0.767	0.859	0.741	0.818	0.696	0.779	0.682	0.797	0.698	0.784	0.686	0.776	0.682	0.766	0.676
1	0.088	0.081	0.100	0.095	0.115	0.094	0.117	0.072	0.107	0.079	0.112	0.083	0.114	0.088	0.117
2	0.021	0.030	0.035	0.031	0.037	0.027	0.034	0.019	0.029	0.019	0.029	0.020	0.031	0.021	0.031
3	0.006	0.014	0.016	0.011	0.014	0.008	0.012	0.006	0.009	0.007	0.010	0.006	0.010	0.005	0.009
4	0.002	0.007	0.007	0.005	0.006	0.004	0.005	0.002	0.003	0.002	0.004	0.002	0.003	0.002	0.003
5	0.001	0.004	0.004	0.002	0.003	0.001	0.002	0.001	0.002	0.001	0.001	0.001	0.001	0.001	0.001
6	0.000	0.002	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.000	0.001	0.000	0.000	0.000	0.001
7	0.000	0.001	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
11	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
12	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
13	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
14	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
15	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
16	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
17	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
18	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
19	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 3.16. Theoretical distribution of the mle and the simulated distributions of the change-point mle and cmle under normal, t_v, and χ_v^2 distributions when n = 100 and τ = 50 and for η = 2.5, k = -20, …, 20.

		chi-square						central t						normal	
		1 *df*		5 *df*		20 *df*		5 *df*		10 *df*		20 *df*			
	Theor.	mle	c mle	mle	c mle	mle	c mle	mle	c mle	mle	c mle	mle	c mle	mle	c mle
-20	0.001	0.000	0.001	0.000	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
-19	0.001	0.000	0.001	0.001	0.001	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.002
-18	0.001	0.000	0.001	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.002
-17	0.002	0.000	0.001	0.001	0.002	0.001	0.002	0.002	0.002	0.001	0.002	0.001	0.002	0.001	0.002
-16	0.002	0.000	0.002	0.001	0.002	0.002	0.003	0.002	0.003	0.002	0.003	0.002	0.003	0.002	0.003
-15	0.002	0.000	0.002	0.001	0.003	0.002	0.003	0.002	0.003	0.002	0.003	0.002	0.003	0.002	0.003
-14	0.003	0.001	0.003	0.002	0.003	0.002	0.004	0.003	0.004	0.003	0.004	0.002	0.004	0.002	0.004
-13	0.003	0.001	0.003	0.003	0.004	0.003	0.005	0.003	0.005	0.003	0.005	0.004	0.005	0.003	0.005
-12	0.004	0.001	0.004	0.003	0.005	0.004	0.006	0.004	0.006	0.003	0.006	0.004	0.006	0.004	0.006
-11	0.005	0.001	0.005	0.004	0.007	0.004	0.007	0.005	0.007	0.005	0.007	0.005	0.007	0.004	0.007
-10	0.006	0.002	0.007	0.005	0.008	0.005	0.009	0.006	0.009	0.006	0.009	0.006	0.009	0.006	0.009
-9	0.008	0.003	0.009	0.006	0.010	0.007	0.011	0.007	0.011	0.008	0.011	0.007	0.011	0.008	0.011
-8	0.010	0.005	0.011	0.008	0.013	0.009	0.014	0.010	0.014	0.010	0.014	0.010	0.014	0.010	0.014
-7	0.013	0.007	0.015	0.011	0.017	0.012	0.017	0.013	0.017	0.012	0.017	0.013	0.017	0.013	0.018
-6	0.017	0.011	0.020	0.016	0.022	0.017	0.022	0.015	0.022	0.016	0.022	0.017	0.022	0.017	0.023
-5	0.023	0.016	0.027	0.023	0.029	0.021	0.029	0.020	0.028	0.022	0.029	0.021	0.029	0.023	0.029
-4	0.032	0.024	0.036	0.029	0.038	0.032	0.039	0.029	0.037	0.029	0.038	0.031	0.038	0.031	0.038
-3	0.045	0.039	0.051	0.044	0.052	0.044	0.052	0.041	0.050	0.044	0.052	0.043	0.052	0.045	0.052
-2	0.069	0.071	0.074	0.071	0.073	0.068	0.073	0.062	0.072	0.066	0.072	0.065	0.072	0.067	0.073
-1	0.118	0.134	0.114	0.118	0.108	0.117	0.107	0.109	0.108	0.114	0.107	0.114	0.107	0.116	0.106
0	0.280	0.313	0.191	0.281	0.172	0.281	0.170	0.320	0.178	0.292	0.171	0.287	0.170	0.280	0.169
1	0.118	0.108	0.107	0.108	0.105	0.111	0.105	0.109	0.107	0.111	0.106	0.113	0.106	0.113	0.106
2	0.069	0.062	0.070	0.068	0.071	0.067	0.071	0.061	0.071	0.067	0.072	0.066	0.072	0.068	0.072
3	0.045	0.042	0.049	0.044	0.051	0.043	0.051	0.041	0.050	0.043	0.051	0.044	0.051	0.043	0.052
4	0.032	0.028	0.035	0.030	0.038	0.030	0.038	0.029	0.037	0.030	0.038	0.031	0.038	0.030	0.038
5	0.023	0.024	0.027	0.023	0.029	0.023	0.029	0.021	0.028	0.020	0.029	0.022	0.029	0.023	0.029
6	0.017	0.018	0.021	0.017	0.022	0.018	0.022	0.017	0.022	0.017	0.022	0.017	0.023	0.017	0.022
7	0.013	0.014	0.017	0.014	0.018	0.013	0.017	0.012	0.017	0.014	0.018	0.013	0.018	0.012	0.017
8	0.010	0.012	0.014	0.011	0.015	0.010	0.014	0.009	0.013	0.010	0.014	0.010	0.014	0.009	0.014
9	0.008	0.009	0.012	0.009	0.012	0.009	0.012	0.007	0.011	0.008	0.011	0.008	0.011	0.008	0.011
10	0.006	0.007	0.009	0.007	0.010	0.007	0.009	0.006	0.009	0.006	0.009	0.006	0.009	0.006	0.009
11	0.005	0.007	0.008	0.006	0.008	0.006	0.008	0.005	0.007	0.005	0.007	0.005	0.007	0.006	0.007
12	0.004	0.006	0.007	0.004	0.006	0.005	0.006	0.004	0.006	0.004	0.006	0.004	0.006	0.004	0.006
13	0.003	0.005	0.006	0.004	0.005	0.003	0.005	0.003	0.005	0.003	0.005	0.003	0.005	0.003	0.005
14	0.003	0.004	0.005	0.003	0.004	0.003	0.004	0.002	0.004	0.002	0.004	0.002	0.004	0.002	0.004
15	0.002	0.003	0.004	0.003	0.004	0.003	0.004	0.002	0.003	0.002	0.003	0.002	0.003	0.002	0.003
16	0.002	0.003	0.004	0.002	0.003	0.002	0.003	0.002	0.003	0.002	0.003	0.002	0.003	0.002	0.003
17	0.002	0.003	0.003	0.002	0.003	0.002	0.003	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
18	0.001	0.002	0.003	0.002	0.002	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.002
19	0.001	0.001	0.002	0.002	0.002	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.002
20	0.001	0.002	0.002	0.001	0.002	0.001	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001

Table 3.17. Theoretical distribution of the mle and the simulated distributions of the change-point mle and cmle under normal, t_v, and χ_v^2 distributions when n = 100 and τ = 50 and for η = 1.0, k = -20, …, 20.

The following Figures 3.2 (a) –(e) show the theoretical distribution of the mle (under normal distribution) and the simulated distributions of the change-point mle and cmle under normal, t_v, and χ_v^2 distributions when n = 100 and τ = 50 and for η=2.5, k = -8, …, 8.

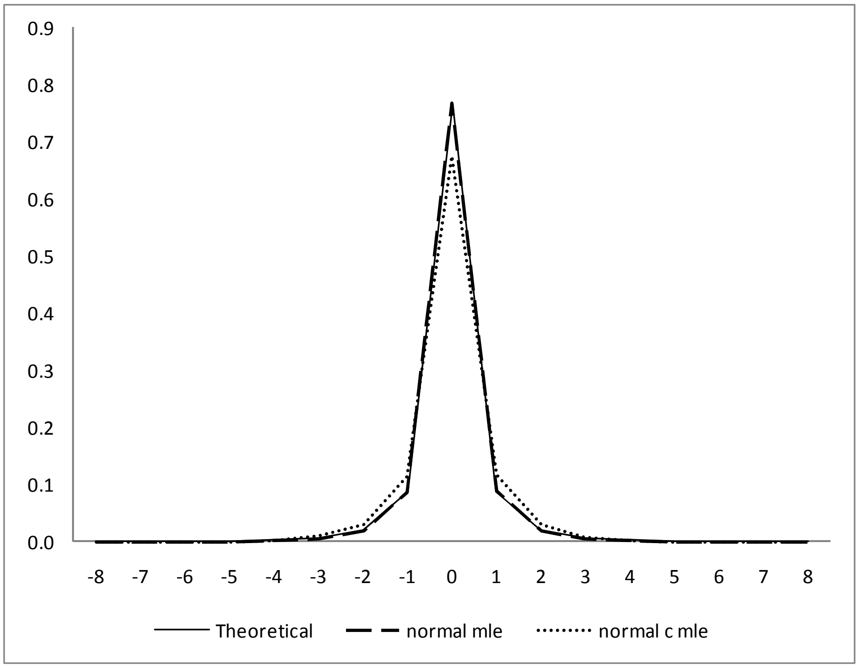


Figure 3.8 (a). Theoretical distribution of the mle (under normal distribution) and the simulated distributions of the change-point mle and cmle under normal distribution when n = 100 and τ = 50 and for η=2.5.

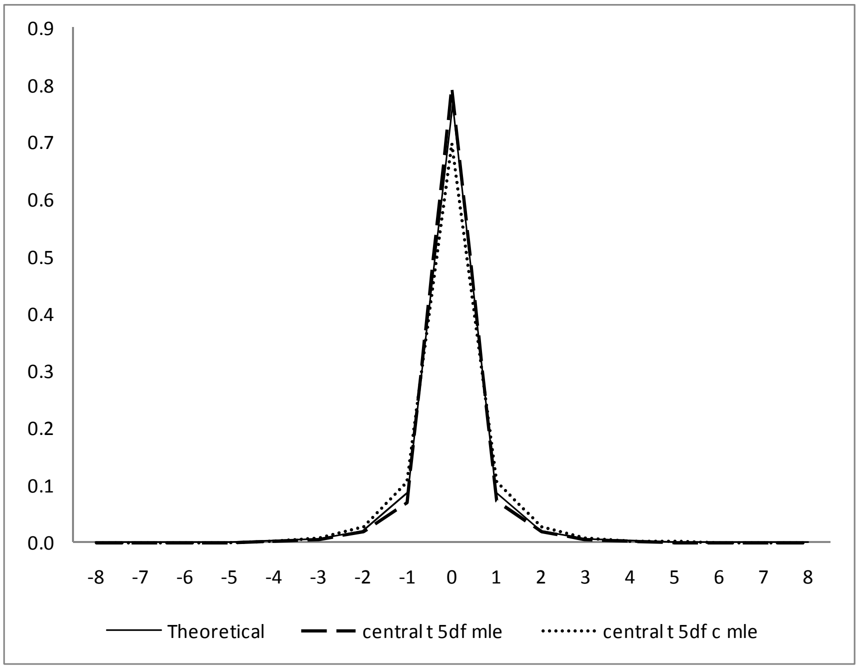


Figure 3.8 (b). Theoretical distribution of the mle (under normal distribution) and the simulated distributions of the change-point mle and cmle under t_ν when n = 100 and τ = 50 and for η=2.5

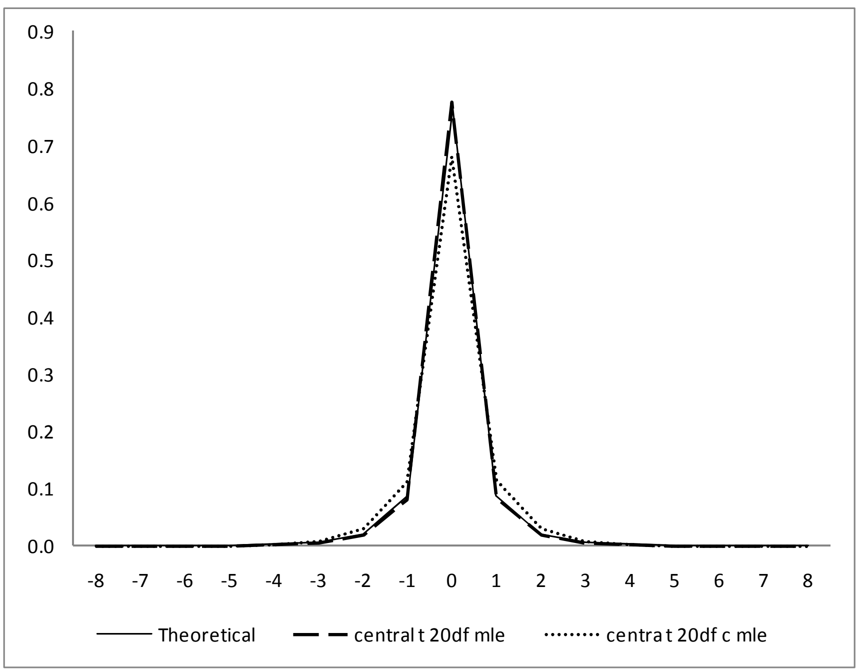


Figure 3.8 (c). Theoretical distribution of the mle (under normal distribution) and the simulated distributions of the change-point mle and cmle under t_ν when n = 100 and τ = 50 and for η=2.5

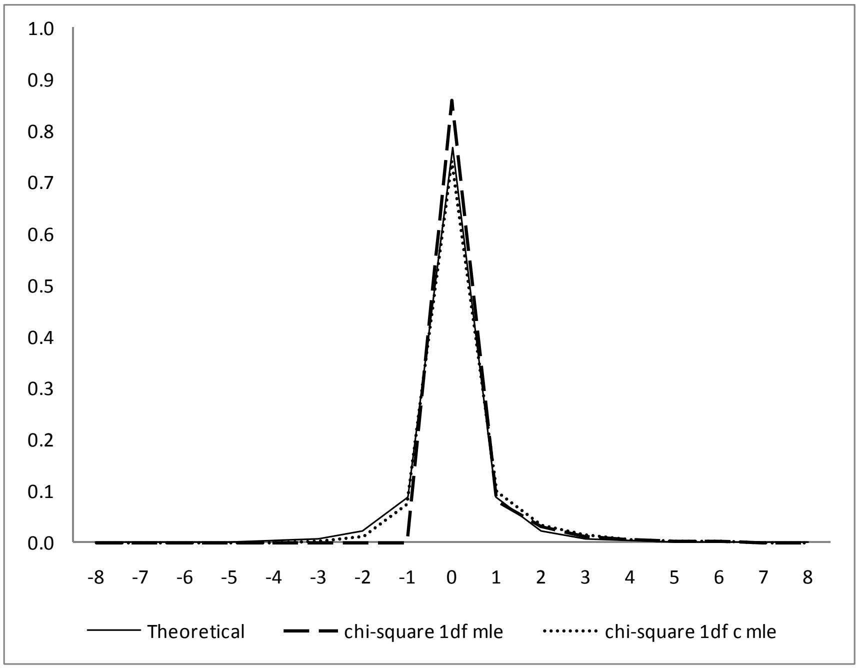


Figure 3.8 (d). Theoretical distribution of the mle (under normal distribution) and the simulated distributions of the change-point mle and cmle under χ_v^2 when n = 100 and τ = 50 and for η=2.5

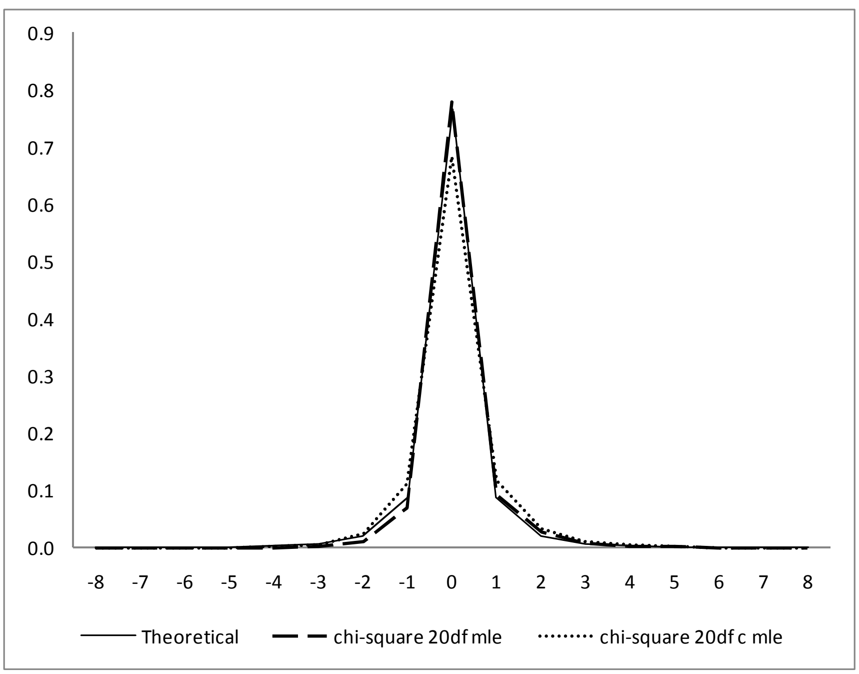


Figure 3.8 (e). Theoretical distribution of the mle (under normal distribution) and the simulated distributions of the change-point mle and cmle under χ_v^2 when n = 100 and τ = 50 and for η=2.5

The following Figures 3.3(a) –(e) show the theoretical distribution of the mle (under normal distribution) and the simulated distributions of the change-point mle and cmle under normal, t_v, and χ_v^2 distributions when n = 100 and τ = 50 and for η=1.0, k = -8, …, 8.

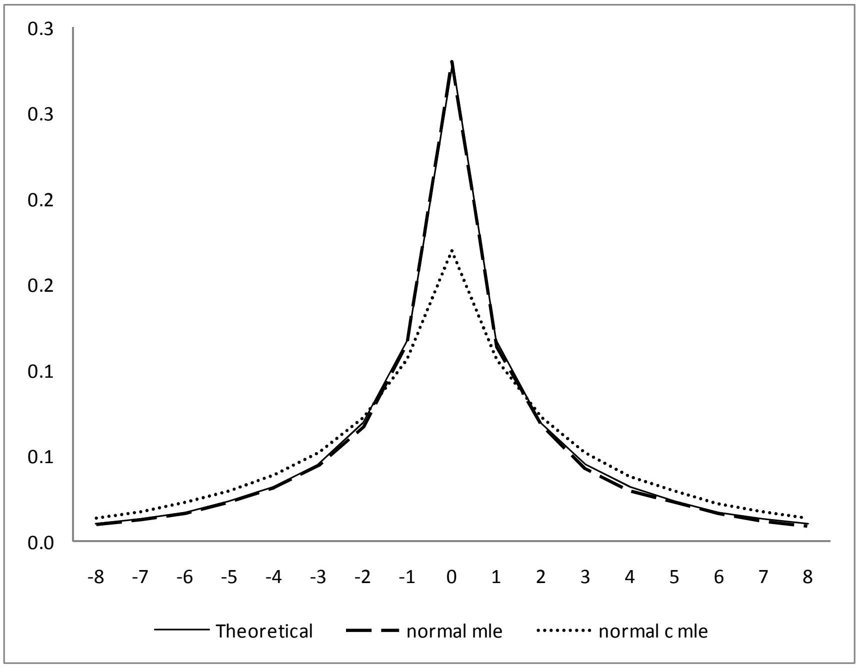


Figure 3.9 (a). Theoretical distribution of the mle (under normal distribution) and the simulated distributions of the change-point mle and cmle under normal distributions when n = 100 and τ = 50 and for η=1.0

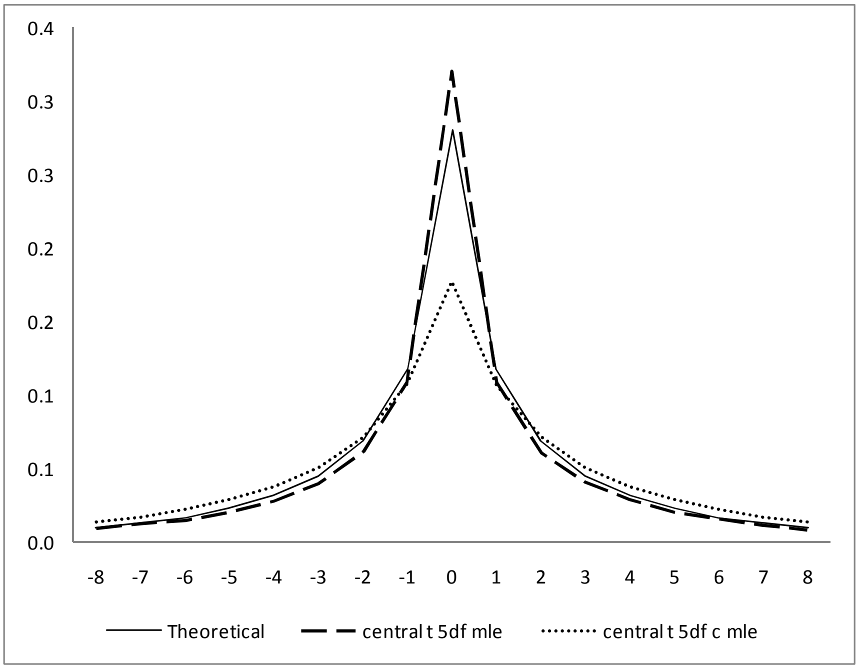


Figure 3.9 (b). Theoretical distribution of the mle (under normal distribution) and the simulated distributions of the change-point mle and cmle under t_{υ} when n = 100 and τ = 50 and for η=1.0

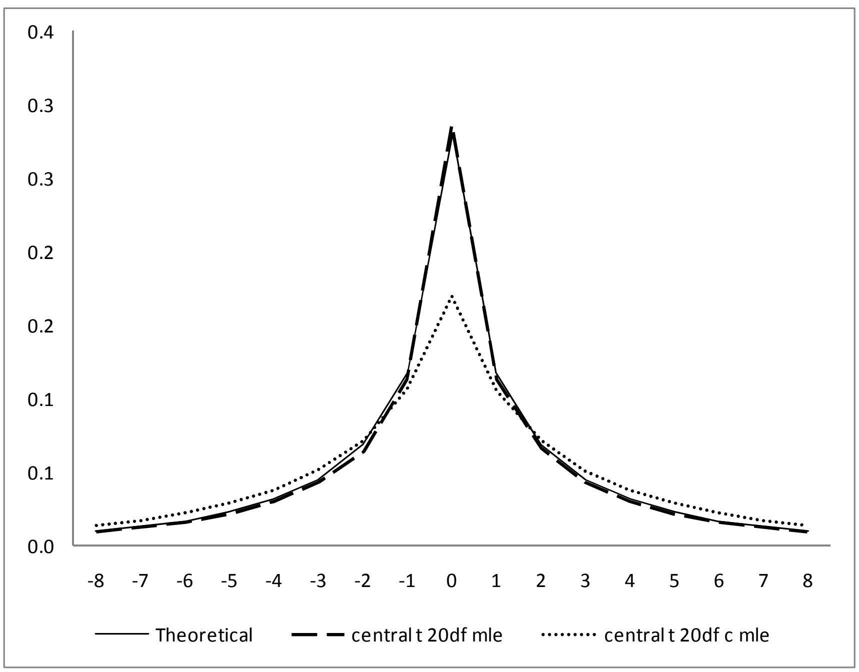


Figure 3.9 (c). Theoretical distribution of the mle (under normal distribution) and the simulated distributions of the change-point mle and cmle under t_{υ} when n = 100 and τ = 50 and for η=1.0

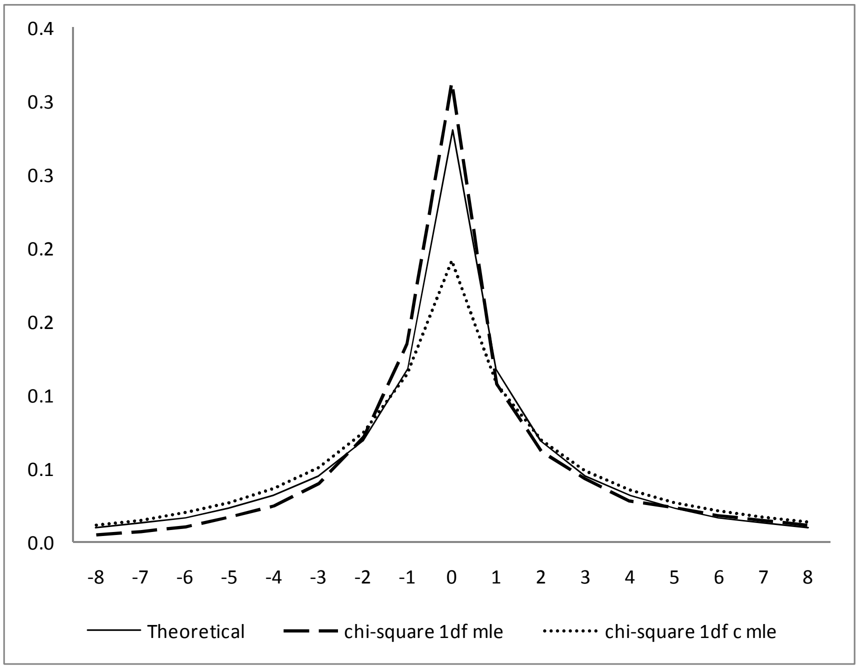


Figure 3.9 (d) Theoretical distribution of the mle (under normal distribution) and the simulated distributions of the change-point mle and cmle under χ_v^2 when n = 100 and τ = 50 and for η=1.0

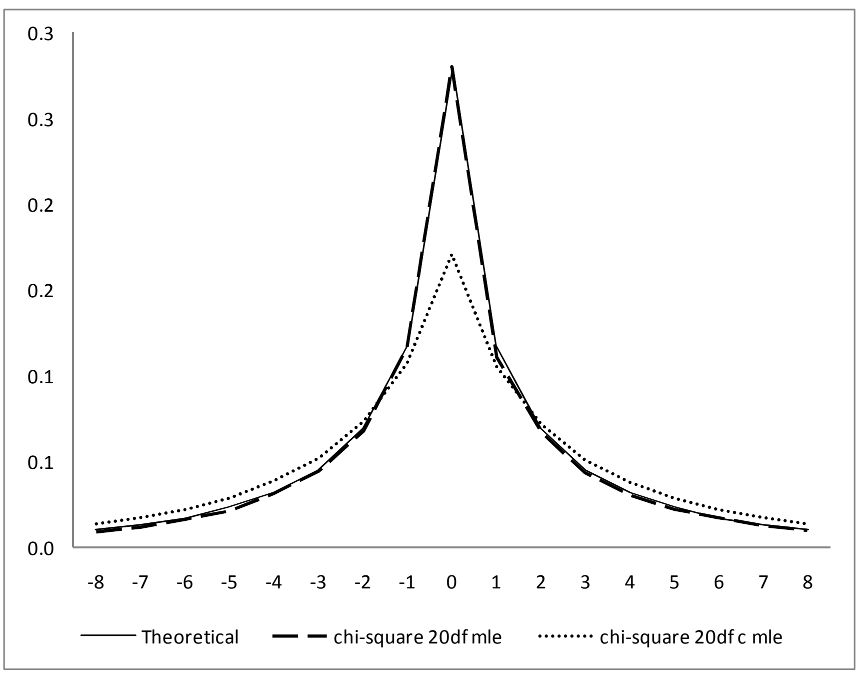


Figure 3.9 (e). Theoretical distribution of the mle (under normal distribution) and the simulated distributions of the change-point mle and cmle under χ_v^2 when n = 100 and τ = 50 and for η=1.0

APPENDIX

A.1 J-fold convolution densities

The following is based on Chapter 7 of the "Introduction to Probability" by Charles M. Grinstead & J. Laurie Snell, and on Killmann & Collani (2001).

Let us consider j independent continuous random variables $X_1, X_2, \dots, X_j$ with the same densities, $f_{X_1} = f_{X_2} = \cdots = f_{X_j}$, then the sum $S_j = X_1 + X_2 + \cdots + X_j$ will have the following density

$$f_{S_j}(x) = \left(f_{X_1} * f_{X_2} * \dots * f_{X_j}\right)(x) = \left(f_{X_1}\right)^{j*}(x)$$

that is a j-fold convolution.

Exponential distribution case.

In the case where $f_{X_1} = \frac{v_1}{v_1 - v_2} e^{\frac{-v_1}{v_1 - v_2}x}$ we have an exponential probability density function with $E(X_1) = \frac{v_1 - v_2}{v_1}$.

Then the j-fold convolution density $f_{S_j}(x)$ will be as follows

$$f_{S_j}(x) = \frac{\frac{v_1}{v_1 - v_2} e^{\frac{-v_1}{v_1 - v_2}x} \left(\frac{v_1}{v_1 - v_2} x\right)^{j-1}}{(j-1)!}$$

For example, in the case where $j = 2$ we will have the following density

$$f_{S_2}(x) = \int_{-\infty}^{\infty} f_{X_1}(x - x_2) f_{X_2}(x_2)\, dx_2$$

$$= \int_0^x \frac{v_1}{v_1 - v_2} e^{\frac{-v_1}{v_1 - v_2}(x - x_2)} \frac{v_1}{v_1 - v_2} e^{\frac{-v_1}{v_1 - v_2} x_2}\, dx_2$$

$$= \int_0^x \left(\frac{v_1}{v_1 - v_2}\right)^2 e^{\frac{-v_1}{v_1 - v_2} x}\, dx_2$$

$$= \left(\frac{v_1}{v_1 - v_2}\right)^2 e^{\frac{-v_1}{v_1 - v_2} x} x = \frac{\frac{v_1}{v_1 - v_2} e^{\frac{-v_1}{v_1 - v_2} x} \left(\frac{v_1}{v_1 - v_2} x\right)^1}{1!}$$

Note that

$$f_{S_j} = f_{S_{j-1}} * f_X$$

Uniform $[a, b]$ distribution case.

In the case where $f_{X_1} \sim Uniform\ [a, b]$ is has been determined that the j-fold convolution density $f_{S_j}(x)$ will be as follows

$$f_{S_j}(x) = \begin{cases} \frac{1}{(j-1)!\,(b-a)^j} \sum_{k=0}^{\tilde{n}(j,x)} (-1)^k \binom{j}{k} \left(x - ja - k(b-a)\right)^{j-1}, & ja < x < jb \\ 0, & otherwise \end{cases}$$

where $\tilde{n}(j, x) := \left\lceil \frac{x - ja}{b - a} \right\rceil$ is the largest integer less than $\frac{x - ja}{b - a}$.

If we let $a = 0$ then we will have the case where $f_{X_1} \sim Uniform\ [0, a]$ and the j-fold convolution density $f_{S_j}(x)$ will be

$$f_{S_j}(x) = \begin{cases} \dfrac{1}{(j-1)!\,(b)^j} \displaystyle\sum_{k=0}^{\tilde{n}(j,x)} (-1)^k \binom{j}{k} (x - kb)^{j-1}, & 0 < x < jb \\ 0, & otherwise \end{cases}$$

Giving us the probability

$$F_{S_j}(x) = \begin{cases} 0, & x < 0 \\ \dfrac{1}{j!\,b^j} \displaystyle\sum_{k=0}^{\tilde{n}(j,x)} (-1)^k \binom{j}{k} (x - kb)^j, & 0 < x < jb \\ 1, & x \geq jb \end{cases}$$

where $\tilde{n}(j, x) := \left\lceil \frac{x}{b} \right\rceil$ is the largest integer less than $\frac{x}{b}$.

If we let $a = 0, b = 1$, we will have the case where $f_{X_1} \sim Uniform\ [0,1]$, and the density $f_{X_1} = \begin{cases} 1, 0 \leq x \leq 1 \\ 0, otherwise \end{cases}$, then the j-fold convolution density $f_{S_j}(x)$ will be as follows

$$f_{S_j}(x) = \begin{cases} \dfrac{1}{(j-1)!} \displaystyle\sum_{k=0}^{x} (-1)^k \binom{j}{k} (x - k)^{j-1}, & 0 < x < j \\ 0, & otherwise. \end{cases}$$

A.2 Determination of $u_k(dx)$ for $P(\xi_\infty = -k), k > 0$

Let

$$u_k(dx) = P(S_k \in dx, T_1^- > k)$$

Then

$$\bar{u}_k(x) = P(S_k > x, T_1^- > k)\,, \qquad x > 0$$

is the probability that the first time the random walk S_k enters $(-\infty, 0)$ is after the k[th] step, with ladder height at that k[th] step being greater than x, and $k \geq 1$. For $k \geq 2$, and $u_k(dx) = P(S_k \in dx, T_1^- > k)$ as defined in $(2.3.15)$ we can obtain the following expression

$$u_k(dx) = \frac{v_1}{v_1 - v_2} dx \left[\bar{u}_k(x) - \bar{u}_{k-1}\left(x + \log\frac{v_2}{v_1}\right)\right] \qquad (A.2.1)$$

Once we have determined an expression for $\bar{u}_1(x)$, we just need to find the subsequent differences $\bar{u}_k(x) - \bar{u}_{k-1}\left(x + \log\frac{v_2}{v_1}\right)$ to obtain the needed expressions for $u_k(dx)$, $k \geq 2$. We can see that $\bar{u}_k(x)$ in above is obtained via the following steps

$$\bar{u}_k(x) = E\left[I\left(X_k > x - S_{k-1}, x - S_{k-1} > log\frac{v_2}{v_1}, T_1^- > k-1\right)\right]$$

$$+E\left[I\left(X_k > x - S_{k-1}, x - S_{k-1} < log\frac{v_2}{v_1}, T_1^- > k-1\right)\right]$$

$$= e^{\frac{-v_1}{v_1-v_2}\left(x - log\frac{v_2}{v_1}\right)} \int_0^{x - log\frac{v_2}{v_1}} e^{\frac{-v_1}{v_1-v_2}y}\left(\bar{u}_{k-1}(dy)\right) + \bar{u}_{k-2}\left(x - log\frac{v_2}{v_1}\right)$$

Taking the derivative will give us $(A.2.1)$.

Expressions for $\bar{u}_1(x)$ and $u_1(dx)$

Let $k = 1$, then recalling $(2.2.4)$

$$P(X > x) = \left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1 - v_2}} e^{\frac{-v_1 x}{v_1 - v_2}} = e^{\frac{-v_1}{v_1 - v_2}\left[x - log\frac{v_2}{v_1}\right]}, \quad x > log\,\frac{v_2}{v_1}$$

we have

$$\bar{u}_1(x) = P(S_1 > x, T_1^- > 1) = P(S_1 > x) = P(X_1 > x) = e^{\frac{-v_1}{v_1 - v_2}\left[x - log\frac{v_2}{v_1}\right]}$$

$$P(X_1 \leq x) = 1 - P(X_1 > x) = 1 - \bar{u}_1(x)$$

$$\frac{d}{dx}\left(1 - \bar{u}_1(x)\right) = -\frac{d}{dx}u_1(x)$$

$$= \frac{-v_1}{v_1 - v_2} e^{\frac{-v_1}{v_1 - v_2}\left[x - log\frac{v_2}{v_1}\right]}$$

and thus

$$u_1(dx) = dx\,\frac{v_1}{v_1 - v_2} e^{\frac{-v_1}{v_1 - v_2}\left[x - log\frac{v_2}{v_1}\right]}$$

Expressions for $\bar{u}_k(x)$, $\bar{u}_k(x) - \bar{u}_{k-1}\left(x + \log\frac{v_2}{v_1}\right)$, and $u_k(dx)$ for $k \geq 2$

$k = 2$

$$\bar{u}_2(x) = P(S_2 > x, T_1^- > 2) = P(S_1 > 0,\ S_2 > x) = P\left(\bigwedge_{j<2} S_j > 0, S_2 > x\right)$$

$S_2 = S_1 + X_2$ and thus $(S_2 > x) = (S_1 + X_2 > x) = (X_2 > x - S_1)$

$$= E\left[I\left(X_2 > x - S_1, x - S_1 > log\frac{v_2}{v_1}, S_1 > 0\right)\right]$$

$$+E\left[I\left(X_2 > x - S_1, x - S_1 < log\frac{v_2}{v_1}, S_1 > 0\right)\right]$$

$$= e^{\frac{-v_1}{v_1 - v_2}\left[x - log\frac{v_2}{v_1}\right]} \int_0^{x - log\frac{v_2}{v_1}} e^{\frac{-v_1}{v_1 - v_2}y}\left(-u_1(dy)\right) + \bar{u}_1\left(x - log\frac{v_2}{v_1}\right)$$

$$= e^{\frac{-v_1}{v_1 - v_2}\left[x - 2log\frac{v_2}{v_1}\right]}\left\{\frac{v_1}{v_1 - v_2}\left(x - log\frac{v_2}{v_1}\right)\right\} + \bar{u}_1\left(x - log\frac{v_2}{v_1}\right)$$

Then after taking the derivative, or applying $(A.2.1)$ directly we obtain the form

$$u_2(dx) = dx\frac{v_1}{v_1 - v_2}\left\{\bar{u}_2(x) - \bar{u}_1\left(x - log\frac{v_2}{v_1}\right)\right\}$$

Then to find the expression needed to determine $P(\xi_\infty = k), k < 0$ we calculate

$$\bar{u}_2(x) - \bar{u}_1\left(x - log\frac{v_2}{v_1}\right) =$$

$$= e^{\frac{-v_1}{v_1-v_2}\left[x-2log\frac{v_2}{v_1}\right]}\left\{\frac{v_1}{v_1-v_2}\left(x-log\frac{v_2}{v_1}\right)\right\}$$

And thus

$$u_2(dx) = dx\,\frac{v_1}{v_1-v_2}\left[e^{\frac{-v_1}{v_1-v_2}\left[x-2log\frac{v_2}{v_1}\right]}\left\{\frac{v_1}{v_1-v_2}\left(x-log\frac{v_2}{v_1}\right)\right\}\right]$$

$k = 3$

$$u_3(x) = P(S_3 > x, T_1^- > 3) = P\left(\bigwedge_{j<3} S_j > 0\,, S_3 > x\right)$$

$$= E\left[I\left(X_3 > x - S_2, x - S_2 > log\left(\frac{v_2}{v_1}\right), \bigwedge_{j<3} S_j > 0\right)\right]$$

$$+E\left[I\left(X_3 > x - S_2, x - S_2 < log\left(\frac{v_2}{v_1}\right), \bigwedge_{j<3} S_j > 0\right)\right]$$

$$= e^{\frac{-v_1}{v_1-v_2}\left[x-log\frac{v_2}{v_1}\right]}\int_0^{x-log\frac{v_2}{v_1}} e^{\frac{-v_1}{v_1-v_2}y}\left(-u_2(dy)\right) + \bar{u}_2\left(x-log\frac{v_2}{v_1}\right)$$

$$= e^{\frac{-v_1}{v_1-v_2}\left[x-3log\frac{v_2}{v_1}\right]}\left\{\frac{\left(\frac{v_1}{v_1-v_2}\left[x-2log\frac{v_2}{v_1}\right]\right)^2}{2!} - \frac{\left(\frac{-v_1}{v_1-v_2}log\frac{v_2}{v_1}\right)^2}{2!}\right\} + \bar{u}_2\left(x-log\frac{v_2}{v_1}\right)$$

Then

$$\bar{u}_3(x) - \bar{u}_2\left(x - log\frac{v_2}{v_1}\right) = e^{\frac{-v_1}{v_1 - v_2}\left(x - 3log\frac{v_2}{v_1}\right)}\left\{\frac{\left(\frac{v_1}{v_1 - v_2}\left[x - 2log\frac{v_2}{v_1}\right]\right)^2}{2!} - \frac{\left(\frac{-v_1}{v_1 - v_2}log\frac{v_2}{v_1}\right)^2}{2!}\right\}$$

And thus

$$u_3(dx) = dx\frac{v_1}{v_1 - v_2}\left[e^{\frac{-v_1}{v_1 - v_2}\left(x - 3log\frac{v_2}{v_1}\right)}\left\{\frac{\left(\frac{v_1}{v_1 - v_2}\left[x - 2log\frac{v_2}{v_1}\right]\right)^2}{2!} - \frac{\left(\frac{-v_1}{v_1 - v_2}log\frac{v_2}{v_1}\right)^2}{2!}\right\}\right]$$

$k = 4$

$$\bar{u}_4(x) = P(S_4 > x, T_1^- > 4) = P\left(\bigwedge_{j<4} S_j > 0\,, S_4 > x\right)$$

$$= E\left[I\left(X_4 > x - S_3, x - S_3 > log\left(\frac{v_2}{v_1}\right), \bigwedge_{j<4} S_j > 0\right)\right]$$

$$+E\left[I\left(X_4 > x - S_3, x - S_3 < log\left(\frac{v_2}{v_1}\right), \bigwedge_{j<4} S_j > 0\right)\right]$$

$$= e^{\frac{-v_1}{v_1 - v_2}\left[x - log\frac{v_2}{v_1}\right]}\int_0^{x - log\frac{v_2}{v_1}} e^{\frac{-v_1}{v_1 - v_2}y}\left(-u_3(dy)\right) + \bar{u}_3\left(x - log\frac{v_2}{v_1}\right) =$$

$$= e^{\frac{-v_1}{v_1-v_2}\left[x-4log\frac{v_2}{v_1}\right]}\left\{\frac{\left(\frac{v_1}{v_1-v_2}\left[x-3log\frac{v_2}{v_1}\right]\right)^3}{3!}-\frac{\left(\frac{-2v_1}{v_1-v_2}log\frac{v_2}{v_1}\right)^3}{3!}\right.$$

$$\left.-\frac{\left(\frac{-v_1}{v_1-v_2}log\frac{v_2}{v_1}\right)^2}{2!}\left(\frac{v_1}{v_1-v_2}\left[x-log\frac{v_2}{v_1}\right]\right)\right\}+\bar{u}_3\left(x-log\frac{v_2}{v_1}\right)$$

Then

$$\bar{u}_4(x)-\bar{u}_3\left(x-log\frac{v_2}{v_1}\right)$$

$$= e^{\frac{-v_1}{v_1-v_2}\left(x-4log\frac{v_2}{v_1}\right)}\left\{\frac{\left(\frac{v_1}{v_1-v_2}[x-3logv]\right)^3}{3!}-\frac{\left(\frac{-2v_1}{v_1-v_2}log\frac{v_2}{v_1}\right)^3}{3!}\right.$$

$$\left.-\frac{\left(\frac{-v_1}{v_1-v_2}log\frac{v_2}{v_1}\right)^2}{2!}\left(\frac{v_1}{v_1-v_2}\left[x-log\frac{v_2}{v_1}\right]\right)\right\}$$

And

$$u_4(dx) = dx\frac{v_1}{v_1 - v_2}\left[e^{\frac{-v_1}{v_1-v_2}\left(x-4log\frac{v_2}{v_1}\right)}\left\{\frac{\left(\frac{v_1}{v_1 - v_2}\left[x - 3log\frac{v_2}{v_1}\right]\right)^3}{3!} - \frac{\left(\frac{-2v_1}{v_1 - v_2}log\frac{v_2}{v_1}\right)^3}{3!}\right.\right.$$

$$\left.\left. - \frac{\left(\frac{-v_1}{v_1 - v_2}log\frac{v_2}{v_1}\right)^2}{2!}\left(\frac{v_1}{v_1 - v_2}\left[x - log\frac{v_2}{v_1}\right]\right)\right\}\right]$$

$k = 5$

$$\bar{u}_5(x) = P(S_5 > x, T_1^- > 5) = P\left(\bigwedge_{j<5} S_j > 0, S_5 > x\right)$$

$$= E\left[I\left(X_5 > x - S_4, x - S_4 > log\left(\frac{v_2}{v_1}\right), \bigwedge_{j<5} S_j > 0\right)\right]$$

$$+E\left[I\left(X_5 > x - S_4, x - S_4 < log\left(\frac{v_2}{v_1}\right), \bigwedge_{j<5} S_j > 0\right)\right]$$

$$= e^{\frac{-v_1}{v_1-v_2}\left(x-5log\frac{v_2}{v_1}\right)}\left\{\frac{\left(\frac{v_1}{v_1 - v_2}\left[x - 4log\frac{v_2}{v_1}\right]\right)^4}{4!} - \frac{\left(\frac{-3v_1}{v_1 - v_2}log\frac{v_2}{v_1}\right)^4}{4!}\right.$$

$$- \frac{\left(\frac{-v_1}{v_1 - v_2}log\frac{v_2}{v_1}\right)^2}{2!}\frac{\left(\frac{v_1}{v_1 - v_2}\left[x - 2log\frac{v_2}{v_1}\right]\right)^2}{2!} + \frac{\left(\frac{-v_1}{v_1 - v_2}log\frac{v_2}{v_1}\right)^4}{2!\,2!}$$

$$\left. - \frac{\left(\frac{-2v_1}{v_1 - v_2}log\frac{v_2}{v_1}\right)^3}{3!}\left(\frac{v_1}{v_1 - v_2}\left[x - log\frac{v_2}{v_1}\right]\right)\right\} + u_4\left(x - log\frac{v_2}{v_1}\right)$$

$$\bar{u}_5(x) - \bar{u}_4\left(x - log\frac{v_2}{v_1}\right)$$

$$= e^{\frac{-v_1}{v_1-v_2}\left(x-5log\frac{v_2}{v_1}\right)}\left\{\frac{\left(\frac{v_1}{v_1-v_2}\left[x-4log\frac{v_2}{v_1}\right]\right)^4}{4!} - \frac{\left(\frac{-3v_1}{v_1-v_2}log\frac{v_2}{v_1}\right)^4}{4!}\right.$$

$$-\frac{\left(\frac{-v_1}{v_1-v_2}log\left(\frac{v_2}{v_1}\right)\right)^2}{2!}\frac{\left(\frac{v_1}{v_1-v_2}\left[x-2log\frac{v_2}{v_1}\right]\right)^2}{2!} + \frac{\left(\frac{-v_1}{v_1-v_2}log\frac{v_2}{v_1}\right)^4}{2!\,2!}$$

$$\left.-\frac{\left(\frac{-2v_1}{v_1-v_2}log\frac{v_2}{v_1}\right)^3}{3!}\left(\frac{v_1}{v_1-v_2}\left[x-log\frac{v_2}{v_1}\right]\right)\right\}$$

Thus,

$$u_5(dx) = dx\frac{v_1}{v_1-v_2}\left[ee^{\frac{-v_1}{v_1-v_2}\left(x-5log\frac{v_2}{v_1}\right)}\left\{\frac{\left(\frac{v_1}{v_1-v_2}\left[x-4log\frac{v_2}{v_1}\right]\right)^4}{4!} - \frac{\left(\frac{-3v_1}{v_1-v_2}log\frac{v_2}{v_1}\right)^4}{4!}\right.\right.$$

$$-\frac{\left(\frac{-v_1}{v_1-v_2}log\frac{v_2}{v_1}\right)^2}{2!}\frac{\left(\frac{v_1}{v_1-v_2}\left[x-2log\frac{v_2}{v_1}\right]\right)^2}{2!} + \frac{\left(\frac{-v_1}{v_1-v_2}log\frac{v_2}{v_1}\right)^4}{2!\,2!}$$

$$\left.\left.-\frac{\left(\frac{-2v_1}{v_1-v_2}log\frac{v_2}{v_1}\right)^3}{3!}\left(\frac{v_1}{v_1-v_2}\left[x-log\frac{v_2}{v_1}\right]\right)\right\}\right]$$

$k = 6$

$$\bar{u}_6(x) = P(S_6 > x, T_1^- > 6) = P\left(\bigwedge_{j<6} S_j > 0\,, S_6 > x\right) =$$

$$= E\left[I\left(X_6 > x - S_5, x - S_5 > log\left(\frac{v_2}{v_1}\right), \bigwedge_{j<6} S_j > 0\right)\right]$$

$$+E\left[I\left(X_6 > x - S_5, x - S_5 < log\left(\frac{v_2}{v_1}\right), \bigwedge_{j<6} S_j > 0\right)\right]$$

$$= e^{\frac{-v_1}{v_1-v_2}\left(x-6log\frac{v_2}{v_1}\right)} \times$$

$$\left\{\frac{\left(\frac{v_1}{v_1-v_2}\left[x-5log\frac{v_2}{v_1}\right]\right)^5}{5!} - \frac{\left(\frac{-v_1}{v_1-v_2}log\left(\frac{v_2}{v_1}\right)\right)^2}{2!}\frac{\left(\frac{v_1}{v_1-v_2}\left[x-3log\frac{v_2}{v_1}\right]\right)^3}{3!}\right.$$

$$-\frac{\left(\frac{-2v_1}{v_1-v_2}log\frac{v_2}{v_1}\right)^3}{3!}\frac{\left(\frac{v_1}{v_1-v_2}\left[x-2log\frac{v_2}{v_1}\right]\right)^2}{2!}$$

$$-\frac{\left(\frac{-3v_1}{v_1-v_2}log\frac{v_2}{v_1}\right)^4}{4!}\frac{\left(\frac{v_1}{v_1-v_2}\left[x-log\frac{v_2}{v_1}\right]\right)}{1!}$$

$$-\frac{\left(\frac{-v_1}{v_1-v_2}log\frac{v_2}{v_1}\right)^4}{2!\,2!}\left(\frac{v_1}{v_1-v_2}\left[x-log\frac{v_2}{v_1}\right]\right) - \frac{\left(\frac{-4v_1}{v_1-v_2}log\frac{v_2}{v_1}\right)^5}{5!}$$

$$\left.+\frac{2\left(\frac{v_1}{v_1-v_2}log\frac{v_2}{v_1}\right)^2}{2!}\frac{\left(\frac{-2v_1}{v_1-v_2}log\frac{v_2}{v_1}\right)^3}{3!}\right\} + \bar{u}_5\left(x - log\frac{v_2}{v_1}\right)$$

Note that

$$u_6(x) - u_5\left(x - log\frac{v_2}{v_1}\right) =$$

$$
\begin{aligned}
&= e^{\frac{-v_1}{v_1-v_2}\left(x-6log\frac{v_2}{v_1}\right)} \left\{ \frac{\left(\frac{v_1}{v_1-v_2}\left[x-5log\left(\frac{v_2}{v_1}\right)\right]\right)^5}{5!} \right. \\
&\quad - \frac{\left(\frac{-v_1}{v_1-v_2}log\left(\frac{v_2}{v_1}\right)\right)^2}{2!} \frac{\left(\frac{v_1}{v_1-v_2}\left[x-3log\left(\frac{v_2}{v_1}\right)\right]\right)^3}{3!} \\
&\quad - \frac{\left(\frac{-2v_1}{v_1-v_2}log\left(\frac{v_2}{v_1}\right)\right)^3}{3!} \frac{\left(\frac{v_1}{v_1-v_2}\left[x-2log\left(\frac{v_2}{v_1}\right)\right]\right)^2}{2!} \\
&\quad - \frac{\left(\frac{-3v_1}{v_1-v_2}log\left(\frac{v_2}{v_1}\right)\right)^4}{4!} \frac{\left(\frac{v_1}{v_1-v_2}\left[x-log\left(\frac{v_2}{v_1}\right)\right]\right)}{1!} \\
&\quad - \frac{\left(\frac{-v_1}{v_1-v_2}log\left(\frac{v_2}{v_1}\right)\right)^4}{2!\,2!} \left(\frac{v_1}{v_1-v_2}\left[x-log\left(\frac{v_2}{v_1}\right)\right]\right) - \frac{\left(\frac{-4v_1}{v_1-v_2}log\left(\frac{v_2}{v_1}\right)\right)^5}{5!} \\
&\quad \left. + \frac{2\left(\frac{v_1}{v_1-v_2}log\left(\frac{v_2}{v_1}\right)\right)^2}{2!} \frac{\left(\frac{-2v_1}{v_1-v_2}log\left(\frac{v_2}{v_1}\right)\right)^3}{3!} \right\}
\end{aligned}
$$

Thus,

$$u_6(dx) = dx\frac{v_1}{v_1 - v_2}\left[e^{\frac{-v_1}{v_1-v_2}\left(x-6log\frac{v_2}{v_1}\right)}\left\{\frac{\left(\frac{v_1}{v_1 - v_2}\left[x - 5log\frac{v_2}{v_1}\right]\right)^5}{5!}\right.\right.$$

$$-\frac{\left(\frac{-v_1}{v_1 - v_2}log\frac{v_2}{v_1}\right)^2}{2!}\frac{\left(\frac{v_1}{v_1 - v_2}\left[x - 3log\frac{v_2}{v_1}\right]\right)^3}{3!}$$

$$-\frac{\left(\frac{-2v_1}{v_1 - v_2}log\frac{v_2}{v_1}\right)^3}{3!}\frac{\left(\frac{v_1}{v_1 - v_2}\left[x - 2log\frac{v_2}{v_1}\right]\right)^2}{2!}$$

$$-\frac{\left(\frac{-3v_1}{v_1 - v_2}log\frac{v_2}{v_1}\right)^4}{4!}\frac{\left(\frac{v_1}{v_1 - v_2}\left[x - log\frac{v_2}{v_1}\right]\right)}{1!}$$

$$-\frac{\left(\frac{-v_1}{v_1 - v_2}log\frac{v_2}{v_1}\right)^4}{2!\,2!}\left(\frac{v_1}{v_1 - v_2}\left[x - log\frac{v_2}{v_1}\right]\right) - \frac{\left(\frac{-4v_1}{v_1 - v_2}log\frac{v_2}{v_1}\right)^5}{5!}$$

$$\left.\left.+\frac{2\left(\frac{v_1}{v_1 - v_2}log\frac{v_2}{v_1}\right)^2}{2!}\frac{\left(\frac{-2v_1}{v_1 - v_2}log\frac{v_2}{v_1}\right)^3}{3!}\right\}\right]$$

A.3 Calculations of $P(\xi_\infty = n),\ n < 0$

We wish to evaluate the following

$$P(\xi_\infty = n) = (1 - \|G_+\|)\left(q_{|n|} - \int_{0^+}^{\infty} P(M_\infty^* > x)P\big(S_{|n|} \in dx, T_1^- > |n|\big)\right), \qquad n < 0$$

$$P(\xi_\infty = -n) = (1 - \|G_+\|)\left(q_n - \int_{0^+}^{\infty} P(M_\infty^* > x)P(S_n \in dx, T_1^- > n)\right)$$

$$= (1 - \|G_+\|)\left(q_n - \int_{0^+}^{\infty} P(M_\infty^* > x)u_n(dx)\right), \qquad n > 0$$

We know that

$$\|G_+\| = \frac{v_2}{v_1} \quad and \quad (1 - \|G_+\|) = 1 - \frac{v_2}{v_1} = \frac{v_1 - v_2}{v_1}$$

and

$$q_n - \int_{0^+}^{\infty} P(M_\infty^* > x)u_n(dx) =$$

$$= \left(1 + \frac{v_2}{v_1 - v_2}\log\frac{v_2}{v_1}\right)\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}\int_{-i\log\frac{v_2}{v_1}}^{\infty}\left[\left(\frac{v_2}{v_1 - v_2}\left(x + i\log\frac{v_2}{v_1}\right)\right)^i e^{\frac{v_2}{v_1 - v_2}\left(x + i\log\frac{v_2}{v_1}\right)}\right]u_n(dx)$$

The following is then the expression for the asymptotic distribution of ξ_∞ when $n < 0$, also shown in (2.3.19)

$$P(\xi_\infty = -n) = \frac{v_1 - v_2}{v_1}\Bigg(1 + \frac{v_2}{v_1 - v_2}\log\frac{v_2}{v_1}\Bigg)\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}\int_{-i\log\frac{v_2}{v_1}}^{\infty}\left[\left(\frac{v_2}{v_1 - v_2}\left(x + i\log\frac{v_2}{v_1}\right)\right)^i e^{\frac{v_2}{v_1 - v_2}\left(x + i\log\frac{v_2}{v_1}\right)}\right]u_n(dx), \qquad n > 0 \qquad (A.3.1)$$

Below we will show the calculations used to obtain the final form for the probabilities $P(\xi_\infty = -n),\ n = 1, 2, 3, 4, 5, and\ 6$. See Appendix [A.3] for calculations of $u_n(dx)$.

$n = 1$

$$P(\xi_\infty = -1) = \frac{v_1 - v_2}{v_1}\Bigg(1 + \frac{v_2}{v_1 - v_2}\log\frac{v_2}{v_1}\Bigg)\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}\int_{-i\log\frac{v_2}{v_1}}^{\infty}\left[\left(\frac{v_2}{v_1 - v_2}\left(x + i\log\frac{v_2}{v_1}\right)\right)^i e^{\frac{v_2}{v_1 - v_2}\left(x + i\log\frac{v_2}{v_1}\right)}\right]u_1(dx) =$$

Note that $u_1(dx) = dx\,\frac{v_1}{v_1 - v_2}\,e^{\frac{-v_1}{v_1 - v_2}\left[x - log\left(\frac{v_2}{v_1}\right)\right]}$, see Appendix [A.2] for details.

$$= \frac{v_1 - v_2}{v_1}\left(1 + \frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}\right)\sum_{\mathrm{i}=0}^{\infty}\frac{(-1)^{\mathrm{i}}}{\mathrm{i}!}\left(\frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2}\right)^{\mathrm{i}}\int_{-\mathrm{i}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}}^{\infty}\left[\left(\mathrm{x}\right.\right.$$

$$\left.\left. + \mathrm{i}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}\right)^{\mathrm{i}} \mathrm{e}^{\frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2}\left(\mathrm{x} + \mathrm{i}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}\right)}\right]\frac{v_1}{v_1 - v_2} e^{\frac{-v_1}{v_1 - v_2}\left[x - log\left(\frac{v_2}{v_1}\right)\right]}dx$$

$$= \left(1 + \frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}\right)\sum_{\mathrm{i}=0}^{\infty}\frac{(-1)^{\mathrm{i}}}{\mathrm{i}!}\left(\frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2}\right)^{\mathrm{i}}\int_{-\mathrm{i}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}}^{\infty}\left(\mathrm{x}\right.$$

$$\left. + \mathrm{i}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}\right)^{\mathrm{i}} \mathrm{e}^{\frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2}\left(\mathrm{x} + \mathrm{i}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}\right) + \frac{-v_1}{v_1 - v_2}\left[x - log\left(\frac{v_2}{v_1}\right)\right]}dx$$

$$= \left(1 + \frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}\right)\sum_{\mathrm{i}=0}^{\infty}\frac{(-1)^{\mathrm{i}}}{\mathrm{i}!}\left(\frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2}\right)^{\mathrm{i}}\int_{-\mathrm{i}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}}^{\infty}\left(\mathrm{x}\right.$$

$$\left. + \mathrm{i}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}\right)^{\mathrm{i}} \mathrm{e}^{\frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2}\left(\mathrm{x} + \mathrm{i}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}\right) + \frac{-v_1}{v_1 - v_2}\left[x + ilog\left(\frac{v_2}{v_1}\right) - (i+1)log\left(\frac{v_2}{v_1}\right)\right]}dx$$

$$let\ u = \mathrm{x} + \mathrm{i}\log\frac{\mathrm{v}_2}{\mathrm{v}_1} \qquad du = dx$$

$$\mathrm{x} = -\mathrm{i}\log\frac{\mathrm{v}_2}{\mathrm{v}_1} \quad \mathrm{u} = -\mathrm{i}\log\frac{\mathrm{v}_2}{\mathrm{v}_1} + \mathrm{i}\log\frac{\mathrm{v}_2}{\mathrm{v}_1} = 0$$

$$= \left(1 + \frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}\right)\sum_{\mathrm{i}=0}^{\infty}\frac{(-1)^{\mathrm{i}}}{\mathrm{i}!}\left(\frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2}\right)^{\mathrm{i}}\int_{0}^{\infty}(\mathrm{u})^{\mathrm{i}}\mathrm{e}^{\frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2}(\mathrm{u}) + \frac{-v_1}{v_1 - v_2}\left[u - (i+1)log\left(\frac{v_2}{v_1}\right)\right]}du$$

$$= \left(1 + \frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}\right)\sum_{\mathrm{i}=0}^{\infty}\frac{(-1)^{\mathrm{i}}}{\mathrm{i}!}\left(\frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2}\right)^{\mathrm{i}}\int_{0}^{\infty}\mathrm{u}^{\mathrm{i}}\mathrm{e}^{-\frac{v_1 - \mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2}(\mathrm{u})} e^{log\left(\frac{v_2}{v_1}\right)^{(i+1)\frac{v_1}{v_1 - v_2}}}du$$

$$= \left(1 + \frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2} \log \frac{\mathrm{v}_2}{\mathrm{v}_1}\right) \sum_{\mathrm{i}=0}^{\infty} \frac{(-1)^{\mathrm{i}}}{\mathrm{i}!} \left(\frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2}\right)^{\mathrm{i}} \left(\frac{v_2}{v_1}\right)^{(i+1)\frac{v_1}{v_1 - v_2}} \int_0^{\infty} \mathrm{u}^{\mathrm{i}} \mathrm{e}^{-\mathrm{u}} du$$

Note that $\int_0^{\infty} \mathrm{x}^{\mathrm{n}-1} \mathrm{e}^{-\mathrm{x}} dx = (\mathrm{n} - 1)! = \Gamma(\mathrm{n})$

$$= \left(1 + \frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2} \log \frac{\mathrm{v}_2}{\mathrm{v}_1}\right) \left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1 - v_2}} \sum_{\mathrm{i}=0}^{\infty} \frac{(-1)^{\mathrm{i}}}{\mathrm{i}!} \left(\frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2} \left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1 - v_2}}\right)^{\mathrm{i}} (\mathrm{i}!)$$

Applying $\sum_{j=0}^{\infty} (-1)^j x^j = \frac{1}{1+x}$ (geometric series)

$$= \left(1 + \frac{v_2}{v_1 - v_2} \ln \frac{v_2}{v_1}\right) \left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1 - v_2}} \frac{1}{1 + \left(\frac{v_2}{v_1 - v_2} \left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1 - v_2}}\right)}$$

Thus finally we have

$$P(\xi_{\infty} = -1) = \frac{\left(1 + \frac{v_2}{v_1 - v_2} \ln \frac{v_2}{v_1}\right) \left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1 - v_2}}}{\left(1 + \frac{v_2}{v_1 - v_2} \left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1 - v_2}}\right)}$$

$n = 2$

From Appendix [A.2] we know

$$u_2(dx) = dx \frac{v_1}{v_1 - v_2} \left[e^{\frac{-v_1}{v_1 - v_2}\left[x - 2log\frac{v_2}{v_1}\right]} \left\{ \frac{v_1}{v_1 - v_2} \left(x - log \frac{v_2}{v_1}\right) \right\} \right]$$

which we then substitute into $(A.3.1)$ to obtain

$$P(\xi_\infty = -2) = \frac{v_1 - v_2}{v_1}\Bigg(1$$

$$+ \frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2} \log \frac{\mathrm{v}_2}{\mathrm{v}_1}\Bigg) \sum_{\mathrm{i}=0}^{\infty} \frac{(-1)^{\mathrm{i}}}{\mathrm{i}!} \left(\frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2}\right)^{\mathrm{i}} \int_{-\mathrm{i}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}}^{\infty} \Bigg[\Bigg(\mathrm{x}$$

$$+ \mathrm{i} \log \frac{\mathrm{v}_2}{\mathrm{v}_1}\Bigg)^{\mathrm{i}} \mathrm{e}^{\frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2}\left(\mathrm{x} + \mathrm{i}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}\right)}\Bigg] \frac{v_1}{v_1 - v_2} \Bigg[e^{\frac{-v_1}{v_1 - v_2}\left[x - 2log\frac{v_2}{v_1}\right]} \Bigg\{\frac{v_1}{v_1 - v_2}\Bigg(x$$

$$- log\frac{v_2}{v_1}\Bigg)\Bigg\}\Bigg] dx$$

$$= \left(1 + \frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2} \log \frac{\mathrm{v}_2}{\mathrm{v}_1}\right) \sum_{\mathrm{i}=0}^{\infty} \frac{(-1)^{\mathrm{i}}}{\mathrm{i}!} \left(\frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2}\right)^{\mathrm{i}} \int_{-\mathrm{i}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}}^{\infty} \Bigg(\mathrm{x}$$

$$+ \mathrm{i} \log \frac{\mathrm{v}_2}{\mathrm{v}_1}\Bigg)^{\mathrm{i}} \mathrm{e}^{-\left(\mathrm{x} + \mathrm{i}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}\right)} e^{\frac{v_1}{v_1 - v_2}(i+2)log\frac{v_2}{v_1}} \Bigg[\Bigg\{\frac{v_1}{v_1 - v_2}\Bigg(x + \mathrm{i}\log\frac{\mathrm{v}_2}{\mathrm{v}_1} - \mathrm{i}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}$$

$$- log\frac{v_2}{v_1}\Bigg)\Bigg\}\Bigg] dx$$

$$= \left(1 + \frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2} \log \frac{\mathrm{v}_2}{\mathrm{v}_1}\right) \left(\frac{v_1}{v_1 - v_2}\right) \sum_{\mathrm{i}=0}^{\infty} \frac{(-1)^{\mathrm{i}}}{\mathrm{i}!} \left(\frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2}\right)^{\mathrm{i}} \left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1 - v_2}(i+2)} \int_{-\mathrm{i}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}}^{\infty} \Bigg(\mathrm{x}$$

$$+ \mathrm{i} \log \frac{\mathrm{v}_2}{\mathrm{v}_1}\Bigg)^{\mathrm{i}} \mathrm{e}^{-\left(\mathrm{x} + \mathrm{i}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}\right)} \left(x + \mathrm{i}\log\frac{\mathrm{v}_2}{\mathrm{v}_1} - (\mathrm{i}+1)\log\frac{\mathrm{v}_2}{\mathrm{v}_1}\right) dx$$

$$let\ u = \mathrm{x} + \mathrm{i}\log\frac{\mathrm{v}_2}{\mathrm{v}_1} \qquad du = dx$$

$$= \left(1 + \frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2} \log \frac{\mathrm{v}_2}{\mathrm{v}_1}\right) \left(\frac{v_1}{v_1 - v_2}\right) \left(\frac{v_2}{v_1}\right)^{\frac{2v_1}{v_1 - v_2}} \sum_{\mathrm{i}=0}^{\infty} \frac{(-1)^{\mathrm{i}}}{\mathrm{i}!} \left(\frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2}\left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1 - v_2}}\right)^{\mathrm{i}} \int_0^{\infty} (\mathrm{u})^{\mathrm{i}} \mathrm{e}^{-\mathrm{u}} \Bigg(u$$

$$- (\mathrm{i}+1)\log\frac{\mathrm{v}_2}{\mathrm{v}_1}\Bigg) du =$$

Note that $\int_0^\infty \mathrm{x}^{\mathrm{n}-1}\mathrm{e}^{-\mathrm{x}}dx = (\mathrm{n}-1)! = \Gamma(\mathrm{n})$ and $\sum_{j=0}^{\infty}(-1)^j x^j (j+1) = \frac{1}{(1+x)^2}$ and the previous geometric series argument

$$= \left(1 + \frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}\right)\left(\frac{v_1}{v_1 - v_2}\right)\left(\frac{v_2}{v_1}\right)^{\frac{2v_1}{v_1 - v_2}}\sum_{\mathrm{i}=0}^{\infty}(-1)^{\mathrm{i}}\left(\frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2}\left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1 - v_2}}\right)^{\mathrm{i}}(\mathrm{i}+1)\left(1 - \ln\frac{v_2}{v_1}\right)$$

Thus the final expression for $P(\xi_\infty = -2)$ follows

$$P(\xi_\infty = -2) = \frac{\left(1 + \frac{v_2}{v_1 - v_2}\ln\frac{v_2}{v_1}\right)\left(\frac{v_1}{v_1 - v_2}\right)\left(\frac{v_2}{v_1}\right)^{\frac{2v_1}{v_1 - v_2}}\left(1 - \ln\frac{v_2}{v_1}\right)}{\left(1 + \frac{v_2}{v_1 - v_2}\left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1 - v_2}}\right)^2}$$

$n = 3$

From Appendix [A.2] we know

$$u_3(dx) = dx\,\frac{v_1}{v_1 - v_2}\left[e^{\frac{-v_1}{v_1 - v_2}\left(x - 3log\frac{v_2}{v_1}\right)}\left\{\frac{\left(\frac{v_1}{v_1 - v_2}\left[x - 2log\,\frac{v_2}{v_1}\right]\right)^2}{2!} - \frac{\left(\frac{-v_1}{v_1 - v_2}\,log\,\frac{v_2}{v_1}\right)^2}{2!}\right\}\right]$$

which we then substitute into $(A.3.1)$ to obtain

$P(\xi_\infty = -3)$

$$= \frac{v_1 - v_2}{v_1}\left(1 + \frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}\right)\sum_{\mathrm{i}=0}^{\infty}\frac{(-1)^{\mathrm{i}}}{\mathrm{i}!}\left(\frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2}\right)^{\mathrm{i}}$$

$$\times \int_{-\mathrm{i}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}}^{\infty}\left[\left(\mathrm{x} + \mathrm{i}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}\right)^{\mathrm{i}} \mathrm{e}^{\frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2}\left(\mathrm{x} + \mathrm{i}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}\right)}\right]\frac{v_1}{v_1 - v_2}\left[e^{\frac{-v_1}{v_1 - v_2}\left(x - 3log\frac{v_2}{v_1}\right)}\left\{\frac{\left(\frac{v_1}{v_1 - v_2}\left[x - 2log\frac{v_2}{v_1}\right]\right)^2}{2!}\right.\right.$$

$$\left.\left. - \frac{\left(\frac{-v_1}{v_1 - v_2}log\frac{v_2}{v_1}\right)^2}{2!}\right\}\right]dx$$

$$let\, u = \mathrm{x} + \mathrm{i}\log\frac{\mathrm{v}_2}{\mathrm{v}_1} \qquad du = dx$$

$$= \left(1 + \frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}\right)\frac{\left(\frac{v_1}{v_1 - v_2}\right)^2}{2!}\left(\frac{v_2}{v_1}\right)^{\frac{3v_1}{v_1 - v_2}}\sum_{\mathrm{i}=0}^{\infty}\frac{(-1)^{\mathrm{i}}}{\mathrm{i}!}\left(\frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2}\left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1 - v_2}}\right)^{\mathrm{i}}$$

$$\times \int_0^{\infty}(\mathrm{u})^{\mathrm{i}}\mathrm{e}^{-(\mathrm{u})}\left\{u^2 + 2(i+2)\left(-log\frac{v_2}{v_1}\right)u + (i+1)(i+2)\left(-log\frac{v_2}{v_1}\right)^2\right.$$

$$\left. + (i+1)\left(-log\frac{v_2}{v_1}\right)^2\right\}du$$

$$= \left(1 + \frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}\right)\frac{\left(\frac{v_1}{v_1 - v_2}\right)^2}{2!}\left(\frac{v_2}{v_1}\right)^{\frac{3v_1}{v_1 - v_2}}\sum_{\mathrm{i}=0}^{\infty}(-1)^{\mathrm{i}}\left(\frac{\mathrm{v}_2}{\mathrm{v}_1 - \mathrm{v}_2}\left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1 - v_2}}\right)^{\mathrm{i}}\left(\frac{(i+2)!}{\mathrm{i}!}\right.$$

$$\left. + \frac{2(i+2)!\left(-log\frac{v_2}{v_1}\right)}{\mathrm{i}!} + \frac{(i+2)!\left(-log\frac{v_2}{v_1}\right)^2}{\mathrm{i}!} + \frac{(i+1)\left(-log\frac{v_2}{v_1}\right)^2}{\mathrm{i}!}\right)$$

Apply $\sum_{j=0}^{\infty}(-1)^j x^j (j+1)(j+2) = \frac{2}{(1+x)^3}$ and the previous geometric series arguments

$$P(\xi_\infty = -3) = \left(1 + \frac{\nu_2}{\nu_1 - \nu_2} \ln \frac{\nu_2}{\nu_1}\right) \left(\frac{\nu_2}{\nu_1}\right)^{\frac{3\nu_1}{\nu_1 - \nu_2}} \frac{\left(\frac{\nu_1}{\nu_1 - \nu_2}\right)^2}{2!} \left\{ \frac{2\left(1 - \ln \frac{\nu_2}{\nu_1}\right)^2}{\left(1 + \frac{\nu_2}{\nu_1 - \nu_2}\left(\frac{\nu_2}{\nu_1}\right)^{\frac{\nu_1}{\nu_1 - \nu_2}}\right)^3} + \frac{\left(-\ln \frac{\nu_2}{\nu_1}\right)^2}{\left(1 + \frac{\nu_2}{\nu_1 - \nu_2}\left(\frac{\nu_2}{\nu_1}\right)^{\frac{\nu_1}{\nu_1 - \nu_2}}\right)^2} \right\}$$

$n = 4$

From Appendix [A.2] we know

$$u_4(dx) = dx \frac{v_1}{v_1 - v_2} \left[e^{\frac{-v_1}{v_1 - v_2}\left(x - 4 log \frac{v_2}{v_1}\right)} \left\{ \frac{\left(\frac{v_1}{v_1 - v_2}\left[x - 3 log \frac{v_2}{v_1}\right]\right)^3}{3!} - \frac{\left(\frac{-2v_1}{v_1 - v_2} log \frac{v_2}{v_1}\right)^3}{3!} - \frac{\left(\frac{-v_1}{v_1 - v_2} log \frac{v_2}{v_1}\right)^2}{2!} \left(\frac{v_1}{v_1 - v_2}\left[x - log \frac{v_2}{v_1}\right]\right) \right\} \right]$$

which we then substitute into $(A.3.1)$ to obtain

$(\xi_\infty = -4)$

$$= \Bigg(1 + \frac{v_2}{v_1 - v_2}\log\frac{v_2}{v_1}\Bigg)\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}\int_{-i\log\frac{v_2}{v_1}}^{\infty}\Bigg[\Bigg(\frac{v_2}{v_1 - v_2}\Bigg(x + i\log\frac{v_2}{v_1}\Bigg)\Bigg)^i e^{\frac{v_2}{v_1 - v_2}\left(x + i\log\frac{v_2}{v_1}\right)}\Bigg]\frac{v_1}{v_1 - v_2}\Bigg[e^{\frac{-v_1}{v_1-v_2}\left(x - 4log\frac{v_2}{v_1}\right)}\Bigg\{\frac{\left(\frac{v_1}{v_1 - v_2}\left[x - 3log\frac{v_2}{v_1}\right]\right)^3}{3!}$$

$$- \frac{\left(\frac{-2v_1}{v_1 - v_2}log\frac{v_2}{v_1}\right)^3}{3!} - \frac{\left(\frac{-v_1}{v_1 - v_2}log\frac{v_2}{v_1}\right)^2}{2!}\left(\frac{v_1}{v_1 - v_2}\left[x - log\frac{v_2}{v_1}\right]\right)\Bigg\}\Bigg]dx$$

$$let\ u = x + i\log\frac{v_2}{v_1} \qquad du = dx$$

$$= \Bigg(1 + \frac{v_2}{v_1 - v_2}\log\frac{v_2}{v_1}\Bigg)\frac{\left(\frac{v_1}{v_1 - v_2}\right)^3}{3!}\left(\frac{v_2}{v_1}\right)^{\frac{4v_1}{v_1 - v_2}}\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}\left(\frac{v_2}{v_1 - v_2}\left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1-v_2}}\right)^i$$

$$\times \int_0^{\infty}(u)^i e^{-(u)}\Bigg\{u^3 + 3(i+3)\left(-log\frac{v_2}{v_1}\right)u^2 + 3u(i+3)^2\left(-log\frac{v_2}{v_1}\right)^2 + (i+3)^3\left(-log\frac{v_2}{v_1}\right)^3 - 8\left(-log\frac{v_2}{v_1}\right)^3\Bigg\}du =$$

Expanding $(i+3)^k, k = 2, 3$ and noting $\int_0^\infty x^{n-1}e^{-x}dx = (n-1)! = \Gamma(n)$

$$= \left(1+\frac{\mathrm{v}_2}{\mathrm{v}_1-\mathrm{v}_2}\log\frac{\mathrm{v}_2}{\mathrm{v}_1}\right)\frac{\left(\frac{v_1}{v_1-v_2}\right)^3}{3!}\left(\frac{v_2}{v_1}\right)^{\frac{4v_1}{v_1-v_2}}\sum_{\mathrm{i}=0}^{\infty}(-1)^{\mathrm{i}}\left(\frac{\mathrm{v}_2}{\mathrm{v}_1-\mathrm{v}_2}\left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1-v_2}}\right)^{\mathrm{i}}\left\{(i+1)(i+2)(i\right.$$

$$+3)\left(1-\ln\frac{v_2}{v_1}\right)^3+3(i+1)(i+2)\left(-log\frac{v_2}{v_1}\right)^2\left(1-\ln\frac{v_2}{v_1}\right)$$

$$\left.+4(i+1)\left(-\ln\frac{v_2}{v_1}\right)^3\right\}$$

Then applying $\sum_{j=0}^{\infty}(-1)^j x^j (j+1)(j+2)(j+3)=\frac{6}{(1+x)^4}$ and the previous geometric series arguments we get the final form

$$(\xi_\infty=-4)=\left(1+\frac{v_2}{v_1-v_2}\ln\frac{v_2}{v_1}\right)\frac{\left(\frac{v_1}{v_1-v_2}\right)^3}{3!}\left(\frac{v_2}{v_1}\right)^{\frac{4v_1}{v_1-v_2}}\left\{\frac{6\left(1-\ln\frac{v_2}{v_1}\right)^3}{\left(1+\frac{v_2}{v_1-v_2}\left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1-v_2}}\right)^4}\right.$$

$$\left.+\frac{6\left(-\ln\frac{v_2}{v_1}\right)^2\left(1-\ln\frac{v_2}{v_1}\right)}{\left(1+\frac{v_2}{v_1-v_2}\left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1-v_2}}\right)^3}+\frac{4\left(-\ln\frac{v_2}{v_1}\right)^3}{\left(1+\frac{v_2}{v_1-v_2}\left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1-v_2}}\right)^2}\right\}$$

$n = 5$

From Appendix [A.2] we know

$$u_5(dx) = dx\frac{v_1}{v_1 - v_2}\left[ee^{\frac{-v_1}{v_1-v_2}\left(x-5log\frac{v_2}{v_1}\right)}\left\{\frac{\left(\frac{v_1}{v_1 - v_2}\left[x - 4log\frac{v_2}{v_1}\right]\right)^4}{4!} - \frac{\left(\frac{-3v_1}{v_1 - v_2}log\frac{v_2}{v_1}\right)^4}{4!}\right.\right.$$

$$- \frac{\left(\frac{-v_1}{v_1 - v_2}log\frac{v_2}{v_1}\right)^2}{2!}\frac{\left(\frac{v_1}{v_1 - v_2}\left[x - 2log\frac{v_2}{v_1}\right]\right)^2}{2!} + \frac{\left(\frac{-v_1}{v_1 - v_2}log\frac{v_2}{v_1}\right)^4}{2!\,2!}$$

$$\left.\left.- \frac{\left(\frac{-2v_1}{v_1 - v_2}log\frac{v_2}{v_1}\right)^3}{3!}\left(\frac{v_1}{v_1 - v_2}\left[x - log\frac{v_2}{v_1}\right]\right)\right\}\right]$$

which we then substitute into

$$P(\xi_\infty = -5) = \frac{v_1 - v_2}{v_1}\left(1 + \frac{v_2}{v_1 - v_2}\log\frac{v_2}{v_1}\right)\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}\int_{-i\log\frac{v_2}{v_1}}^{\infty}\left[\left(\frac{v_2}{v_1 - v_2}\left(x + i\log\frac{v_2}{v_1}\right)\right)^i e^{\frac{v_2}{v_1 - v_2}\left(x + i\log\frac{v_2}{v_1}\right)}\right]u_5(dx)$$

to obtain

$$P(\xi_\infty = -5) = \left(1 + \frac{v_2}{v_1 - v_2}\log\frac{v_2}{v_1}\right)\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}\left(\frac{v_2}{v_1 - v_2}\right)^i$$

$$\times \int_{-i\log\frac{v_2}{v_1}}^{\infty} \left[\left(\left(x + i\log\frac{v_2}{v_1}\right)\right)^{i} e^{\frac{v_2}{v_1-v_2}\left(x+i\log\frac{v_2}{v_1}\right)}\right]\left[e^{\frac{-v_1}{v_1-v_2}\left(x-5log\left(\frac{v_2}{v_1}\right)\right)}\left\{\frac{\left(\frac{v_1}{v_1-v_2}\left[x-4log\left(\frac{v_2}{v_1}\right)\right]\right)^4}{4!}\right.\right.$$

$$-\frac{\left(\frac{-3v_1}{v_1-v_2}log\left(\frac{v_2}{v_1}\right)\right)^4}{4!}-\frac{\left(\frac{-v_1}{v_1-v_2}log\left(\frac{v_2}{v_1}\right)\right)^2}{2!}\frac{\left(\frac{v_1}{v_1-v_2}\left[x-2log\left(\frac{v_2}{v_1}\right)\right]\right)^2}{2!}$$

$$\left.\left.+\frac{\left(\frac{-v_1}{v_1-v_2}log\left(\frac{v_2}{v_1}\right)\right)^4}{2!\,2!}-\frac{\left(\frac{-2v_1}{v_1-v_2}log\left(\frac{v_2}{v_1}\right)\right)^3}{3!}\left(\frac{v_1}{v_1-v_2}\left[x-log\left(\frac{v_2}{v_1}\right)\right]\right)\right\}\right]dx$$

note that $\left(\frac{v_1-v_2}{v_1}\times\frac{v_1}{v_1-v_2}\right)$ cancel out.

Taking care of the middle parenthesis { } and using $\pm ilog\left(\frac{v_2}{v_1}\right)$ we have

$$\frac{1}{4!}\left\{\left(\frac{v_1}{v_1-v_2}\left[x-4log\left(\frac{v_2}{v_1}\right)\pm ilog\left(\frac{v_2}{v_1}\right)\right]\right)^4-\left(\frac{-3v_1}{v_1-v_2}log\left(\frac{v_2}{v_1}\right)\right)^4\right.$$

$$-6\left(\frac{-v_1}{v_1-v_2}log\left(\frac{v_2}{v_1}\right)\right)^2\left(\frac{v_1}{v_1-v_2}\left[x-2log\left(\frac{v_2}{v_1}\right)\pm ilog\left(\frac{v_2}{v_1}\right)\right]\right)^2$$

$$+6\left(\frac{-v_1}{v_1-v_2}log\left(\frac{v_2}{v_1}\right)\right)^4$$

$$\left.-4\left(\frac{-2v_1}{v_1-v_2}log\left(\frac{v_2}{v_1}\right)\right)^3\left(\frac{v_1}{v_1-v_2}\left[x-log\left(\frac{v_2}{v_1}\right)\pm ilog\left(\frac{v_2}{v_1}\right)\right]\right)\right\}=$$

$$let\ u = x + i\log\frac{v_2}{v_1}$$

$$= \Bigg\{u^4 + 4u^3(i+4)\left(-log\left(\frac{v_2}{v_1}\right)\right) + 6\boldsymbol{u^2(i+4)^2}\left(-log\left(\frac{v_2}{v_1}\right)\right)^2 + 4\boldsymbol{u(i+4)^3}\left(-log\left(\frac{v_2}{v_1}\right)\right)^3$$

$$+ \boldsymbol{(i+4)^4}\left(-log\left(\frac{v_2}{v_1}\right)\right)^4$$

$$- 6\left(-log\left(\frac{v_2}{v_1}\right)\right)^2\left[u^2 + 2u(i+2)\left(-log\left(\frac{v_2}{v_1}\right)\right) + \boldsymbol{(i+2)^2}\left(-log\left(\frac{v_2}{v_1}\right)\right)^2\right]$$

$$- 32\left(-log\left(\frac{v_2}{v_1}\right)\right)^3\left[u + (i+1)\left(-log\left(\frac{v_2}{v_1}\right)\right)\right] - 75\left(-log\left(\frac{v_2}{v_1}\right)\right)^4\Bigg\}$$

Expanding the selections (shown in bold above) we obtain

$$(j+4)^4 = (j+1)(j+2)(j+3)(j+4) + 6(j+1)(j+2)(j+3) + 25(j+1)(j+2) + 65(j+1) + 81$$

$$u(j+4)^3 = u[(j+2)(j+3)(j+4) + 3(j+2)(j+3) + 7(j+2) + 8]$$

$$u^2(j+4)^2 = u^2\big[\big((j+3+1)(j+4)\big)\big]$$

$$= u^2\big[\big((j+3)(j+4) + (j+4)\big)\big] = u^2\big[\big((j+3)(j+4) + (j+3) + 1\big)\big]$$

$$(j+2)^2 = (j+1)(j+2) + (j+1) + 1$$

Continuing with calculations we now have

$$P(\xi_\infty = -5) = \left(1 + \frac{v_2}{v_1 - v_2}\ln\frac{v_2}{v_1}\right)\left(\frac{v_2}{v_1}\right)^{\frac{5v_1}{v_1 - v_2}}\frac{\left(\frac{v_1}{v_1 - v_2}\right)^4}{4!}\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}\left(\frac{v_2}{v_1 - v_2}\left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1 - v_2}}\right)^i \times$$

$$\times\int_0^\infty u^i e^{-u}\Bigg\{u^4+4u^3(i+4)\left(-log\left(\frac{v_2}{v_1}\right)\right)+6u^2(i+4)^2\left(-log\left(\frac{v_2}{v_1}\right)\right)^2$$

$$+4u(i+4)^3\left(-log\left(\frac{v_2}{v_1}\right)\right)^3+(i+4)^4\left(-log\left(\frac{v_2}{v_1}\right)\right)^4$$

$$-6\left(-log\left(\frac{v_2}{v_1}\right)\right)^2\left[u^2+2u(i+2)\left(-log\left(\frac{v_2}{v_1}\right)\right)+(i+2)^2\left(-log\left(\frac{v_2}{v_1}\right)\right)^2\right]$$

$$-32\left(-log\left(\frac{v_2}{v_1}\right)\right)^3\left[u+(i+1)\left(-log\left(\frac{v_2}{v_1}\right)\right)\right]-75\left(-log\left(\frac{v_2}{v_1}\right)\right)^4\Bigg\}du$$

Note that $\int_0^\infty \mathrm{x}^{\mathrm{n}-1}\mathrm{e}^{-\mathrm{x}}dx=(\mathrm{n}-1)!=\Gamma(\mathrm{n})$

$$=\left(1+\frac{v_2}{v_1-v_2}\ln\frac{v_2}{v_1}\right)\left(\frac{v_2}{v_1}\right)^{\frac{5v_1}{v_1-v_2}}\frac{\left(\frac{v_1}{v_1-v_2}\right)^4}{4!}\sum_{\mathrm{i}=0}^{\infty}(-1)^{\mathrm{i}}\left(\frac{\mathrm{v}_2}{\mathrm{v}_1-\mathrm{v}_2}\left(\frac{v_2}{v_1}\right)^{\frac{v_1}{v_1-v_2}}\right)^{\mathrm{i}}\times$$

$$\times\Bigg\{(j+1)(j+2)(j+3)(j+4)\left[1-log\left(\frac{v_2}{v_1}\right)\right]^4$$

$$+(j+1)(j+2)(j+3)6\left(-log\left(\frac{v_2}{v_1}\right)\right)^2\left[1-log\left(\frac{v_2}{v_1}\right)\right]^2$$

$$+(j+1)(j+2)\left(-log\left(\frac{v_2}{v_1}\right)\right)^3\left[16-19log\left(\frac{v_2}{v_1}\right)\right]+(j+1)27\left(-log\left(\frac{v_2}{v_1}\right)\right)^4\Bigg\}$$

Now we use the following geometric series arguments

$$\sum_{j=0}^{\infty}(-1)^j x^j=\frac{1}{1+x}$$

$$\sum_{j=0}^{\infty}(-1)^j x^j (j+1) = \frac{1}{(1+x)^2}$$

$$\sum_{j=0}^{\infty}(-1)^j x^j (j+1)(j+2) = \frac{2}{(1+x)^3}$$

$$\sum_{j=0}^{\infty}(-1)^j x^j (j+1)(j+2)(j+3) = \frac{6}{(1+x)^4}$$

$$\sum_{j=0}^{\infty}(-1)^j x^j (j+1)(j+2)(j+3)(j+4) = \frac{24}{(1+x)^5}$$

to obtain the final form

$$P(\xi_\infty = -5) = \left(1 + \frac{\nu_2}{\nu_1 - \nu_2}\ln\frac{\nu_2}{\nu_1}\right)\left(\frac{\nu_2}{\nu_1}\right)^{\frac{5\nu_1}{\nu_1-\nu_2}} \frac{\left(\frac{\nu_1}{\nu_1-\nu_2}\right)^4}{4!}\left\{\frac{24\left(1-\ln\frac{\nu_2}{\nu_1}\right)^4}{\left(1+\frac{\nu_2}{\nu_1-\nu_2}\left(\frac{\nu_2}{\nu_1}\right)^{\frac{\nu_1}{\nu_1-\nu_2}}\right)^5}\right.$$

$$+\frac{36\left(-\ln\frac{\nu_2}{\nu_1}\right)^2\left(1-\ln\frac{\nu_2}{\nu_1}\right)^2}{\left(1+\frac{\nu_2}{\nu_1-\nu_2}\left(\frac{\nu_2}{\nu_1}\right)^{\frac{\nu_1}{\nu_1-\nu_2}}\right)^4} + \frac{2\left(-\ln\frac{\nu_2}{\nu_1}\right)^3\left(16-19\ln\frac{\nu_2}{\nu_1}\right)}{\left(1+\frac{\nu_2}{\nu_1-\nu_2}\left(\frac{\nu_2}{\nu_1}\right)^{\frac{\nu_1}{\nu_1-\nu_2}}\right)^3}$$

$$\left. + \frac{27\left(-\ln\frac{\nu_2}{\nu_1}\right)^4}{\left(1+\frac{\nu_2}{\nu_1-\nu_2}\left(\frac{\nu_2}{\nu_1}\right)^{\frac{\nu_1}{\nu_1-\nu_2}}\right)^2}\right\}$$

$n = 6$

From Appendix [A.2] we know

$$u_6(dx) = dx\frac{v_1}{v_1-v_2}\left[e^{\frac{-v_1}{v_1-v_2}\left(x-6log\frac{v_2}{v_1}\right)}\left\{\frac{\left(\frac{v_1}{v_1-v_2}\left[x-5log\frac{v_2}{v_1}\right]\right)^5}{5!}\right.\right.$$

$$-\frac{\left(\frac{-v_1}{v_1-v_2}log\frac{v_2}{v_1}\right)^2}{2!}\frac{\left(\frac{v_1}{v_1-v_2}\left[x-3log\frac{v_2}{v_1}\right]\right)^3}{3!}$$

$$-\frac{\left(\frac{-2v_1}{v_1-v_2}log\frac{v_2}{v_1}\right)^3}{3!}\frac{\left(\frac{v_1}{v_1-v_2}\left[x-2log\frac{v_2}{v_1}\right]\right)^2}{2!}$$

$$-\frac{\left(\frac{-3v_1}{v_1-v_2}log\frac{v_2}{v_1}\right)^4}{4!}\frac{\left(\frac{v_1}{v_1-v_2}\left[x-log\frac{v_2}{v_1}\right]\right)}{1!}$$

$$-\frac{\left(\frac{-v_1}{v_1-v_2}log\frac{v_2}{v_1}\right)^4}{2!\,2!}\left(\frac{v_1}{v_1-v_2}\left[x-log\frac{v_2}{v_1}\right]\right)-\frac{\left(\frac{-4v_1}{v_1-v_2}log\frac{v_2}{v_1}\right)^5}{5!}$$

$$\left.\left.+\frac{2\left(\frac{v_1}{v_1-v_2}log\frac{v_2}{v_1}\right)^2}{2!}\frac{\left(\frac{-2v_1}{v_1-v_2}log\frac{v_2}{v_1}\right)^3}{3!}\right\}\right]$$

which we then substitute into $(A.3.1)$. Then similar arguments as in above apply, note that here we use

$$\sum_{j=0}^{\infty}(-1)^j x^j (j+1)(j+2)(j+3)(j+4)(j+5) = \frac{120}{(1+x)^6}$$

in addition to the above mentioned geometric series arguments and the expansion of $(j+5)^5$ is shown below

$$(j+5)^5 = (j+1)(j+2)(j+3)(j+4)(j+5) + 10(j+1)(j+2)(j+3)(j+4)$$
$$+ 65(j+1)(j+2)(j+3) + 299(j+1)(j+2) + 781(j+1) + 1024$$

The final form is as follows

$$P(\xi_\infty = -6) = \left(1 + \frac{\nu_2}{\nu_1 - \nu_2}\ln\frac{\nu_2}{\nu_1}\right)\left(\frac{\nu_2}{\nu_1}\right)^{\frac{6\nu_1}{\nu_1-\nu_2}} \frac{\left(\frac{\nu_1}{\nu_1-\nu_2}\right)^5}{5!} \left\{ \frac{120\left(1-\ln\frac{\nu_2}{\nu_1}\right)^5}{\left(1+\frac{\nu_2}{\nu_1-\nu_2}\left(\frac{\nu_2}{\nu_1}\right)^{\frac{\nu_1}{\nu_1-\nu_2}}\right)^6} \right.$$
$$+ \frac{240\left(-\ln\frac{\nu_2}{\nu_1}\right)^2\left(1-\ln\frac{\nu_2}{\nu_1}\right)^3}{\left(1+\frac{\nu_2}{\nu_1-\nu_2}\left(\frac{\nu_2}{\nu_1}\right)^{\frac{\nu_1}{\nu_1-\nu_2}}\right)^5} + \frac{30\left(-\ln\frac{\nu_2}{\nu_1}\right)^3\left(8-19\ln\frac{\nu_2}{\nu_1}+11\left(-\ln\frac{\nu_2}{\nu_1}\right)^2\right)}{\left(1+\frac{\nu_2}{\nu_1-\nu_2}\left(\frac{\nu_2}{\nu_1}\right)^{\frac{\nu_1}{\nu_1-\nu_2}}\right)^4}$$
$$\left. + \frac{54\left(-\ln\frac{\nu_2}{\nu_1}\right)^4\left(5-7\ln\frac{\nu_2}{\nu_1}\right)}{\left(1+\frac{\nu_2}{\nu_1-\nu_2}\left(\frac{\nu_2}{\nu_1}\right)^{\frac{\nu_1}{\nu_1-\nu_2}}\right)^3} + \frac{256\left(-\ln\frac{\nu_2}{\nu_1}\right)^5}{\left(1+\frac{\nu_2}{\nu_1-\nu_2}\left(\frac{\nu_2}{\nu_1}\right)^{\frac{\nu_1}{\nu_1-\nu_2}}\right)^2} \right\}$$

A.4 The likelihood approach to hypothesis testing

Based on Csörgö & Horváth (1997, p.1 The Likelihood approach).

The likelihood approach to hypothesis testing: at most one change-point model (AMOC) where parameter ϑ changes at some unknown point in time τ. We have a two sample hypothesis test, where the parameters are unknown and must be estimated.

$$H_0:\ \vartheta_1 = \cdots = \vartheta_n \qquad vs. \qquad H_1:\ \vartheta_1 = \cdots = \vartheta_\tau \neq \vartheta_{\tau+1} = \cdots = \vartheta_n$$

If the change occurs at $\tau = k$, then we reject H_0 for small values of Λ_k

$$\Lambda_k = \frac{\sup_{\vartheta\in\Theta} \prod_{i=1}^{n} f(X_i;\vartheta)}{\sup_{(\vartheta,\vartheta^*)\in\Theta} \prod_{i=1}^{\tau} f(X_i;\vartheta) \prod_{i=\tau+1}^{n} f(X_i;\vartheta^*)}$$

Here $\{X_i: i = 1, \dots, n\}$ have probability density functions $f(x;\vartheta_i)$. For each possible value of $1 \leq k \leq n$ we find unique (mle) estimators $\hat{\vartheta}_k$, $\hat{\vartheta}_k^*$, $\hat{\vartheta}_n$, note that if $k = n$ then there is no change, and $\hat{\vartheta}^*$ cannot be found.

Rewriting the likelihood ratio we get

$$-2\log\Lambda_k = 2\left\{L_k\left(\hat{\vartheta}\right) + L_k^*\left(\hat{\vartheta}^*\right) - L_n\left(\hat{\vartheta}\right)\right\}$$

Therefore we will reject H_0 if the maximum of the above likelihood ratio, the test statistic U_n, is large. We use the maximum since τ is unknown.

$$U_n = \max_{1\leq k<n} \left(-2\log\Lambda_k\right)$$

The asymptotic distribution of the above test statistic is as follows

$$\lim_{n\to\infty} P\left\{(2\log\log n\, U_n)^{1/2} - \left(2\log\log n + \frac{d}{2}\log\log\log n - \log\Gamma\left(\frac{d}{2}\right)\right) \le t\right\} = e^{-2e^{-t}} \quad \forall\, t$$

where d is the number of parameters that change under the H_1 hypothesis.

Exponential distribution case

In the case of the exponential distribution where $Y \sim Exp(\vartheta)$, the density function is $f(y,\vartheta) = \frac{1}{\vartheta}e^{-y/\vartheta}$, $y \ge 0$, the mle of ϑ is $\hat{\vartheta} = \frac{1}{n}\sum_{i=1}^{n} Y_i$, note that $v = \frac{1}{\vartheta}$, $\tau = k$, and the likelihood ratio is then as follows

$$-2\log\Lambda_k = 2\left\{k\log\frac{1}{\hat{\vartheta}_k} + (n-k)\log\frac{1}{\hat{\vartheta}_k^*} - n\log\frac{1}{\hat{\vartheta}_n}\right\}$$

where

$$\hat{\vartheta}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i \qquad \hat{\vartheta}_k = \frac{1}{k}\sum_{i=1}^{k} Y_i \qquad \hat{\vartheta}_k^* = \frac{1}{n-k}\sum_{i=k+1}^{n} Y_i$$

Note that the asymptotic theory of U_n is quite robust to departures from the assumptions of exponentiality and independence.

As an example let us use one of the datasets provided in Proschan (1963). The dataset we chose contains 27 observations of times between successive failures of the air conditioning system of a Boeing 720 jet airplane (7913), see Figure A.4.1 below. We assume the data are independent and exponential.

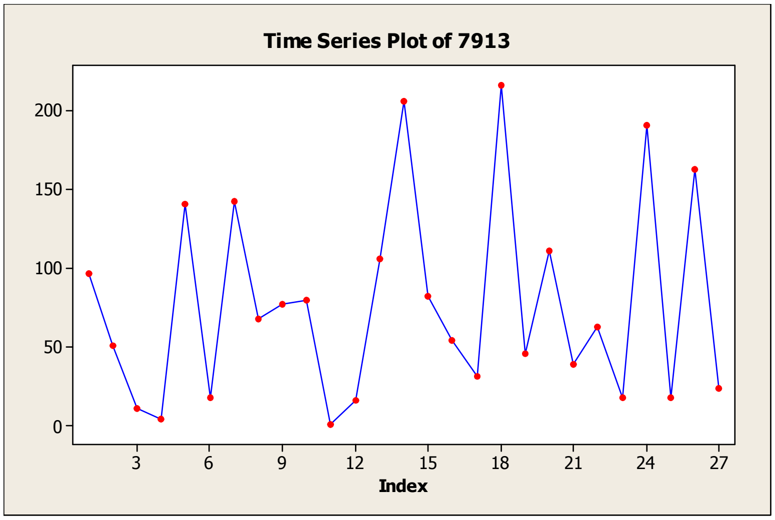


Figure A.4.1. Time series plot of intervals between failures of equipment.

We calculate the test statistic $U_{27} = 1.46$ which gives us $\hat{\tau} = 4$, and thus $\hat{\vartheta}_{27} = 76.8$, $\hat{\vartheta}_4 = 40.75$, and $\hat{\vartheta}_4^* = 83.14$.

Then based on the asymptotic distribution (double exponential) below

$$P\left\{W \leq (2\log\log n\, U_n)^{1/2} - \left(2\log\log n + \frac{d}{2}\log\log\log n - \log\Gamma\left(\frac{d}{2}\right)\right) = w\right\} \sim e^{-2e^{-w}}$$

We obtain

$$w = \left((2\log\log 27)1.46\right)^{1/2} - \left(2\log\log 27 + \frac{1}{2}\log\log\log 27 - \log\Gamma\left(\frac{1}{2}\right)\right) = -0.581$$

And thus the

$$p - value = P\{W > 0.581\} + P\{W \leq -0.581\} = \left(1 - e^{-2e^{-0.581}}\right) + e^{-2e^{0.581}} = 0.7012$$

Since the $p - value = 0.7012$ is quite large, we do not reject the “no change” null hypothesis, and we do not detect a change in the value of the parameter ϑ.

Note that the assumptions were not violated. We checked for independence by looking at autocorrelation and partial autocorrelation plots, there were no significant lags. And the exponentiality assumption is validated by Figure A.4.3 below.

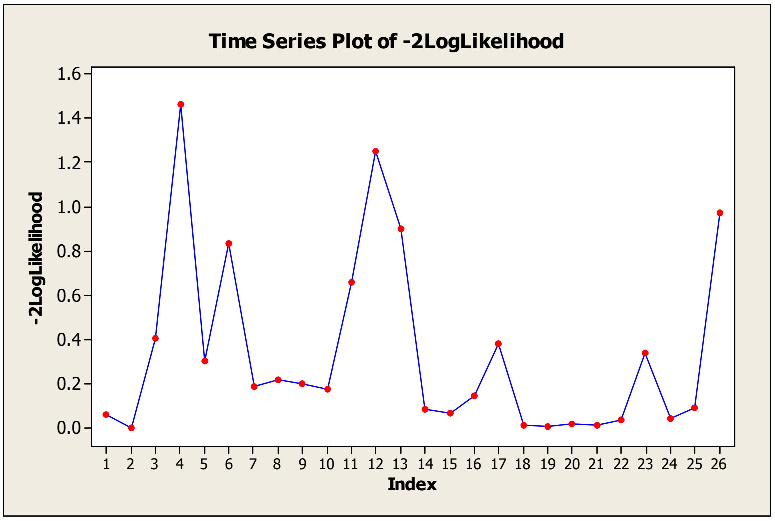


Figure A.4.2. Graph of $-2\log\Lambda_{\mathrm{k}}$, you can see the highest peak at 4.

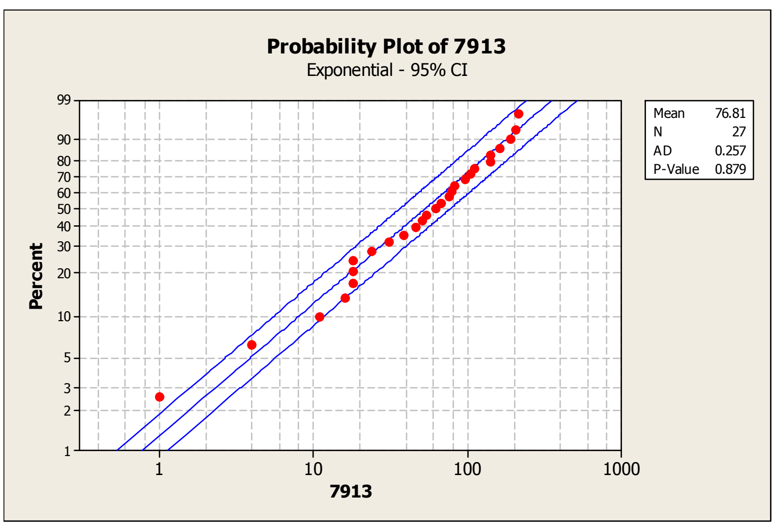


Figure A.4.3. Test for exponentiality. Anderson-Darling statistic = 0.257, p-value = 0.879.

Multivariate normal distribution case

In the case of the multivariate normal distribution we consider a sequence of independent time-ordered random vectors $\boldsymbol{Y}_i \in \boldsymbol{R}^d,\ i = 1, \dots, n$. When testing for a change in the mean only, we have the following

$$\boldsymbol{Y}_i \sim \begin{cases} f(\boldsymbol{\mu_1}, \boldsymbol{\Sigma}) = \dfrac{1}{\left(\sqrt{2\pi}\right)^d |\boldsymbol{\Sigma}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{\mu_1})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{Y}-\boldsymbol{\mu_1})}, & i = 1, \dots, \tau \\ f(\boldsymbol{\mu_2}, \boldsymbol{\Sigma}) = \dfrac{1}{\left(\sqrt{2\pi}\right)^d |\boldsymbol{\Sigma}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{\mu_2})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{Y}-\boldsymbol{\mu_2})}, & i = \tau + 1, \dots, n \end{cases}$$

Let $\tau = k$, then the test statistic is as follows

$$U_n = \max_{1 \le k < n} (-2 \log \Lambda_{\mathrm{k}}) = \max_{1 \le k < n} n \log \left(\frac{|\widehat{\boldsymbol{\Sigma}}_{\mathrm{n}}|}{|\widehat{\boldsymbol{\Sigma}}_{\mathrm{k}}|} \right)$$

where the estimators of the mean are

$$\widehat{\boldsymbol{\mu}}_{\mathrm{n}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{Y}_i \qquad \widehat{\boldsymbol{\mu}}_{\mathrm{k}} = \frac{1}{k} \sum_{i=1}^{k} \boldsymbol{Y}_i \qquad \widehat{\boldsymbol{\mu}}_{\mathrm{k}}^* = \frac{1}{n-k} \sum_{i=k+1}^{n} \boldsymbol{Y}_i$$

and the estimators of the variance-covariance matrix are

$$\widehat{\boldsymbol{\Sigma}}_{\mathrm{k}} = \frac{1}{\mathrm{n}} \left\{ \sum_{i=1}^{k} (\boldsymbol{Y}_i - \widehat{\boldsymbol{\mu}}_{\mathrm{k}})(\boldsymbol{Y}_i - \widehat{\boldsymbol{\mu}}_{\mathrm{k}})^T + \sum_{i=k+1}^{n} (\boldsymbol{Y}_i - \widehat{\boldsymbol{\mu}}_{\mathrm{k}}^*)(\boldsymbol{Y}_i - \widehat{\boldsymbol{\mu}}_{\mathrm{k}}^*)^T \right\}$$

$$\widehat{\boldsymbol{\Sigma}}_{\mathrm{n}} = \frac{1}{\mathrm{n}} \left\{ \sum_{i=1}^{n} (\boldsymbol{Y}_i - \widehat{\boldsymbol{\mu}}_{\mathrm{n}})(\boldsymbol{Y}_i - \widehat{\boldsymbol{\mu}}_{\mathrm{n}})^T \right\}$$

Note that the asymptotic theory of U_n is quite robust to departures from the assumptions of normality and independence. Then based on the asymptotic distribution (double exponential) below

$$P\left\{W \leq (2\log\log n\, U_n)^{1/2} - \left(2\log\log n + \frac{d}{2}\log\log\log n - \log\Gamma\left(\frac{d}{2}\right)\right) = w\right\} \sim e^{-2e^{-w}}$$

We can obtain the p-value and draw our conclusion.

BIBLIOGRAPHY

Akman, V.E. and Raftery, A.E. (1986) Asymptotic inference for a change-point Poisson process. *Ann. Stat*., **14**, 1583-1590.

Annual mean temperature anomalies in .01 C selected zonal means (2009). Source: *GHCN 1880-03/2009*. WEB: http://data.giss.nasa.gov/gistemp/tabledata/ZonAnn.Ts.txt

Asmussen, S. (1987) *Applied Probability and Queues*. New York: Wiley.

Barry, D. and Hartigan, J. A. (1993) A Bayesian analysis for change point problems. *Journal of the American Statistical Association*, **88**, 309-319.

Belardinelli, M. E., Bizzarri, A. and Cocco, M. (2003) Earthquakes triggering by static and dynamic stress changes. *J. Geophys. Res*. **108**, doi:10.1029/2002JB001779.

Bhattacharya, P. K. (1987) Maximum likelihood estimation of a change-point in the distribution of independent random variables: general multiparameter case. *Journal of Multivariate Analysis*, **23**, 183-208.

Bhattacharya, P. K. and Frierson Jr., D. (1981) A nonparametric control chart for detecting small disorders. *The Annals of Statistics*, **9**, 544-554.

Brodsky, E. E., Karakostas, V. and Kanamori, H. (2000) A new observation of dynamically triggered regional seismicity: Earthquakes following the August 1999 Izmit, Turkey earthquake. *Geophys. Res. Lett*., **27**, 2741-2744.

Brown, J. V. and Muller, H. (1998) Statistical methods for DNA sequence segmentation. *Statistical Science*, **13**, 142-162.

Borovkov, A. A. (1999) Asymptotically optimal solutions in the change-point problem. Theory *of Probability and Its Applications*, **43**, 539-561.

Borovkov, K. (1999) On records and related processes for sequences with trends. *J. Appl. Prob.*, **36**, 668-681.

Carlstein, E. (1988) Nonparametric change-point estimation. *The Annals of Statistics*, **16**, 188-197.

Carter, R. L. and Blight, B. J. N. (1981) A Bayesian change-point problem with an application to the prediction and detection of ovulation in women. *Biometrics*, **37**, 743-751.

Chen, J., and Gupta, A. K. (2000) *Parametric statistical change point analysis*, New York: Birkhäuser.

Chernoff, H. and Rubin, H. (1956) The estimation of the location of a discontinuity in density. *Proc. 3rd. Berkeley Symp. Math. Stat. Prob.*, **1**, 19-37.

Chernoff, H. and Zacks, S. (1964) Estimating the current mean of a normal distribution which is subjected to changes in time. *The Annals of Mathematical Statistics,* **35**, 999-1018.

Chover, J., Ney, P., and Wainger, S. (1973) Functions on probability measures. *Journal d'Analyse Mathématique*, **26**, 255-302.

Cobb, G. W. (1978) The problem of the Nile: Conditional solution to a changepoint problem. *Biometrika*, **65**, 243-251.

Csörgö, M and Horváth, L. (1997). *Limit theorems in change-point analysis*. New York: Wiley.

Deshayes, J. (1984) Rupteres de modeles pour des processus de Poisson. *Ann. Sci. Univ. Clermont-Ferrand II*, **78**, 1-7.

Feller, W. R. (1971) An *Introduction to probability theory and its applications, Vol. II*. New York: Wiley.

Ferger, D. (1995) Nonparametric tests for nonstandard change-point problems. *The Annals of Statistics*, **23**, 1848-1861.

Fotopoulos, S. B. (2009) The geometric convergence rate of the classical change-point estimate. *Stat. Prob. Letters*., **79**, 131-137.

Fotopoulos, S. B. and Jandhyala, V.K. (2001) Maximum likelihood estimation of a change point for exponentially distributed random variables. *Stat. Prob. Letters*., **51**, 423-429.

Fotopoulos, S. B. and Jandhyala, V.K. (2007) On Hinkley's estimator: inference about the change-point. *Stat. Prob. Letters*., **77**, 1449-1458.

Fotopoulos, S. B., Jandhyala, V.K. and Tan, L. (2009) Asymptotic study of the change point mle in multivariate Gaussian families under contiguous alternatives. *Journal of Statistical Planning and Inference*, **139**, 1190-1202.

Gombay, E. and Horvath, L. (1990) Asymptotic distributions of maximum likelihood tests for change in mean. *Biometrika*, **77**, 411-414.

Gomberg, J., Reasenberg, P. A., Bodin, P. and Harris, R. A. (2001) Earthquake triggering by seismic waves following the Landers and Hector Mine earthquakes. *Nature*, **411**, 462-466.

Grinstead, C. M. and Snell, J. L. (2008) *Introduction to probability*. WEB: http://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/amsbook.mac.pdf

Groisman, P. V., Karl, T. R., Easterling, D. R., Knight, R. W., Jamason, P. F., Hennessy, K. J., Suppiah, R., Page, C. M., Wibig, J., Fortuniak, K., Razuvaev, V. N., Douglas, A., Førland, E., and Zhai, P. M. (1999) Changes in the probability of heavy precipitation: Important indicators of climate change. *Climatic Change*, **42**, 243-283.

Haccou, P. and Meelis, E. (1988) Testing for the number of change points in a sequence of exponential random variables. *Journal of Statistical Computations and Simulation*, **30**, 285-298.

Haccou, P., Meelis, E. and Geer, S. (1988) The likelihood ratio test for the change point problem for exponentially distributed random variables. *Stochastic Processes and Their Applications*, **27**, 121-139.

Hansen, J. and Lebedeff, S. (1987) Global trends of measured surface air temperature. *Journal of Geophysical Research*, **92**, 13345-13372.

Henze, N., and Zirkler, B. (1990) A class of invariant and consistent tests for multivariate normality. *Communications in Statistics: Theory and Methodology*, **19**, 3595-3617.

Hinkley, D.V. (1970) Inference about the change-point in a sequence of random variables. *Biometrika*, **57**, 1-17.

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

Hinkley, D. V. (1972) Time ordered classification. *Biometrika*, **59**, 509-523.

Hsu, D. A. (1979) Detecting shifts of parameter in Gamma sequences with applications to stock price and air traffic flow analysis. *Journal of the American Statistical Association*, **74**, 31-40.

Hu, I. and Rukhin, A. L. (1995) A lower bound for error probability in change-point estimation. *Stat. Sinic.*, **5**, 319-331.

International Seismological Centre (2008). International Seismological Centre On-line Bulletin, Thatcham, United Kingdom WEB: http://www.isc.ac.uk

Iwata, T. and Nakanishi, I. (2004) Hastening of occurrences of earthquakes due to dynamic triggering: The observation at Matsushiro, central Japan. *J. Seismol.*, **8**: 165–177.

Jain, S. (2001) Estimating the change point of Erlang interarrival time distribution. *INFOR*, **39**, 200-207.

Jandhyala, V. K. and Fotopoulos, S. B. (1999) Capturing the distributional behavior of the maximum likelihood estimator of a change-point. *Biometrika*, **86**, 129-140.

Jandhyala, V. K. and Fotopoulos, S. B. (2001) Rate of convergence of the maximum likelihood estimate of a change-point. *Sankhyă,* **63**, 277-85.

Jandhyala, V. K., Fotopoulos, S. B., and Hawkins, D. M. (2002) Detection and estimation of abrupt changes in the variability of a process. *Computational Statistics & Data Analysis*, **40**, 1-19.

Karasoy, D. S. and Kadilar, C. (2007) A new Bayes estimate of the change-point in the hazard function. *Computational Statistics & Data Analysis*, **51**, 2993-3001.

Khaliq, M. N., Ouarda, T. B. M. J., St-Hilaireb, A. and Gachonc, P. (2007) Bayesian change-point analysis of heat spell occurrences in Montreal, Canada. *Int. J. Climatol*., **27**, 805–818

Killmann, F. and Collani, E. (2001) A note on the convolution of the uniform and related distributions and their use in quality control. *Econ. Qual. Contr.* **16**, 17-41.

Kim, H. and Siegmund, D. (1989) The likelihood ratio test for a change-point in simple linear regression. *Biometrika*, **76**, 409-423.

Koenker, R. and Schorfheide, F. (1994) Quantile spline models for global temperature change. *Climatic Change*, **28**, 395-404.

Lau, K.-M., and Wu, H.-T. (2007) Detecting trends in tropical rainfall characteristics, 1979-2003. *International Journal of Climatology*, **27**, 979-988.

Loader, C. R. (1991) Inference for a hazard rate change point. *Biometrika*, **78**, 749-757.

Lombard, F. (1987) Rank tests for changepoint problems. *Biometrika*, **74**, 615-624.

Mardia, K. V. (1970) Measures of multivariate skewness and kurtosis with applications. *Biometrika*, **57**, 519-530.

Matthews, D. E. and Farewell, V. T. (1982) On testing for a constant hazard against a change-point alternative. *Biometrics*, **38**, 463-468.

Mitchell, T. D., Carter, T. R., Jones, P. D., Hulme, M., and New, M. (2004) A Comprehensive set of high-resolution grids of monthly climate for Europe and the globe: The observed record (1901–2000) and 16 scenarios (2001–2100). *Tyndall Working Paper*, **55**, 1–30.

Muller, H. and Wang, J. (1990) Nonparametric analysis of changes in hazard rates for censored survival data: An alternative to change-point models. *Biometrika*, **77**, 305-314.

New, M., Todd, M., Hulme, M., and Jones, P. (2001) Precipitation measurements and trends in the twentieth century. *International Journal of Climatology*, **21**, 1899-1922.

Page, E. S. (1955) A test for a change in a parameter occurring at an unknown point. *Biometrika*, **42**, 523-526.

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

Perry, M. B., Pignatiello Jr., J. J., and Simpson, J. R. (2007) Change-point estimation for monotonically changing Poisson rates in SPC. *Int. J. of Production Research*, **45**, 1971-1813.

Plummer, N., Salinger, M. J., Nicholls, N., Suppiah, R., Hennessy, K. J., Leighton, R. M., Trewin, B., Page, C. M., and Lough, J. M. (1999) Changes in climate extremes over the Australian region and New Zealand Over the twentieth century. *Climatic Change*, **42**, 183-202.

Prabhu, N. U. (1965) *Stochastic Processes*. New York: Macmillan.

Proschan, F. (1963) Theoretial explanation of observed decreasing failure rate. *Technometrics*, **5**, 375-383.

Qian, W., and Zhu, Y. (2001) Climate change in China from 1880 to 1998 and its impact on the environmental condition. *Climatic Change*, **50**, 419-444.

Raftery, A. E. and Akman, V. E. (1986) Bayesian analysis of a Poisson process with a change-point. *Biometrika*, **73**, 85-89.

Ramanayake, A. and Gupta, A. K. (2003) Tests for an epidemic change in a sequence of exponentially distributed random variables. *Biometrical Journal*, **45**, 946-958.

Ritov, Y. (1990) Asymptotic efficient estimation of the change-point with unknown distributions. *The Annals of Statistics*, **18**, 1829-1839.

Robeck, E. D. (2005) Correlation between increased earthquake frequency and static stress changes induced by the 2004 Sumatra-Andaman and 2005 Sumatra earthquakes. *Geol. Soc. Amer. Abstr. Prog.,* **37**, No. 7, p. 521.

Rubin, H. (1961) The estimation of discontinuities in multivariate densities, and related problems in stochastic process. *Proc. 4th Berkeley Symp. Math. Stat. Prob*., **1**, 563-574.

Sen, A.K. and Srivastava, M. S. (1975) On tests for detecting change in mean. *Annals of Statistics*, **3**, 98-108.

Shiryaev, A. N., Kabanov, Y. M., Kramkov, D. O., and Melnikov, A.V. (1993) Towards the theory of pricing of options of both European and American types, II, continuous time. *Theory of Probability and Its Applications*, **39**, 61-102.

Siegmund, D. (1988) Confidence sets in change-point problems. *International Statistical Review*, **56**, 31-48.

Siegmund, D. (1975) Error probabilities and average sample number of the sequential probability ratio test. *Journal of the Royal Statistical Society, Series B*, **37**, 394-401.

Siegmund, D. and Venkatraman, E. S. (1995) Using the generalized likelihood ratio statistic for sequential detection of a change-point. *The Annals of Statistics*, **23**, 255-271.

Smith, A. F. M. (1975) A Bayesian approach to inference about a change-point in a sequence of random variables. *Biometrika*, **62**, 407-416.

Srivastava, M. S. and Worsley K. J. (1986) Likelihood ratio tests for a change in the multivariate normal mean. *Journal of the American Statistical Association*, **81**, 199-204.

Topographic locator map of Sumatra (2007) *Created with GMT from SRTM data by Sadalmelik* WEB: http://en.wikipedia.org/wiki/File:Sumatra_Locator_Topography.png

Vardeman, S. B. and Jobe, J. M. (1999) *Statistical quality assurance methods for engineers*. New York: Wiley.

Worsley, K. J. (1986) Confidence regions and tests for a change-point in a sequence of exponential family random variables. *Biometrika*, **73**, 91-104.

Yao, Y. C. (1987) Approximating the distribution of the maximum likelihood estimate of the change-point in a sequence of independent random variables. *Ann. Stat.*, **15**, 1321-1328.