The Use of Extreme Value Theory for Making Statistical Inference About Endpoints of Distributions, with Applications in Global Optimization and Meteorology

Emily Hamilton
Cardiff University

December 2008

A thesis submitted for the degree of Doctor of Philosophy

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

SUMMARY

We use extreme value theory to make statistical inference about the endpoint of distributions. First we compare estimators of the endpoint of several distributions, including a distribution that appears in problems of global optimization. These estimators use a fixed number of order statistics (k) from a sample of fixed size (n). Two of the estimators investigated are the optimal linear estimator and the maximum likelihood estimator. We find that the optimal linear estimator often outperforms the maximum likelihood estimator.

We next investigate how the order statistics change as sample size increases. In order to do this, we define record times: the sample size at which the set of k smallest order statistics changes. We give the distributions of several statistics related to order statistics and record times, in particular we show that records occur according to a nonhomogeneous Poisson process. We show that order statistics can be modeled using a Markov chain, and use this Markov chain to investigate estimators of the endpoint of a distribution. Two additional estimators are derived and investigated using the Markov chain model.

Finally, we consider a meteorological application of extreme value theory. In particular, we estimate the maximum and minimum sea level at several ports in the Netherlands. This is done using a combination of record theory, singular spectrum decomposition and known estimators of the endpoint of a distribution.

ACKNOWLEDGEMENTS

It is a pleasure to thank the many people who made this thesis possible. Great thanks are owed to my supervisor Anatoly Zhigljavsky. His enthusiasm, advice and knowledge have been invaluable.

I would also like to thank my friends at Cardiff University for making being a student here so enjoyable. Four people deserve special thanks; James Hindmarsh and Jane Bartlett for their friendship and generous offer of accommodation; my office-mate Mari Jones for being a continual source of fun, support and coffee; and finally Maxwell Wallace, for his emotional support, and for motivating and distracting me at all the right times.

An enormous gratitude is owed to my family for the love they have always given me. They have been a constant source of encouraging words and have always been ready to lend an attentive ear.

Finally I would like to thank EPSRC for providing me with the grant that made this research financially viable. For this I am very grateful.

Table of Contents

Table of Contents iv

1 Introduction 1
- 1.1 Summary of Thesis . . . 1
  - 1.1.1 Using Order Statistics to Estimate Endpoint . . . 1
  - 1.1.2 Records, Poisson Processes and Markov chains . . . 6
  - 1.1.3 Meteorological Applications . . . 10
- 1.2 Results Concerning the Distributions of Extremes . . . 12
  - 1.2.1 Asymptotic Distribution of Minimum Order Statistic . . . 12
  - 1.2.2 Asymptotic Distribution of Extreme Order Statistics . . . 15
  - 1.2.3 Asymptotic Behavior of $\kappa_n - m$ . . . 17
- 1.3 Defining the Estimators . . . 17
  - 1.3.1 Linear Estimators . . . 18
  - 1.3.2 Other Linear Estimators . . . 21
  - 1.3.3 Maximum Likelihood Estimator . . . 21

2 Comparison of Estimators: Static Sample 23
- 2.1 Defining More Estimators . . . 23
  - 2.1.1 One Non-Zero Coefficient . . . 23
  - 2.1.2 Two Non-Zero Coefficients . . . 24
  - 2.1.3 Three Non-Zero Coefficients . . . 25
- 2.2 Density of Normalized Estimators . . . 29
  - 2.2.1 Method for Deriving the Densities . . . 30
  - 2.2.2 Density of a Normalized General Linear Estimator . . . 33
  - 2.2.3 Density of Estimator $(m^{(2)} - m)/\omega$ with $F(\cdot)$ Weibull . . . 34
  - 2.2.4 Density of $m^{(2)} - m/\omega$ where $\omega = \omega_3$ or ω_4 and $F(\cdot)$ is Weibull 37
  - 2.2.5 Density of $(m^* - m)/\omega$ with $F(\cdot)$ Weibull and $k = 2$ . . . 41
  - 2.2.6 Density of $(m^* - m)/\omega$ where $\omega = \omega_3$ or ω_4 with $F(\cdot)$ Weibull and $k = 2$ . . . 43
- 2.3 Analysis Using Densities . . . 44

2.3.1 Comparing m° and m^* . . . 44
2.3.2 Investigation into the Effect of Increasing Sample Size . . . 44
2.4 Simulation study: Known α . . . 49
2.4.1 Histograms of Normalized Estimators . . . 49
2.4.2 Efficiency, Mean Square Error and Bias of Estimators . . . 55
2.5 Comparison of Estimators: Unknown α . . . 69
2.5.1 Efficiency: Wrong Tail Index . . . 69
2.5.2 Simulations Study . . . 72

3 Comparison of Estimators: Increasing Sample Size **76**
3.1 Introduction to Records and Poisson Process . . . 77
3.1.1 Notation and Definitions . . . 77
3.1.2 Weak Records . . . 81
3.1.3 Maximal Records . . . 82
3.1.4 Moments and Distributions . . . 82
3.1.5 Altering Time Scale . . . 86
3.2 Modeling Order Statistics and Estimators at Records . . . 89
3.2.1 Modeling Order Statistics at Record Values . . . 89
3.2.2 Modeling Normalized Order Statistics at Records . . . 91
3.2.3 Modeling Normalized Estimators at Records . . . 93
3.3 Analysis of Markov Chain . . . 96
3.3.1 Transition densities of Markov Chain . . . 96
3.3.2 Autocorrelation Functions of Markov Chain . . . 98
3.4 Comparison of the Estimators and Their Related Functionals on the Markov Chain . . . 105
3.4.1 Density Function of $\zeta_k^{(2)}$ . . . 105
3.4.2 Density Function of $\zeta_k^{(3)}$ . . . 106
3.4.3 Density Function of ζ_k^*, when $k = 2$, $\alpha \geq 2$ . . . 107
3.4.4 Histograms . . . 110
3.4.5 Efficiency . . . 112
3.5 Functional Model: Unknown Value of the tail index Study . . . 115
3.5.1 Analytical Comparison of Estimators . . . 115
3.5.2 Simulation results . . . 116

4 Two More Estimators of m **118**
4.1 Weighted Estimator . . . 118
4.1.1 Expected Wait for Next Update . . . 118
4.1.2 A New Estimator: Weighted Average Estimator . . . 122
4.1.3 Weighted Functional . . . 123
4.1.4 Simulation Study . . . 124

4.2 Estimating m, c_0 known . . . 133
4.2.1 Defining estimator and analysis . . . 133
4.2.2 Using the Markov Chain to Model Trajectory . . . 135
4.2.3 Simulation Study . . . 137

5 Meteorological Applications **143**
5.1 Reported Records: How Likely Are They? . . . 144
5.1.1 Expected Number of Reported Records . . . 144
5.1.2 The Number of Records in a Non-stationary time series . . . 148
5.2 Sea level data . . . 152
5.2.1 Sea level in The Netherlands . . . 152
5.2.2 Occurrences of Records . . . 155
5.3 Using CaterpillarSSA on Sea level Data . . . 159
5.3.1 Introduction to CaterpillarSSA . . . 159
5.3.2 Analysis . . . 161
5.3.3 Removing Trend and Periodicities From Mean Sea level in April in Harlingen . . . 166
5.3.4 Change Point . . . 177
5.4 Estimating Tail Index and Endpoint . . . 177
5.4.1 Methodology . . . 177
5.4.2 Results . . . 179

6 Conclusion **188**
6.1 Summary . . . 188
6.1.1 The Estimators . . . 190
6.1.2 Estimators when $k = 2$ . . . 191
6.1.3 Summary of Densities . . . 191
6.1.4 Efficiency . . . 193
6.1.5 Transition Densities . . . 194
6.1.6 Autocorrelation Function . . . 194
6.2 Conclusions . . . 195
6.2.1 Densities and Histograms . . . 195
6.2.2 Efficiency, Mean Square Error and Bias . . . 196
6.2.3 Wrong α . . . 198
6.2.4 Waiting times . . . 198
6.2.5 Simulating order statistics at record times . . . 199
6.2.6 The Weighted Estimator . . . 199
6.2.7 The estimator $m_{k,n}^{\times}$ . . . 200
6.2.8 Meteorological Applications . . . 201
6.2.9 Using CaterpillarSSA . . . 202

Chapter 1

Introduction

1.1 Summary of Thesis

1.1.1 Using Order Statistics to Estimate Endpoint

One of the main topics of the thesis is the estimation of the endpoint of a cumulative distribution function (c.d.f.), $F(x)$, using the k smallest order statistics of a sample drawn from $F(x)$. Here k is some positive integer, much smaller than the sample size. The lower endpoint, denoted m, is defined as $m = \inf\{a : F(a) > 0\}$. We consider estimators of m (denoted $\hat{m}$) that have been derived using extreme value theory. Throughout the majority of this thesis we consider the problem of estimating the lower endpoint of a distribution, however the problem of estimation of the upper endpoint (M) is an almost identical problem.

A full literature review will not be conducted here as [37] reviews extensively the subject of global random search, and in particular the estimation of the endpoint of a distribution using order statistics.

The problem of the estimation of the minimum or maximum of a function can be reduced (by pure random search) to the problem of the estimation of the endpoint of a c.d.f.. Being able to estimate the minimum or maximum of a function is of great

importance in many fields. For example, engineers building a structure may need to estimate the strongest wind that the structure will have to endure; manufacturers of the steel bars that make such a structure may need to estimate, from a sample of tested bars, the minimum force that can break these bars. Another major application is in meteorology, in particular the estimation of the maximum sea level that could occur at a particular location. This application is of vital importance when designing sea defences. A wealth of other applications can be found in most books on the subject; for example, [28], [13] and [2].

In applications very little may be known about the distribution whose endpoint is to be estimated, even in these cases the theory discussed in this thesis can be applied. We make assumptions on $F(x)$ that are easily met by most common distributions. In particular the assumptions are met when $F(x)$ is the distribution arising from pure random search (under some nonrestrictive conditions). We define pure random search (PRS) and show how the distribution $F(x)$ arises in PRS, shortly.

The assumptions on $F(x)$ are, that $F(x)$ can be approximated close to its endpoint by

$$F(x) = c_0(x-m)^\alpha + o((t-m)^\alpha), \quad x \downarrow m. \tag{1.1.1}$$

Here c_0 is a function of $v = 1/(x-m)$ that varies slowly at infinity as $v \to \infty$. In particular c_0 can be any positive constant. The value of m is the endpoint that we wish to estimate, we must have $m > -\infty$. Another important parameter is the tail index α. We must have $0 < \alpha < \infty$. In the majority of estimators that we consider throughout this thesis the tail index, α, will be assumed to be known. These estimates are defined in Sections 1.3 and 2.1.

All of the estimators of m in this thesis are based on taking a random sample of size n from the distribution $F(x)$. This sample is denoted $Y_n = \{y_1, \ldots, y_n\}$. If we sort this sample into ascending order and relabel, we obtain the order statistics $y_{1,n}, \ldots, y_{n,n}$, where $y_{1,n} \leq y_{2,n} \leq \ldots \leq y_{n,n}$. Clearly we could estimate m using the minimum order statistic, $y_{1,n}$. This statistic converges monotonically to m as

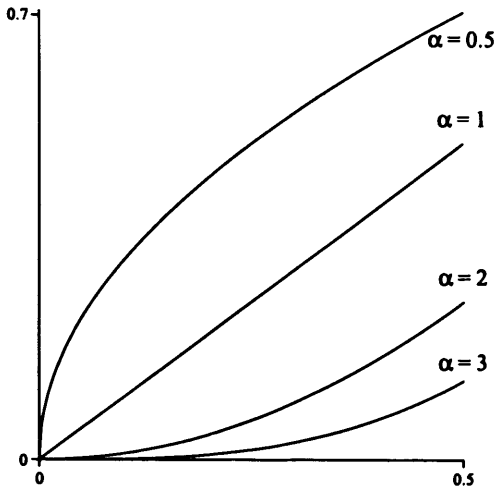


Figure 1.1: This figure shows $F(x) = c_0(x-m)^\alpha$ where $c_0 = 1$, $m = 0$ and $\alpha = 0.5$, 1, 2 and 3 (as labeled).

$n \to \infty$. However, it is a poor estimate, it has large bias and the convergence is extremely slow. Indeed, consider the following situation: Draw one random variable from $F(x)$ and labeled it $y_{1,1}$. Then draw random variables, y_i, $i = 2, 3, \ldots$, one at a time from $F(x)$. The expected value of n at which $y_n < y_{1,1}$ occurs is infinite.

In order to reduce bias (and mean square error) and improve convergence, we derive estimators that use the k smallest order statistics from the sample of size n. We do not use the entire sample for two main reasons. Firstly, the higher order statistics contain very little information about m. Secondly, we can only use the limit theorem for extreme order statistics if k is such that $k/n \to 0$ as $n \to \infty$. Discussions on the limiting behavior of the smallest k order statistics (extreme order statistics) can be found in Section 1.2. Theoretically the optimal choice of k is $k \to \infty$. In practice however, n will never be large enough to allow for very large k. [37] shows that small values of k (say $k = 5$) are almost as efficient as values of k twice or three times as large. We now describe a very important application of the estimates that we consider in Chapter 2: pure random search.

Applications to PRS

Pure random search is part of the wider class of optimization techniques, global random search (GRS). [35] and [37] are excellent reviews of GRS, they give an overview of a variety of algorithms and are a rich source of references for this subject. They point out that global random search is very popular with both practitioners and theoreticians, saying that this popularity is due to the following advantages. Global random search algorithms are often simple for practitioners to implement, as they are easily written as subroutines. Convergence can be guaranteed. Algorithms are successful under a wide variety of situations: when the objective function or feasible region is irregular, in the situation of a 'black box' objective function (where deterministic approaches almost certainly fail), when the dimension of the objective function is moderately high, or if the objective function cannot be evaluated without noise. For theoreticians global random search can be an appealing topic. Indeed, existing techniques often have simple structures that have provided researchers with interesting work. These existing techniques can be easily and intuitively extended, giving an abundance of theory to be investigated. The advantages listed above are in particular (and on some points, especially) true of PRS. We now formulate the PRS problem.

Let A be a feasible region over which an objective function $f : A \to \mathbb{R}$ is defined and let $m = \min_{x \in A} f(x)$ be its global minimum. We define the global minimizer to be the point $x^* \in A$ such that $f(x^*) = m$. We assume that; A is a compact subset of $\mathbb{R}^d$ for some $d \geq 1$; $\text{vol}(A) > 0$ (where $\text{vol}(\cdot)$ stands for 'volume'); $m > -\infty$; there is at least one global minimizer; and the objective function f is continuous in the neighborhood of this minimizer x^*. For simplicity we also assume that the objective function f is bounded from above. This last condition is made for technical reasons and can be relaxed. In order to apply the estimators found in this thesis we must assume that $f(x)$ can be evaluated without noise. [38] discusses methods, based on random search, that deal with estimating the maximum of a function in the case

where the function cannot be evaluated without noise.

The PRS algorithm can be described as follows. Let x_i, $i = 1, \ldots, n$, be random variables drawn from probability measure P, where P is defined on A. P must be such that there is a positive probability that a random point x_i will be in the vicinity of x^*.

Under these assumptions the set of n points $X_n = \{x_1, \ldots, x_n\}$ will be an independent and identically distributed random vector (i.i.d.r.v.). Let $Y_n = \{f(x_1), \ldots, f(x_n)\} = \{y_1, \ldots, y_n\}$ be the i.i.d.r.v. obtained by computing $f(\cdot)$ at the elements of the sample X_n. The values y_i will have common cumulative distribution function (c.d.f.)

$$F(t) = \mathrm{P}\left(x \in A \,:\, f(x) \leq t\right) = \int_{f(x) \leq t} P(dx)\,. \tag{1.1.2}$$

This c.d.f. is very important when studying the PRS algorithm. The minimum value of $f(\cdot)$ is the essential infimum of the r.v. η with the c.d.f. (1.1.2):

$$m = \min_{x \in A} f(x) = \operatorname{ess\,inf} \eta = \inf\{a : F(a) > 0\}\,.$$

Note that in PRS the choice of k has another important implication, this is discussed in [37]. Here it is pointed out that k must not be chosen to be too large. Indeed if there exist one or more sample points x_i such that $f(x_i) \leq y_{k,n}$ and x_i falls outside the vicinity of the global minimizer, it could lead to the over estimation of m (as inference may be made about a local minimum).

Comparison of Estimators

In Chapter 2 we compare two well known estimators of m and some simple estimators of m. The well known estimators are the maximum likelihood estimator [19], and the optimal linear estimator [8]. These are defined in Section 1.3. The density and efficiency of the estimators are the main points compared. We first do this under the assumption that the value of α is known. In this study we take $F(x)$ to be Weibull, beta and the c.d.f. derived from PRS, (1.1.2). A second comparison of

the estimators is undertaken where the value of α is unknown. We do not consider estimators of α here. We draw samples from a Weibull distribution (where the value of α is known) and then derive the estimators using the wrong value of tail index (ϑ). We study the efficiency of these estimators with respect to the optimal linear estimator using the correct value of tail index (see Sections 2.5.1 and 2.5.2).

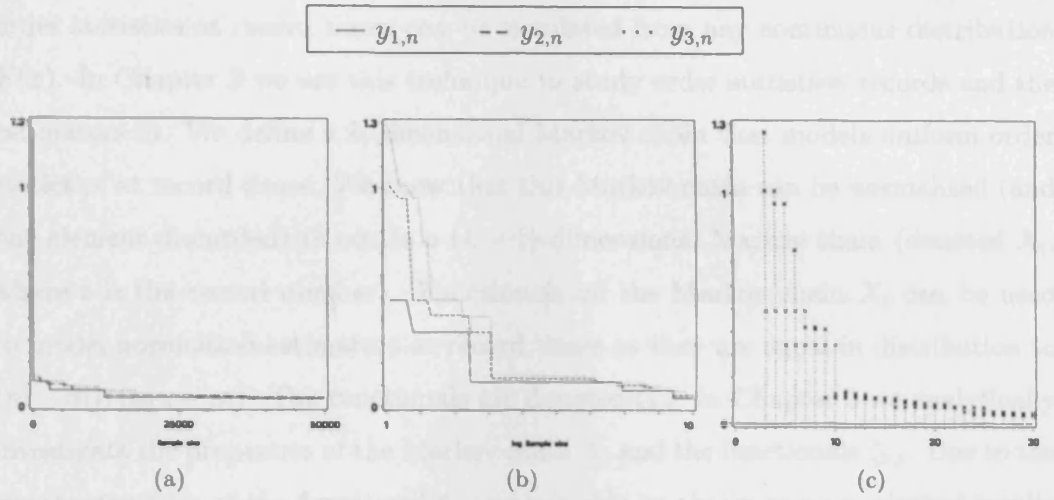
Choice of n

In the investigations in Chapter 2 sample size is fixed. In real-life situations the choice of n will usually depend on a balance between the level of accuracy demanded in the estimate of m and the resources available. n could be determined beforehand and set to some fixed number. This fixed number could be limited by time, computing power, cost, or the available data (for example in meteorological applications only a limited number of observations will be available). The other way to select n is by some stopping rule. The stopping rule may depend on an estimate of how close the smallest observation, ($y_{1,n}$), is to the minimum (for example [11]), or the expected number of random variables needed to improve an estimate of the minimum (as discussed in [37]). [34] discuss a variety of methods used to create stopping rules for global random search algorithms. In this thesis we investigate the frequency with which improvements to the estimator are expected. The estimators of m are based on the k smallest order statistics, as time increases the order statistics get closer to m, indeed, $y_{i,n} - m \leq y_{j,n} - m$ where $i < j$. When one or more of the k smallest order statistics moves closer to m, the estimates also usually, but not always, move closer to m (improve). It is therefore very important to be able to understand how the order statistics change with time.

1.1.2 Records, Poisson Processes and Markov chains

Figure 1.2 shows the three smallest order statistics related to a sample of size n, as n goes from 3 to 50 000. (a) shows the order statistics plotted against n. It can

be seen from this figure that order statistics change very infrequently, even at only moderately large sample sizes. This makes studying the behavior of the estimators through simulation very difficult. Figure (b) shows the same order statistics as (a) plotted against $\log(n)$. This change in time scale makes it much easier to see when each of the order statistics changes (updates). In fact it will be shown that the number of times that the k^{th} order statistic changes during the time that the sample size is increased from k to n, is approximately equal to $k\log(n)$ as $n \to \infty$. In both plots (a) and (b) a lot of repeated observations occur. We can discard these repeated observations and consider the order statistics just at the times where at least one of the order statistics changes. The order statistics that have changed are called record values. The new value of the i^{th} order statistic is called the (type 2) i^{th} record value. The sample size at which a record occurs is called a record time, and the number of records that have occurred at a particular sample size is called the record number. Together these statistics are called record statistics, they will be defined precisely in Sections 3.1.1, 3.1.2 and 3.1.3. Figure 1.2 (c) shows the record values plotted against record number. The k smallest order statistics at sample size n form a k-dimensional Markov chain whose members become closer to m as time increases. The record values also form a Markov chain, it is an embedded Markov chain of points visited by the Markov chain of order statistics. These two Markov chains have very different properties.



(a) (b) (c)

Figure 1.2: (a) shows the first three order statistics $y_{1,n}$, $y_{2,n}$ and $y_{3,n}$ from a sample whose size is increasing. (b) shows the same first three order statistics, $y_{1,n}$, $y_{2,n}$ and $y_{3,n}$, plotted against $\log n$. (c) shows the order statistics $y_{1,n}$, $y_{2,n}$ and $y_{3,n}$ at record times plotted against the record number.

There are many excellent and very readable articles and books dedicated to properties of order statistics and records, for example [1, 5, 10, 16, 26, 27, 31]. In Section 3.1 we extract definitions and properties from these sources that are useful in understanding the behavior of the estimators defined in Sections 1.3 and 2.1 as n increases. Many of the properties of records do not depend on the distribution $F(x)$, requiring only that $F(x)$ is continuous. One of the important properties of records is that asymptotically the occurrence of records follows a nonhomogeneous Poisson process with intensity $r\log(n)$, where r is a constant that depends on the precise definition of records. A homogeneous Poisson process with intensity r can be created by transforming the time scale. This can be seen in Figure 1.2 (b) and the figures in Section 3.1.5.

Another important property of order statistics that we make use of, is that the order statistics from a sample drawn from any continuous distribution $F(x)$ can be modeled using the order statistics from a uniform sample of the same size. It is easy to simulate uniform order statistics at record times (record values) and so

order statistics at record times can be simulated from any continuous distribution $F(x)$. In Chapter 3 we use this technique to study order statistics, records and the estimators $\hat{m}$. We define a k-dimensional Markov chain that models uniform order statistics at record times. We show that this Markov chain can be normalized (and one element discarded) to create a $(k-1)$-dimensional Markov chain (denoted X_t, where t is the record number). Functionals on the Markov chain X_t can be used to model normalized estimators at record times as they are equal in distribution to $(\hat{m}-m)/(y_{k,n}-m)$. The functionals are denoted $\zeta_{k,t}$. In Chapter 3 we analytically investigate the properties of the Markov chain X_t and the functionals $\zeta_{k,t}$. Due to the simple structure of the functional $\zeta_{k,t}$ we are able to obtain some analytical results more easily than would have been possible by considering the estimator $\hat{m}$. Next we use the functionals to carry out simulation studies similar to the ones carried out in Chapter 2. Throughout Chapter 3 we verify that the distribution of the functionals $\zeta_{k,t}$ is indeed equal to that of the normalized estimators $(\hat{m}-m)/(y_{k,n}-m)$.

In Chapter 4 we continue to consider the sample Y_n as a time series. Here we define two more estimators of m. We define and study the first one, called the weighted estimator (WE), in Section 4.1. It is based on the entire trajectory of an estimator as sample size increases. Indeed, as the sample Y_n increases from $n=k$ to $n=N$, the order statistics are recorded. At the record times the estimator $\hat{m}$ is calculated, the WE is a weighted average of these estimates. The weight depends on the difference between the previous record time and the current one (known as waiting time, see Section 3.1). In Section 4.1.1 we give some interesting properties of the conditional distributions and conditional expectations of waiting times of records. In Section 4.1.2 we define the WE and in Section 4.1.3 we define a functional on X_t that has the same distribution as the WE. This has been made possible by using a result from Section 4.1.1 that allows us to model the record times using record values. The simulation study in Section 4.1.4 (which relies heavily on the functional model) shows that the WE is not as efficient as the standard estimator

$\hat{m}$. We give reasons for the inefficiency of the WE and suggest similar estimators that may be more efficient.

The second estimator introduced in this chapter is an estimate based on the expected value of the k^{th} order statistic at sample size n. This estimator assumes that the value of parameter c_0 is known (the value of α is also assumed to be known). It is shown theoretically and through simulation that this extra knowledge produces dramatic improvements in efficiency for some distributions $F(x)$.

1.1.3 Meteorological Applications

In Chapter 5 we consider meteorological applications of the estimators of m and the record theory discussed in Section 3.1. First, in Section 5.1 we discuss the reporting of records in the media and show that extreme care must be taken when using record numbers to draw conclusions about whether, or by how much, the climate is changing. [14] proposes a test statistic based on the frequency of maximal and minimal records that is designed to test the alternative hypothesis that a time series contains a trend (a second test statistic is designed to test whether a time series has increasing/decreasing variance against the null hypothesis that variance is constant). We show how the number of records can be affected by a variety of transformations of a stationary time series.

In Section 5.2 we consider mean monthly sea level data from six locations in The Netherlands - Delfzijl, Harlingen, Den Helder, Ijmuiden, Maassluis and Hellevoetsluis. We calculate the expected number of records and estimate probability mass functions for the number of records that we would expect to see in such a time series if it were stationary. [3] considers the number of record highs and the number of record lows in 17 independent time series of mean monthly temperature worldwide. He found that the frequency of record highs was unexpectedly high compared with 17 stationary time series, and the number of record lows was unexpectedly low.

In Section 5.3 we attempt to forecast the maximum and minimum (mean monthly)

sea level in Harlingen. This kind of estimation is of vital importance for many coastal towns when considering flood risks. In reality it is the estimation of the maximum sea level that is required in order to determine say, seawall height. However, flood defenses are usually built with large safety factors [12] allowing for the under estimates that using a monthly average (rather than daily or hourly) data creates.

A major problem with estimating extreme maximal sea level is that sea level data does not form a stationary time series. It is generally thought that sea level is composed of three components; a mean sea level trend, tidal variation and variation associated with surges. The mean sea level reflects long term changes in global water levels and land height. The main tidal cycle has a period of approximately 12 hours and 26 minutes. This cycle has a modulating amplitude of 14 days: the spring/neap tide cycle. As we are considering monthly average data we would expect these variations to be almost entirely smoothed out. [12] suggest that for practical purposes a long-term cycle of 18.61 years can be expected, we may expect to see this cycle in our data. Variability due to surges is mainly a seasonal effect, thus has a 12 month period. This can be clearly seen in the data (see Section 5.4 especially). The amount that the three components affect the overall sea level is governed by location. For further reading on factors affecting sea level see [29]. Many methods can been applied to estimate extreme sea levels, such as annual maxima method [18], joint probability method [30, 33] and non-statistical methods that aim to model the physical processes.

In Section 5.3.1 we introduce the Singular Spectrum Analysis (SSA) algorithm and the program CaterpillarSSA. SSA is an algorithm that can be used to decompose a time series into the sum of a small number of independent and interpretable components such as a slowly varying trend, oscillatory components and a structureless noise [21]. The program CaterpillarSSA can be used to forecast the trend and oscillatory components and it hypothesized that the distribution of the noise can be made to meet the conditions of Theorem 1, and so allows the use of the estimators

defined in this thesis. The aim of Sections 5.3 and 5.4 is to use SSA to split the Harlingen time series into the two components; one, a forecastable time series consisting of trend and regular oscillation (the reconstruction); the other, a stationary time series of the remaining noise (the residual). We then can use the standard techniques already discussed in this thesis to make estimates of the upper and lower endpoints of this residual. Finally we add these estimates to the forecast (made from the reconstruction using CaterpillarSSA) to obtain forecasts of the endpoints of the distribution of monthly sea level.

In Section 5.3 we use different methods within the CaterpillarSSA framework to create reconstructions and residuals from the time series of mean monthly sea level in Harlingen. We aim to capture the trend and periodicities of the original time series in the reconstruction, and create a residual of structureless noise. We calculate the number of records in each of the residuals and asses whether it is likely that the residual is a stationary time series. In Section 5.4 we use separation methods from Section 5.3 and forecast the upper and lower endpoints of the mean monthly sea level for each month of three different years.

1.2 Results Concerning the Distributions of Extremes

The estimators that we will be considering in this thesis are semiparametric estimators derived by considering the distribution of the smallest members of the sample Y_n. Before we can define these estimators, we must give results concerning the distribution of these extreme statistics.

1.2.1 Asymptotic Distribution of the Minimum Order Statistic

Let the sample size n be fixed and $y_{1,n} \leq \ldots \leq y_{n,n}$ be the order statistics corresponding to the independent random sample Y_n. Here $y_{i,n}$ represents the i^{th} smallest

member within the sample $Y_n = \{y_1, \ldots, y_n\}$.

Consider the asymptotic distribution of the sequence of (normalized) minimum order statistics $y_{1,n}$, as $n \to \infty$. Generally, in the case $m = \operatorname{ess\,inf} \eta > -\infty$ (where the random variable η has c.d.f. $F(x)$) there are two limiting distributions possible. However, (as discussed earlier) [11] and [35] show us that in global random search applications we can assume (1.1.1), this means that the following theorem, Theorem 1, can be applied. Theorem 1 shows us only one asymptotic distribution can arise; specifically, the Weibull distribution with the c.d.f.

$$\Psi_\alpha(z) = \begin{cases} 0 & \text{for } z < 0 \\ 1 - \exp(-z^\alpha) & \text{for } z \geq 0\,. \end{cases} \tag{1.2.1}$$

This c.d.f. has only one parameter, α, which is called the 'tail index'. The mean of the Weibull distribution with tail index α is $\Gamma(1 + 1/\alpha)$; the density corresponding to the c.d.f. (1.2.1) is

$$\psi_\alpha(t) = (\Psi_\alpha(t))' = \alpha\, t^{\alpha-1} \exp(-t^\alpha)\,, \quad t > 0\,. \tag{1.2.2}$$

We now formulate this classical result from the theory of extreme order statistics. For proofs, discussions and generalizations see [10, 13, 15, 25, 27].

Theorem 1. Assume $\operatorname{ess\,inf} \eta = m > -\infty$, where η has c.d.f. $F(t)$, and the function

$$V(v) = F\left(m + \frac{1}{v}\right), \quad v > 0,$$

regularly varies at infinity with some exponent $(-\alpha)$, $0 < \alpha < \infty$; that is,

$$\lim_{v \to \infty} \frac{V(tv)}{V(v)} = t^{-\alpha}, \quad \text{for each } t > 0\,. \tag{1.2.3}$$

Then

$$\lim_{n \to \infty} F_{1,n}(m + (\kappa_n - m)z) = \Psi_\alpha(z)\,, \tag{1.2.4}$$

where $F_{1,n}$ is the c.d.f. of the minimum order statistics $y_{1,n}$, the c.d.f. $\Psi_\alpha(z)$ is defined in (1.2.1) and κ_n is the $(\frac{1}{n})$-quantile of $F(\cdot)$; that is, $\kappa_n = \inf\{u | F(u) \geq 1/n\}$.

The asymptotic relation (1.2.4) means that the distribution of the sequence of random variables $(y_{1,n} - m)/(\kappa_n - m)$ converges (as $n \to \infty$) to the random variable with c.d.f. $\Psi_\alpha(z)$.

The c.d.f. $\Psi_\alpha(z)$, along with its limiting case $\Psi_\infty(z) = \lim_{\alpha\to\infty} \Psi_\alpha(1 + z/\alpha)$ $= 1 - \exp(-\exp(z))$, $-\infty < z < \infty$, are the only nondegenerate limits of the c.d.f.'s of the sequences $(y_{1,n}-a_n)/b_n$, where $\{a_n\}$ and $\{b_n\}$ are arbitrary sequences of positive numbers.

If there exist numerical sequences $\{a_n\}$ and $\{b_n\}$ such that the c.d.f.'s of $(y_{1,n}-a_n)/b_n$ converge to Ψ_α, then we say that $F(\cdot)$ belongs to the domain of attraction of $\Psi_\alpha(\cdot)$ and express this as $F \in D(\Psi_\alpha)$. The conditions stated in Theorem 1 are necessary and sufficient for $F \in D(\Psi_\alpha)$. There are two conditions: $m = \operatorname{ess\,sup} \eta < \infty$ and the condition (1.2.3). The first one is easily met, in particular it is always valid in global random search applications. The condition (1.2.3) demands more attention. For example, it is never valid in discrete optimization problems as in these problems the c.d.f. $F(\cdot)$ is not continuous (whereas, (1.2.3) implies that $F(\cdot)$ has to be continuous in the vicinity of m). In fact, for a c.d.f. with a jump at its lower end-point no non-degenerate asymptotic distribution for $y_{1,n}$ exists, whatever the normalization (that is, sequences $\{a_n\}$ and $\{b_n\}$).

The condition (1.2.3) can be written as

$$F(t) = c_0(t - m)^\alpha + \mathrm{o}((t - m)^\alpha) \quad \text{as } t \downarrow m\,, \tag{1.2.5}$$

where c_0 is a function of $v = 1/(t - m)$, slowly varying at infinity as $v \to \infty$. In particular c_0 may be any positive constant, but the actual range of eligible functions is much wider.

Notice that the condition (1.2.5) can be met by many distributions, for example, by using a Taylor expansion. Therefore, Theorem 1 can be applied to the minimum order statistic from a wide class of distributions that are not derived as a result of global random search. Indeed many papers, including [8] and [9], use an equivalent assumption to justify their estimators of the endpoint of a distribution. The follow-

ing condition is stronger than the condition (1.2.5) and is often used for justifying properties of the maximum likelihood estimators (MLE):

$$F(t) = c_0(t-m)^\alpha\,(1+\ \mathrm{O}((t-m)^\gamma)) \quad \text{as } t \downarrow m \tag{1.2.6}$$

for some positive constants c_0, α and γ. This condition is sometimes called Hall's condition as it has appeared in Hall's paper [19] which was devoted to the MLE estimators of m.

1.2.2 Asymptotic Distribution of Extreme Order Statistics

Consider the distribution of order statistics $y_{k(n),n}$ as $n \to \infty$. Here $y_{k(n),n}$ is the k^{th} order statistics from a sample of size n where k is a function of n. If $k(n)$ is constant as $n \to \infty$ the order statistic $y_{k(n),n}$ is called the k^{th} extreme order statistic. Notice that the minimum order statistic is an example of an extreme order statistic. The asymptotic distribution of extreme order statistics when suitably centered and normalized is well known, for example see [27]. Indeed, under the same conditions as Theorem 1, the asymptotic distribution function of $y_{k,n}$ is given by:

$$\lim_{n\to\infty} F_{k,n}(a_n x - b_n) = 1 - \exp(-x^\alpha)\sum_{i=0}^{k-1}\frac{x^{\alpha i}}{i!},\ \ x > 0. \tag{1.2.7}$$

Here $F_{k,n}$ is the c.d.f. of the k^{th} order statistic from a sample of size n. The normalizing constants a_n and b_n are unchanged from those in Theorem 1.

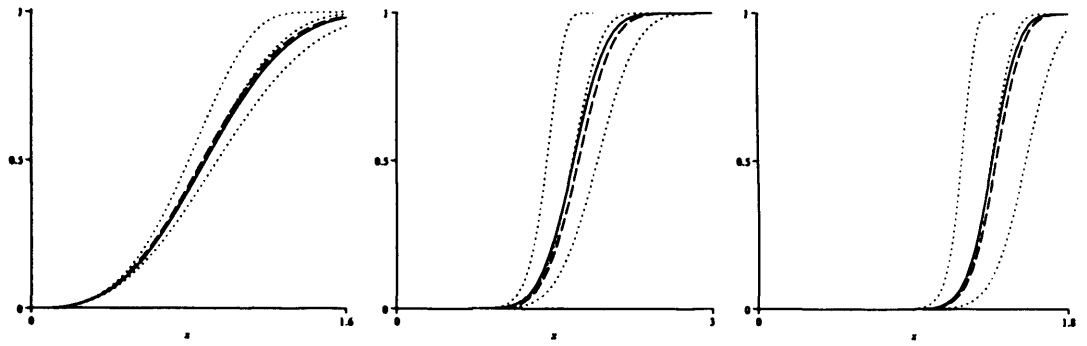
Consider the distribution of $y_{1,n}b_n + a_n = y_{1,n}(\kappa_n - m) + m$ where $y_{1,n}$ is the minimum order statistic from a sample of size n drawn from the Weibull distribution, m is the lower endpoint of the Weibull distribution ($m = 0$) and κ_n is the $\left(\frac{1}{n}\right)^{\text{th}}$ quantile of the Weibull distribution. This random variable has the following distribution:

$$F_{1,n}(x) = 1 - \left(\exp\left(-nx^\alpha \log\left(\frac{n}{n-1}\right)\right)\right),\ \ x > 0 \tag{1.2.8}$$

As $\log\left(\frac{n}{n-1}\right) = \frac{1}{n} + o(1/n^2)$ we can see that (1.2.8) can be written as

$$1 - \exp(-x^\alpha)\exp(-x^\alpha/2n)\exp(-x^\alpha/3n^2)\ldots$$

and hence converges very quickly to (1.2.7) as $n \to \infty$.



(a) $k = 1, \alpha = 3, n = 10$ (b) $k = 5, \alpha = 3, n = 20$ (c) $k = 5, \alpha = 5, n = 20$

Figure 1.3: Figure show the distribution of $y_{k,n}(\kappa_n - m) + m$ where $y_{k,n}$ is the k^{th} order statistic from the Weibull distribution (dashed line) and the beta distribution (dotted lines). The asymptotic distribution (1.2.7) is plotted with a solid line. The distribution of the minimum order statistic from the beta distribution is plotted with three different values for the parameter β: $\beta = 0.5, 1$ and 2. The highest dashed line on the graph corresponds to $\beta = 0.5$, the next is $\beta = 1$ and finally $\beta = 2$. The tail index of all of the distributions in (a) and (b) is $\alpha = 3$ and the sample size in (a) is $n = 10$ and in (b) is $n = 20$. In (c) the tail index is $\alpha = 5$ and $n = 20$.

Figure 1.3 shows the cumulative distribution function (c.d.f.) of $y_{k,n}(\kappa_n - m) + m$ where $y_{k,n}$ is the k^{th} order statistic from a sample of n Weibull random variables and κ_n is the $\left(\frac{1}{n}\right)^{\text{th}}$ quantile of the Weibull distribution. It also shows the c.d.f.s of $y_{k,n}(\kappa_n - m) + m$ where $y_{k,n}$ is the k^{th} order statistic from a sample of n beta random variables (with various parameters) and κ_n is the $\frac{1}{n}^{\text{th}}$ quantile of the same beta distributions. Also plotted is the asymptotic distribution (1.2.7). The beta c.d.f. is given by

$$F_\beta(t; \alpha, \beta) = \frac{B_t(\alpha, \beta)}{B(\alpha, \beta)}$$

where $B_t(\alpha, \beta)$ is the incomplete beta function $\int_0^t x^{\alpha-1}(1-x)^{\beta-1}dx$ and $B(\alpha, \beta)$ is the beta function. We can approximate this c.d.f. by (1.2.5) where $c_0 = \frac{1}{B(\alpha,\beta)\alpha}$ and the tail index is equal to α and $m = 0$. By setting $\beta = 1$ we have that $c_0 = 1$. Then the beta distribution is exactly equal to $F(x) = c_0(t - m)^\alpha$ (see (1.2.5)). We can see from this figure that the minimal order statistics from all of the distributions

converge very quickly to the asymptotic distribution (1.2.7). The 5^{th} order statistics from the distributions $F(x;\alpha,0.5)$ and $F(x;\alpha,2)$ are not very close to the asymptotic distribution when the sample size is 20. The distributions of the Weibull extreme order statistics and the beta extreme order statistics when $\beta = 1$ are almost exactly equal to the asymptotic one.

1.2.3 Asymptotic Behavior of $\kappa_n - m$

The quantity $\kappa_n - m$, where $m = \operatorname{ess\,inf} \eta$ and κ_n is the $\left(\frac{1}{n}\right)^{\text{th}}$-quantile of $F(\cdot)$, enters many formulae below and therefore its asymptotic behaviour is very important. Fortunately, the asymptotic behaviour of $\kappa_n - m$ is clear. Note that since the objective function $f(\cdot)$ is continuous in a neighborhood of x^*, the c.d.f. $F(\cdot)$ is continuous in the vicinity of m and for n large enough we have $F(\kappa_n) = 1/n$. Indeed, as long as (1.2.5) holds with some c_0, we have

$$\frac{1}{n} = F(\kappa_n) \sim c_0 (\kappa_n - m)^{\alpha} \quad \text{as } n \to \infty$$

implying

$$(\kappa_n - m) \sim (c_0 n)^{-1/\alpha} \quad \text{as } n \to \infty . \tag{1.2.9}$$

We say $a_n \sim b_n$ for $b_n \neq 0$, a_n and b_n random variables if we have $\frac{a_n}{b_n} \to 1$ in distribution.

1.3 Defining the Estimators

Let $y_{1,n} \leq \ldots \leq y_{n,n}$ be the order statistics corresponding to the sample Y_n, drawn from a c.d.f. that meets the conditions of Theorem 1. Here we will assume that the value of the tail index, α, is known. The consequences of using the wrong value of the tail index is dealt within Section 2.5, although estimators of α will not be. Most estimates of m where α is unknown estimate α and then substitute this estimate into $\hat{m}$ in place of the exact value of α, see [19, 22, 36, 11].

In the current section, following the exposition of [35], we review existing estimators of m. Two of these estimators will be thoroughly defined and used in the remainder of this thesis (along with three more, simple estimators that will be defined in Chapter 2). Note that there are some other estimators available; we only consider the estimators which we believe are the most important ones.

1.3.1 Linear Estimators

A general linear estimator of m can be written as

$$\hat{m}_{k,n}(a) = \sum_{i=1}^{k} a_i y_{i,n} ,$$

where $a = (a_1, \ldots, a_k)' \in \mathbb{R}^k$ is a vector of coefficients. Linear estimators first appeared in this general form in [7]. It can be shown (see [35], Section 7.3) that as $n \to \infty$ we have

$$E\hat{m}_{k,n}(a) = m \sum_{i=1}^{k} a_i - (\kappa_n - m) a'b + \ o(\kappa_n - m) = m \sum_{i=1}^{k} a_i + \ o(1) . \qquad (1.3.1)$$

Here $b = (b_1, \ldots b_k)' \in \mathbb{R}^k$, where $b_i = \Gamma(i + 1/\alpha) \, / \, \Gamma(i)$.

From (1.3.1) it is clear that (as the objective function f is bounded and therefore the variances of all $y_{i,n}$ are finite), a necessary and sufficient condition for an estimator with vector of coefficients a to be consistent is:

$$\sum_{i=1}^{k} a_i = 1 . \qquad (1.3.2)$$

This consistency condition can be found in [37]. The main property that we wish to impose on estimators is a small mean square error (MSE). The MSE of a linear estimator is given by

$$MSE(\hat{m}_{k,n}) = E(\hat{m}_{k,n}(a) - m)^2 \sim (\kappa_n - m)^2 \, a' \Lambda a, \ \ n \to \infty , \qquad (1.3.3)$$

where $\Lambda = \|\lambda_{ij}\|_{i,j=1}^{k}$ is a symmetric $k \times k$-matrix with elements λ_{ij} defined for $i \geq j$ by the formula

$$\lambda_{ij} = \frac{\Gamma(i+2/\alpha) \, \Gamma(j+1/\alpha)}{\Gamma(i+1/\alpha) \, \Gamma(j)} . \qquad (1.3.4)$$

The (non-asymptotic) mean square error of a linear estimator of m from the distribution $F(x) = c_0(x-m)^\alpha$, $m \le x \le M$, is given by

$$\begin{aligned} \mathrm{E}(\hat{m}_{k,n} - m)^2 &= \mathrm{E}\left(\sum_{i=1}^{k} a_i y_{i,n} - m\right)^2 \\ &= \sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j \mathrm{E}(y_{i,n} y_{j,n}) \\ &= \frac{\Gamma(n+1)}{\Gamma(n+1+2/\alpha)} a'\Lambda a. \end{aligned} \tag{1.3.5}$$

Here $M = \left(\frac{1}{c_0}\right)^{1/\alpha} + m$ is the upper endpoint of $F(x)$. (1.3.5) can be calculated using the following well known expression for the joint density of two order statistics, $y_{i,n}$ and $y_{j,n}$, from a sample of size n drawn from some general distribution $F(x)$ (see for example [27]);

$$p(x,y) = \frac{n!}{(i-1)!(j-i-1)!(n+1-j-1)!}(F(x))^{i-1}(F(y)-F(x))^{j-i-1} \times$$

$$(1-F(y))^{n-j} f(x) f(y). \tag{1.3.6}$$

The expression (1.3.6) allows us to calculate $\mathrm{E}(y_{i,n}y_{j,n})$ as

$$\mathrm{E}(y_{i,n}y_{j,n}) = \frac{\Gamma(n+1)\Gamma(i+2/\alpha)\Gamma(j+1/\alpha)}{\Gamma(n+1+2/\alpha)\Gamma(i+1/\alpha)\Gamma(j)}, \quad i<j.$$

The above density is valid for $0 < i < j \le n$ and $m \le x < y \le M$ where m and M are the lower and upper endpoints of the distribution $F(x)$. Here $f(x)$ is the density related to the c.d.f. $F(x)$.

We can impose an unbiasedness condition on linear estimators. Indeed, using (1.3.1) we have that,

$$m - \mathrm{E}\hat{m}_{k,n}(a) \sim (\kappa_n - m)a'b, \quad n \to \infty.$$

It follows from this that an unbiasedness criteria is $a'b = 0$.

For the following estimator we choose the r.h.s. of (1.3.3) as our optimality criterion for selecting a. To ensure that the estimator is consistent, we also impose the condition (1.3.2).

Optimal Linear Estimator

The optimal consistent linear estimator $m^\circ = \hat{m}_{k,n}(a^\circ)$, we shall call it *the optimal linear estimator*, is determined by the vector of coefficients

$$a^\circ = \arg \min_{a:a'\,\mathbf{1}=1} a'\Lambda a = \frac{\Lambda^{-1}\,\mathbf{1}}{\mathbf{1}'\Lambda^{-1}\,\mathbf{1}}\,. \tag{1.3.7}$$

The estimator m° has been suggested in [8], where the form (1.3.7) for the vector of coefficients was obtained. Solving the quadratic programming problem in (1.3.7) is straightforward. In the process of doing that, we obtain

$$\min_{a:a'\,\mathbf{1}=1} a'\Lambda a = (a^\circ)'\Lambda a^\circ = 1/\;\mathbf{1}'\Lambda^{-1}\,\mathbf{1}\,. \tag{1.3.8}$$

Lemma 7.3.4 in [35] gives the following expression for the r.h.s. of (1.3.8):

$$\mathbf{1}'\Lambda^{-1}\,\mathbf{1} \;=\; \begin{cases} \frac{1}{\alpha-2}\left(\frac{\alpha\Gamma(k+1)}{\Gamma(k+2/\alpha)} - \frac{2}{\Gamma(1+2/\alpha)}\right) & \text{for } \alpha \neq 2, \\ \sum_{i=1}^{k} 1/i & \text{for } \alpha = 2\,; \end{cases}$$

this expression is valid for all $\alpha > 0$ and $k = 1, 2, \ldots$

The components a_i° $(i = 1, \ldots, k)$ of the vector a° can be evaluated explicitly:

$$a_i^\circ = u_i/\;\mathbf{1}'\Lambda^{-1}\,\mathbf{1} \quad \text{for } i = 1, \ldots, k \tag{1.3.9}$$

with

$$\begin{aligned} u_1 &= (\alpha+1)\,/\,\Gamma(1+2/\alpha), \\ u_i &= (\alpha-1)\,\Gamma(i)/\Gamma(i+2/\alpha) \qquad \text{for } i = 2, \ldots, k-1, \\ u_k &= -(\alpha k-\alpha+1)\Gamma(k)\,/\,\Gamma(k+2/\alpha). \end{aligned}$$

Deriving this expression for the coefficients of the vector a° is far from trivial, see [35], Section 7.3.3.

If the conditions (1.2.6), $\alpha \geq 2$, $k \to \infty$, $k/n \to 0$ (as $n \to \infty$) are satisfied then the optimal linear estimators of m are asymptotically normal and asymptotically efficient in the class of asymptotically normal estimators and their mean square error $E(\hat{m} - m)^2$ is asymptotically (as $n \to \infty$)

$$E(\hat{m} - m)^2 \sim \begin{cases} (1 - \frac{2}{\alpha})(\kappa_n - m)^2\, k^{-1+2/\alpha} & \text{for } \alpha > 2, \\ (\kappa_n - m)^2 \log k & \text{for } \alpha = 2\,. \end{cases} \tag{1.3.10}$$

1.3.2 Other Linear Estimators

Hall [19] uses (1.3.3) along with the consistency and unbiasedness criteria, to define the linear estimator $m^+ = m^+_{k,n}(a) = \hat{m}_{k,n}(a^+)$, with vector of coefficients

$$a^+ = \arg \min_{a:\, a'\mathbf{1}=1, a'b=0} a'\Lambda a.$$

Here $\mathbf{1}$ is the k-dimensional vector $\mathbf{1} = (1, \ldots, 1)$. The estimator m^+ was investigated by [36] and [20]. [36] shows that asymptotically $\text{MSE}(m^+) \sim \text{MSE}(m^\circ)$ as $k \to \infty$. [20] showed that for finite samples m^+ has larger bias than m°.

[9] investigated a linear estimator that was designed to have the same properties as Hall's maximum likelihood estimator (defined in Section 1.3.3 below). Unlike Hall's maximum likelihood estimator that has no closed form (except when $k = 2$), Csörgő and Mason's linear estimator, m^{CM}, can be written in explicit form. This estimator can be found in [9]. Here it is shown to be strongly consistent and it is shown that asymptotically, $\text{MSE}(m^{CM}) \sim \text{MSE}(m^\circ)$ as $k \to \infty$, where m^{CM} is the linear estimator defined in [9]. [20] found that for finite samples and small α, m^{CM} has large MSE.

1.3.3 Maximum Likelihood Estimator

In defining the maximum likelihood estimator (MLE) we have to assume the condition (1.2.6) which is stronger than (1.2.5). Additionally, we have to assume $\alpha \geq 2$.

Taking the asymptotic form of the likelihood function as exact (for details see [19] and [35], Chapter 7), we obtain that the maximum likelihood estimator of the minimum, m^*, is the smallest solution z to the following likelihood equation:

$$(\alpha - 1) \sum_{j=1}^{k-1} \frac{(y_{k,n} - y_{j,n})}{(y_{j,n} - z)} = k \tag{1.3.11}$$

conditionally $z < y_{1,n}$; if there are no solutions to this equation for $z < y_{1,n}$, then we set $m^* = y_{1,n}$; if there is more than one solution in the region $z \in (-\infty, y_{1,n})$, then we take the smallest of these solutions.

The asymptotic properties of the maximum likelihood estimators coincide with the properties of the optimal linear estimators and hold under the same regularity conditions (we again refer to [35], Section 7.3.3). The maximum likelihood estimators $m^* = \hat{m}^*_{k,n}$ of m are asymptotically normal and their MSE $E(m^* - m)^2$ asymptotically behaves like the r.h.s. of (1.3.10). Unlike the optimal linear estimator, this estimator is not defined for $\alpha < 2$ (m° is defined for all $\alpha > 0$). From [20] it can be seen that the MLE behaves poorly for small α (close to, but larger than 2), whereas the optimal linear estimator behaves well for small α.

It should be noted that for many distributions, $F(x)$, the likelihood function for the estimation of m has a discontinuity at the point m. This irregularity is discussed by [6]. In particular the case where $F(x)$ is the Weibull distribution is considered. In cases where a discontinuity of the likelihood function occurs, the maximum likelihood estimator () is a local maximum of the likelihood function. This is discussed in [32], who show that when $\alpha > 2$ the MLE is consistent.

Chapter 2

Comparison of Estimators: Static Sample

2.1 Defining More Estimators

In this section we define a further three linear estimators. These estimators use the same optimality criterion as the optimal linear estimator, but have additional constraints imposed upon them. Each of them minimize the MSE given by the r.h.s. of (1.3.3), subject to the condition that a satisfies (1.3.2). Additionally we impose constraints on the number (and for the latter two estimators, position) of elements in the vector of coefficients, a, free to be non-zero.

2.1.1 One Non-Zero Coefficient

The first estimator that we define in this section is the optimal consistent linear estimator with one coefficient free to be nonzero. The remaining $(k-1)$ elements must be equal to zero. This estimator denoted, $m^\bullet = \hat{m}(a^\bullet)$, is simply the minimum order statistic, and hence is determined by the following vector of coefficients:

$$a^{\bullet\prime} = (1, 0, \ldots, 0).$$

This estimator is optimal as it has a lower MSE than any other estimator with one nonzero coefficient; $\hat{m} = a_i y_{i,n}$, where $i \neq 1$. The estimator $m^\bullet$ does not depend on α or k and so is useful for comparison.

2.1.2 Two Non-Zero Coefficients

We now define an estimator with two coefficients in a free to be non-zero, $\hat{m}(a^{(2)}) = m^{(2)}$. The MSE of this estimator is given by:

$$(\kappa_n - m)^2((a_i^{(2)})^2\lambda_{ii} + 2a_i^{(2)}a_j^{(2)}\lambda_{ij} + (a_j^{(2)})^2\lambda_{jj})$$

where $1 \leq i < j \leq k$ and i and j are integers denoting the positions of the two non-zero coefficients within a. Although it is not optimal, for simplicity it has been decided that $i = 1$ and $j = k$. It can be seen below that this estimator is not the most efficient estimator in the wider class of all estimators with two coefficients free to be nonzero, i.e. the class with no restrictions on position of the nonzero elements. Due to the consistency constraint (1.3.2), $a_1^{(2)} + a_k^{(2)} = 1$, so letting $C_k = -a_k^{(2)}$ we can rewrite the MSE above as

$$(\kappa_n - m)^2((1 + C_k)^2\lambda_{11} - 2C_k(1 + C_k)\lambda_{k1} + C_k^2\lambda_{kk}) \tag{2.1.1}$$

We can minimize this MSE to obtain that the MSE of the estimator $m^{(2)}$ is given by:

$$\frac{(\kappa_n - m)^2}{\mathbf{1}'_{(2)}\Lambda_{(2)}^{-1}\mathbf{1}_{(2)}}$$

where $\mathbf{1}_{(2)} = (1, 1)'$ and

$$\Lambda_{(2)} = \begin{pmatrix} \lambda_{11} & \lambda_{k1} \\ \lambda_{k1} & \lambda_{kk} \end{pmatrix}. \tag{2.1.2}$$

In the process we find that the value of C_k that achieves this minimum is

$$C_k = \frac{\lambda_{k1} - \lambda_{11}}{\lambda_{11} - 2\lambda_{k1} + \lambda_{kk}}. \tag{2.1.3}$$

It can be seen that $a^\circ = a^{(2)}$ when $k = 2$ or when $\alpha = 1$.

A Wider Class of Estimators - The Choice of i

Figure 2.1 shows that $m^{(2)}$ is not the most efficient estimator in the class of consistent linear estimators with two non-zero coefficients. We define an estimator with two

non-zero coefficients, $m^{\Delta_i} = (1 - a_k^{\Delta_i})y_{i,n} + a_k^{\Delta_i} y_{k,n}$. Here i is free to be any positive integer less than or equal to k and $a_k^{\Delta_i}$ is chosen to minimize the MSE. Notice that $m^{\Delta_1} = m^{(2)}$.

It can easily be shown that the MSE of this estimator is given by

$$\text{MSE}(m^{\Delta_i}) = \frac{(\kappa_n - m)^2}{\mathbf{1}'\Lambda_{\Delta_i}^{-1}\mathbf{1}}.$$

Here

$$\Lambda_{\Delta_i} = \begin{pmatrix} \lambda_{ii} & \lambda_{ki} \\ \lambda_{ki} & \lambda_{kk} \end{pmatrix}.$$

In finding this MSE we obtain that the coefficient $a_k^{\Delta_i}$ is given by:

$$\frac{\lambda_{i,i} - \lambda_{k,i}}{\lambda_{i,i} + \lambda_{k,k} - 2\lambda_{k,i}}$$

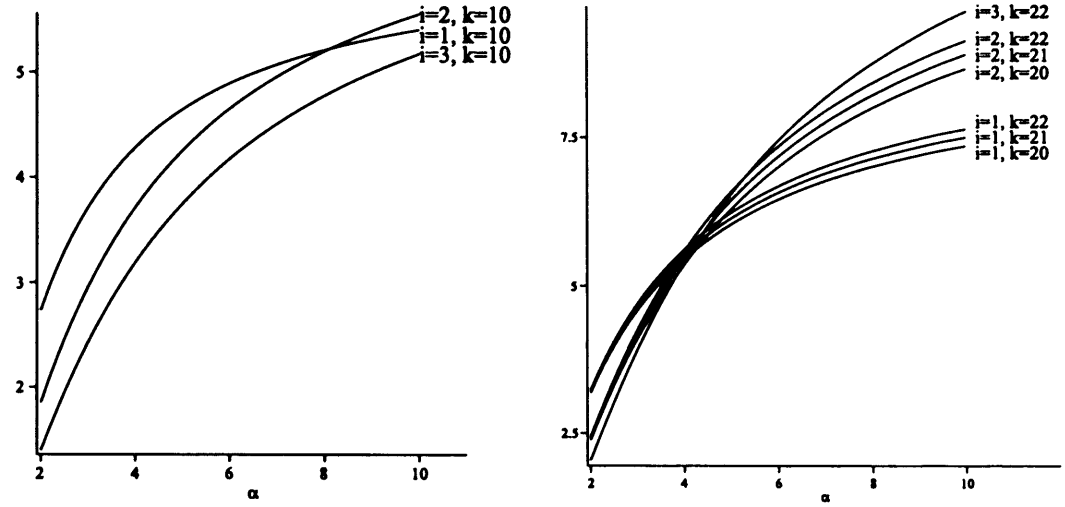
Figure 2.1 shows $(\kappa_n - m)^2/\text{MSE}(m^{\Delta_i})$ plotted against α for different values of i. The figure shows that when either α or k (or both) are large, m^{Δ_2} has a smaller MSE than $m^{\Delta_1} = m^{(2)}$.

A Wider Class of Estimators - The Choice of k

From Figure 2.1 it also appear that for any particular i or α, as k increases, $\text{MSE}/(\kappa_n - m)^2$ decreases. This means that there is no need to consider an estimator made with any two order statistics less than or equal to $y_{k,n}$: the estimator with smallest MSE will always be made with $y_{k,n}$. This does not mean that taking k to equal n would actually result in an estimator with very small MSE, as an assumption made in the derivation of this theoretical MSE is that $k/n \to 0$. Notice also that the estimator based on k order statistics, $\hat{m} = a_i y_{i,n} + a_j y_{j,n}$ where $i < j < k$, is the same as the estimator m^{Δ_i} based on j order statistics.

2.1.3 Three Non-Zero Coefficients

The vector of coefficients of our final estimator, $m^{(3)} = \hat{m}_{k,n}(a^{(3)})$, has at most three non-zero elements. For simplicity we set the three coefficients free to be non-zero



(a) $k = 10$ (b) $k = 20, 21, 22$

Figure 2.1: The plots above show $(\kappa_n - m)^2/\text{MSE} = 1'\Lambda_{\Delta_i}1$ against α. In (a), $k = 10$ and in (b), $k = 20$, 21, 22 for both i takes values $i = 1$, 2, 3.

as a_1, a_i and a_k where $1 < i < k$. The position, i, of the middle coefficient is free and therefore is chosen to minimize MSE. Due to the constraint (1.3.2) we can write $a_i^{(3)} = 1 - a_1^{(3)} - a_k^{(3)}$. This means $m^{(3)}$ is defined to be equal to

$$a_1^{(3)} y_{1,n} + (1 - a_1^{(3)} - a_k^{(3)}) y_{i,n} + a_k^{(3)} y_{k,n}$$

where $a_1^{(3)}$, $a_k^{(3)}$ and i are chosen to minimize the MSE of $m^{(3)}$ which is given by:

$$(\kappa_n - m)^2 (a_1^{(3)},\ 1 - a_1^{(3)} - a_k^{(3)},\ a_k^{(3)}) \begin{pmatrix} \lambda_{11} & \lambda_{i1} & \lambda_{k1} \\ \lambda_{i1} & \lambda_{ii} & \lambda_{ki} \\ \lambda_{k1} & \lambda_{ki} & \lambda_{kk} \end{pmatrix} \begin{pmatrix} a_1^{(3)} \\ 1 - a_1^{(3)} - a_k^{(3)} \\ a_k^{(3)} \end{pmatrix} \tag{2.1.4}$$

Table 2.1 and Figure 2.2, below, demonstrate the value of i that minimizes (2.1.4) for any particular choice of k and α. The graphs in Figure 2.2 show the efficiency with respect to m° of the estimator

$$m^{\circ_i} = a_1^{\circ_i} y_{1,n} + (1 - a_1^{\circ_i} - a_k^{\circ_i}) y_{i,n} + a_k^{\circ_i} y_{k,n},$$

where the values of $a_1^{\diamond i}$ and $a_k^{\diamond i}$ are chosen to minimize the MSE for the particular choice of α, k and i. Here efficiency is defined to be

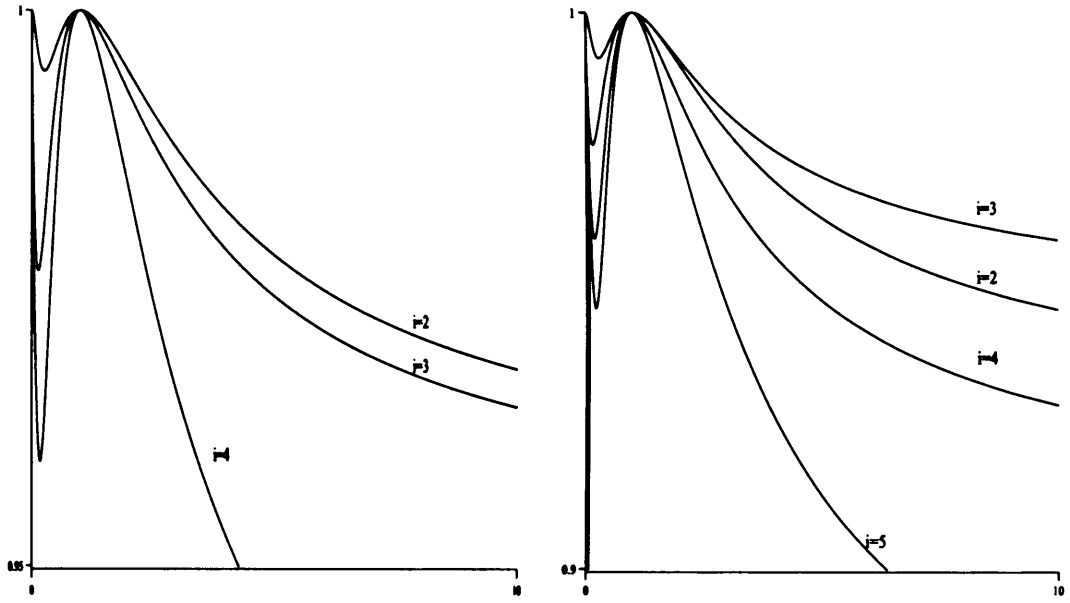
$$\frac{\mathrm{E}(m^\circ - m)^2}{\mathrm{E}(m^{\diamond i} - m)^2}.$$

The estimator $m^{\diamond i}$, $i \in 2, \ldots, k-1$ that exhibits the highest efficiency for given value of k and α coincides with the estimator $m^{(3)}$, therefore the position of the middle non-zero coefficient in $m^{(3)}$ can easily be seen in the graphs.

k	$\alpha = 2$				$\alpha = 3$				$\alpha = 5$			
	i	$a_1^{(3)}$	$a_k^{(3)}$	$\frac{MSE}{(\kappa_n - m)^2}$	i	$a_1^{(3)}$	$a_k^{(3)}$	$\frac{MSE}{(\kappa_n - m)^2}$	i	$a_1^{(3)}$	$a_k^{(3)}$	$\frac{MSE}{(\kappa_n - m)^2}$
3	2	1.64	-0.91	0.55	2	1.95	-1.54	0.44	2	2.60	-2.84	0.38
4	2	1.44	-0.75	0.48	2	1.66	-0.30	0.38	2	2.14	-2.45	0.32
5	2	1.32	-0.65	0.44	2	1.48	-1.15	0.33	2	1.85	-2.20	0.27
6	3	1.31	-0.65	0.41	3	1.48	-1.24	0.31	3	1.86	-2.48	0.24
7	3	1.24	-0.59	0.39	3	1.37	-1.14	0.28	3	1.68	-2.30	0.22
8	3	1.19	-0.55	0.37	3	1.29	-1.06	0.27	3	1.55	-2.15	0.20
9	3	1.14	-0.51	0.36	3	1.22	-1.00	0.25	3	1.45	-2.04	0.19
10	3	1.10	-0.48	0.35	3	1.16	-0.94	0.24	4	1.48	-2.26	0.18

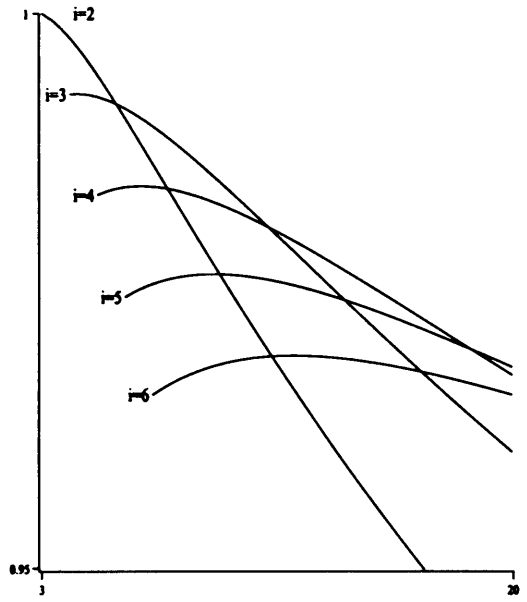
Table 2.1: A table showing the value of i used in $m^{(3)}$, the values of $a_1^{(3)}$ and $a_k^{(3)}$ ($a_i^{(3)} = 1 - a_1^{(3)} - a_k^{(3)}$). Also shown is the normalised theoretical MSE of the estimator for various k and α.

Let i^* be the value of i that minimizes the MSE of $m^{\diamond i}$ for any particular choice of α and k. For $\alpha = 2$, 3, 5, 10 and $k \leq 5$, $i^* = 2$. Generally as k increases i^* also increases. When $k = 3$, the estimator $m^{(3)}$ coincides with the estimator m° and so clearly $i^* = 2$ and efficiency is equal to one. When $\alpha = 1$, $a^{(3)}$ has just two non-zero coefficients and coincides with both a° and $a^{(2)}$. When $\alpha = 1$, $a_i^\circ = 0$ ($1 < i < k$), this is why at $\alpha = 1$ in Figures 2.2(a) and 2.2(b) the efficiency is equal to 1 for all values of k and i. It can be seen from both the graphs and the table that as k increases the efficiency of $m^{(3)}$ decreases. This is because the number of

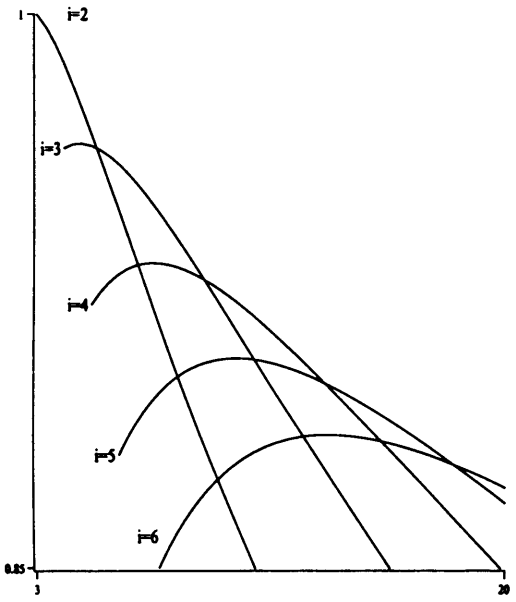


(a) $k = 5$

(b) $k = 6$



(c) $\alpha = 2$



(d) $\alpha = 10$

Figure 2.2: Efficiency with respect to m° of estimator m°_i}. In (a) and (b) this is plotted as a function of α and in (c) and (d) this is plotted as a function of k.

order statistics being used in the optimal linear estimator is increasing, whereas the

number of order statistics in $m^{(3)}$ does not increase. For $\alpha > 1$ the efficiency of $m^{(3)}$ decreases as α increases.

2.2 Density of Normalized Estimators

In this chapter we derive the densities of the estimators defined above, under particular conditions.

First, in Section 2.2.2, we find joint densities of normalized linear estimators and the order statistics that they are made up of. We explain how this joint density may be used to find the univariate density of the normalized estimator. Throughout this section we consider four different normalizations of our estimators (see below). The densities found at this stage can be applied to any consistent linear estimator based on k order statistics, where k is any integer greater than 1. Here the c.d.f. of the random variables y_i is any valid c.d.f..

In Section 2.2.3 we consider the joint densities found in Section 2.2.2, however now we look at the specific case of $F(\cdot)$ being Weibull and the linear estimator being $m^{(2)}$, again considering all four normalizations.

In Section 2.2.4, under the same conditions as in Section 2.2.3, we concentrate on just two different normalizations. These two normalized estimators have univariate densities that can be written in a simple form. The asymptotic form (as $n \to \infty$) of these densities has also been found.

Section 2.2.5 considers the density of m^* (normalized) when $k = 2$ and $F(\cdot)$ is Weibull. All four normalizations are considered in this section.

Considering the same two normalizations from Section 2.2.4 and again letting $F(\cdot)$ be the c.d.f. of the Weibull distribution and $k = 2$, in section 2.2.6, we find both the exact and the asymptotic form of the density of the normalized estimator m^*.

Finally, in Section 2.3 we study the densities found in Sections 2.2.4 and 2.2.6. We compare the densities of m° and m^* (normalized) and investigate the effect that

sample size has on the distributions.

In Section 2.4 we investigate the estimators described above for finite samples of size n by plotting these densities on histograms of the corresponding estimators.

2.2.1 Method for Deriving the Densities

Four Normalizations

Throughout this section and Sections 2.3 and 2.4, we consider four ways of normalizing the estimators defined above.

The normalizations are of the form $(\hat{m} - m)/\omega$:

(1) $(\hat{m} - m)/\omega_1 = (nc_0)^{1/\alpha}(\hat{m} - m)$

(2) $(\hat{m} - m)/\omega_2 = \eta(\hat{m} - m)/(y_{1,n} - m)$

(3) $(\hat{m} - m)/\omega_3 = (\hat{m} - m)/(y_{1,n} - m)$

(4) $(\hat{m} - m)/\omega_4 = (\hat{m} - m)/(y_{k,n} - m)$

Here η is an independent Weibull random variable.

In order to find the densities of the normalized estimators (1), (3) and (4), we can make a transformation on the (k-dimensional) joint density of the smallest k order statistics (a well known result) to obtain the (k-dimensional) joint density of the $2^{\text{nd}}, \ldots, k^{\text{th}}$ order statistics with the normalized estimator. This joint density can be integrated with respect to the $2^{\text{nd}}, \ldots, k^{\text{th}}$ order statistics to obtain the univariate density of the normalized linear estimator. The density of the normalized estimator (2) is very similar except the joint densities described above must be $(k+1)$-dimensional in order to include the independent (of the order statistics) Weibull random variable, η.

Let $\Phi : \mathbb{R}^k \to \mathbb{R}^k$ be a one-to-one differentiable function mapping $(y_{1,n}, \ldots, y_{k,n})$

onto $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_k)$. For normalizations (1), (3) and (4) define Φ as:

$$\begin{aligned} &\Phi: \\ \xi_1 &= \frac{\hat{m} - m}{\omega} \\ \xi_2 &= y_{2,n} \\ &\vdots \quad \vdots \quad \vdots \\ \xi_k &= y_{k,n} \end{aligned}$$

For normalization (2) we must define a slightly different function, $\Phi' : \mathbb{R}^{k+1} \rightarrow \mathbb{R}^{k+1}$. This function is a one-to-one differentiable function, mapping $(y_{1,n}, \ldots, y_{k,n}, \eta)$ onto $\boldsymbol{\xi}' = \{\xi_1, \ldots \xi_{k+1}\}$, where $\xi_1, \ldots, \xi_k$ are defined as above and $\xi_{k+1} = \eta$, indeed;

$$\begin{aligned} &\Phi': \\ \xi_1 &= \frac{\hat{m} - m}{\omega} \\ \xi_2 &= y_{2,n} \\ &\vdots \quad \vdots \quad \vdots \\ \xi_k &= y_{k,n} \\ \xi_{k+1} &= \eta \end{aligned} \tag{2.2.1}$$

We can invert $\Phi(y_1, \ldots, y_k)$ and $\Phi'(y_1, \ldots, y_k, \eta)$ to get Φ^{-1} and Φ'^{-1} respectively. Φ^{-1} is of the form:

$$\begin{aligned} y_{1,n} &= z(\xi_1, \ldots, \xi_k) \\ y_{2,n} &= \xi_2 \\ &\vdots \quad \vdots \\ y_{k,n} &= \xi_k. \end{aligned}$$

Here $z(\xi_1, \ldots, \xi_k) : \mathbb{R}^k \to \mathbb{R}$ is some function of $\boldsymbol{\xi}$. Φ'^{-1} is similar:

$$\begin{aligned} y_{1,n} &= z(\xi_1, \ldots, \xi_k, \xi_k + 1) \\ y_{2,n} &= \xi_2 \\ &\vdots \quad\quad \vdots \\ y_{k,n} &= \xi_k \\ \eta &= \xi_{k+1} \end{aligned}$$

Here $z(\xi_1, \ldots, \xi_{k+1}) : \mathbb{R}^{k+1} \to \mathbb{R}$ is some function of $\boldsymbol{\xi}'$.

Now we can say that, for normalizations (1), (3) and (4), the joint density of vector $\boldsymbol{\xi}$ is given by:

$$p_{\xi_1,\xi_2,\ldots,\xi_k}(x_1, x_2, \ldots, x_k) = f_{1,2,\ldots k:n}(\Phi^{-1}(x_1, x_2, \ldots, x_k)) \left| \frac{D\Phi^{-1}}{d\boldsymbol{\xi}} \right| \tag{2.2.2}$$

and for normalization (2), the joint density of vector $\boldsymbol{\xi}'$ is given by:

$$p_{\xi_1,\xi_2,\ldots,\xi_{k+1}}(x_1, x_2, \ldots, x_{k+1}) = f_{1,2,\ldots k:n,\eta}(\Phi'^{-1}(x_1, x_2, \ldots, x_{k+1})) \left| \frac{D\Phi'^{-1}}{d\boldsymbol{\xi}} \right|. \tag{2.2.3}$$

Here $f_{1,2,\ldots,k:n}(y_1, y_2, \ldots, y_k)$ denotes the joint density of the first k order statistics, $y_{1,n}, y_{2,n}, \ldots, y_{k,n}$, where $1 \leq k \leq n$. Letting $f(y) = \frac{dF(y)}{dy}$ be the density of y_i, then

$$f_{1,2,\ldots,k:n}(y_1, y_2, \ldots, y_k) = \begin{cases} \frac{n!}{(n-k)!}(1 - F(y_k))^{n-k} \prod_{i=1}^{k} f(y_i) \\ \quad \text{if } m < y_1 < y_2 < \ldots < y_k < M \\ 0 \quad \text{otherwise.} \end{cases}$$

We also define the joint density of the first k order statistics and an independent Weibull random variable.

$$f_{1,2,\ldots,k:n,\eta}(y_1, y_2, \ldots, y_k, \eta) = \begin{cases} \frac{n!}{(n-k)!}\alpha x^{\alpha-1} \exp(-x^\alpha)(1 - F(y_k))^{n-k} \prod_{i=1}^{k} f(y_i) \\ \quad \text{if } m < y_1 < y_2 < \ldots < y_k < M,\ 0 < x < \infty \\ 0 \quad \text{otherwise.} \end{cases}$$

Here in both cases, M is the upper support of $F(x)$. $\left| \frac{D\Phi^{-1}}{d\boldsymbol{\xi}} \right|$ is the Jacobian of Φ^{-1}, and $\left| \frac{D\Phi'^{-1}}{d\boldsymbol{\xi}'} \right|$ is the Jacobian of Φ'^{-1}. The densities (2.2.2) and (2.2.3) have supports $m < z(x_2, \ldots, x_k) < x_2 < \ldots < x_k < M$ and $m < z(x_2, \ldots, x_{k+1}) < x_2 < \ldots < x_k < M$, $0 < x_{k+1} < \infty$ respectively.

The density of the normalized estimators can now be found by integrating over $\xi_2, \ldots, \xi_k$ or $\xi_2, \ldots, \xi_{k+1}$, depending on the normalization.

2.2.2 Density of a Normalized General Linear Estimator

Table 2.2 shows $y_{1,n}$ from the functions Φ^{-1} or Φ'^{-1} (see above) and the Jacobians, $\left|\frac{d\Phi^{-1}}{d\xi}\right|$ or $\left|\frac{d\Phi'^{-1}}{d\xi}\right|$. We can substitute these expressions into (2.2.2) and (2.2.3) in order to obtain joint densities of $\boldsymbol{\xi}$ and $\boldsymbol{\xi}'$ respectively.

Normalization	$y_{1,n}$	Jacobian
$\omega_1 = (nc_o)^{-1/\alpha}$	$\frac{(nc_0)^{-1/\alpha}\xi_1 - \sum_{i=2}^k a_i\xi_i + m}{a_1}$	$\frac{1}{(nc_0)^{1/\alpha}a_1}$
$\omega_2 = (y_{1,n} - m)/\eta$	$\frac{\xi_{k+1}\sum_{i=2}^k a_i(\xi_i - m)}{\xi_1 - a_1\xi_{k+1}} + m$	$\frac{\xi_{k+1}\sum_{i=2}^k a_i(\xi_i - m)}{(\xi_1 - a_1\xi_{k+1})^2}$
$\omega_3 = (y_{1,n} - m)$	$\frac{\sum_{i=2}^k a_i(\xi_i - m)}{\xi_1 - a_1} + m$	$\frac{\sum_{i=2}^k a_i(\xi_i - m)}{(\xi_1 - a_1)^2}$
$\omega_4 = (y_{k,n} - m)$	$\frac{\xi_1(\xi_k - m) - \sum_{i=2}^k a_i\xi_i + m}{a_1}$	$\frac{\xi_k - m}{a_1}$

Table 2.2: Summary table for derivation of the joint density of the vector of random variables $\boldsymbol{\xi}$ and $\boldsymbol{\xi}'$ that can be used to find the density of a normalized general linear estimator $\hat{m}$.

Densities

For $\omega = \omega_1 = (nc_0)^{1/\alpha}$ the joint density of $\xi_1, \ldots, \xi_k$, $p_1(z, x_2, \ldots, x_k)$, is

$p_1(z, x_2, \ldots x_k) =$

$$= \left|\frac{1}{(nc_0)^{1/\alpha}a_1}\right| \frac{n!}{(n-k)!}(1 - F(x_k))^{n-k} f\left(\frac{(nc_0)^{-1/\alpha}z - \sum_{i=2}^k a_i x_i + m}{a_1}\right) \prod_{i=2}^k f(x_i) \tag{2.2.4}$$

For $\omega = \omega_2 = (y_{1,n} - m)/\eta$ the joint density of $\xi_1, \ldots, \xi_{k+1}$ is given by

$p_2(z, x_2, \ldots x_k, \eta) =$

$$= \left|\frac{\eta\sum_{i=2}^k a_i(x_i - m)}{(x_1 - a_1\eta)^2}\right| \frac{n!\alpha\eta_1^{\alpha-1}\exp(-x^\alpha)}{(n-k)!}(1 - F(x_k))^{n-k} f\left(\frac{\eta\sum_{i=2}^k a_i(x_i - m)}{x_1 - a_1\eta} + m\right) \prod_{i=2}^k f(x_i) \tag{2.2.5}$$

For $\omega = \omega_3 = (y_{1,n} - m)$ the joint density of $\xi_1, \ldots, \xi_k$ is given by

$$p_3(z, x_2, \ldots x_k) =$$

$$= \left| \frac{\sum_{i=2}^{k} a_i(x_i - m)}{(x_1 - a_1)^2} \right| \frac{n!}{(n-k)!}(1 - F(x_k))^{n-k} f\left(\frac{\sum_{i=2}^{k} a_i(x_i - m)}{x_1 - a_1} + m \right) \prod_{i=2}^{k} f(x_i) \tag{2.2.6}$$

Finally, for $\omega = \omega_4 = (y_{k,n} - m)$ the joint density of $\xi_1, \ldots \xi_k$ is given by

$$p_4(z, x_2, \ldots x_k) =$$

$$= \left| \frac{x_k - m}{a_1} \right| \frac{n!}{(n-k)!}(1 - F(x_k))^{n-k} f\left(\frac{x_1(x_k - m) - \sum_{i=2}^{k} a_i x_i + m}{a_1} \right) \prod_{i=2}^{k} f(x_i) \tag{2.2.7}$$

The above densities (2.2.4), (2.2.6) and (2.2.7) are defined as above when $m \leq y_{1,n} \leq x_2 \leq \ldots \leq x_k \leq M$, and are equal to zero elsewhere. The density (2.2.5) is defined on the support $m \leq y_{1,n} \leq x_2 \leq \ldots \leq x_k \leq M$, $0 \leq \eta < \infty$ and is equal to zero elsewhere. Here m and M are upper and lower support of distribution $F(\cdot)$. $y_{1,n}$ is defined as in Table 2.2.

We can integrate the multivariate densities (2.2.4), (2.2.6) and (2.2.7) with respect to $x_2, \ldots, x_k$ (and η for (2.2.5)) to find the univariate density of the normalized linear estimator $\xi_1 = (\hat{m} - m)/\omega$, where $\omega = \omega_1$, ω_2, ω_3 and ω_4. These expressions can get very complicated as k increases.

2.2.3 Density of Estimator $(m^{(2)} - m)/\omega$ with $F(\cdot)$ Weibull

Let us now consider finding the density of normalized estimators for the particular case when the distribution of the random variables, y_i, is Weibull, and the estimator is the linear estimator, $m^{(2)}$.

We follow the method of derivation described in Section 2.2.1. Table 2.3 shows $y_{1,n}$ from Φ^{-1} (or Φ'^{-1}) for each of the normalizaions.

We could now substitute $y_{1,n}, \ldots, y_{k,n}$ (and η) and the Jacobian into (2.2.2) or (2.2.3). As the estimator $m^{(2)}$ only depends on the first and k^{th} order statistics,

Normalization	$y_{1,n}$	Jacobian
$\omega_1 = (nc_o)^{-1/\alpha}$	$\frac{(nc_0)^{-1/\alpha}\xi_1 + C_k\xi_2}{(1+C_k)}$	$\frac{1}{(nc_0)^{1/\alpha}(1+C_k)}$
$\omega_2 = (y_{1,n} - m)/\eta$	$\frac{\xi_2\xi_3 C_k}{\xi_3(1+C_k)-\xi_1}$	$\frac{\xi_2\xi_3 C_k}{(\xi_3(1+C_k)-\xi_1)^2}$
$\omega_3 = (y_{1,n} - m)$	$\frac{\xi_2 C_k}{1+C_k-\xi_1}$	$\frac{\xi_2 C_k}{(1+C_k-\xi_1)^2}$
$\omega_4 = (y_{k,n} - m)$	$\frac{\xi_2(\xi_1+C_k)}{1+C_k}$	$\frac{\xi_2}{1+C_k}$

Table 2.3: Summary table for derivation of the joint density of the vector of random variables $\boldsymbol{\xi}$ and $\boldsymbol{\xi}'$ that can be used to find the density of a normalized linear estimator $m^{(2)}$.

some simplifications can be made to (2.2.2) and (2.2.3): By integrating these two expressions with respect to $\xi_2, \ldots, \xi_{k-1}$ over their support, we obtain the joint density of just the normalized estimator and the k^{th} order statistic. Indeed, we can say that, for normalizations (1), (3) and (4), the joint density of vector (ξ_1, ξ_k) is given by:

$$p_{\xi_1,\xi_k}(x_1, x_k) = f_{W(1,k:n)}(\Phi^{-1}_{(2)}(x_1, x_k)) \left| \frac{D\Phi^{-1}}{d\boldsymbol{\xi}} \right|. \tag{2.2.8}$$

For normalization (2), the joint density of vector $(\xi_1, \xi_k, \xi_{k+1})$ is given by:

$p_{\xi_1,\xi_k,\xi_{k+1}}(x_1, x_k, x_{k+1}) =$

$$= \alpha x_{k+1}^{\alpha-1} \exp(-x_{k+1}^{\alpha}) f_{W(1,k:n,\eta)}({\Phi'_{(2)}}^{-1}(x_1, x_k, x_{k+1})) \left| \frac{D\Phi'^{-1}}{d\boldsymbol{\xi}} \right|. \tag{2.2.9}$$

$\Phi^{-1}_{(2)}$ and ${\Phi'_{(2)}}^{-1}$ are two and three dimensional functions respectively made up of just the elements $y_{1,n}, y_{k,n}$ and η from Φ^{-1} and Φ'^{-1} respectively.

The function $f_{W(1,k:n)}$ is the joint density of the 1^{st} and k^{th} order statistics where random variables come from a Weibull distribution with parameter α. For $k = 2$ this is given by the following expression:

$$f_{W(1,k:n)}(x_1, x_k) = n(n-1)\alpha^2 x_1^{\alpha-1} x_2^{\alpha-1} \mathrm{e}^{(-x_1^{\alpha} - (n-1)x_2^{\alpha})}, \quad 0 \le x_1 \le x_2 < \infty. \tag{2.2.10}$$

For $k > 2$ it is given by:

$$f_{W(1,k:n)}(x_1, x_k) = \\ = \frac{n!\alpha^2 x_1^{\alpha-1} x_2^{\alpha-1}}{(k-2)!(n-k)!} \left(e^{\left(\frac{-1}{k-2}((k-1)x_1^\alpha + (n-k+1)x_2^\alpha)\right)} - e^{\left(\frac{-1}{k-2}(x_1^\alpha + (n-1)x_2^\alpha)\right)} \right)^{k-2}, \\ 0 \le x_1 \le x_2 < \infty. \qquad (2.2.11)$$

The function $f_{W(1,k:n,\eta)}$ is the joint density of the 1^{st} and k^{th} order statistics and an independent Weibull random variable. Here the order statistics are obtained from a random sample of size n from a Weibull distribution with parameter α. For $k = 2$, $f_{W(1,k:n,\eta)}$ is given by the following expression

$$f_{W(1,k:n,\eta)}(x_1, x_k, \eta) = n(n-1)\alpha^3 x_1^{\alpha-1} x_2^{\alpha-1} \eta^{\alpha-1} e^{(-x_1^\alpha - (n-1)x_2^\alpha - \eta^\alpha)}, \\ 0 \le x_1 \le x_2 < \infty.$$

For $k > 2$, $f_{W(1,k:n,\eta)}$ is given by

$$f_{W(1,k:n,\eta)}(x_1, x_k, \eta) = \\ = \frac{n!\alpha^3 x_1^{\alpha-1} x_2^{\alpha-1} \eta^{\alpha-1}}{(k-2)!(n-k)!} \exp(-\eta^\alpha) \left(e^{\left(\frac{-1}{k-2}((k-1)x_1^\alpha + (n-k+1)x_2^\alpha)\right)} - e^{\left(\frac{-1}{k-2}(x_1^\alpha + (n-1)x_2^\alpha)\right)} \right)^{k-2}, \\ 0 \le x_1 \le x_2 < \infty.$$

Densities

The densities of the normalized estimators $(m^{(2)} - m)/\omega$ when the distribution of the random variables y_i is Weibull are as shown below.

For $\omega = \omega_1 = (nc_0)^{1/\alpha}$ the joint density of ξ_1 and ξ_k, $p_1(z, x_k)$ is

$$p_1(z, x_k) = \frac{1}{(nc_0)^{1/\alpha}(1 + C_k)} f_{W(1,2:n)} \left(\frac{(nc_0)^{-1/\alpha} z + C_k x_2}{(1 + C_k)}, x_2 \right) \qquad (2.2.12)$$

For $\omega = \omega_2 = (y_{1,n} - m)/\eta$ the joint density of ξ_1, ξ_k and ξ_{k+1} is given by

$$p_2(z, x_k, \eta) = \frac{x_2 \eta C_k}{\eta(1 + C_k) - z} f_{W(1,2:n,\eta)} \left(\frac{x_2 \eta C_k}{(\eta(1 + C_k) - z)^2}, x_2, \eta \right) \qquad (2.2.13)$$

For $\omega = \omega_3 = (y_{1,n} - m)$ the joint density of ξ_1 and ξ_k is given by

$$p_3(z, x_k) = \frac{x_2 C_k}{(1 + C_k - z)^2} f_{W(1,2:n)} \left(\frac{x_2 C_k}{1 + C_k - z}, x_2 \right) \qquad (2.2.14)$$

Finally, for $\omega = \omega_4 = (y_{k,n} - m)$ the joint density of ξ_1 and ξ_k is given by

$$p_4(z, x_k) = \frac{x_2}{1 + C_k} f_{W(1,2:n)} \left(\frac{x_2(z + C_k)}{1 + C_k}, x_2 \right) \tag{2.2.15}$$

The above densities (2.2.12), (2.2.14) and (2.2.15) are defined as above when $0 \leq y_{1,n} \leq x_k < \infty$, and are equal to zero elsewhere. The density (2.2.13) is defined on the support $0 \leq y_{1,n} \leq x_k < \infty$, $0 \leq \eta < \infty$ and is equal to zero elsewhere. $y_{1,n}$ is defined as in Table 2.3.

We can find the univariate density of the normalized linear estimator $\xi_1 = (\hat{m} - m)/\omega$, in a similar way as for the general linear estimator $(\hat{m} - m)/\omega$ (in both cases $\omega = \omega_1$, ω_2, ω_3 and ω_4). Indeed, integrating the multivariate densities (2.2.12), (2.2.14) and (2.2.15) with respect to $x_2, \ldots, x_k$ (and η for (2.2.13)), results in the univariate densities of the normalized linear estimator $\xi_1 = (\hat{m} - m)/\omega$, where $\omega = \omega_1$, ω_2, ω_3 and ω_4.

2.2.4 Density of $(m^{(2)} - m)/\omega$ where $\omega = \omega_3$ or ω_4 and $F(\cdot)$ is Weibull

Normalizations 3 and 4 have convenient forms that allow the explicit forms of densities of $(m^{(2)} - m)/\omega_3$ and $(m^{(2)} - m)/\omega_4$ to be found for any k.

Lemma 2.2.1. *The density of the normalized estimators, $(m^{(2)} - m)/(y_{1,n} - m)$ and $(m^{(2)} - m)/(y_{k,n} - m)$, for all integer $k \geq 2$, is given by*

$$p^{(2)}(z) = \frac{n! h'(z)}{(n-k)!} \sum_{r=0}^{k-2} \frac{(-1)^r}{(k-2-r)! r! [n - (k-r-1)(1-h(z))]^2} \qquad -l \leq z \leq 1. \tag{2.2.16}$$

Here we define; for normalization 3;

$$\begin{aligned} h(z) &= \left(\frac{C_k}{1 + C_k - z} \right)^{\alpha}, \\ h'(z) &= \frac{\alpha C_k^{\alpha}}{(1 + C_k - z)^{\alpha+1}} \text{ and} \\ l &= \infty \end{aligned}$$

and for normalization 4;

$$\begin{aligned} h(z) &= \left(\frac{z+C_k}{1+C_k}\right)^{\alpha}, \\ h'(z) &= \frac{\alpha(z+C_k)^{\alpha-1}}{(1+C_k)^{\alpha}} \text{ and} \\ l &= C_k. \end{aligned}$$

Proof. By substituting (2.2.10) into (2.2.14) or (2.2.15) we obtain that when $k = 2$ the general form of the density of the normalized estimator $(m^{(2)} - m)/\omega$, where $\omega = \omega_3$ or ω_4 and $F(\cdot)$ is Weibull is as follows:

$$p^{(2)}(z) = n(n-1)\alpha h'(z) \int_{x=0}^{\infty} x^{2\alpha-1} e^{(-x^{\alpha}((n-1)+h(z))} dx, \quad -l \leq z \leq 1.$$

Similarly, by substituting (2.2.11) into (2.2.14) or (2.2.15), for $k > 2$ we have that the density of the normalized estimators is of the form:

$$\begin{aligned} p^{(2)}(z) &= \frac{n!\alpha h'(z)}{(n-k)!(k-2)!} \int_{x=0}^{\infty} x^{2\alpha-1} \left(\exp\left(\frac{-x^{\alpha}}{k-2}(n-k+1+(k-1)h(z))\right)\right. \\ &\qquad \left. - \exp\left(\frac{-x^{\alpha}}{k-2}(n-1+h(z))\right)\right)^{k-2} dx, \\ &\quad -l \leq z \leq 1. \end{aligned}$$

When $k = 2$ the density, $p^{(2)}(z)$, can be simplified easily by noting that

$$\int_{x=0}^{\infty} x^{2\alpha-1} \exp(-x^{\alpha}B) dx = \frac{1}{\alpha B^2}, \tag{2.2.17}$$

where B is a positive constant. So we get for $k = 2$

$$p^{(2)}(z) = n(n-1)h'(z)\left[(n-1+h(z))\right]^{-2}, \quad -l \leq z \leq 1. \tag{2.2.18}$$

and

$$\lim_{n\to\infty} p^{(2)}(z) = h'(z), \quad -l \leq z \leq 1. \tag{2.2.19}$$

The case when $k > 2$ needs slightly more work to get it into a simple form. Using a binomial expansion and simplifying

$$\left(\exp\left(\frac{-x^{\alpha}}{k-2}(n-k+1+(k-1)h(z))\right) - \exp\left(\frac{-x^{\alpha}}{k-2}(n-1+h(z))\right)\right)^{k-2}$$

$$= \sum_{r=0}^{k-2} \frac{(k-2)!(-1)^r}{(k-2-r)!r!} \exp\left(-x^\alpha(n-(k-r-1)(1-h(z)))\right)$$

Substituting this into $p^{(2)}(z)$ and simplifying we obtain

$$p^{(2)}(z) = \frac{n!\alpha h'(z)}{(n-k)!} \sum_{r=0}^{k-2} \frac{(-1)^r}{(k-2-r)!r!} \int_0^\infty x^{2\alpha-1} \exp(-x^\alpha[n-(k-r-1)(1-h(z))])dx$$
$$-l \leq z \leq 1$$

Now, noting that for both normalizations $h'(z) \leq 1$ which implies $(n-(k-r-1)(1-h(z)) > 0$, we can again make use of (2.2.17) and so obtain

$$p^{(2)}(z) = \frac{n!h'(z)}{(n-k)!} \sum_{r=0}^{k-2} \frac{(-1)^r}{(k-2-r)!r![n-(k-r-1)(1-h(z))]^2}$$
$$-l \leq z \leq 1$$

□

Lemma 2.2.2. *The asymptotic density of* $(m^{(2)}-m)/\omega$, $\omega = \omega_3$ *or* ω_4 *as* $n \to \infty$ *is given by:*

$$\lim_{n\to\infty} p^{(2)}(z) = (k-1)h'(z)(1-h(z))^{k-2}, \quad -l \leq z \leq 1. \tag{2.2.20}$$

Here h *and* h' *are defined as above.*

Proof. As $p^{(2)}(z)$ is a density, integrating over the support gives 1. Make a change of variable; $u = h(z)$. For normalization 3 and 4 this results in the following equation:

$$\begin{aligned} & \tfrac{n!}{(n-k)!} \textstyle\int_{u=0}^{1} \sum_{r=0}^{k-2} \frac{(-1)^r}{(k-2-r)!r![n-(k-r-1)(1-u)]^2} du && = 1 \\ \Rightarrow \quad & \tfrac{n!}{(n-k)!} \textstyle\sum_{r=0}^{k-2} \frac{(-1)^r}{(k-2-r)!r!n(n-(k-r-1))} && = 1 \\ \Rightarrow \quad & \textstyle\sum_{r=0}^{k-2} \frac{(-1)^r}{(k-2-r)!r!n(n-(k-r-1))} && = \frac{(n-k)!}{n!} \end{aligned} \tag{2.2.21}$$

Now consider the c.d.f. associated with $p^{(2)}(z)$, $P^{(2)}(z) = \mathrm{P}(Z \leq z)$, obtained by integrating the density $p^{(2)}(Z)$ from the lower support to z:

$$\mathrm{P}(Z \leq z) = \int_{Z=-l}^{z} \frac{n!h'(Z)}{(n-k)!} \sum_{r=0}^{k-2} \frac{(-1)^r}{(k-2-r)!r![n-(k-r-1)(1-h(Z))]^2} dZ$$

For simplicity we can make the change of variable $U = h(Z)$ to obtain;

$$\begin{aligned} \mathrm{P}(Z \leq z) &= \frac{n!}{(n-k)!} \int_{U=0}^{h(z)} \sum_{r=0}^{k-2} \frac{(-1)^r}{(k-2-r)!r![n-(k-r-1)(1-U)]^2} dU \\ &= \frac{n!u}{(n-k)!} \sum_{r=0}^{k-2} \frac{(-1)^r}{(k-2-r)!r![n-(k-r-1)][n-(k-r-1)(1-u)]} \end{aligned}$$

where $u = h(z)$.

As

$$\begin{aligned} &\frac{1}{(n-(k-r-1))(n-(k-r-1)(1-u))} \\ &\quad = \frac{1}{un(n-(k-r-1))} + \frac{1-u}{un(n-(k-r-1)(1-u))} \end{aligned}$$

we have that

$$\begin{aligned} \mathrm{P}(Z \leq z) &= \frac{n!}{(n-k)!} \left(\sum_{r=0}^{k-2} \frac{(-1)^r}{(k-2-r)!r!n(n-(k-r-1))} \right. \\ &\qquad \left. - \frac{(-1)^r(1-u)}{(k-2-r)!r!n(n-(k-r-1)(1-u))} \right) \end{aligned}$$

by (2.2.21) we have

$$\mathrm{P}(Z \leq z) = 1 - \frac{(-1)^r}{(1-u)(k-2-r)!r!\frac{n}{1-u}\left(\left(\frac{n}{1-u}\right)-(k-r-1)\right)}$$

using (2.2.21) again

$$\mathrm{P}(Z \leq z) = 1 - \frac{n!\Gamma\left(\frac{n-(k-1)(1-u)}{1-u}\right)}{(n-k)!(1-u)\Gamma\left(\frac{n+1-u}{1-u}\right)}$$

We can now take the limit of the above expression to obtain

$$\lim_{n\to\infty} \mathrm{P}(Z \leq z) = 1 - (1-u)^{k-1}$$

Substituting $u = h(z)$ and differentiating with respect to z:

$$\lim_{n\to\infty} p^{(2)}(z) = (k-1)h'(z)(1-h(z))^{k-2}.$$

□

The density of $(m^{(2)} - m)(nc_0)^{(1/\alpha)}$ where y_i come from a Weibull distribution can be found but a simple form cannot. The density of this normalized random variable can be found by integrating (2.2.12) with respect to x_k from l to ∞. Here $l = \frac{-z}{n^{1/\alpha}C_k}$ when $z < 0$ and $l = n^{-1/\alpha}z$ when $z \geq 0$. Notice that for the Weibull distribution $c_0 = 1$. Although it has not been possible to evaluate this integral to obtain the density explicitly even for particular values of k and α, it has been possible to obtain numerical solutions of integrals of the density over small ranges of z for particular values of k and α. By doing this it has been possible to plot the density of $(m^{(2)} - m)/\omega_1$ on the histograms below.

Notice that for any linear estimator defined by $m^{(\Delta)} = (1 + C_k^{(\Delta)})y_{1,n} - C_k^{(\Delta)}y_{k,n}$, $C^{(\Delta)} > 0$, the density of $(m^{(\Delta)} - m)/\omega$ where $\omega = \omega_1$, ω_2, ω_3 and ω_4 can be found by substituting $C_k^{(\Delta)}$ in place of C_k in any of the densities in Sections 2.2.3 and 2.2.4.

2.2.5 Density of $(m^* - m)/\omega$ with $F(\cdot)$ Weibull and $k = 2$ Amendment to MLE

Above, the definition of the MLE depends on the solution to the likelihood equation, (1.3.3), being less than the minimum order statistic. It is easy to show that there is always a solution to the likelihood equation less than the minimum order statistic.

Lemma 2.2.3. *There is exactly one solution to the likelihood equation in the region* $(-\infty, y_{1,n}]$ *for any integer* k *and* $\alpha > 1$.

Proof. In the region $(-\infty, y_{1,n})$ the left hand side of the likelihood equation is continuous in z. We have

$$\lim_{z \uparrow y_{1,n}} (\alpha - 1) \sum_{i=1}^{k-1} \frac{y_{k,n} - y_{i,n}}{y_{i,n} - z} = \infty.$$

Also,

$$\lim_{z \downarrow -\infty} (\alpha - 1) \sum_{i=1}^{k-1} \frac{y_{k,n} - y_{i,n}}{y_{i,n} - z} = -\infty.$$

The function $(\alpha-1)\sum_{i=1}^{k-1}\frac{y_{k,n}-y_{i,n}}{y_{i,n}-z}$ is strictly increasing for $z \in (-\infty, y_{1,n})$. Therefore as $k \in (0,\infty)$ there is exactly one solution to the likelihood equation in the region $(-\infty, y_{1,n}]$. □

We can also find explicit expressions for the density of the normalized MLE when $k = 2$ and $F(x)$ is Weibull. In this case the MLE is the solution to the linear equation

$$\begin{aligned} m^* &= y_{1,n} - \frac{\alpha-1}{2}(y_{2,n} - y_{1,n}) \\ &= \frac{\alpha+1}{2}y_{1,n} - \frac{\alpha-1}{2}y_{2,n}. \end{aligned}$$

This is simply a linear estimator on two order statistics with $C_k^{(\Delta)} = \frac{\alpha-1}{2}$. Its density can therefore easily be found from Sections 2.2.3 and 2.2.4.

Notice the similarity between the densities of $(m^{(2)} - m)/\omega$ and $(m^* - m)/\omega$. For the estimator $m^{(2)}$, $C_2 = \alpha/2$ for all α. Let us define an estimator to be $m^\Delta = (1 + C_2^\Delta)y_{1,n} - C_2^\Delta y_{k,n}$, where $C_2^\Delta = (\alpha - 1)/2$, then we see the density of $(m^\Delta - m)/\omega$ is equal to that of $(m^* - m)/\omega$, where either $\omega = \omega_3$ or $\omega = \omega_4$.

Let us derive expressions for the densities of the normalized estimators $(m^* - m)/\omega$ when $k = 2$ and the distribution of the random variables y_i is Weibull. For $\omega = \omega_1 = (nc_0)^{1/\alpha}$ the joint density of ξ_1 and ξ_2, $p_1(z, x_k)$ is

$$p_1(z,x_2) = \frac{2}{(nc_0)^{1/\alpha}(\alpha+1)} f_{W(1,2:n)}\left(\frac{2(nc_0)^{-1/\alpha}z + (\alpha+1)x_2}{\alpha+1}, x_2\right) \quad (2.2.22)$$

For $\omega = \omega_2 = (y_{1,n} - m)/\eta$ the joint density of ξ_1, ξ_k and ξ_{k+1} is given by

$$p_2(z,x_k,\eta) = \frac{x_2\eta(\alpha-1)}{\eta(\alpha+1) - 2z} f_{W(1,2:n,\eta)}\left(\frac{2x_2\eta(\alpha-1)}{(\eta(\alpha+1)-2z)^2}, x_2, \eta\right) \quad (2.2.23)$$

For $\omega = \omega_3 = (y_{1,n} - m)$ the joint density of ξ_1 and ξ_k is given by

$$p_3(z,x_k) = \frac{2x_2(\alpha-1)}{(\alpha+1-2z)^2} f_{W(1,2:n)}\left(\frac{x_2(\alpha-1)}{\alpha+1-2z}, x_2\right) \quad (2.2.24)$$

Finally for $\omega = \omega_4 = (y_{k,n} - m)$ the joint density of ξ_1 and ξ_k is given by

$$p_4(z,x_k) = \frac{2x_2}{\alpha+1} f_{W(1,2:n)}\left(\frac{x_2(2z+\alpha-1)}{\alpha+1}, x_2\right) \quad (2.2.25)$$

The above densities (2.2.22), (2.2.24) and (2.2.25) are defined as above when $0 \leq y_{1,n} \leq x_2 < \infty$, and are equal to zero elsewhere. The density (2.2.23) is defined on the support $0 \leq y_{1,n} \leq x_2 < \infty$, $0 \leq \eta < \infty$ and is equal to zero elsewhere. $y_{1,n}$ is defined as in Table 2.3 with $(\alpha - 1)/2$ substituted in place of C_k.

2.2.6 Density of $(m^* - m)/\omega$ where $\omega = \omega_3$ or ω_4 with $F(\cdot)$ Weibull and $k = 2$

As with the linear estimator $m^{(2)}$, the joint densities (2.2.24) and (2.2.25) are in convenient form so that an explicit form of the density can easily be found.

Lemma 2.2.4. *The densities of* $(m^* - m)/(y_{1,k} - m)$ *and* $(m^* - m)/(y_{k,n} - m)$ *are given by:*

$$p^*(z) = \frac{n(n-1)h^{*\prime}(z)}{(n-1+h^*(z))^2}, \quad -l \leq z \leq 1. \tag{2.2.26}$$

The associated asymptotic densities as $n \to \infty$ *are given by*

$$\lim_{n\to\infty} p^*(z) = h^{*\prime}(z), \quad -l \leq z \leq 1. \tag{2.2.27}$$

Here for normalization 3 we define

$$\begin{aligned} h^*(z) &= \left(\frac{\alpha - 1}{\alpha + 1 - 2z}\right)^{\alpha}, \\ h^{*\prime}(z) &= \frac{2\alpha(\alpha-1)^{\alpha}}{(\alpha+1-2z)^{\alpha+1}} \text{ and} \\ l &= \infty \end{aligned}$$

and for normalization 4,

$$\begin{aligned} h^*(z) &= \left(\frac{2z + \alpha - 1}{\alpha + 1}\right)^{\alpha}, \\ h^{*\prime}(z) &= \frac{2\alpha(2z+\alpha-1)^{\alpha-1}}{(\alpha+1)^{\alpha}} \text{ and} \\ l &= \frac{(\alpha-1)}{2}. \end{aligned}$$

Proof. The results follow by substituting $C_k = (\alpha - 1)/2$ into (2.2.16) and (2.2.20)

□

2.3 Analysis Using Densities

2.3.1 Comparing m° and m^*

Let us now compare the densities and asymptotic densities of $(m^\circ - m)/\omega$ and $(m^* - m)/\omega$ for $k = 2$, where $\omega = \omega_3$ or $\omega = \omega_4$.

ξ	$p(z)$	$\lim_{n\to\infty} p(z)$	Support
$\frac{m^\circ - m}{\omega_3}$	$\frac{n(n-1)\alpha C_2^\alpha}{(1+C_2-z)^{\alpha+1}}\left(n-1+\left(\frac{C_2}{1+C_2-z}\right)^\alpha\right)^{-2}$	$\frac{\alpha C_2^\alpha}{(1+C_2-z)^{\alpha+1}}$	$-\infty < z \leq 1$
$\frac{m^\circ - m}{\omega_4}$	$\frac{n(n-1)\alpha(z+C_2)^{\alpha-1}}{(1+C_2)^\alpha}\left(n-1+\left(\frac{z+C_2}{1+C_2}\right)^\alpha\right)^{-2}$	$\frac{\alpha(z+C_2)^{\alpha-1}}{(1+C_2)^\alpha}$	$-C_2 < z \leq 1$
$\frac{m^* - m}{\omega_3}$	$\frac{2n(n-1)\alpha(\alpha-1)^\alpha}{(\alpha+1-2z)^{\alpha+1}}\left(n-1+\left(\frac{\alpha-1}{\alpha+1-2z}\right)^\alpha\right)^{-2}$	$\frac{2\alpha(\alpha-1)^\alpha}{(\alpha+1-2z)^{\alpha+1}}$	$-\infty < z < 1$
$\frac{m^* - m}{\omega_4}$	$\frac{2n(n-1)\alpha(2z+\alpha-1)^{\alpha-1}}{(\alpha+1)^\alpha}\left(n-1+\left(\frac{2z+\alpha-1}{\alpha+1}\right)^\alpha\right)^{-2}$	$\frac{2\alpha(2z+\alpha-1)^{\alpha-1}}{(\alpha+1)^\alpha}$	$-\frac{\alpha-1}{2} < z < 1$

Table 2.4: The density and asymptotic density of normalized estimators m° and m^* when $k = 2$.

Notice the similarity between the densities of $(m^{(2)} - m)/\omega$ and $(m^* - m)/\omega$. From (2.1.3) we can see that $C_2 = \alpha/2$. If we define an estimator to be $m^{(\Delta)} = (1 + C_2^{(\Delta)})y_{1,n} - C_2^{(\Delta)}y_{k,n}$, where $C_2^{(\Delta)} = (\alpha - 1)/2$. Then we see the density of $(m^\Delta - m)/\omega$ is equal to that of $(m^* - m)/\omega$, where either $\omega = \omega_3$ or $\omega = \omega_4$. This similarity is explored again in Section 2.5.

2.3.2 Investigation into the Effect of Increasing Sample Size

Here we investigate the effect of increasing sample size on the normalized k^{th} order statistic from a Weibull distribution and on the normalized estimators $(m^{(2)} -$

$m)(nc_0)^{1/\alpha}$, $(m^{(2)} - m)/(y_{1,n} - m)$ and $(m^{(2)} - m)/(y_{k,n} - m)$ where y_i come from a Weibull distribution. We also briefly discuss the convergence of the order statistics from the beta distribution to the asymptotic distribution.

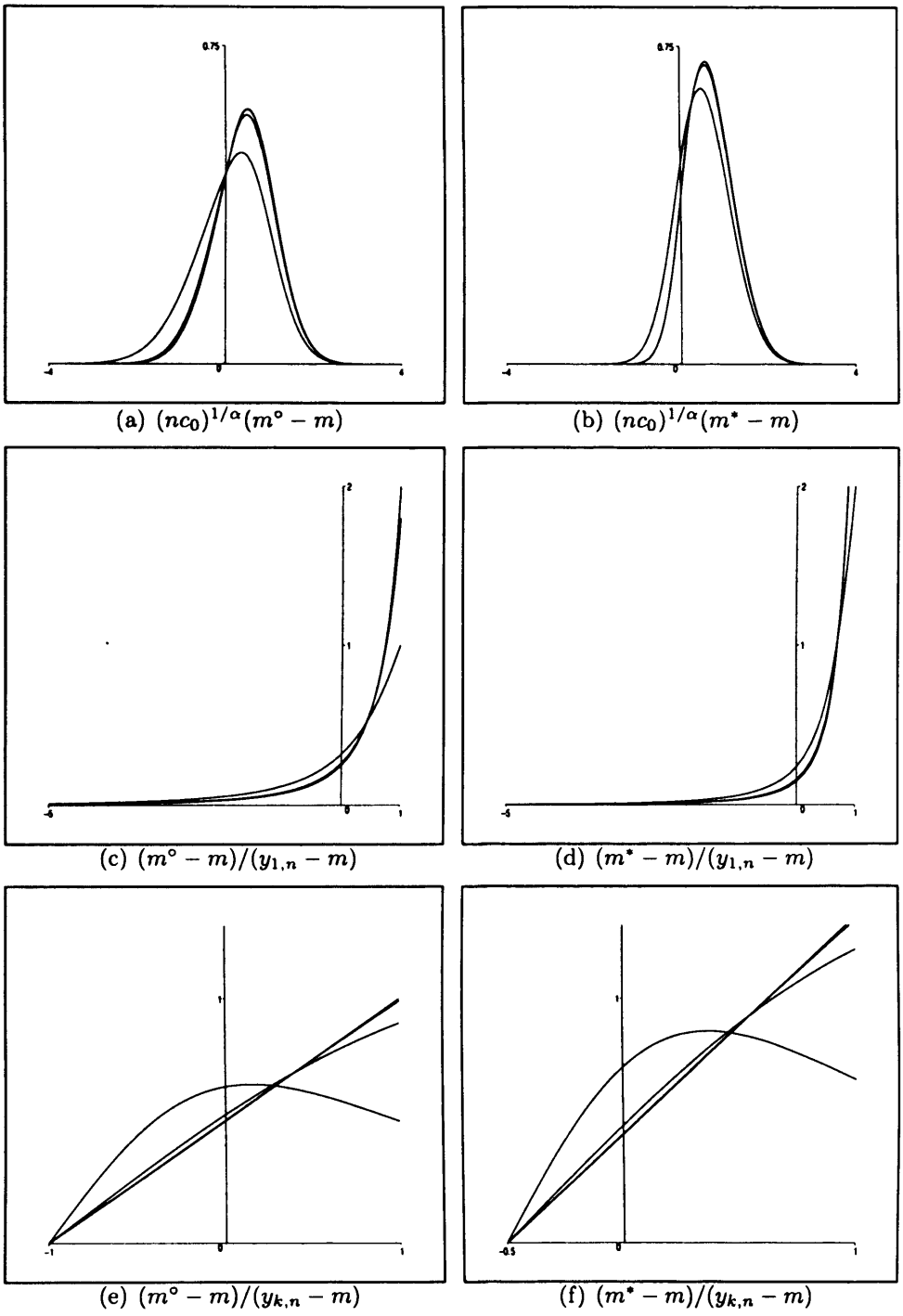
Figure 2.3 shows the densities of normalized estimators m° and m^*. The normalization used are as labeled. Here $F(\cdot)$ is the Weibull distribution with $\alpha = 2$ and $k = 2$. It can be seen that the density of the normalized estimator $(m^{(2)} - m)/\omega$ reaches its asymptotic density very quickly. The densities of the normalized estimators when $n = 10$ are quite close to that of the normalized estimators when $n = 1\,000$. The densities of the normalized estimators when $n = 100$ and $n = 1\,000$ are almost indistinguishable.

Figure 2.4 (a) and (b) show (with a dashed line) the distribution of $y_{k,n}(\kappa_n - m) + m$ where $y_{k,n}$ is the k^{th} order statistic from the Weibull distribution. This has been denoted $p_{W(k:n)}(x)$. For the Weibull distribution $m = 0$ and $\kappa_n - m = \left(\log\left(\frac{n}{n-1}\right)\right)^{1/\alpha}$. $p_{W(k:n)}(x)$ is given by the following expression:

$$\sum_{i=k}^{n} \frac{n!}{(n-i)!i!} \left(1 - \exp\left(-x^\alpha \log\left(\frac{n}{n-1}\right)\right)\right)^i \left(\exp\left(-x^\alpha \log\left(\frac{n}{n-1}\right)\right)\right)^{n-i}$$

In Figure 2.4, (a) and (b) show (using a solid line) the asymptotic distribution of the normalized k^{th} order statistic from a general distribution satisfying the conditions of Theorem 1. The equation for this is given by (1.2.7). In (a), $\alpha = 3$ and $k = 5$ and in (b), $k = 10$ and $\alpha = 3$. The sizes of the samples of Weibull random variables are $n = 10$, 15, 20, 50 and 100. As n increases the samples plotted converge monotonically to the asymptotic one. From these two plots it can be seen that when $\alpha = 3$, the normalized 5^{th} order statistic converges more quickly to the asymptotic distribution than the normalized 10^{th} order statistic. For both $k = 5$ and $k = 10$ the density of the normalized k^{th} order statistic when $n = 100$ is close to the asymptotic one.

Figure 2.4 (c)-(h) show the density (2.2.16) of $(m^{(2)} - m)/(y_{k,n} - m)$ for $n = 10$ and $n = 100$. Also shown is the asymptotic density (2.2.20). It can be seen that



(a) $(nc_0)^{1/\alpha}(m^\circ - m)$ (b) $(nc_0)^{1/\alpha}(m^* - m)$

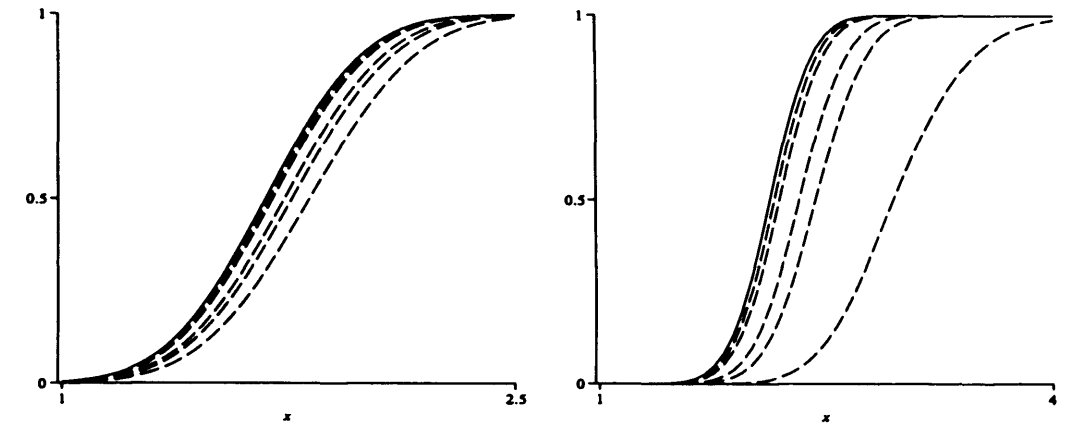
(c) $(m^\circ - m)/(y_{1,n} - m)$ (d) $(m^* - m)/(y_{1,n} - m)$

(e) $(m^\circ - m)/(y_{k,n} - m)$ (f) $(m^* - m)/(y_{k,n} - m)$

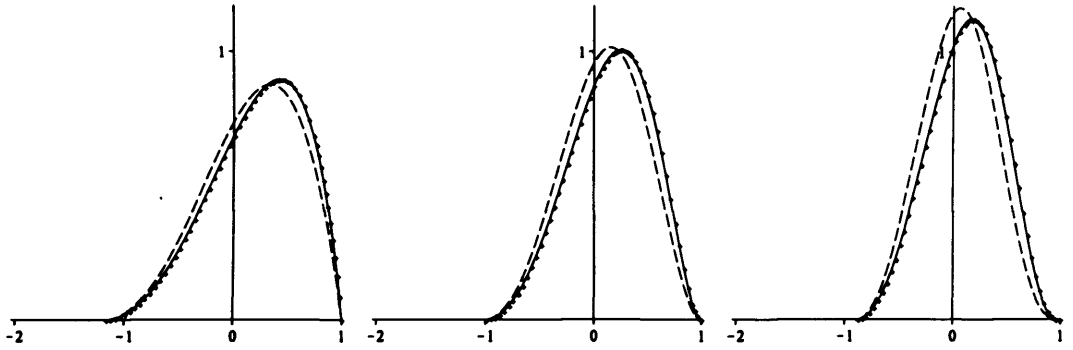
Figure 2.3: The plots on the left show the densities of normalized $m^\circ = m^{(2)}$ estimators, the plot on the right show the densities of normalized m^* estimators. Normalizations are as labeled. In all plots the paramters $\alpha = 2$, $k = 2$ and n =2, 10, 100 and 1 000 have been used.

for all α and k shown, the density of $(m^{(2)} - m)/(y_{k,n} - m)$ is very close to the asymptotic density. This is because the normalized order statistics from a Weibull distribution converge quickly to the asymptotic distribution (as shown in (a) and (b)).

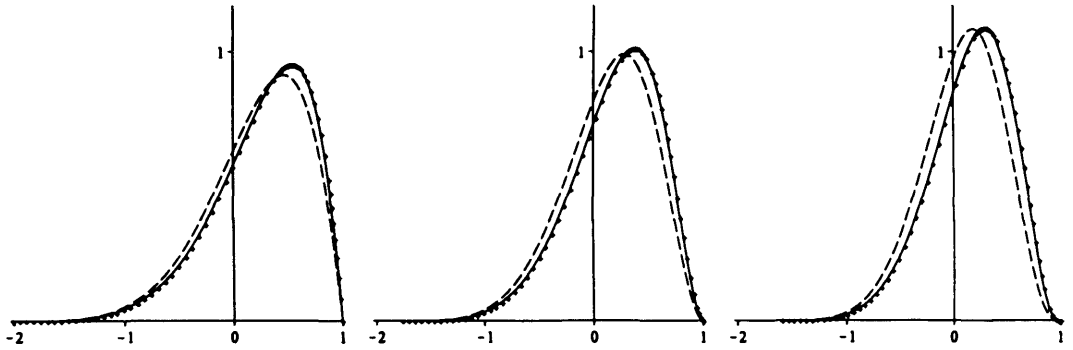
This convergence to the asymptotic distribution of the normalized linear estimators can be explained by considering the distributions of the extreme order statistics from the Weibull distribution. It was shown in Section 1.2.2 that correctly normalized extreme order statistics from the Weibull distribution (and the beta distribution, $F_\beta(x; \alpha, 1)$) converge quickly to the asymptotic distribution (especially for small k). This has the direct implication that the linear estimators (normalized using normalization (1)) must also converge to an asymptotic distribution. This means that for any further simulation studies involving the Weibull distribution, a small sample.size (of say $n = 100$) may be used to reduce simulation time without affecting results. The quick convergence is not true of every distribution, for example, extreme order statistics from the beta distribution, $F_\beta(x; 3, 2)$, do not converge very quickly to the asymptotic distribution (especially for large k), and thus the normalized estimator based on extreme order statistics from the beta distribution, $F_\beta(x; 3, 2)$, distribution will not converge quickly to the asymptotic distribution.



(a) $k = 5, \alpha = 3$ (b) $k = 10, \alpha = 3$



(c) $k = 3, \alpha = 3$ (d) $k = 4, \alpha = 3$ (e) $k = 5, \alpha = 3$



(f) $k = 3, \alpha = 5$ (g) $k = 4, \alpha = 5$ (h) $k = 5, \alpha = 5$

Figure 2.4: Figures (a) shows $p_{W(k:n)}(x)$ for $k = 5$, $\alpha = 3$ and $n = 10$, 15, 20, 50 and 100. Also shown is (1.2.7) for $k = 5$ and $\alpha = 3$. Figures (b) shows $p_{W(k:n)}(x)$ for $k = 10$, $\alpha = 3$ and $n = 10$, 15, 20, 50 and 100. Also shown is (1.2.7) for $k = 10$ and $\alpha = 3$. Figures (c)-(h) shows the density (2.2.16) and asymptotic density (2.2.20) of the normalized estimator $\xi_1 = (m^{(2)} - m)/(y_{k,n} - m)$. The density (2.2.16) has parameters $n = 50$ (dashed line) and $n = 100$ (solid line). The asymptotic density is marked with $+$.

2.4 Simulation Study: Known α

Below we use a finite sample simulation study to compare the estimators defined above. First, we compare the distributions of the estimators using histograms. We then consider the efficiency, bias and MSE of the estimators.

2.4.1 Histograms of Normalized Estimators

The histograms plotted below are of normalized estimates of m, the lower endpoint of a Weibull distribution. Plotted on the relevant histograms are the densities calculated in Section 2.2. Densities (2.2.18), (2.2.16) have been plotted. It was also possible to obtain numerical solutions to integrals of the density of the normalized estimator $(m^{(2)} - m)(nc_0)^{(1/\alpha)}$ for particular values of k and α over finite ranges. This has been plotted on the relevant histogram.

Note that the c.d.f. of the Weibull distribution can be represented in the form (1.2.6) with $m = 0$ and $c_0 = 1$, since $1 - \exp(-t^\alpha) = t^\alpha(1 + O(t^\alpha))$ as $t \to 0$. This means that the normalized minimum order statistic of a sample taken from $F(x)$ will also be distributed Weibull with parameter α. Sections 1.2.2 and 2.3.2 show us that with $F(x)$ as Weibull, the densities of the extreme order statistics and the distribution of the normalized estimators converge to their corresponding asymptotic densities at small sample sizes.

Normalization

The estimators shown in the histograms in Figure 2.5 have been normalized in the four different ways discussed in Section 2.2; $(nc_0)^{1/\alpha}(\hat{m} - m)$, $(\hat{m} - m)\eta/(y_{1,n} - m)$, $(\hat{m} - m)/(y_{1,n} - m)$ and $(\hat{m} - m)/(y_{k,n} - m)$, where η is an independent random variable drawn from the Weibull distribution with parameter α. We will now study the different effects that these normalizations have on the distribution of the corresponding normalized estimators. It can easily be seen from Figure 2.5 that none of the normalizations produce a normalized estimator with the same density as another.

As $E(y_{1,n} - m) = (\kappa_n - m) \sim (nc_0)^{-1/\alpha}$, you might expect $(nc_0)^{1/\alpha}(\hat{m} - m)$ and $(\hat{m} - m)/(y_{1,n} - m)$ to be equal in distribution, this is not the case.

For large n we can say that if η' is a variable chosen such that the following statement holds

$$\frac{(\hat{m} - m)\eta'}{(y_{1,n} - m)} = (nc_0)^{1/\alpha}(\hat{m} - m),$$

then η' will have a Weibull distribution with parameter α. This can be seen by noting that $(y_{1,n} - m) \stackrel{d}{=} (\kappa_n - m)\eta$ and using (1.2.9). This fact does not mean that $(nc_0)^{1/\alpha}(\hat{m} - m)$ and $(\hat{m} - m)\eta/(y_{1,n} - m)$ will be equal in distribution, as there is dependence between $\hat{m}$ and $y_{1,n}$ whereas $\hat{m}$ and η are independent.

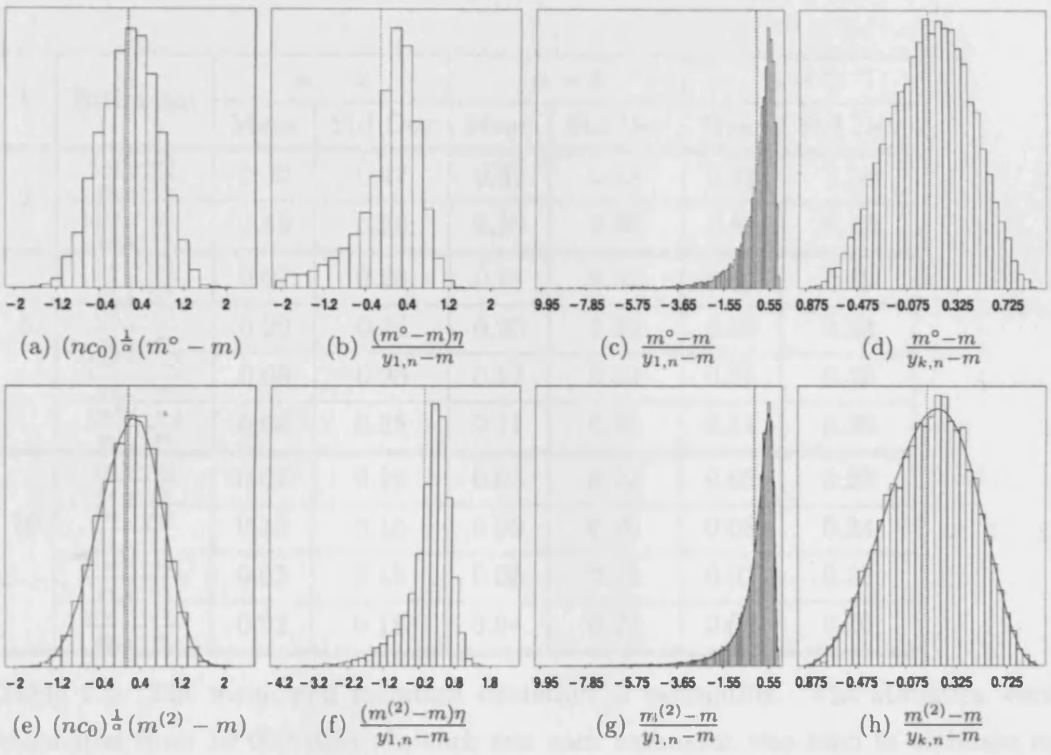
It can be seen from the histograms in Figure 2.5 (as well as Figure 2.3) that the four normalizations used have very different effects on the distribution of the normalized estimators. Using normalization (1) makes the density of the normalized estimators similar to the normal distribution. Normalizations (1) and (2) do not impose any bounds on the size of the normalized estimators. Normalization (3) imposes an upper bound of 1 on the normalized estimator, but no lower bound. Normalization (4) ensures that the resulting normalized estimator is between a_k and 1. In future all discussions about normalized estimators will be considering normalization (4).

Estimators

The different estimators do not appear to vary greatly in distribution. They do however have different means and variances. In general the MLE has a lower variance than the optimal linear estimator (and estimators $m^{(2)}$ and $m^{(3)}$), however the mean of the MLE is greater in general than the linear estimators.

Parameters α and k

From Table 2.5 it can be seen that as α increases, the standard deviation of the estimators increases. The mean of the linear estimators also increases with α. For all estimators both the mean and the standard deviation decrease as k is increased.



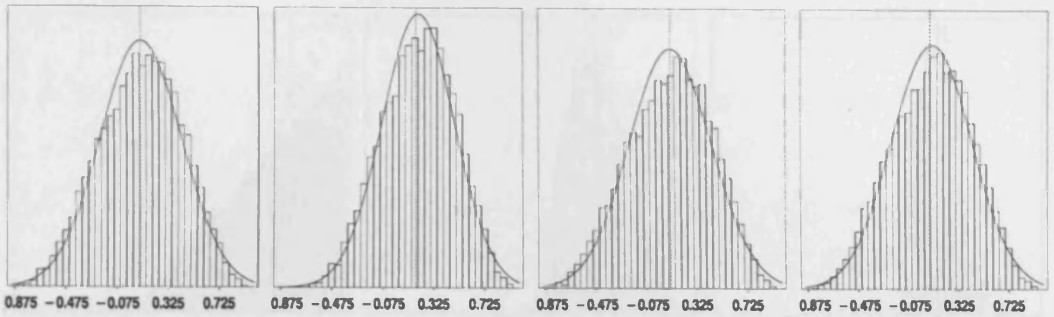
(a) $(nc_0)^{\frac{1}{\alpha}}(m^\circ - m)$ (b) $\frac{(m^\circ - m)\eta}{y_{1,n} - m}$ (c) $\frac{m^\circ - m}{y_{1,n} - m}$ (d) $\frac{m^\circ - m}{y_{k,n} - m}$

(e) $(nc_0)^{\frac{1}{\alpha}}(m^{(2)} - m)$ (f) $\frac{(m^{(2)} - m)\eta}{y_{1,n} - m}$ (g) $\frac{m^{(2)} - m}{y_{1,n} - m}$ (h) $\frac{m^{(2)} - m}{y_{k,n} - m}$

Figure 2.5: Histograms of 10 000 estimates of m from the Weibull distribution with parameter $\alpha = 3$. The estimator in (a), (b), (c) and (d) is the optimal linear estimator and in (e), (f), (g) and (h) it is $m^{(2)}$. All estimators are using $k = 5$ order statistics from random samples of size 100. The different histograms show these estimates normalized different ways. The mean of the estimator has been plotted using a vertical dotted line.

k	Estimator	$\alpha = 2$		$\alpha = 3$		$\alpha = 5$	
		Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
2	$\frac{(m^\circ - m)}{y_{k,n} - m}$	0.32	0.47	0.37	0.48	0.42	0.50
	$\frac{(m^* - m)}{y_{k,n} - m}$	0.49	0.36	0.50	0.39	0.50	0.43
5	$\frac{(m^\circ - m)}{y_{k,n} - m}$	0.07	0.28	0.11	0.32	0.13	0.35
	$\frac{(m^* - m)}{y_{k,n} - m}$	0.20	0.24	0.20	0.29	0.19	0.33
	$\frac{(m^{(2)} - m)}{y_{k,n} - m}$	0.08	0.28	0.12	0.33	0.15	0.37
	$\frac{(m^{(3)} - m)}{y_{k,n} - m}$	0.08	0.28	0.11	0.32	0.14	0.36
10	$\frac{(m^\circ - m)}{y_{k,n} - m}$	0.02	0.18	0.03	0.22	0.05	0.25
	$\frac{(m^* - m)}{y_{k,n} - m}$	0.10	0.16	0.09	0.20	0.08	0.24
	$\frac{(m^{(2)} - m)}{y_{k,n} - m}$	0.03	0.18	0.05	0.23	0.07	0.28
	$\frac{(m^{(3)} - m)}{y_{k,n} - m}$	0.02	0.18	0.04	0.22	0.05	0.26

Table 2.5: The mean and standard deviation of estimators. The statistics were calculated from 10 000 runs. In each run each estimator was used to estimate m from a sample of size $n = 100$ taken from the Weibull distribution (with parameter α as indicated). k order statistics were used to calculate each estimator. The estimators considered are the three linear estimators m°, $m^{(2)}$ and $m^{(3)}$, and the MLE, m^*.



(a) $\frac{m^\circ - m}{y_{k,n} - m}$ (b) $\frac{m^* - m}{y_{k,n} - m}$ (c) $\frac{m^{(2)} - m}{y_{k,n} - m}$ (d) $\frac{m^{(3)} - m}{y_{k,n} - m}$

Figure 2.6: Histograms of 10 000 estimates of m from the Weibull distribution with parameter $\alpha = 3$. The estimates were made using $k = 5$ order statistics from samples of size 100 and normalized using normalization (4). The different histogram show estimates made using different estimators. The mean of the estimator has been plotted using a vertical dotted line. A normal distribution curve has been fitted to each distribution and plotted.

For all α and k the standard deviation of the MLE is smaller than that of the linear estimators, but the mean of the MLE is larger than that of the linear estimators.

From Figure 2.7 it can be seen that the distribution of the estimator when $k = 2$ is very different from the distribution when $k > 2$. First, considering the case when $k = 2$; from Section 2.2 we can see that the density of the normalized estimators $(m^\circ - m)/(y_{k,n} - m)$ and $(m^* - m)/(y_{k,n} - m)$ is a function of $z^{\alpha-1}$. This means that when $\alpha = 2$, the density is a linear function, when $\alpha = 3$ the density is a quadratic and so on. This is verified by the histograms. When $k > 2$ then densities become more symmetric. As α increases the distributions of the estimators become negatively skewed.

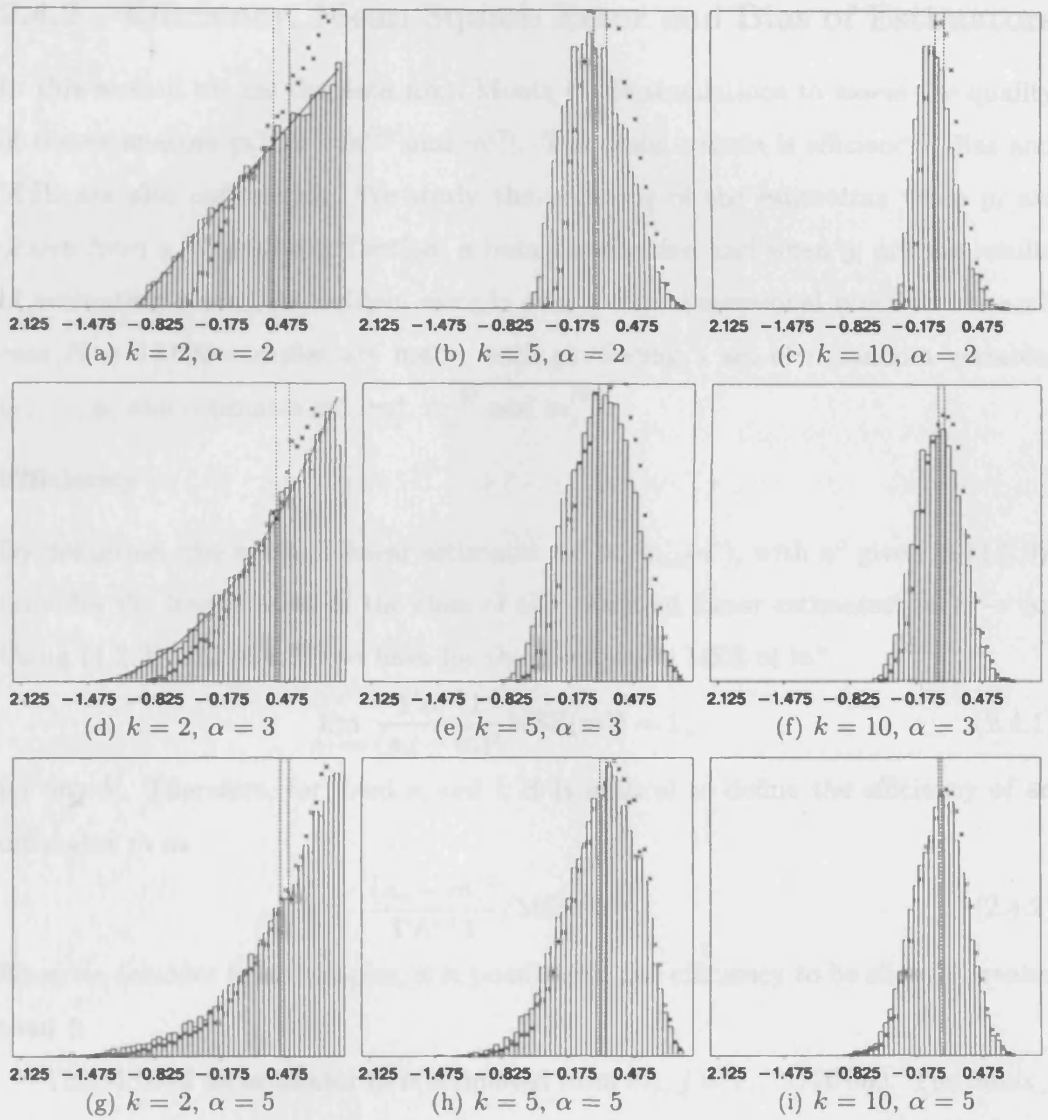


Figure 2.7: Histograms of 10 000 normalized estimates of m from the Weibull distribution with parameter $\alpha = 2$, 3 or 5. The estimates were made using the optimal linear estimator using $k = 2$, 5 or 10 order statistics from random samples of size 100 and normalized using normalization (4). From the same 10 000 samples of 100 random variables the normalized estimator $(m^* - m)/(y_{k,n} - m)$ was calculated. The percentage frequencies of normalized estimators $(m^* - m)/(y_{k,n} - m)$ (crosses) were plotted at each of the midpoints of the original histogram. The sample means for $(m^\circ - m)/(y_{k,n} - m)$ and $(m^* - m)/(y_{k,n} - m)$ are plotted using vertical dashed lines. The sample mean of the MLE is greater than that of the optimal linear estimator in all of the above plots. The density of $(m^\circ - m)/(y_{k,n} - m)$ (derived in section 2.2.4) has been plotted on the histograms relating to $k = 2$ with a solid line.

2.4.2 Efficiency, Mean Square Error and Bias of Estimators

In this section we use the data from Monte Carlo simulations to assess the quality of the estimators m^*, m°, $m^{(3)}$ and $m^{(2)}$. The main criteria is efficiency. Bias and MSE are also considered. We study the behavior of the estimators when y_i are drawn from a Weibull distribution, a beta distribution and when y_i are the results of evaluating a uniform random sample over a four dimensional function. In each case $R = 10\,000$ samples are made, each producing a set of n random variables $y_1, \ldots, y_n$ and estimates m_j^*, m_j°, $m_j^{(3)}$ and $m_j^{(2)}$.

Efficiency

By definition, the optimal linear estimator $m^\circ = \hat{m}_{n,k}(a^\circ)$, with a° given in (1.3.9), provides the lowest MSE in the class of all consistent linear estimators as $n \to \infty$. Using (1.3.3) and (1.3.8), we have for the asymptotic MSE of m°:

$$\lim_{n\to\infty} \frac{\mathbf{1}'\Lambda^{-1}\mathbf{1}}{(\kappa_n - m)^2}\mathrm{MSE}(m^\circ) = 1\,, \tag{2.4.1}$$

for any k. Therefore, for fixed n and k it is natural to define the efficiency of an estimator $\hat{m}$ as

$$\frac{(\kappa_n - m)^2}{\mathbf{1}'\Lambda^{-1}\mathbf{1}}/\mathrm{MSE}(\hat{m})\,. \tag{2.4.2}$$

Since we consider finite samples, it is possible for the efficiency to be slightly greater than 1.

The MSE of an estimator $\hat{m}$ is estimated from $\hat{m}_j$, $j = 1, \ldots, 10\,000$. The index j indicates that each $\hat{m}_j$ is estimated from a different sample of size n. The estimated mean square error is given by the following expression:

$$\mathrm{MSE}(\hat{m}) \approx \frac{1}{R}\sum_{j=1}^{R}(\hat{m}_j - m)^2\,.$$

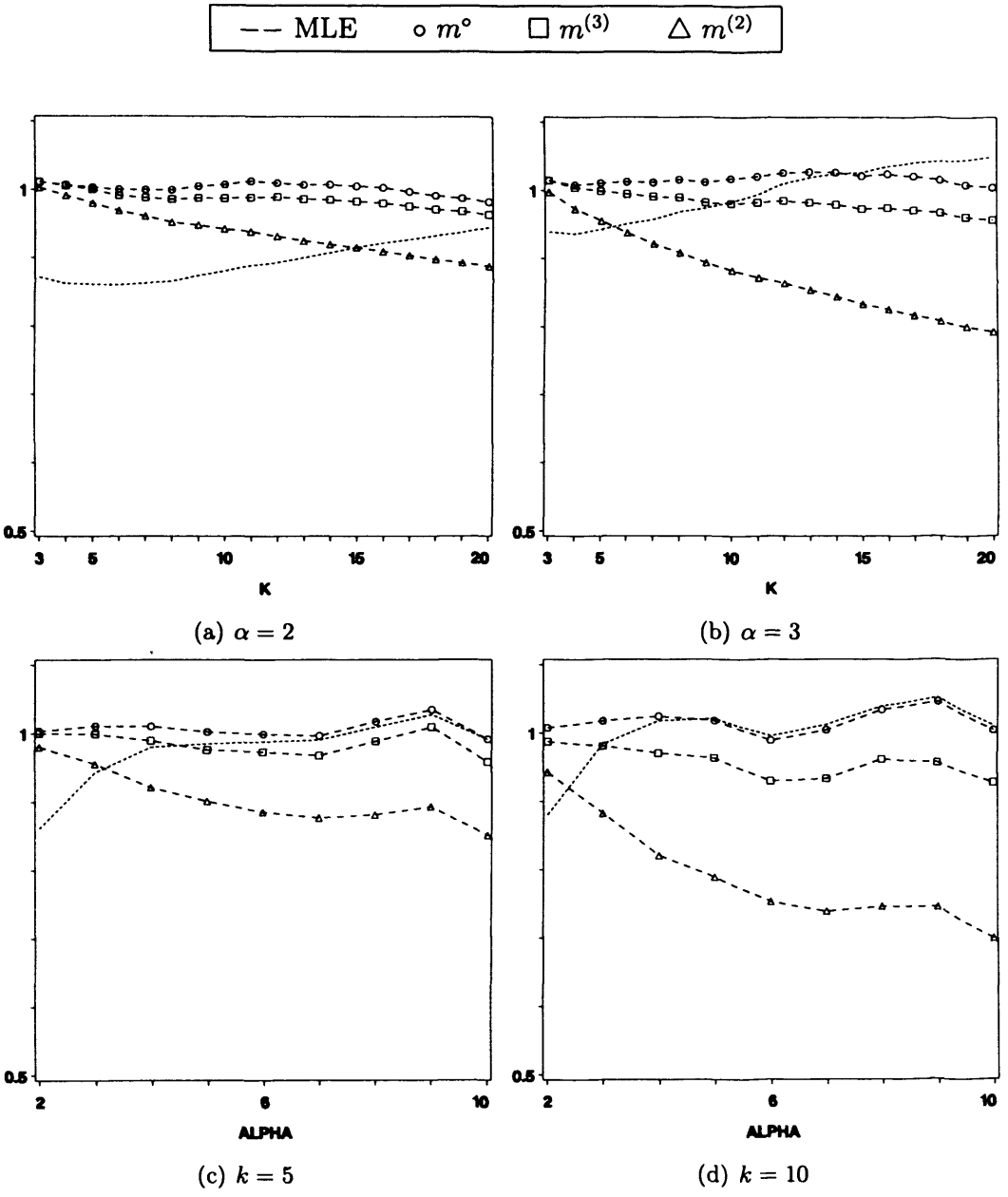
For fixed k, n and R, we use the following definition for efficiency of an estimator $\hat{m}$:

$$\mathrm{eff}(\hat{m}) = \left[\frac{(\kappa_n - m)^2}{\mathbf{1}'\Lambda^{-1}\mathbf{1}}\right] / \left[\frac{1}{R}\sum_{j=1}^{R}(\hat{m}_j - m)^2\right]\,, \tag{2.4.3}$$

where in our case $R = 10\,000$ and k varies. As $R \to \infty$, the finite-sample efficiency (2.4.3) tends to (2.4.2).

Simulation Study: Weibull Distribution

Figures 2.8-2.10 show the results from carrying out the above simulation study with y_i drawn from the Weibull distribution with parameter α. Using this distribution means that the tail index used in the estimators is α, $m = 0$ and $c_0 = 1$. In this simulation study the sample size was $n = 100$. The results that can be seen from these figures are as follows. The MLE does not perform well for small α. In particular, the efficiency of the MLE is lower than the optimal linear estimator for all values of k when $\alpha = 2$. The same is true of the bias except for large k. In general for small k and for small α the MLE is outperformed by the optimal linear estimator. For simultaneously small α and k the simple estimator $m^{(3)}$ also outperforms the MLE. The simple estimator $m^{(2)}$ also outperforms MLE when $\alpha = 2$ and k is not too large.



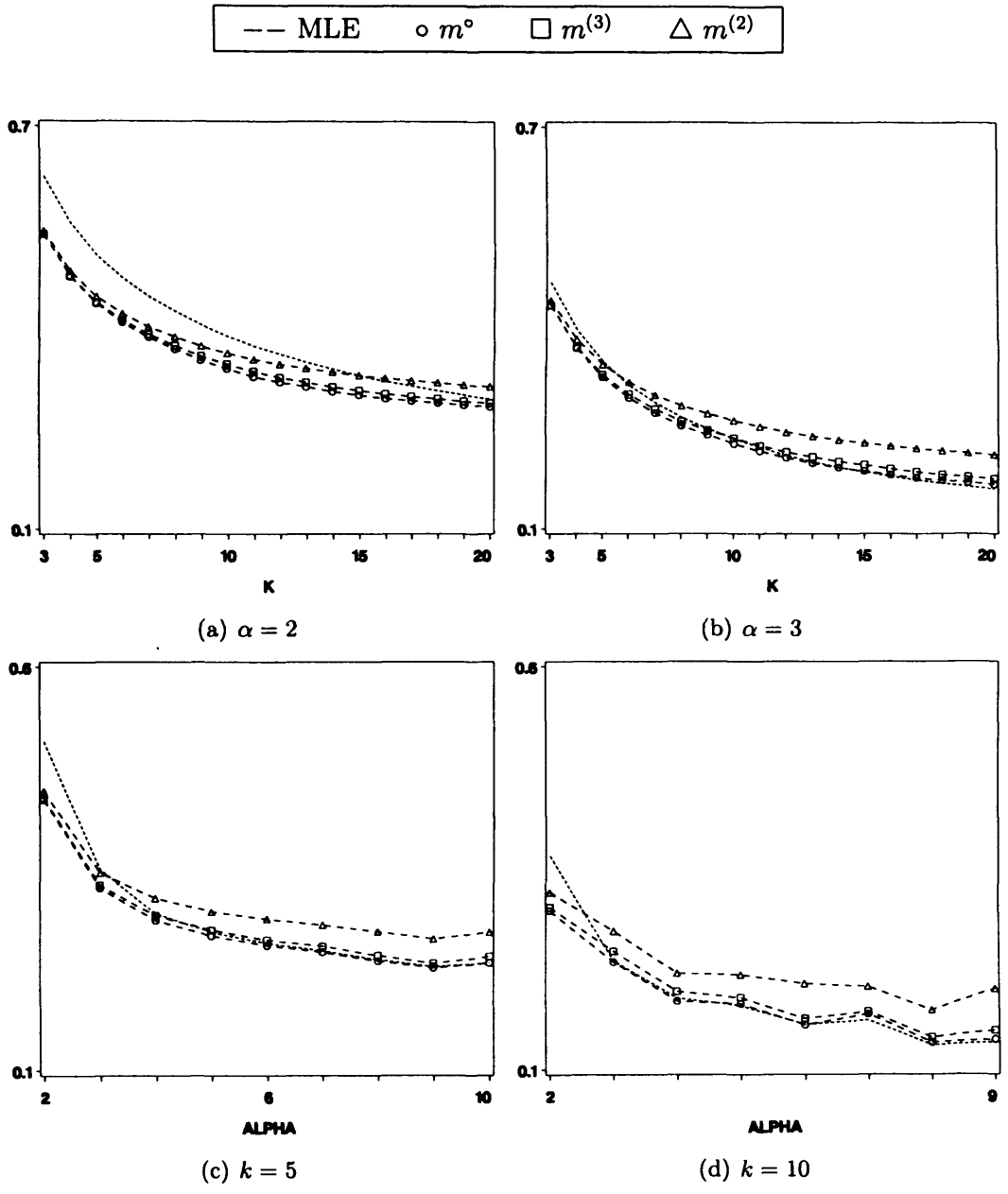
(a) $\alpha = 2$

(b) $\alpha = 3$

(c) $k = 5$

(d) $k = 10$

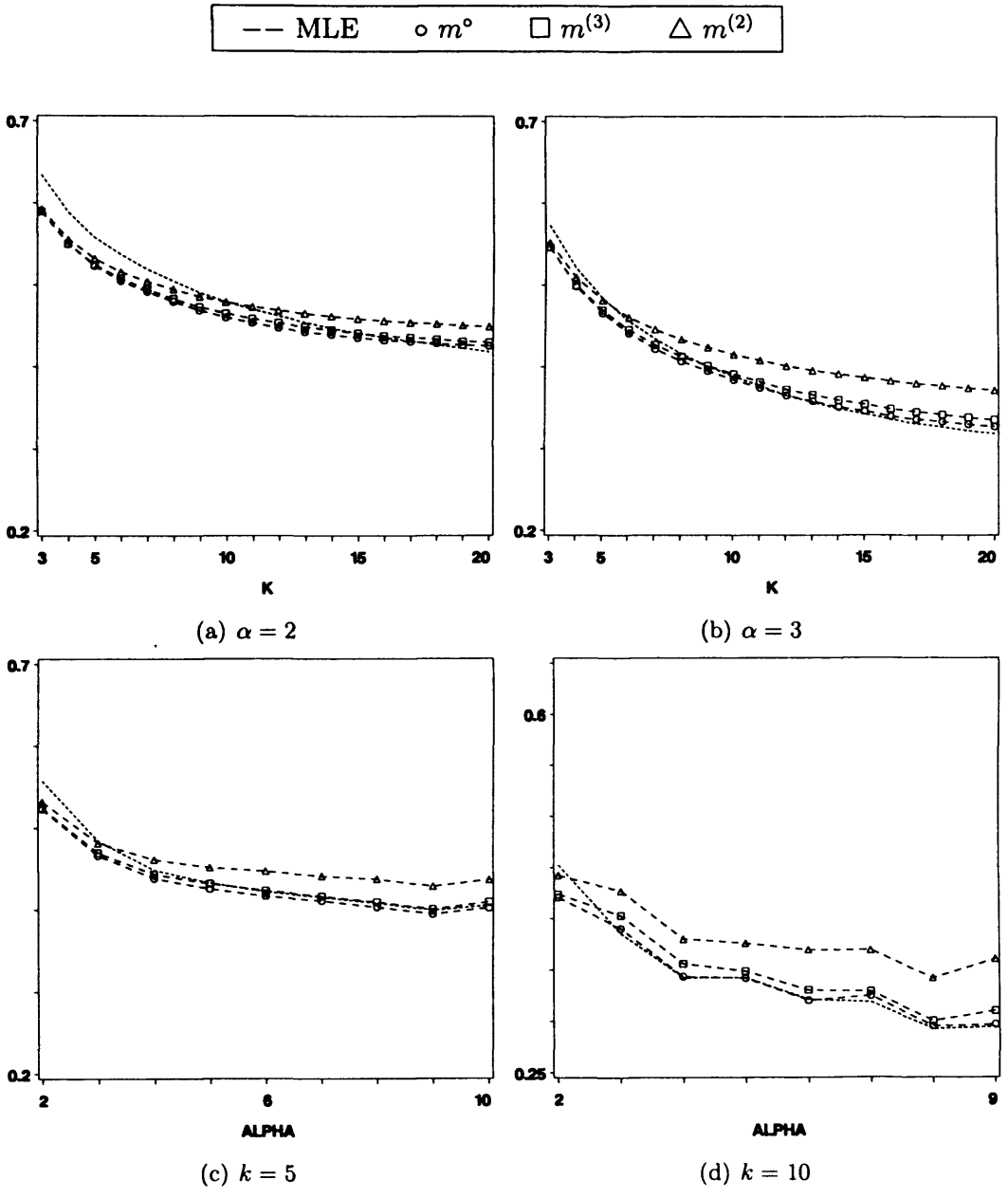
Figure 2.8: Estimated efficiency (2.4.3) of each estimator with respect to the optimal linear estimator (asymptotic). In figures (a) and (b) the estimators are based on $k = 2 \ldots 20$ order statistics. In figures (c) and (d) estimators are based on $k = 5$ and 10 order statistics respectively. The sample size used is $n = 100$. $F(x)$ is Weibull in all figures. The parameter of this distribution in figures (a) and (b) is $\alpha = 2$ and 3 respectively. In figures (c) and (d) α takes integer values in the range $[2, 10]$.



(a) $\alpha = 2$ (b) $\alpha = 3$

(c) $k = 5$ (d) $k = 10$

Figure 2.9: Normalized estimated mean square error $(nc_0)^{1/\alpha}\frac{1}{R}\sum_{j=1}^{R}(\hat{m}_j - m)^2$ of estimators. In figures (a) and (b) the estimators are based on $k = 2 \ldots 20$ order statistics. In figures (c) and (d) estimators are based on $k = 5$ and 10 order statistics respectively. The sample size used is $n = 100$. $F(x)$ is Weibull in all figures. The parameter of this distribution in figures (a) and (b) is $\alpha = 2$ and 3 respectively. In figures (c) and (d) α takes integer values in the range $[2, 10]$.



(a) $\alpha = 2$ (b) $\alpha = 3$

(c) $k = 5$ (d) $k = 10$

Figure 2.10: Normalized bias, $(nc_0)^{1/\alpha}\frac{1}{R}\sum_{j=1}^{R}|\hat{m}_j - m|$ of estimators. In figures (a) and (b) the estimators are based on $k = 2 \ldots 20$ order statistics. In figures (c) and (d) estimators are based on $k = 5$ and 10 order statistics respectively. The sample size used is $n = 100$. $F(x)$ is Weibull in all figures. The parameter of this distribution in figures (a) and (b) is $\alpha = 2$ and 3 respectively. In figures (c) and (d) α takes integer values in the range $[2, 10]$.

Simulation Study: Beta Distribution

In this simulation study the distribution $F(x)$ was taken to be the beta distribution with c.d.f.

$$F_\beta(t;\alpha,\beta) = \frac{B_t(\alpha,\beta)}{B(\alpha,\beta)}, \quad t \in [0,1]$$

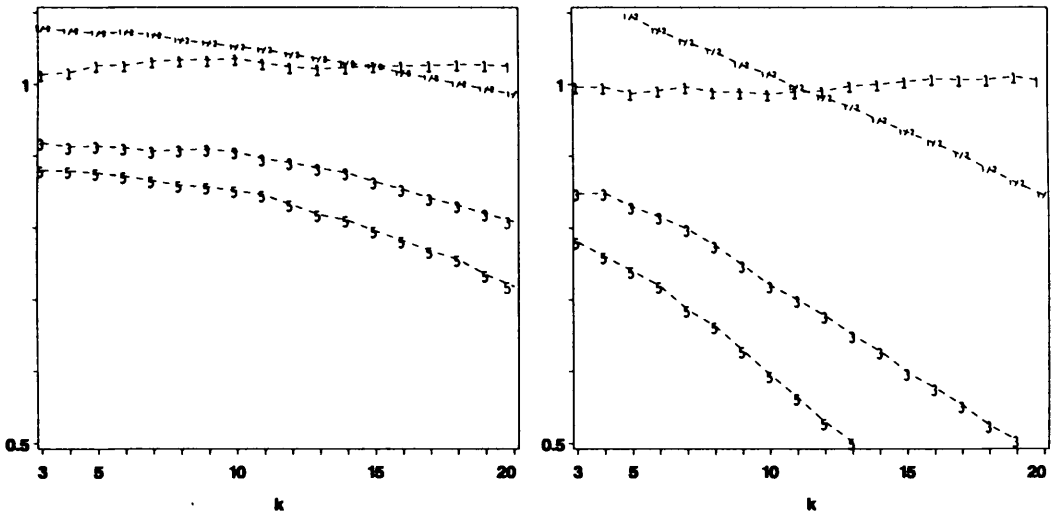
where $B_t(\alpha,\beta)$ is the incomplete beta function $\int_0^t x^{\alpha-1}(1-x)^{\beta-1}dx$ and $B(\alpha,\beta)$ is the beta function. We can approximate this c.d.f. by (1.2.5) where $c_0 = \frac{1}{B(\alpha,\beta)\alpha}$, the tail index is equal to α and $m = 0$. By setting $\beta = 1$ we have that $c_0 = 1$. Then the beta distribution is exactly equal to $F(x) = c_0(t-m)^\alpha$ (see (1.2.5)). It has been shown that for certain selections of the parameter β, extreme order statistics from a beta distribution do not converge quickly to their asymptotic distribution. For this reason sample sizes of both $n = 100$ and $n = 1\,000$ have been considered.

The results from this simulation study are as follows. When $\beta = 1$ the efficiency of the optimal linear estimator is approximately one for all k, α and n shown, implying that the asymptotic MSE has been reached in all these cases. Indeed, it is the case that the optimal linear estimator has reached asymptotic efficiency when $\beta = 1$, the same is also true of any other linear estimator: From (1.3.5) we can see that the analytic efficiency of $\hat{m}$ when $F(x)$ is the beta c.d.f. with $\beta = 1$ is given by

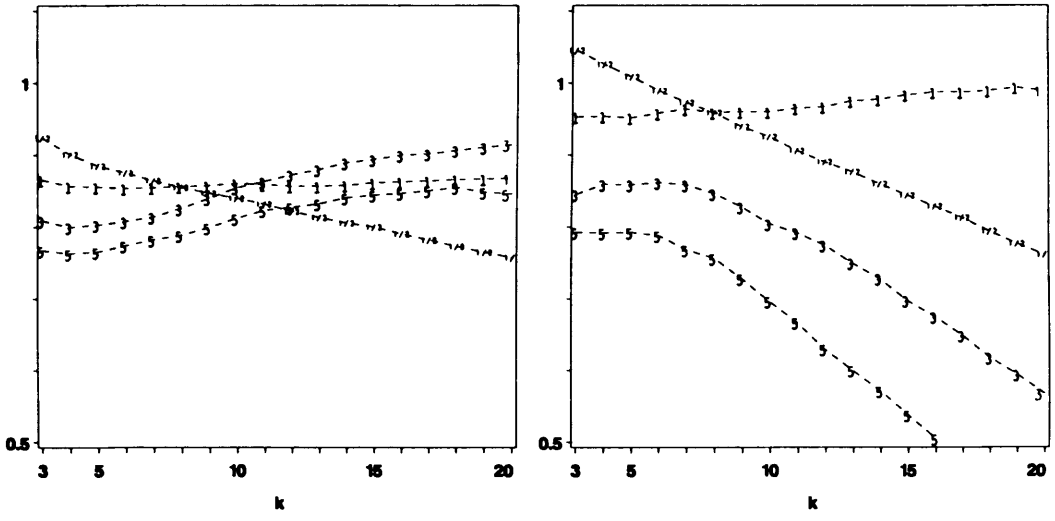
$$\frac{a^{\circ\prime}\Lambda a^\circ}{a'\Lambda a}\frac{\Gamma(n+1+\alpha/2)}{\Gamma(n+1)n^{2/\alpha})},$$

which is a good approximation to the asymptotic efficiency even for small n. The efficiency of the MLE when $\beta = 1$ is approximately constant for all k with fixed α. In all ($\beta = 1$) cases shown, the MLE has slightly worse efficiency than the optimal linear estimator, as α increases the MLE improves. When n is increased from 100 to 1 000 the efficiencies of m° and m^*, when $\beta \neq 1$ appear to converge to the efficiencies when $\beta = 1$ of m° and m^* respectively. When n is small, $\beta \neq 1$ and $\alpha > 2$, as k increases estimators m^* and m° become poor. This is because the k^{th} order statistic no longer within the region where (1.2.5) can be assumed true. It was also seen in Figure 2.4 that higher order statistic do not converge as quickly as lower order

statistics to their asymptotic distribution. For larger n this is still the case but not to such a degree. When $\beta = 3$ or 5 and $\alpha = 2$ the efficiency of m^* increases with k. For $n = 1\,000$ and $\beta = 0.5$, $\beta = 1$ or $\alpha = 2$ m° outperforms m^*.

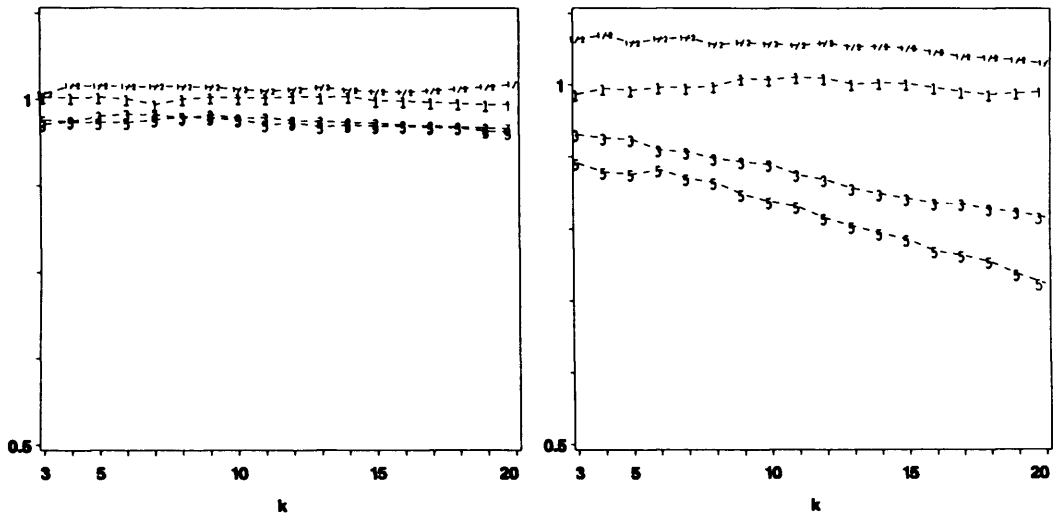


(a) m° $\alpha = 2$ (b) m° $\alpha = 4$



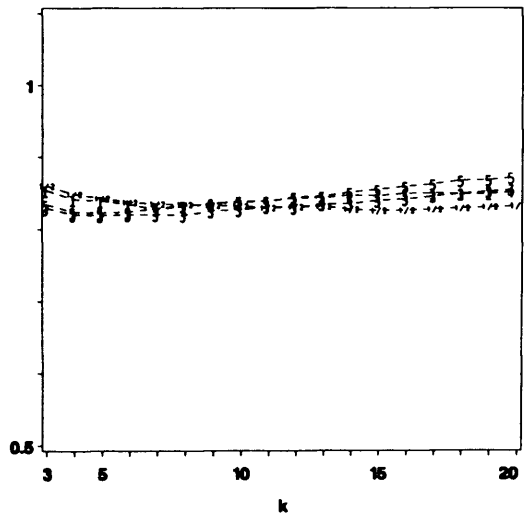
(c) m^* $\alpha = 2$ (d) m^* $\alpha = 4$

Figure 2.11: The estimated efficiency (2.4.3) with respect to the optimal linear estimator (asymptotic) of the estimators m° and m^*. 10 000 estimates were made from samples of size $n = 100$ that were drawn from the beta distribution with parameters $\beta = 0.5$, 1, 3 and 5 (as marked on different lines on plot) and $\alpha = 2$, and 4 (as marked below each plot).

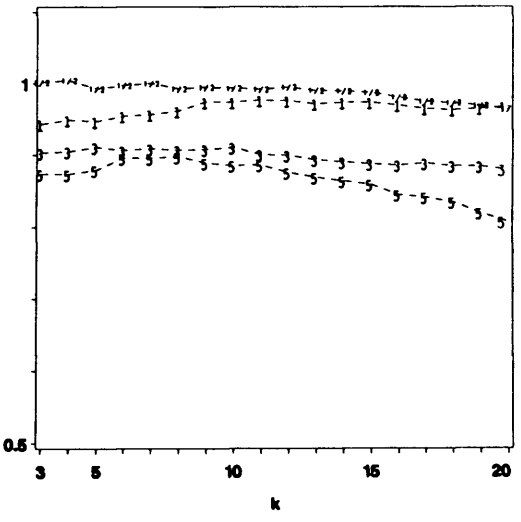


(a) m° $\alpha = 2$

(b) m° $\alpha = 4$



(c) m^* $\alpha = 2$



(d) m^* $\alpha = 4$

Figure 2.12: The estimated efficiency (2.4.3) with respect to the optimal linear estimator (asymptotic) of the estimators m° and m^*. 10 000 estimates were made from samples of size $n = 1\,000$ that were drawn from the beta distribution with parameters $\beta = 0.5$, 1, 3 and 5 (as marked on different lines on plot) and $\alpha = 2$, and 4 (as marked below each plot).

Simulation Study: Optimizing an Objective Function

A simulation study similar to the ones described for $F(x)$ Weibull and beta was conducted. The objective function

$$f(w,x,y,z) = 10\cos(2w+2.3) + \cos(12w) + 10\sin(2x-2.5) + \\ + \sin(14x) + \sin(4y+3.1) + \cos(8z-0.4) \qquad (2.4.4)$$

was considered over the feasible region $(w,x,y,z) \in [0,1]^4$. This function has four local minima. $f(0.70,0.73,0.40,0.44) = -20.3$, $f(0.70,0.36,0.40,0.44) = -21.73$, $f(0.30,0.73,0.40,0.44) = -21.95$ and $f(0.30,0.36,0.40,0.44) = -23.34$. The latter, $m^* = -23.34$ at $w = w^* = 0.30$, $x = x^* = 0.36$, $y = y^* = 0.40$, $z = z^* = 0.44$ is the global minimum. A random sample of n uniform random vectors on $[0,1]^4$ space was made. The objective function was evaluated at each of these, creating independent identically distributed random variables (i.i.d.r.v.) y_i with c.d.f. $F(x)$, where $F(x)$ is approximately given by the following expression:

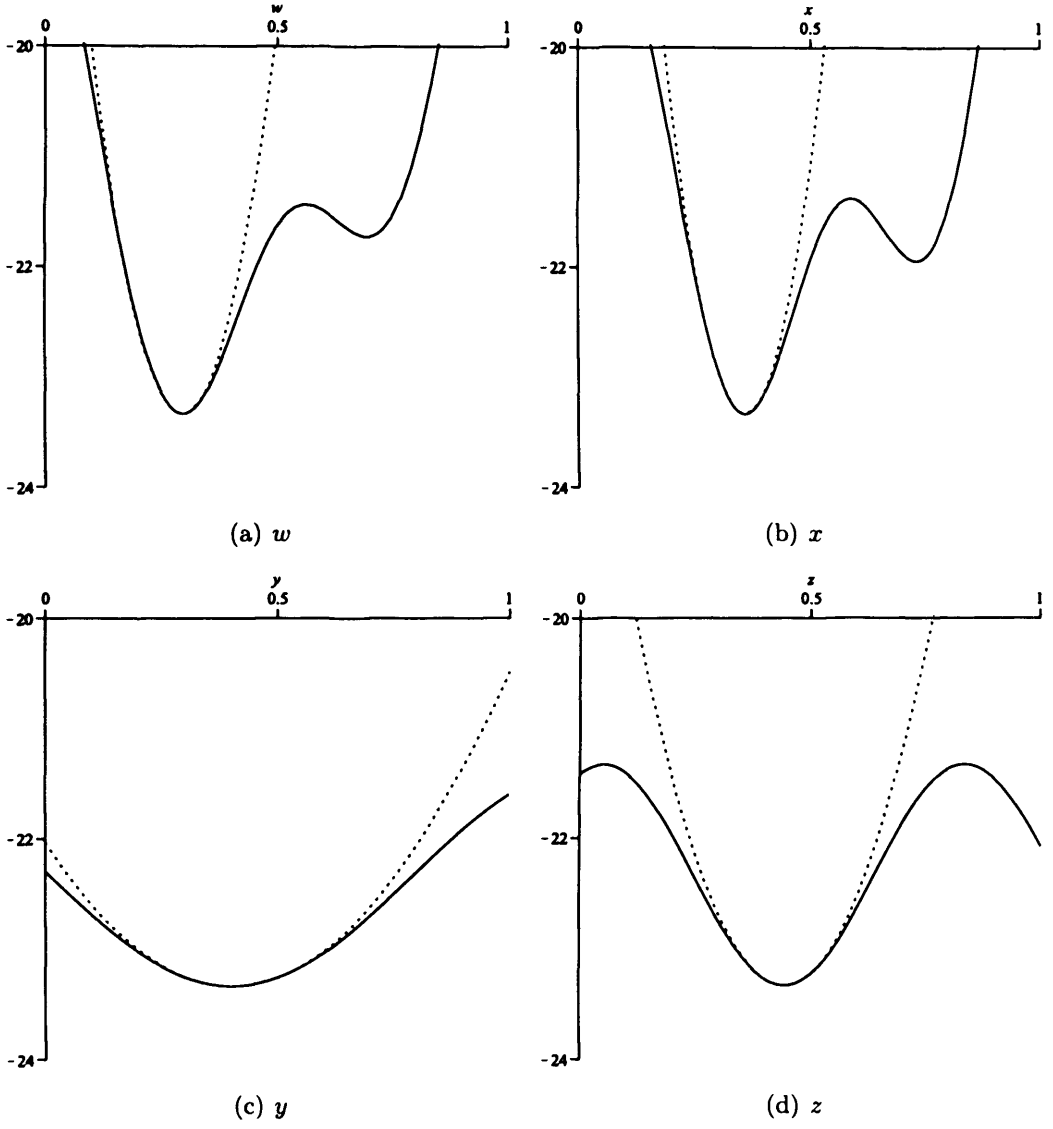
$$\frac{\pi^2\sqrt{9605}}{307360}(x+23.34)^2. \qquad (2.4.5)$$

This approximate c.d.f. is in the form $F(x) = c_0(x-m)^\alpha$, where $c_0 = \frac{\pi^2\sqrt{9605}}{307360} \approx 0.001$, $m = -23.34$ and $\alpha = 2$. The approximation has been found by approximating the original function $f(w,x,y,z)$ by $f_a(w,x,y,z)$ by performing a 4 dimensional Taylor expansion. The quadratic form $f_a(w,x,y,z)$ is given by

$$f_a(w,x,y,z) = \frac{170(w-w^*)^2 + 226(x-x^*)^2 + 16(y-y^*)^2 + 64(z-z^*)^2}{2} + m.$$

This approximation is valid close to the global minimum (see Figure 2.13).

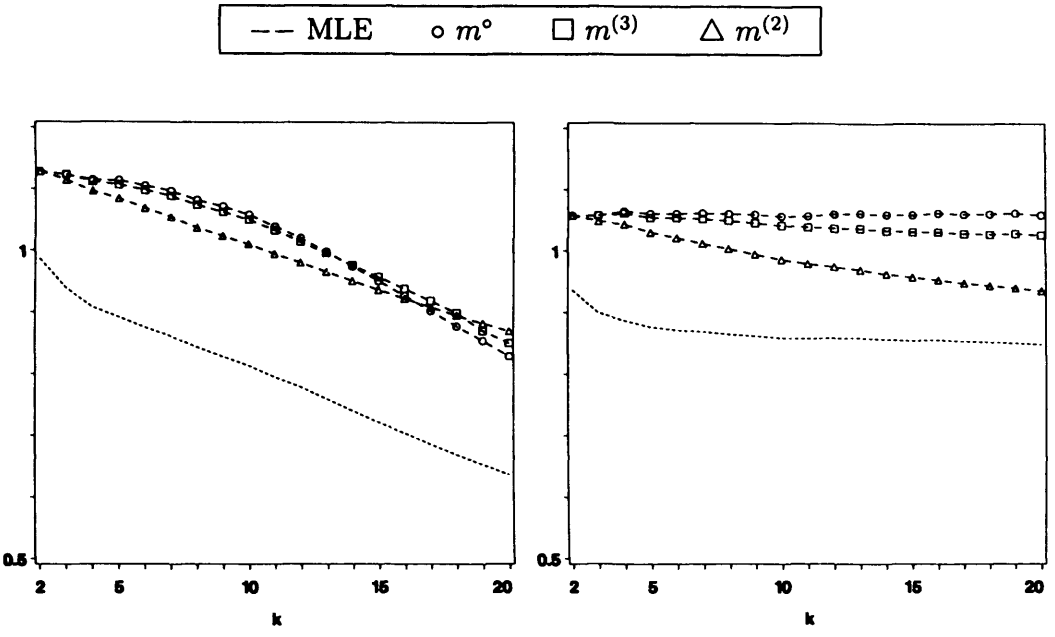
Using Theorem 2.2 from [37] we can conclude that the c.d.f. $F(x)$ meets the assumptions of Theorem 1, and that the value of the tail index, α, is $\alpha = d/2 = 4/2 = 2$.



(a) w (b) x

(c) y (d) z

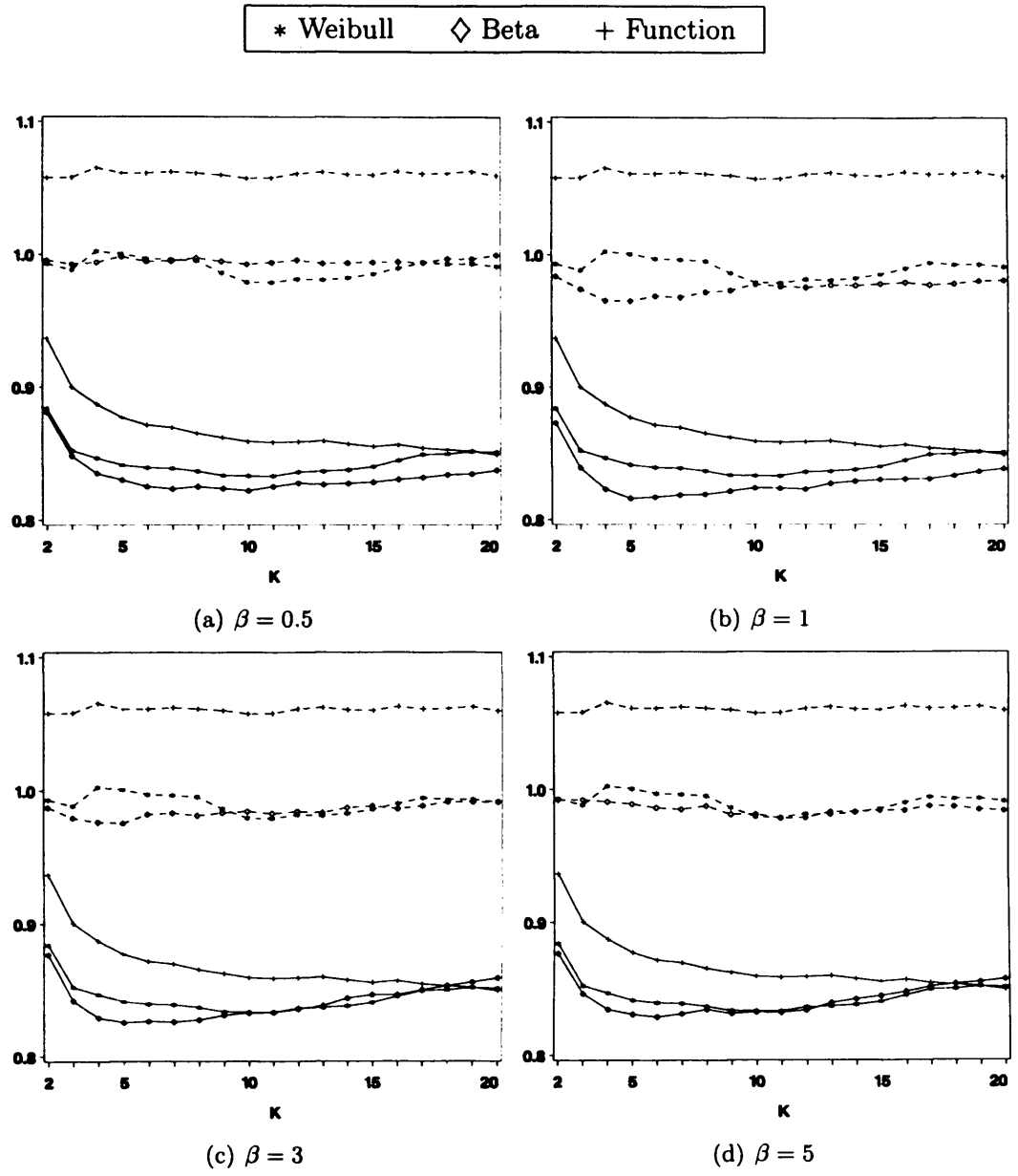
Figure 2.13: Projection of function (2.4.4). Three of the parameters are set to their optimal values; $w = w^*$, $x = x^*$, $y = y^*$, $z = z^*$, and the remaining value is allowed to range between 0 and 1. The free variable is labeled beneath each plot. Also plotted on the above graphs (dotted line) are the projections of the approximation $f_a(w, x, y, z)$.



(a) $n = 1000$ (b) $n = 10000$

Figure 2.14: Estimated efficiency (2.4.3) of each estimator with respect to the optimal linear estimator as defined above. The estimators are based on $k = 2 \ldots 20$ order statistics. Estimators assume approximation (2.4.5); $\alpha = 2$ and $c_0 \approx 0.001$. In plot (a) the sample size used is 1000, and 10 000 in plot (b).

It can be seen from Figure 2.14 that the MLE performs much worse than the optimal linear estimator when the c.d.f. $F(x)$ is the result of evaluating a sample of objective function values. This is particularly noticeable when n is small. This suggests that the MLE does not work well when the minimum order statistic is not Weibull. When the c.d.f. was set to be Weibull or beta with tail index $\alpha = 2$ the MLE was often outperformed by the optimal linear estimator, but this effect is seen to a much greater degree when the c.d.f. is the result of evaluating a function.



(a) $\beta = 0.5$ (b) $\beta = 1$

(c) $\beta = 3$ (d) $\beta = 5$

Figure 2.15: The estimated efficiency with respect to the asymptotic optimal linear estimator (2.4.3) of the optimal linear estimator (broken lines) and the MLE (solid lines). 10 000 estimates were made from samples of size $n = 10\,000$. The distributions used to make the estimators are either Weibull, beta or the distribution approximated by (2.4.5) (obtained by evaluating (2.4.4) at the points of a uniform random sample). Beta distribution has parameters $\beta = 0.5$, 1, 3, and 5 (as marked below plots). All distributions have tail index $\alpha = 2$.

From Figure 2.15 we can see that the optimal linear estimator performs much better than the MLE in estimating m from the Weibull distribution, the beta distributions and the four-dimensional function. When k is large compared to n the estimates of m from the four-dimensional function become very poor. This is because the k^{th} order statistic is likely to fall in the neighborhood of a local minima rather than the global minimum. It can be seen from Figure 2.15 that when n is large, m° and m^* have larger efficiency when estimating m from the four-dimensional function than from the Weibull distribution or the beta distribution. This is not because it is easier to estimate the minimum of a function than the endpoint of a distribution. The reason is likely to be that the tail index and c_0 that are used in the estimators m^* and m° (and the calculation of efficiency) are estimates from the approximation (2.4.5). We can obtain different estimates of these parameters by considering that the un-normalized minimum order statistic comes from a Weibull distribution with scale parameter $\sigma = \kappa_n - m$ and shape parameter α. Indeed $y_{1,n}$ has density $\psi(x;\sigma,\alpha)$ given by

$$\psi(x;\sigma,\alpha) = \frac{\alpha}{\sigma}\left(\frac{x}{\sigma}\right)^{\alpha-1} \exp\left(-\left(\frac{x}{\sigma}\right)\right), \quad x > 0$$

Using the sample of $y_{1,10\,000}$ (of size 10 000 obtained by the Monte Carlo simulation described above), we made maximum likelihood estimates of σ and α, $\tilde{\sigma} = 0.179$ and $\tilde{\alpha} = 2.105$ respectively. Compare these to the estimates from the approximation (2.4.5) of $\left(\frac{\pi\sqrt{9605}}{307360}10\,000\right)^{-1/2} = 0.178$ and 2. The two estimates of $\kappa_n - m$ are similar, whereas the two values of α are not so similar.

How Close is the Distribution to the Approximation?

In deriving the estimators and their efficiencies we have made the assumption that for t close to m, the c.d.f. $F(t)$ is approximately equal to $c_0(t-m)^\alpha$. It is of interest to investigate how close the distributions that we have investigated are to this assumption.

For the beta distribution we have that the ratio of the beta density to the assumed one, $c_0\alpha(t-m)^{\alpha-1}$, is given by:

$$\frac{t^{\alpha-1}(1-t)^{\beta-1}/B(\alpha,\beta)}{(t-m)^{\alpha-1}/B(\alpha,\beta)} = (1-t)^{\beta-1}.$$

The ratio of the Weibull density to $c_0\alpha(t-m)^{\alpha-1}$ is given by

$$\frac{\alpha t^{\alpha-1}\exp(-x^{\alpha})}{\alpha t^{\alpha-1}} = \exp(-t^{\alpha}).$$

For the beta and Weibull distributions, $m = 0$ and for the beta distribution $c_0 = \frac{1}{B(\alpha,\beta)\alpha}$. The supports of the beta distribution, the Weibull distribution and the assumed distribution are [0,1], $[1,\infty)$ and $[m,(c_0)^{1/\alpha}+m]$ respectively. The ratios are only valid within these supports.

The ratios show that (as previously discussed) when $F(x)$ is the beta distribution with parameter $\beta = 1$, beta density is exactly equal to the assumed density. When $\beta > 1$ and t is close to m, the beta density is smaller than the assumed density. When $\beta < 1$ and t is close to m, the beta density is larger than the assumed density. The ratio of the beta distribution to the assumed distribution does not depend on α. When $F(x)$ is Weibull the density is always smaller than the assumed distribution, however, as α increases the Weibull density converges to the assumed density. The ratios are plotted in Figure 2.16 for various β an α.

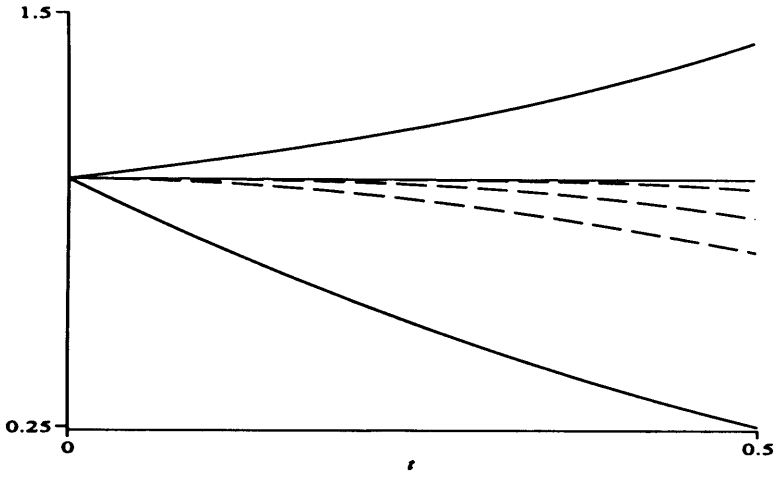


Figure 2.16: The ratio of the beta distribution to the distribution $c_0\alpha t^{\alpha-1}$ and the ratio of the Weibull distribution to the distribution $c_0\alpha t^{\alpha-1}$. Here the beta distribution has parameter $\beta = 0.5$, 1 or 3 (higher, middle and lower solid lines) and the Weibull distribution has parameter $\alpha = 5$, 3 or 2 (higher, middle and lower dash lines respectively).

2.5 Comparison of estimators: Unknown α

In reality it is not always possible to know the correct value of the tail index. Even in the above case (in Section 2.4.2) of optimizing a known objective function, it was not possible to obtain the exact value of the tail index. In this section we give analytical and experimental results concerning how estimators react to using the wrong value of the tail index.

2.5.1 Efficiency: Wrong Tail Index

Let us first introduce some notation. Let $\hat{m}(\vartheta)$ be the estimator that is derived assuming that the tail index is ϑ. For example, when $\hat{m}(\vartheta) = m^\circ(\vartheta)$ this is given by $\hat{m}_{k,n}(a^\circ(\vartheta))$ where $a(\vartheta)$ is the vector of coefficients given by

$$a^\circ(\vartheta) = \frac{\Lambda_\vartheta^{-1}\mathbf{1}}{\mathbf{1}'\Lambda_\vartheta^{-1}\mathbf{1}}.$$

Here Λ_ϑ, is the matrix $\Lambda = ||\lambda_{i,j}||$ as defined in (1.3.4) with ϑ in place of α. The MLE with the wrong value of the tail index is denoted $m^*(\vartheta)$ and is the solution $z \in (\infty, y_{1,n}]$ to the following equation.

$$(\vartheta - 1)\sum_{i=1}^{k-1} \frac{y_{k,n} - y_{i,n}}{y_{i,n} - z} = k.$$

We wish to find the asymptotic efficiency $\text{eff}(\hat{m}(\vartheta))$. For a general consistent linear estimator with vector of coefficients $a(\vartheta)$ this is given by

$$\text{eff}(\hat{m}(\vartheta)) = \frac{1}{\mathbf{1}'\Lambda^{-1}\mathbf{1} \cdot a'(\vartheta)\Lambda a(\vartheta)}. \tag{2.5.1}$$

The efficiency of $m^\circ(\vartheta)$ and $m^\bullet(\vartheta)$ can easily be found analytically for all $k \geq 2$. The minimum order statistic is not affected by the choice of ϑ, so is useful to asses how robust the other estimators are under changes to ϑ. The efficiency of $m^*(\vartheta)$ has only been found analytically for $k = 2$.

Let us first consider the efficiency of the estimator $m^\circ(\vartheta)$. This can easily be derived by substituting $a^\circ(\vartheta)$ into (2.5.1), it can also be found in [37]. For general k we have that

$$\text{eff}(m^\circ(\vartheta)) = \frac{(\mathbf{1}'\Lambda_\vartheta^{-1}\mathbf{1})^2}{\mathbf{1}\Lambda^{-1}\mathbf{1} \cdot \mathbf{1}'\Lambda_\vartheta^{-1}\Lambda\Lambda_\vartheta^{-1}\mathbf{1}}$$

Therefore for $k = 2$ the efficiency of $m^\circ(\vartheta)$ is given by:

$$\frac{\alpha + 2}{\alpha + 2 + \alpha(1 - \frac{\vartheta}{\alpha})}.$$

Notice that this efficiency is maximized when $\vartheta = \alpha$.

The asymptotic efficiency of the minimum order statistic for $k = 2$ is given by:

$$\frac{(\alpha + 2)}{2(\alpha + 1)}.$$

These efficiencies have been plotted on Figure 2.17 of simulated efficiencies.

The asymptotic efficiency of the MLE can be calculated when $k = 2$. It has an interesting property.

Lemma 2.5.1. *For $k = 2$, the efficiency, $eff(m^*(\vartheta))$ of $m^*(\vartheta)$ with respect to $m^\circ(\alpha)$ is given by*

$$eff(m^*(\vartheta)) = \frac{\alpha(\alpha+2)}{2\alpha^2 + 4\alpha - 2\alpha\vartheta - 2\vartheta + \vartheta^2 + 1}. \tag{2.5.2}$$

this is maximized when $\vartheta = \alpha + 1$. The asymptotic efficiency at this maximum is $\max_\vartheta eff(m^*(\vartheta)) = 1$.

Proof. For $k = 2$, there is exactly one solution, z, to the likelihood equation (1.3.3), it is given by:

$$m^*(\alpha) = y_{1,n} - \frac{(\alpha-1)}{2}(y_{2,n} - y_{1,n}).$$

The asymptotic MSE of the estimator $m^\circ(\alpha)$ is given by substituting (1.3.8) into (1.3.3):

$$\mathbf{E}(m^\circ(\alpha) - m)^2 \sim \frac{(\kappa_n - m)^2}{\mathbf{1}'\Lambda^{-1}\mathbf{1}}, \quad n \to \infty$$

Now consider $\mathbf{E}(m^*(\vartheta) - m)^2$:

$$\mathbf{E}(m^*(\vartheta) - m)^2 = \mathbf{E}\left(y_{1,n} - \frac{(\vartheta-1)}{2}(y_{2,n} - y_{1,n}) - m\right)^2 \tag{2.5.3}$$

From Lemma 7.1.3 in [35], we have that under conditions already assumed

$$\mathbf{E}(y_{i,n} - m)(y_{j,n} - m) \sim (\kappa_n - m)^2\lambda_{i,j}, \quad n \to \infty. \tag{2.5.4}$$

Therefore we can find $\text{eff}(m^*(\vartheta))$ when the random variables y_i are drawn from a Weibull distribution by expanding (2.5.3) and using (2.5.4);

$$\begin{aligned}
\mathbf{E}(m^*(\vartheta) - m)^2 &= \frac{(\vartheta+1)^2}{4}\mathbf{E}(y_{1,n}^2) - \frac{(\vartheta^2-1)}{2}\mathbf{E}(y_{1,n}y_{2,n}) + \frac{(\vartheta-1)^2}{4}\mathbf{E}(y_{2,n}^2) \\
&= (\kappa_n - m)^2\frac{(2\alpha^2 + 4\alpha - 2\alpha\vartheta - 2\vartheta + \vartheta^2 + 1)\Gamma(1+2/\alpha)}{2\alpha(\alpha+2)}
\end{aligned}$$

The efficiency is then found by calculating $\frac{\mathbf{E}(m^\circ(\alpha)-m)^2}{\mathbf{E}(m^*(\vartheta)-m)^2}$. Maximization by differentiating with respect to ϑ confirms that maximum efficiency is achieved at $\vartheta = \alpha + 1$. Substitution of $\vartheta = \alpha + 1$ into $\text{eff}(m^*(\vartheta))$ gives $\text{eff}(m^*(\alpha+1)) = 1$. □

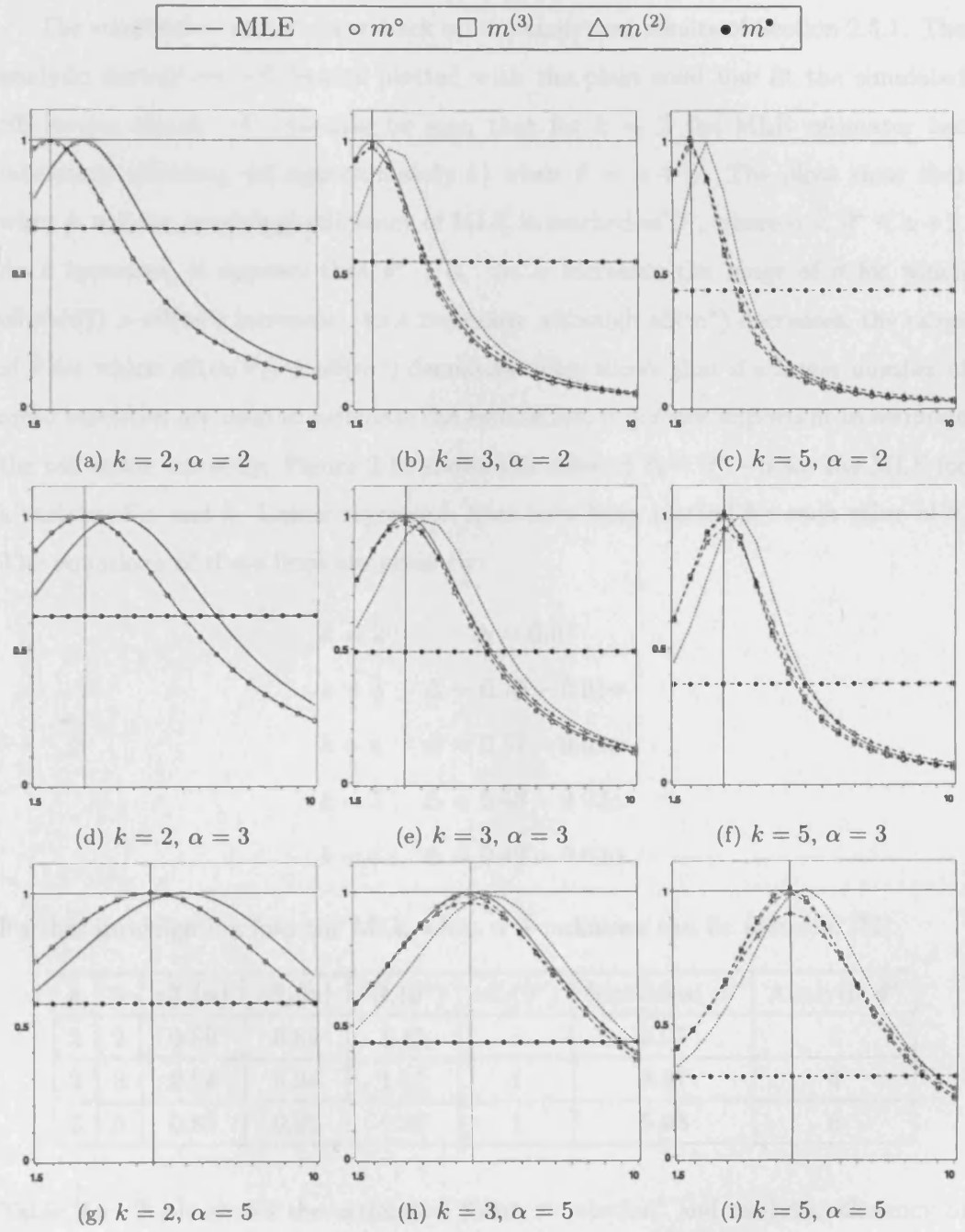
Notice that it was shown in Table 2.4 and the discussion that followed (Section 2.3.1), that under the conditions that $k = 2$ and $F(x)$ is Weibull with parameter α, the normalized estimator $m^*(\alpha + 1)$ has the same density as the normalized estimator $m^{(2)}(\alpha) = m^\circ(\alpha)$.

2.5.2 Simulations Study

In this section we find the estimated efficiency of each estimator when the value of the tail index is incorrectly assumed to be ϑ (instead of the correct tail index, α). $R = 10\,000$ samples of size $n = 100$ were drawn from a Weibull distribution with parameter $\alpha = 2$, 3, 4 and 5. Estimates of m were made using each of the five estimators defined in Sections 1.3 and 2.1. When making these estimates the value of the tail index of the distribution was assumed to be $\vartheta = 2, 2.5, \ldots, 10$. The estimates made using the wrong value of the tail index have been denoted $m^*(\vartheta)$, $m^\circ(\vartheta)$, $m^{(3)}(\vartheta)$ and $m^{(2)}(\vartheta)$. The simplest estimator, the minimum order statistic, make no assumption about the tail index, it has been used for comparison.

The efficiency of each of these estimators has been calculated with respect to the optimal linear estimator using correct value of the tail index; $m^\circ(\alpha)$.

$$\text{eff}(\hat{m}(\vartheta)) = \left[\frac{(\kappa_n - m)^2}{\mathbf{1}'\Lambda^{-1}\mathbf{1}}\right] \Big/ \left[\frac{1}{R}\sum_{j=1}^{R}(\hat{m}_j(\vartheta) - m)^2\right], \tag{2.5.5}$$



(a) $k = 2$, $\alpha = 2$ (b) $k = 3$, $\alpha = 2$ (c) $k = 5$, $\alpha = 2$

(d) $k = 2$, $\alpha = 3$ (e) $k = 3$, $\alpha = 3$ (f) $k = 5$, $\alpha = 3$

(g) $k = 2$, $\alpha = 5$ (h) $k = 3$, $\alpha = 5$ (i) $k = 5$, $\alpha = 5$

Figure 2.17: $\text{eff}(\hat{m})(\vartheta)$ plotted against ϑ. Efficiency is estimated from 10 000 samples of size 100 from a Weibull distribution with parameter α. The number of order statistics used in each estimator is k.

The simulation results above back up the analytical results of Section 2.5.1. The analytic asymptotic efficiencies plotted with the plain solid line fit the simulated efficiencies closely. It can also be seen that for $k = 2$ the MLE estimator has maximum efficiency (of approximately 1) when $\vartheta = \alpha + 1$. The plots show that when $k > 2$ the maximum efficiency of MLE is reached at ϑ^*, where $\alpha < \vartheta^* < \alpha + 1$. As k increases, it appears that $\vartheta^* \downarrow \alpha$. As α increases the range of ϑ for which $\text{eff}(\hat{m}(\vartheta)) > \text{eff}(m^\bullet)$ increases. As k increases, although $\text{eff}(m^\bullet)$ decreases, the range of ϑ for which $\text{eff}(\hat{m}(\vartheta)) > \text{eff}(m^\bullet)$ decreases. This shows that if a larger number of order statistics are used to calculate the estimators, it is more important to estimate the tail index correctly. Figure 2.18 shows the value of $\Delta = \vartheta^* - \alpha$ for the MLE for a variety of α and k. Linear regression lines have been plotted for each value of k. The equations of these lines are given by:

$$
\begin{aligned}
k = 2: &\quad \Delta = 0.97 \\
k = 3: &\quad \Delta = 0.70 - 0.01\alpha \\
k = 4: &\quad \Delta = 0.57 - 0.01\alpha \\
k = 5: &\quad \Delta = 0.48 - 0.02\alpha \\
k = 6: &\quad \Delta = 0.42 - 0.03\alpha
\end{aligned}
$$

Further investigation into the MLE when α is unknown can be found in [32].

k	α	$\text{eff}_s(\alpha)$	$\text{eff}_a(\alpha)$	$\text{eff}_s(\vartheta^*)$	$\text{eff}_a(\vartheta^*)$	Simulated ϑ^*	Analytic ϑ^*
2	2	0.89	0.89	1.11	1	2.97	3
2	3	0.94	0.94	1.01	1	3.97	4
2	5	0.98	0.97	1.01	1	5.98	6

Table 2.6: Table shows the estimated (from simulation) and analytic efficiency of $m^*(\alpha)$ ($\text{eff}_s(\alpha)$ and $\text{eff}_a(\alpha)$ respectively), estimated and analytic efficiency of $m^*(\vartheta^*)$ ($\text{eff}_s(\vartheta^*)$ and $\text{eff}_a(\vartheta^*)$ respectively). Also shown is ϑ^* found through simulation and analytically. ϑ^* is the value of ϑ that maximizes the efficiency.

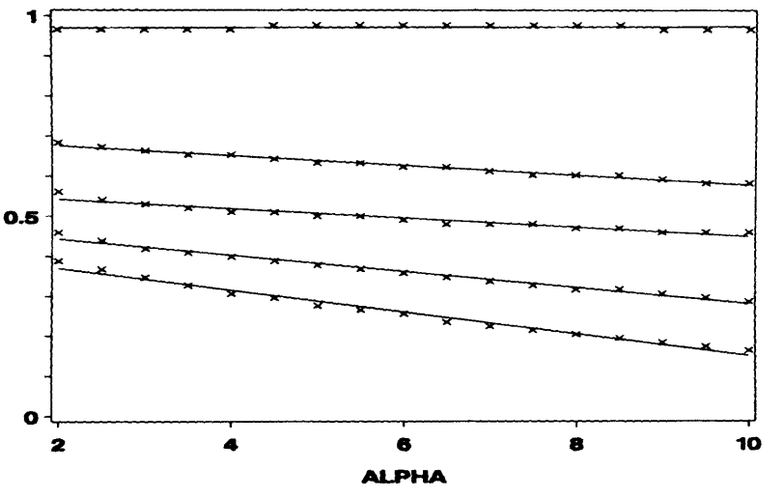


Figure 2.18: $\vartheta^* - \alpha$ plotted against α for MLE when $k = 2, \ldots, 6$. $\vartheta^* - \alpha$ decreases as k increases. v^* is estimated by taking 10 000 sample of size 100 from a Weibull distribution with parameter α and making estimates of m assuming that the tail index of the distribution is ϑ. In the simulation values of ϑ were tested in increments of 0.01.

Chapter 3

Comparison of Estimators: Increasing Sample Size

In this Chapter we consider the sample $Y_n = \{y_1, \ldots, y_n\}$ as a time series. In Chapter 2 we analyzed estimators of m at a fixed sample size, here we analyze the estimators as n increases.

As n increases, the k smallest order statistics related to the sample change. When there is a change in the order statistics we say that a record has occurred. In Sections 3.1.1, 3.1.2 and 3.1.3 we formally define records and some related statistics. In Section 3.1.4 we give some well known results about the frequency with which we expect to see records. We show that records occur very infrequently. We show links between record occurrence and Poisson processes. In Section 3.2 we show that we can model order statistics from a uniform distribution at record times. We create a stationary Markov chain that can be used to model normalized estimators of m at record times. This is very useful due to the fact that records occur infrequently, so this model enables us to use simulation to observe estimators at sample sizes that would not be possible through direct simulation. We analyze the Markov chain and the random variables that model estimators. We then conduct a simulation study using this model.

3.1 Introduction to Records and Poisson Process

3.1.1 Notation and Definitions

An important consideration in the estimation of the endpoint of a distribution is how large to make the sample size n. In general as n increases the estimates improve. However, as n increases the order statistics on which they are based update very infrequently. This means that the sample size can be increased dramatically without seeing any improvement in the estimate. In order to describe the process of the updating of order statistics (and hence changes of the estimate) we must consider the random sample $y_1, \ldots y_n$ as a stochastic process, rather than as n random variables observed simultaneously. So observation y_n is the random variable that occurs at time n (when the sample size is n).

The definitions below give a systematic means of recording when and how often records occur. For consistency, notation is mainly taken from [26]. We define a type 1 record and a type 2 record, both of these definitions are used in literature.

In Chapter 5 we will consider a discrete time series. In this situation new observations can occur that are equal to the current record value. This is known as a weak record. Weak records are defined in Section 3.1.2. When dealing with discrete data, records as defined in Section 3.1.1 are often called strong records. In Section 3.1.3 we consider maximal records. Table 3.1 shows notation relevant to these three sections.

Definitions

First we define the type 1 and type 2 k^{th} indicator functions. A type 1 k^{th} indicator function is equal to 1 if a type 1 k^{th} record has occurred. Similarly, a type 2 k^{th} indicator function is equal to 1 if a type 2 k^{th} record has occurred.

Name	(Strong) records		Weak records		Maximal records
	Type 1	Type 2	Type 1	Type 2	
k^{th} indicator function	$I^*_{k,n}$	$I_{k,n}$	$Iw^*_{k,n}$	$Iw_{k,n}$	Add + superscript
k^{th} record count	$N^*_{k,n}$	$N_{k,n}$	$Nw^*_{k,n}$	$Nw_{k,n}$	
k^{th} record Time	$T^*_{k,t}$	$T_{k,t}$	$Tw^*_{k,n}$	$Tw_{k,n}$	
k^{th} record value	$R^*_{k,t}$	$R_{k,t}$	$Rw^*_{k,n}$	$Rw_{k,n}$	
k^{th} record Value vector	$S^*_{k,n}$	$S_{k,n}$	$Sw^*_{k,n}$	$Sw_{k,n}$	
k^{th} waiting time	$W^*_{k,t}$	$W_{k,t}$	$Ww^*_{k,n}$	$Ww_{k,n}$	
Random variable	y_n				
k^{th} order statistic	$y_{k,n}$				

Table 3.1: Table of notation. Notation and definitions for records in continuous data are the same as those for strong records in discrete data.

Type 1 k^{th} indicator function:

$$I^*_{k,k} = \begin{cases} 1, & y_k > y_{k-1,k-1}, \\ 0, & \text{otherwise} \end{cases}$$

$$I^*_{k,n} = \begin{cases} 1, & y_{k-1,n-1} < y_n < y_{k,n-1} \\ 0, & y_n \geq y_{k,n-1} \text{ or } y_n \leq y_{k-1,n-1} \end{cases}$$

Type 2 k^{th} indicator function:

$$I_{k,k} = 1$$

$$I_{k,n} = \begin{cases} 1, & y_n < y_{k,n-1} \\ 0, & y_n > y_{k,n-1} \end{cases}$$

The requirements for a type 1 k^{th} indicator function to equal 1 are stricter than those for a type 2 k^{th} indicator function. The type 2 k^{th} indicator function, $I_{k,n}$, is equal to 1 if the new random variable y_n is less than or equal to the existing k^{th} order statistics from the sample at size $n-1$. However, the type 1 k^{th} indicator function, $I^*_{k,n}$, is only equal to 1 if y_n is between the k^{th} and $(k-1)^{\text{th}}$ order statistics from the sample at size $n-1$.

The number of times that the set of k smallest order statistics changes when increasing the sample size from k to n is given by the type 2 k^{th} record count, $N_{k,n}$ (defined below). Notice that an equivalent way to describe type 2 k^{th} record count is as the number of times that the k^{th} order statistic changes. Also defined below is the type 1 k^{th} record count, $N^*_{k,n}$; this is the number of times that a new random variable y_j causes the k^{th} order statistic to change without affecting order statistics $y_{1,j}, \ldots y_{k-1,j}$. Equivalently, this is the number of type 1 k^{th} records observed between times k and n.

Type 1 k^{th} and type 2 k^{th} record counts are given by the following expressions.

Type 1 k^{th} record count:

$$N^*_{k,n} = \sum_{j=k}^{n} I^*_{k,j} \quad n \geq k$$

Type 2 k^{th} records count:

$$N_{k,n} = \sum_{j=k}^{n} I_{k,j} \quad n \geq k$$

Notice that in the record count summations, the index variable j starts at $j = k$ not $j = 1$, this is because the indicator functions $I_{k,n}$ and $I^*_{k,n}$ are not defined for $n < k$.

A type 1 k^{th} record time, $T^*_{k,t}$, is the sample size at which a type 1 k^{th} record occurs for the t^{th} time. A type 2 k^{th} record time, $T_{k,t}$, is the sample size at which a type 2 k^{th} record occurs for the t^{th} time. See definitions below.

Type 1 k^{th} record time:

$$\begin{aligned} T^*_{k,1} &= \min_s(s > k-1 : I^*_{k,s} = 1) \\ T^*_{k,t} &= \min_s(s > T^*_{k,t-1} : I^*_{k,s} = 1),\ t > 1 \end{aligned}$$

Type 2 k^{th} record time:

$$\begin{aligned} T_{k,1} &= k \\ T_{k,t} &= \min_s(s > T_{k,t-1} : I_{k,n} = 1),\ t > 1 \end{aligned}$$

The t^{th} type 1 k^{th} record value is the value of the k^{th} order statistic at time $T^*_{k,t}$. Similarly, the t^{th} type 2 k^{th} record value is the value of the k^{th} order statistic at time $T_{k,t}$.

Type 1 k^{th} record value:

$$R^*_{k,t} = y_{k,T^*_{k,t}}, t \geq 1$$

Type 2 k^{th} record value:

$$R_{k,t} = y_{k,T_{k,t}}, t \geq 1$$

The record values can be concatenated to form a record values vector these are defined as follows.

Type 1 k^{th} record value vector:

$$S^*_{k,t} = \{R^*_{k,1}, R^*_{k,2}, \ldots R_{k,t}\}$$

Type 2 k^{th} record value vector:

$$S_{k,t} = \{R_{k,1}, R_{k,2}, \ldots R_{k,t}\}$$

We also define the limiting cases;$S^*_{k,\infty} = \lim_{t\to\infty} S^*_{k,t}$ and $S_{k,\infty} = \lim_{t\to\infty} S_{k,t}$.

The time between type 1 or type 2 records is called the waiting time and is defined as follows.

Type 1 k^{th} waiting time:

$$W^*_{k,t} = T^*_{k,t} - T^*_{k,t-1} - 1, t > 1$$

Type 2 k^{th} waiting time:

$$W_{k,t} = T_{k,t} - T_{k,t-1} - 1, t > 1$$

The definitions of waiting time are such that if (type 1 or type 2) k^{th} records occur at both time n and time $n+1$, the waiting time between them defined to be zero.

In the remainder of this text if the type of record is not specified it may be assumed that a type 2 record is being referred to.

3.1.2 Weak Records

Weak records are defined for discrete time series. They occur when a new observation is equal to an existing order statistic. In a discrete time series, if a new observation is strictly less than a current record value, we say that a strong record has occurred. The type 1 and type 2 k^{th}; strong record indicator function, strong record count, strong record time, strong record value and strong record value vector are defined as in Section 3.1.1. The type 1 and type 2 k^{th} weak record indicator function are defined as follows.

Type 1 k^{th} weak indicator function:

$$Iw^*_{k,k} = \begin{cases} 1, & y_k > y_{k-1,k-1}, \\ 0, & \text{otherwise} \end{cases}$$

$$Iw^*_{k,n} = \begin{cases} 1, & y_{k-1,n-1} < y_n \leq y_{k,n-1}, n > k \\ 0, & y_n \geq y_{k,n-1} \text{ or } y_n \leq y_{k-1,n-1}, n > k \end{cases}$$

Type 2 k^{th} weak indicator function:

$$Iw_{k,k} = 1$$

$$Iw_{k,n} = \begin{cases} 1, & y_n \leq y_{k,n-1}, n > k \\ 0, & y_n > y_{k,n-1}, n > k \end{cases}$$

Other definitions relating to weak records are intuitive: Type 1 and type 2 k^{th} weak record counts and record times (denoted $Nw^*_{k,n}$, $Nw_{k,n}$, $Tw^*_{k,t}$ and $Tw_{k,t}$ respectively) are defined as for strong record counts and times but with $Iw^*_{k,n}$ in place of $I^*_{k,n}$, and $Iw_{k,n}$ in place of $I_{k,n}$. Type 1 and type 2 k^{th} record values and waiting times (denoted $Rw^*_{k,t}$, $Rw_{k,t}$, $Ww^*_{k,t}$ and $Ww_{k,t}$ respectively) are defined as for strong record values and waiting times but with $T^*_{k,t}$ replaced with $Tw^*_{k,t}$, and $T_{k,t}$ replaced with $Tw_{k,n}$. Finally type 1 and type 2 k^{th} record value vectors (denoted $Sw^*_{k,t}$ and $Sw_{k,t}$ respectively) are defined as strong record value vectors but with $R^*_{k,t}$ replaced with $Rw^*_{k,t}$ and $R_{k,t}$ replaced with $Rw_{k,t}$.

In Chapter 5 a discrete time series is created by rounding a continuous time series to the nearest integer. In order to estimate the number of records in the underlying

time series, a small noise (a uniform [-0.1,0.1] random variable) was added to each of the members of the time series. The number of records in this time series is an unbiased estimator of the number of records in the underlying continuous time series.

3.1.3 Maximal Records

Instead of considering, as we have done, the record smallest elements in a sample, we can consider the record largest elements in a sample. All of the definitions in Sections 3.1.1 and 3.1.2 can be repeated for maximal records. A maximal record from a time series y_i is equal to a minimal record from the time series $-y_i$. These are denoted with a superscript $+$. For example, the type 1 maximal k^{th} record count is denoted $N^{*+}_{k,n}$ and the i^{th} type 1 maximal weak record $Nw^{*+}_{i,n}$.

3.1.4 Moments and Distributions

This section collects well-known results that are useful for understanding the properties of records. These can be found in literature, for example [5], [27] and [31], [1]. [16] provides an excellent, and very readable introductory review combining examples, results, applications and citations related to 1^{st} records.

The following theorem is useful for discussing properties of record values as it shows that results about the distribution of $S^*_{1,n}$ can equally well be applied to $S^*_{2,n}$, $S^*_{3,n}, \ldots$. It first appeared in [23], but can be found in many sources including [24].

Theorem 3.1.1. ***(Ignatov's Theorem)*** *$S^*_{1,n}, S^*_{2,n}, S^*_{3,n}, \ldots$ are independent and identically distributed random sets.*

A set is a finite or infinite collection of objects in which order has no significance. Elements in a set are distinct. A random set is a set made up of random variables. Notice that although the sets $S^*_{1,n}, S^*_{2,n}, \ldots$ are independent of each other, the elements within each set are a strictly increasing sequence of random variables.

[5] examines Ignatov's theorem, giving excellent figures and further explanations of its implications.

The expected value and the variance of the number of type 1 k^{th} records to occur between time k and time n are given below.

$$\begin{aligned} \mathrm{E}N^*_{k,n} &= \mathrm{E}(I^*_{k,k} + \ldots + I^*_{k,n}) = \textstyle\sum_{j=k}^{n} \frac{1}{j} \\ &= \log n - \textstyle\sum_{j=1}^{k-1} \frac{1}{j} + \gamma + O(\frac{1}{n}), \quad n \to \infty \\ \mathrm{Var}N^*_{k,n} &= \textstyle\sum_{j=k}^{n} \left(\frac{1}{j} - \frac{1}{j^2}\right) \end{aligned}$$

where $\gamma = 0.5772\ldots$ is the Euler constant. This derivation uses the fact that $E(I^*_{k,n}) = \frac{1}{n}$ for all $n \geq k$. This fact is shown to be true in [27], and is a consequence of the fact that when adding a new random variable (y_n) to a sample of size $n-1$ it is equally likely to fall in any of the n sections, $[m, y_{1,n-1}), [y_{2,n-1}), \ldots, [y_{n-1,n-1}, M]$. Here M is the upper end of the support of $F(x)$.

Similarly we can find the expectation and variance of the k^{th} type 2 record count. This is given below.

$$\begin{aligned} \mathrm{E}N_{k,n} &= \mathrm{E}(I_{k,k} + \ldots + I_{k,n}) = \textstyle\sum_{j=k}^{n} \frac{k}{j} \\ &= k\left[\log n - \textstyle\sum_{j=1}^{k-1} \frac{1}{j} + \gamma + O(\frac{1}{n})\right], \quad n \to \infty \\ \mathrm{Var}N_{k,n} &= \textstyle\sum_{j=k}^{n} \left(\frac{k}{j} - \left(\frac{k}{j}\right)^2\right) \end{aligned}$$

[31] shows that if $F(\cdot)$ is continuous then

$$\frac{N^*_{1,n}}{\log n} \to 1 \text{ a.s. as } n \to \infty$$

where a.s. stands for almost surely. We can show similar results for k^{th} type 1 records and k^{th} type 2 records. Indeed, for $n \geq k$ we have that $I^*_{1,n} \stackrel{d}{=} I_{k,n}$. Using this and the definitions in Section 3.1.1 we can say that

$$N^*_{k,n} \stackrel{d}{=} N^*_{1,n} - N^*_{1,k-1}. \tag{3.1.1}$$

Therefore, for finite k we can say that

$$\frac{N^*_{k,n}}{\log n} \to 1 \text{ a.s. as } n \to \infty. \tag{3.1.2}$$

Again by their definitions we can say that

$$\begin{aligned} N_{k,n} &= N^*_{1,n} + \ldots + N^*_{k,n} - N^*_{1,k-1} - \ldots - N^*_{k-1,k-1} \\ &= N^*_{1,n} + \ldots + N^*_{k,n} - (k-1). \end{aligned} \tag{3.1.3}$$

Using this, and the fact that for finite k all $N^*_{k,n}$ converge almost surely to $\log n$ we can say that

$$\frac{N_{k,n}}{k \log n} \rightarrow 1 \text{ a.s. as } n \rightarrow \infty. \tag{3.1.4}$$

These results have been verified by the simulations whose results are shown in Figures 3.2 and 3.3.

Records and Poisson Processes

Records have many links to Poisson processes, these are given below. [31] shows that asymptotically as $n \rightarrow \infty$, $N^*_{1,n}$ is a nonhomogeneous Poisson process with intensity $\lambda(n) = \log n$. Here a nonhomogeneous Poisson process is defined to be a continuous time stochastic process, $\{N(t), t = 0\,1\,\ldots\}$, with the following properties:

- $N(t) \geq 0, \forall t$ and $N(0) = 0$
- $N(s) \leq N(t)\ \forall s, t$ such that $s \leq t$
- $\mathrm{P}(N(t+h) - N(t) = 1) = \lambda(t)h + o(h), \forall t$
- $\mathrm{P}(N(t+h) - N(t) > 1) = o(h), \forall t$

where $\frac{o(h)}{h} = 0$ as $h \rightarrow \infty$. Here $\lambda(t)$ is the time dependent intensity. Notice that this definition implies that for non-overlapping time intervals, $(u, v]$ and $(s, t]$, $\mathrm{P}(N(t) - N(s) = y)$ is independent of $\mathrm{P}(N(v) - N(u) = x)$. It also has the implication that the probability that the number of events between time a and b $(0 \leq a \leq b)$ is equal to some integer k is given by

$$\mathrm{P}(N(b) - N(a) = k) = \frac{\exp(-\lambda_{a,b})\lambda^k_{a,b}}{k!}, \quad k = 0, 1, \ldots$$

where $\lambda_{a,b} = \int_a^b \lambda(t)dt$.

Using (3.1.1) we can say that $N^*_{k,n}$ is also asymptotically a nonhomogeneous Poisson process with intensity $\log n$. It is commonly known that the superposition of k Poisson processes with intensities $\mu_1, \ldots, \mu_2$ is also a Poisson process with intensity $\mu_1 + \ldots + \mu_k$. Therefore we can say, using (3.1.3), that $N_{k,n}$ is asymptotically a nonhomogeneous Poisson process with intensity $k \log n$.

A Markov process is defined to be a stochastic process, $x(t)$, with the property that for every n and $t_1 < t_2 < \ldots < t_n$ the following equality holds:

$$\mathrm{P}\left(x(t_n) \leq x_n | x(t_{n-1}), \ldots, x(t_1)\right) = \mathrm{P}\left(x(t_n) \leq x_n | x(t_{n-1})\right).$$

If the possible values of $x(t)$ form a countable set S (called the state space), the Markov process is called a Markov chain. $\{y_{k,n}, n \geq k\}$ is a Markov process its transition distribution is given by:

$$\mathrm{P}(y_{k,n+t} \leq x_2 | y_{k,n} = x_1) = \begin{cases} 1 - (1 - F(x_2))^t, & x_2 \leq x_1 \\ 0, & x_2 > x_1 \end{cases} \tag{3.1.5}$$

As the right hand side of (3.1.5) does not depend on n, $y_{k,n}$ is said to be a Markov process with stationary transition probabilities. The Markov chain $y_{k,n}$ has a constant path, except at type 2 k^{th} record times (or type 1; $1^{\text{st}}, 2^{\text{nd}}, \ldots, k^{\text{th}}$ record times), here the path jumps to a value closer to m, i.e. $(y_{T_{k,t+1}} - m) < (y_{T_{k,t}} - m)$. The record values $\{R_{k,t}, t \geq 1\}$ are therefore an embedded Markov process of states visited by the Markov chain $\{y_{k,n}, n \geq k\}$. The Markov chain $R_{k,t}$, $t = 1, 2, \ldots$ has stationary transition probabilities given by:

$$\mathrm{P}(R_{k,t+1} < y | R_{k,t} = x) = \begin{cases} \frac{F(y)}{F(x)}, & y \leq x \\ 0, & y > x \end{cases} \quad t \geq 1 \tag{3.1.6}$$

[31] shows us that if $F(\cdot)$ is continuous with $m = \inf\{y : F(y) > 0\}$ and $M = \sup\{y : F(y) < 1\}$, then by letting $H(t) = \log(F(x))$ so that $H : (m, M) \to (-\infty, 0)$, we can say that $\{R_{1,t}\}$ is a Poisson process on (m, M) with intensity

$$H[a, b) = H(b) - H(a). \tag{3.1.7}$$

Finally from [31] we know that if $F(\cdot)$ is again continuous with support (m, M), then we can say that the points $\{R_{1,t}, W_{1,t+1}, t \geq 1\}$ are the points of a two dimensional Poisson process on $(m, M) \times \{1, 2, 3, \ldots\}$ with intensity:

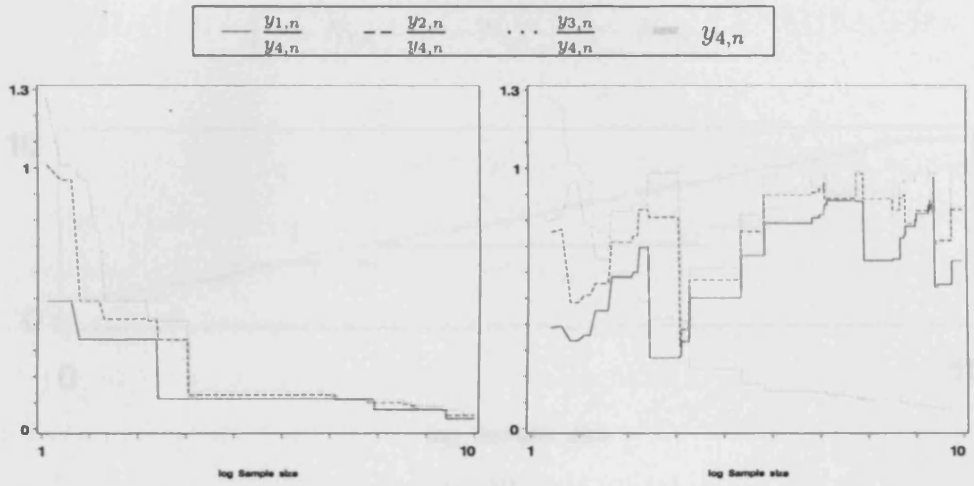
$$\mu([a, b) \times \{j\}) = \frac{(1 - F(a))^j - (1 - F(b))^j}{j}. \tag{3.1.8}$$

Here we use the definition of a two dimensional Poisson process as found in [31].

We know that $R_{1,t} \equiv R^*_{1,t}$ and $W_{1,t} \equiv W^*_{1,t}$, because of this we can extend the two previous results concerning $R_{1,t}$ to say that if $F(\cdot)$ is continuous with support (m, M) then $\{R^*_{1,t}\}$ is a Poisson process on (m, M) with intensity given by (3.1.7) and $\{R^*_{1,t}, W^*_{1,t+1}, t \geq 1\}$ are the points in a two dimensional Poisson process on $(m, M) \times \{1, 2, 3, \ldots\}$ with intensity given by (3.1.8).

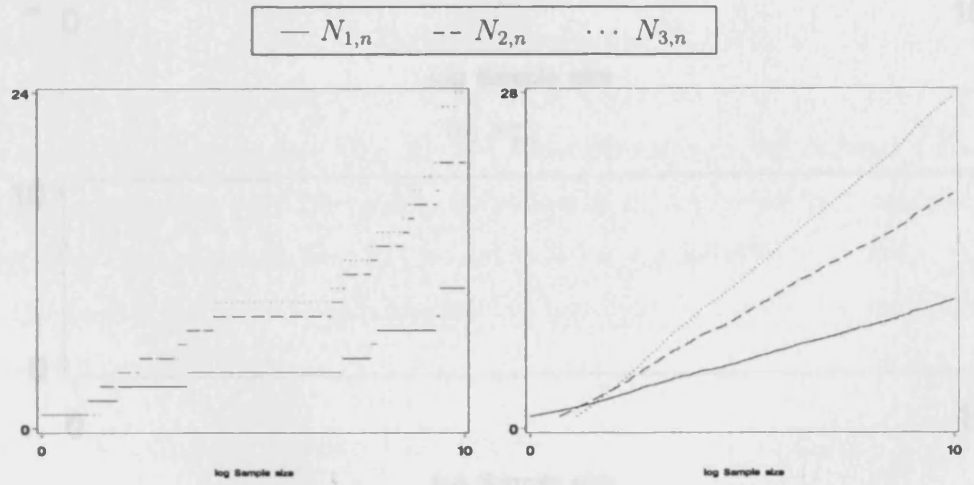
3.1.5 Altering Time Scale

It is stated above that asymptotically $N^*_{i,n}$ is a Poisson process with intensity $\log(n)$ and $N_{i,n}$ is a Poisson process with intensity $i \log(n)$. Therefore by letting $t = \log(n)$ we have, $EN^*_{i,n} = EN^*_{i,e^t} = t$, so N^*_{i,e^t} is a Poisson process with intensity 1, and N_{i,e^t} is a Poisson process with intensity i. Figures 3.2 and 3.3 below show results from a simulation of 100 samples of size 22 000 from a Weibull distribution. Figure 3.1 shows the order statistics from one of these samples. These figures verify that by altering the time scale the number of records becomes a homogeneous Poisson process.



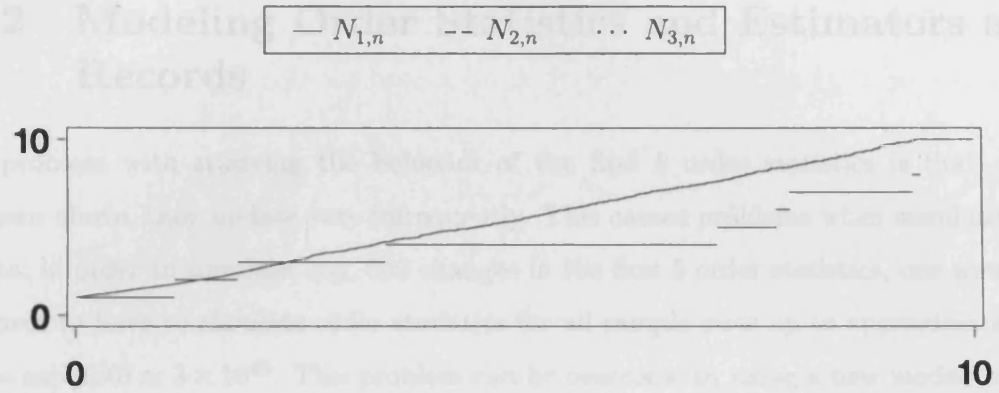
(a) (b)

Figure 3.1: (a) shows the first three order statistics, $y_{1,n}$, $y_{2,n}$ and $y_{3,n}$, from a single sample, plotted against $\log n$. (b) shows the normalized order statistics, $(y_{1,n} - m)/(y_{4,n} - m)$, $(y_{2,n} - m)/(y_{4,n} - m)$ and $(y_{3,n} - m)/(y_{4,n} - m)$, from a single sample, plotted against $\log(n)$. Also plotted on (b) is $y_{4,n}$ against $\log n$.

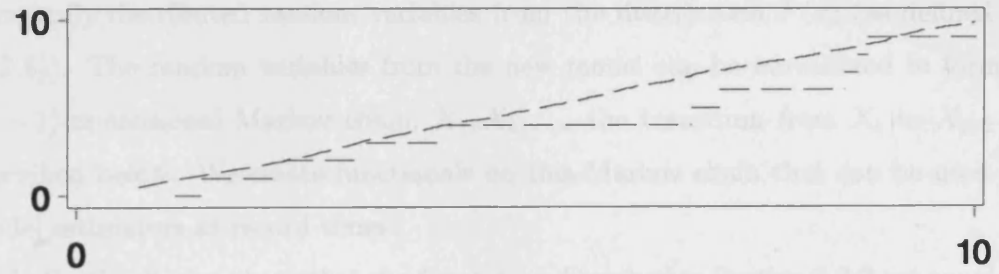


(a) (b)

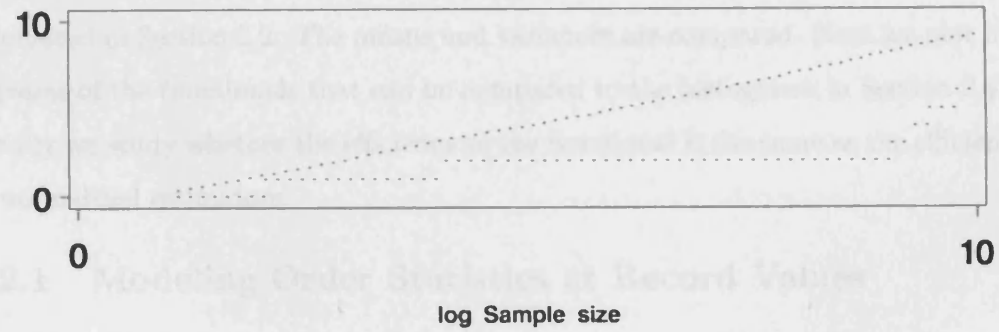
Figure 3.2: (a): Count, $N_{1,n}$, $N_{2,n}$ and $N_{3,n}$ (from a single sample) plotted against $\log(n)$. (b): The mean count over 100 runs, $\frac{1}{100}\sum_{r=1}^{100} N_{k,n,r}$, plotted against $\log(n)$, where each $N_{i,n,r}$ is $N_{i,n}$ from a separate run.



(a) $N^*_{1,n}$



(b) $N^*_{2,n}$



(c) $N^*_{3,n}$

Figure 3.3: Count, $N^*_{i,n}$ (from a single sample) and $\frac{1}{100}\sum_{r=1}^{100} N^*_{i,n,r}$ plotted against $\log(n)$, where each $N^*_{i,n,r}$ is $N^*_{i,n}$ from a separate run.

3.2 Modeling Order Statistics and Estimators at Records

A problem with studying the behavior of the first k order statistics is that, as shown above, they update very infrequently. This causes problems when simulating data; in order to simulate, say, 500 changes in the first 5 order statistics, one would expect to have to simulate order statistics for all sample sizes up to approximately $n = \exp(100) \simeq 3 \times 10^{43}$. This problem can be overcome by using a new model that simulates random variables equal in distribution to order statistics at record times of a time series y_i, $i = 1, \ldots, n$. In the model described below y_i are independent, identically distributed random variables from the distribution $F(x)$ (as defined in (1.2.6)). The random variables from the new model can be normalized to form a $(k-1)$-dimensional Markov chain, $X_1, X_2, \ldots$, the transition from X_t to X_{t+1} is described below. We create functionals on this Markov chain that can be used to model estimators at record times.

In Section 3.4 we check that the functional described in Section 3.2.3 are equal in distribution to the normalized estimators. First we do this by deriving the densities of some of the functionals. We compare these densities to the relevant densities calculated in Section 2.2. The means and variances are compared. Next we plot histograms of the functionals that can be compared to the histograms in Section 2.4.1. Finally we study whether the efficiency of the functional is the same as the efficiency of normalized estimators.

3.2.1 Modeling Order Statistics at Record Values

In order to derive random variables that model order statistics, we require the following representation which can be found in [27]. It shows that order statistics from a general continuous distribution F can be modeled using uniform order statistics.

Representation 3.2.1. Let $y_{1,n} \leq \ldots \leq y_{n,n}$, $n = 1, 2, \ldots$ be order statistics from sample $y_1, \ldots y_n$ of i.i.d.r.v. with any continuous c.d.f. F. Also let $U_{1,n} \leq \ldots \leq$

$U_{n,n}$, $n = 1, 2, \ldots$, be 'Uniform order statistics', i.e. the order statistics from a sample, $U_1, \ldots U_n$, of i.i.d.r.v.s from a uniform distribution on $[0, 1]$. Then for any $n = 1, 2, \ldots$

$$(F(y_{1,n}), \ldots F(y_{n,n})) \stackrel{d}{=} (U_{1,n} \ldots U_{n,n}),$$

where $\stackrel{d}{=}$ means that the two vectors have the same distribution.

If we define the inverse of $F(x)$ (denoted $F^{-1}(s)$) to be

$$F^{-1}(s) = \inf(x : F(x) > s),$$

then Representation 3.2.1 can be rewritten as
$(y_{1,n} \ldots y_{n,n}) \stackrel{d}{=} (F^{-1}(U_{1,n}), \ldots F^{-1}(U_{n,n}))$. In fact, the continuity condition can be dropped for this second form of the representation.

Representation 3.2.1 shows that if we simulate uniform order statistics at record time, $T_{k,t}$, then by applying F^{-1} to the order statistics we obtain random variables with the same distribution as order statistics $y_{1,T_{k,t}}, \ldots, y_{k,T_{k,t}}$. Here $y_{i,T_{k,t}}$ is the i^{th} order statistic from a sample of size $T_{k,t}$, drawn from a distribution with c.d.f. $F(x)$. In this section we denote the random variables that have the same distribution as uniform order statistics at record number t as $U_{1,\mathbf{t}}, \ldots, U_{k,\mathbf{t}}$. The bold font on the second subscript distinguishes these random variables from actual uniform order statistics.

From Representation 3.2.1 we can see that the first step in modeling order statistics at records from a general distribution, is to model order statistics at records from the uniform distribution. The k uniform order statistics at time $T_{k,1} = k$ are k uniform random variables arranged into ascending order. So $U_{1,\mathbf{1}} = U_{1,k}, \ldots, U_{k,\mathbf{1}} = U_{k,k}$ To model the k smallest uniform order statistic after the next update (when $t = 2$ and $n = T_{k,2}$) we can use the fact that, for uniform random variables, the next type 1 i^{th} record ($i \leq k$) is distributed uniformly on $[0, U_{k,k}]$. Let such a random variable be denoted U_2. Let U_2 be such that $U_{j-1,k} < U_2 < U_{j,k}$, for some $j \in [1, \ldots, k]$ and let $U_{0,k} = 0$. Therefore the j^{th} order statistic at time $T_{k,2}$ is modeled by $U_{j,\mathbf{2}} = U_2$.

The order statistics lower than this; $U_{1,2}, \ldots, U_{j-1,2}$, do not change and so are given by $U_{1,2} = U_{1,1}, \ldots, U_{j-1,2} = U_{j-1,1}$. The higher order statistics; $U_{j+1,2}, \ldots, U_{k,2}$ are given by $U_{j+1,2} = U_{j,1}, \ldots, U_{k,2} = U_{k-1,1}$. To model the order statistics at time $T_{k,t}$, $t = 3, 4, \ldots$ this process is repeated; a random variable U_t is drawn from a uniform $[0, U_{k,t-1}]$ distribution. This is added to the set of (modeled) order statistics at record number $t-1$. This set of $k+1$ random variables is reordered and the k smallest order statistics are selected to create the random variables that model the k smallest order statistics at record time t. In doing this we create a k-dimensional Markov chain where each k-dimensional vector is made up of the modeled order statistics. By applying F^{-1} to each element of each k-dimensional vector, we achieve a k-dimensional Markov chain of random variables with the same distribution as order statistics from a sample of increasing size with c.d.f. $F(x)$ at times $T_{k,t}$, $t = 1, 2, \ldots$.

3.2.2 Modeling Normalized Order Statistics at Records

In the analysis and simulations that follow, it was useful to obtain random variables equal in distribution to normalized order statistics; $(y_{i,n} - m)/(y_{k,n} - m)$ at records. These order statistics form a stationary $(k-1)$-dimensional Markov chain (the k^{th} element is always equal to 1, so is dropped). In order to achieve this normalization, at time $T_{k,1} = k$ (record time $t = 1$) modeled order statistics $U_{1,1}, \ldots, U_{k,1}$ are divided by $U_{k,1}$. We define a $(k-1)$-dimensional vector $X_1 = \{x_1^{(1)}, \ldots, x_{k-1}^{(1)}\}$ to be; $x_1^{(1)} = U_{1,1}/U_{k,1}, \ldots, x_{k-1}^{(1)} = U_{k-1,1}/U_{k,1}$. These make the first vector of a $(k-1)$-dimensional stationary Markov chain. We now simulate the next vector (equal in distribution to $k-1$ normalized order statistics at time $n = T_{k,2}$, record time $t = 2$): A uniform [0,1] random variable is added to the vector X_1, denote this random variable as U_2. The resulting k-dimensional vector $\{x_1^{(1)}, \ldots, x_{k-1}^{(1)}, U_2\}$ is normalized by dividing by the largest element (denote this largest element $U_{k,2}$). Re-ordering these random variables and selecting the $k-1$ smallest gives us $X_2 = \{x_1^{(2)}, \ldots, x_{k-1}^{(2)}\}$. From here the next vector of the $(k-1)$-dimensional Markov chain

can be found similarly, by adding a uniform [0,1] random variable, re-ordering, re-normalizing by dividing by the largest element (denote his largest element $U_{k,t}$) and finally discarding the element equal to 1 (the largest). The notation that we have adopted for the Markov chain vector is: $X_1, X_2, \ldots$. The elements that make up each vector, X_t, of the Markov chain are given by $x_j^{(t)}$, $j = 1, \ldots, k-1$:

$$\begin{array}{ccccc} & & X_1 & \rightarrow & X_2 \rightarrow \ldots \\ \begin{pmatrix} U_{1,k}/U_{k,k} \\ \vdots \\ U_{k-1,k}/U_{k,k} \end{pmatrix} & = & \begin{pmatrix} x_1^{(1)} \\ \vdots \\ x_{k-1}^{(1)} \end{pmatrix} & \rightarrow & \begin{pmatrix} x_1^{(2)} \\ \vdots \\ x_{k-1}^{(2)} \end{pmatrix} \rightarrow \ldots \end{array}$$

with $x_1^{(1)} \leq x_2^{(1)} \leq \ldots \leq x_k^{(1)} = 1$.

The transition from X_t to X_{t+1} is described by the following equations. Case 1 describes how the elements of the vector update if the new random variable (U_{t+1}) is smaller than the largest element of the $(k-1)$-dimensional vector ($x_{k-1}^{(t)}$). Case 2 describes how elements of the vector update if the new random variable U_{t+1} is larger than the largest element of the vector at time t.

Case 1: $U_{t+1} < x_{k-1}^{(t)}$

$$\begin{cases} x_i^{(t+1)} = \frac{x_i^{(t)}}{x_{k-1}^{(t)}}, & i \leq l-1 \\ x_l^{(t+1)} = \frac{U_t}{x_{k-1}^{(t)}} & \\ x_i^{(t+1)} = \frac{x_{i-1}^{(t)}}{x_{k-1}^{(t)}}, & l < i < k \end{cases} \tag{3.2.1}$$

where l is such that $x_{l-1}^{(t)} \leq U_{t+1} < x_l^{(t)}$ Case 2: $x_{k-1}^{(t)} \leq U_t$

$$\left\{ x_i^{(t+1)} = \frac{x_i^{(t)}}{U_t}, \quad i \leq k-1 \right. \tag{3.2.2}$$

These transition equations will be used to derive transition densities in Section 3.3.1.

By applying F^{-1} to each element of the vectors of the Markov chain X_t, we obtain random vectors equal in distribution to the order statistics at records from a general distribution.

3.2.3 Modeling Normalized Estimators at Records

The next section shows how the above Markov chain has been used to model estimators of m from a general distribution of the form (1.2.5). First we consider modeling consistent linear estimators.

Modeling a Consistent Linear Estimators

A general linear estimator of m based on a sample of size n is given by:

$$\hat{m}_{k,n} = \sum_{i=1}^{k} a_i y_{i,n},$$

where $y_{i,n}$ are defined as above, i.e. as the i^{th} order statistic from a sample of size n.

Using Representation 3.2.1 we have that

$$\hat{m}_{k,n} = \sum_{i=1}^{k} a_i y_{i,n} \stackrel{d}{=} \sum_{i=1}^{k} a_i F^{-1}(U_{i,n})$$

From (1.2.5) we can say that as $n \to \infty$

$$\begin{aligned}
\sum_{i=1}^{k} a_i F^{-1}(U_{i,n}) &\sim \sum_{i=1}^{k} a_i \left[\left(\frac{U_{i,n}}{c_0} \right)^{1/\alpha} + m \right] \\
&= m + \sum_{i=1}^{k} a_i \left(\frac{U_{i,n}}{c_0} \right)^{1/\alpha} \\
&= m + \left(\frac{U_{k,n}}{c_0} \right)^{1/\alpha} \sum_{i=1}^{k} a_i \left(\frac{U_{i,n}}{U_{k,n}} \right)^{1/\alpha} \\
&\stackrel{d}{=} m + \left(\frac{U_{k,n}}{c_0} \right)^{1/\alpha} \sum_{i=1}^{k} a_i \left(\frac{U_{i,k}}{U_{k,k}} \right)^{1/\alpha} \\
&= m + \left(\frac{U_{k,n}}{c_0} \right)^{1/\alpha} \sum_{i=1}^{k} a_i (x_i^{(t)})^{1/\alpha} \qquad (3.2.3)
\end{aligned}$$

Above we used that for a consistent linear estimator $\sum_{i=1}^{k} a_i = 1$.

Let $\zeta_{k,t} = \sum_{i=1}^{k} a_i (x_i^{(t)})^{1/\alpha}$ for $t = 1, 2 \ldots$ and $x_i^{(t)}$ be defined as in Section 3.2.2. $\zeta_{k,t}$ is a functional of a stationary Markov chain so $\zeta_{k,t}$, $t = 1, 2, \ldots$ is stationary. We will therefore often drop the t index and just use the notation ζ_k. In order to identify the vector of coefficients (a) used in functional ζ_k we will use the similar notation to that used for estimators defined in Chapters 1 and 2: ζ°, ζ_k°, or $\zeta_{k,t}^\circ$ represent $\zeta_{k,t}$ based on vector of coefficients a°; $\zeta^{(2)}$, $\zeta_k^{(2)}$, or $\zeta_{k,t}^{(2)}$ represent $\zeta_{k,t}$ based on $a^{(2)}$ and so on. From above, for large n we have that

$$\zeta_k \sim \left(\frac{c_0}{U_{k,n}}\right)^{1/\alpha} (\hat{m}_{k,n} - m). \tag{3.2.4}$$

By comparing the histograms in Figure 2.7, to those of ζ_k in Figure 3.9, it appears that $\zeta_k \stackrel{d}{=} (\hat{m}_{k,n} - m)/(y_{k,n} - m)$. In fact as $n \to \infty$ this is true, indeed; as

$$U_{k,n} \stackrel{d}{=} F(y_{k,n}) \sim c_0 (y_{k,n} - m)^\alpha \tag{3.2.5}$$

by substituting into (3.2.4) we can say that

$$\begin{aligned} \zeta_k &\sim (\hat{m}_{k,n} - m) \left(\frac{c_0}{c_0 (y_{k,n} - m)^\alpha}\right)^{1/\alpha} \\ &= \frac{(\hat{m}_{k,n} - m)}{(y_{k,n} - m)} \end{aligned}$$

As $\mathrm{E}kU_{1,n} = \mathrm{E}U_{k,n}$ you might also expect that the distribution of ζ_k and $(\hat{m}_{k,n} - m)/(k^{1/\alpha}(y_{1,n} - m))$ to be asymptotically equal. However the distributions of the random variables $kU_{1,n}$ and $U_{k,n}$ are not equal. Also although $\mathrm{E}U_{k,n} = \frac{k}{n}$ the distribution of $\left(\frac{c_0 n}{k}\right)^{1/\alpha} (\hat{m}_{k,n} - m)$ is not equal to the distribution of ζ_k.

Modeling the Maximum Likelihood Estimator

We now consider using the Markov Chain $(X_t,\ t = 1, 2, \ldots)$ to simulate maximum likelihood estimator. First let $\zeta' = \left(\frac{c_0}{U_{k,n}}\right)^{1/\alpha} (m^* - m)$. Now consider the maximum likelihood equation (1.3.3); substitute z for $\zeta' \left(\frac{U_{k,n}}{c_0}\right)^{1/\alpha} + m$. We can say that $\zeta_{k,t}^*$

is the solution z' to the following equation:

$$\begin{aligned} k &= (\alpha-1)\sum_{i=1}^{k-1}\frac{y_{k,n}-y_{i,n}}{y_{i,n}-z'\left(\frac{U_{k,n}}{c_0}\right)^{1/\alpha}-m} \\ &= (\alpha-1)\sum_{i=1}^{k-1}\frac{(y_{k,n}-m)-(y_{i,n}-m)}{(y_{i,n}-m)-z'\left(\frac{U_{k,n}}{c_0}\right)^{1/\alpha}} \end{aligned} \tag{3.2.6}$$

Using Representation 3.2.1:

$$k \stackrel{d}{=} (\alpha-1)\sum_{i=1}^{k-1}\frac{(F^{-1}(U_{k,n})-m)-(F^{-1}(U_{i,n})-m)}{(F^{-1}(U_{i,n})-m)-z'\left(\frac{U_{k,n}}{c_0}\right)^{1/\alpha}}$$

Using (1.2.5) and rearranging:

$$k \sim (\alpha-1)\sum_{i=1}^{k-1}\frac{1-\left(\frac{U_{i,n}}{U_{k,n}}\right)^{1/\alpha}}{\left(\frac{U_{i,n}}{U_{k,n}}\right)^{1/\alpha}-z'}$$

As $\frac{U_{i,n}}{U_{k,n}} \stackrel{d}{=} \frac{U_{i,k}}{U_{k,k}}$ for $i=1,\ldots,k$ we can write

$$k \stackrel{d}{=} (\alpha-1)\sum_{i=1}^{k-1}\frac{1-\left(\frac{U_{i,k}}{U_{k,k}}\right)^{1/\alpha}}{\left(\frac{U_{i,k}}{U_{k,k}}\right)^{1/\alpha}-z'}$$

So finally we have

$$k = (\alpha-1)\sum_{i=1}^{k-1}\frac{1-x_i^{(t)1/\alpha}}{x_i^{(t)1/\alpha}-z'} \tag{3.2.7}$$

We defined ζ' to be such that $\left(\frac{U_{k,n}}{c_0}\right)^{1/\alpha}\zeta'+m=m^*$, and substituted this into the maximum likelihood equation (see (3.2.6)). If we set $\zeta^*_{k,t}$ equal to the solution z' from (3.2.7) we have that

$$\lim_{n\to\infty}\zeta^*_{k,t} \stackrel{d}{=} \zeta' = \left(\frac{c_0}{U_{k,n}}\right)^{1/\alpha}(m^*-m).$$

When referring to $\zeta^*_{k,t}$ we will also use the notation ζ^*_k and ζ^*. If the notation ζ, ζ_k or $\zeta_{k,t}$ is used, we are referring to any of the functionals defined in this section.

3.3 Analysis of Markov Chain

In this section we examine some of the properties of the Markov chain, X_t, and the functional, ζ_k. We find the transition density of the 1-dimensional Markov chain when $k = 2$, so $X_t = x_1^{(t)}$. We also find the transition density of ζ_k when $k = 2$. Notice that for $k = 2$, $x_1^{(t)}$ is a Markov chain. As $\zeta_2 = a_1 x^{1/\alpha} + a_k$ is a one-to-one function, $\zeta_2 = a_1 x^{(t)1/\alpha} + a_k$ is also a Markov chain.

Using these transition densities we were able to calculate the autocorrelation functions of $x_1^{(t)}$ (when $k = 2$) and ζ_2. We have also used simulation to estimate the autocorrelation functions of $x_i^{(t)}$, $i = 1, \ldots, k-1$ and ζ_k° ($\alpha = 3$). In both these cases $k = 2$, 3, 4 and 5.

Throughout this section, for ease of notation we denote $x_1^{(t)}$ as x, $x_1^{(t+1)}$ as x', $\zeta_{2,t}$ as ζ and $\zeta_{2,t+1}$ as ζ'.

3.3.1 Transition densities of Markov Chain

Transition density for Markov Chain, $x_1^{(t)}$, when $k = 2$

Here we calculate the transition density of the 1-dimensional Markov chain $X_1 = x_1^{(t)}$, $k = 2$. In order to simplify notation let $x_1^{(t)} = x$ and $x_1^{(t+1)} = x'$. As shown by the transition equations (3.2.1) and (3.2.2), the value of x' depends on the relative size of x and the uniform order statistic U_{t+1}. Below we consider the two cases; that the new uniform random variable is less than x; and that the uniform random variable is greater than x.

$$\left(x_1^{(t)} \right) = (x) \Rightarrow \begin{cases} \text{Case 1: } U_{t+1} \leq x \\ \\ \text{Case 2: } U_{t+1} > x \end{cases}$$

By the law of total probability the transition c.d.f. of $x_1^{(t+1)}$ is equal to

$$\mathrm{P}(\text{Case 1})\mathrm{P}(x' \leq y | \text{Case 1}) + \mathrm{P}(\text{Case 2})\mathrm{P}(x' \leq y | \text{Case 2}).$$

These probabilities are not difficult to derive. Consider first Case 1. Obviously P(Case 1) = P($U_t \leq x$) = x. In Case 1 x' is given by

$$\left(\frac{U_t}{x} \right) = (x'), \text{ with } 0 \leq x' \leq 1. \tag{3.3.1}$$

So the conditional c.d.f. of x' given Case 1 occurs is

P($x' \leq y$|Case 1) = y, $(0 \leq y \leq 1)$.

For Case 2 we have; P(Case 2) = P($U_{t+1} > x$) = $1 - x$. x' is given by

$$\left(\frac{x}{U_t} \right) = (x'), \text{ where } x \leq x' \leq 1. \tag{3.3.2}$$

So the conditional c.d.f. of x' given Case 2 occurs is

P($x' \leq y$|Case 2) = $\frac{y-x}{y(1-x)}$, $(x \leq y \leq 1)$.

So using the law of total probability, the transition c.d.f of $x_1^{(t+1)}$ for $k = 2$ is

$$\begin{aligned} \mathrm{P}(x' \leq y) &= x\mathrm{P}(x' \leq y|\text{Case 1}) + (1-x)\mathrm{P}(x' \leq y|\text{Case 2}) \\ &= \begin{cases} xy, & 0 \leq y \leq x \\ 1 + xy - \frac{x}{y}, & x \leq y \leq 1 \end{cases} \end{aligned} \tag{3.3.3}$$

We can differentiate (3.3.3) to obtain the following expression for the transition density of $x_1^{(t)}$:

$$p(y;x) = \begin{cases} x, & 0 \leq y \leq x \\ x + \frac{x}{y^2}, & x \leq y \leq 1 \end{cases}$$

Transition density for ζ_k when $k = 2$

The method for finding the transition density of ζ_2 is similar to the method for finding the transition density of $x_1^{(t)}$ when $k = 2$.

As we have that $\zeta_2 = a_1 x^{1/\alpha} + a_2$ we can say that

$$x = \left(\frac{\zeta - a_2}{a_1} \right)^{\alpha}. \tag{3.3.4}$$

Therefore, given that the transition between x and x' is split into two cases ($U_{t+1} < x$ and $x \leq U_{t+1}$) the transition between ζ and ζ' must be split into the same two cases.

We can write these cases in terms of ζ and U_{t+1}. If we do so, the condition for Case 1 becomes $U_{t+1} < \left(\frac{\zeta - a_2}{a_1}\right)^\alpha$ and the condition for Case 2 becomes $\left(\frac{\zeta - a_2}{a_1}\right)^\alpha \leq U_{t+1}$.

The probability of Case 1 occurring is still P(Case 1) $= x$, writing this in terms of ζ we obtain P(Case 1) $= \left(\frac{\zeta - a_2}{a_1}\right)^\alpha$. By substituting (3.3.4) into (3.3.1) we find that the transition from ζ to ζ' can be described as

$$\zeta' = \frac{a_1^2 U_t^{1/\alpha}}{\zeta - a_2} + a_2, \quad a_2 \leq \zeta' \leq 1.$$

The conditional density of ζ' given Case 1, is found by taking p_1, making the transformation $x' = \left(\frac{\zeta' - a_2}{a_1}\right)^\alpha$ then substituting (3.3.4). It is therefore given by

$$p_{\zeta 1}(z;\zeta) = \frac{\alpha(z - a_2)^{\alpha - 1}}{a_1^\alpha}, \quad a_2 \leq z \leq 1.$$

Similarly, the probability of Case 2 occurring in terms of ζ is given by P(Case 2) $= 1 - x = 1 - \left(\frac{\zeta - a_2}{a_1}\right)^\alpha$. The transition is described by

$$\zeta' = \frac{\zeta - a_2}{U_t^{1/\alpha}} + a_2, \quad \zeta < \zeta' \leq 1.$$

Finally, the conditional density of ζ' given Case 2, is given by

$$p_{\zeta 2}(\zeta;\zeta') = \frac{\alpha(\zeta - a_2)^\alpha a_1^\alpha}{(a_1^\alpha - (\zeta - a_2)^\alpha)(\zeta' - a_2)^{\alpha+1}}, \quad \zeta < z \leq 1.$$

Therefore, by the law of total probability, we have that the transition density of ζ for $k = 2$ is given by:

$$p_\zeta(z;\zeta) = \begin{cases} \frac{\alpha(z-a_2)^{\alpha-1}(\zeta-a_2)^\alpha}{a_1^{2\alpha}}, & a_2 \leq z \leq \zeta \\ \frac{\alpha(z-a_2)^{\alpha-1}(\zeta-a_2)^\alpha}{a_1^{2\alpha}} + \frac{\alpha(\zeta-a_2)^\alpha}{(z-a_2)^{\alpha+1}}, & \zeta < z \leq 1 \end{cases}$$

3.3.2 Autocorrelation Functions of Markov Chain

In this section we derive the autocorrelation function of $X_1 = x_1^{(t)}$ when $k = 2$ and the autocorrelation function ζ_2. A trajectory of $x_1^{(t)}$ where $t = 1, \ldots, 1\,000\,000$ and $k = 2$, 3, 4 and 5 is simulated. The autocorrelation function of this trajectory is

estimated and plotted. The autocorrelation function of a time series $x_1, \ldots, x_n$ is estimated by

$$\hat{R}(l) = \frac{\sum_{i=1}^{n-l} (x_i - \bar{x})(x_{i+l} - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}, \tag{3.3.5}$$

where $\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$ is the sample mean. Another estimate of the autocorrelation function is

$$\hat{R}_2(l) = \frac{n \sum_{i=1}^{n-l} (x_i - \bar{x})(x_{i+l} - \bar{x})}{(n-k) \sum_{i=1}^{n} (x_i - \bar{x})^2}. \tag{3.3.6}$$

The sample autocorrelation function (3.3.6) is a better estimate of the autocorrelation function than (3.3.5) in cases where the time series is a long memory process. Indeed the estimate (3.3.5) has large bias in these cases. However, $x_k^{(t)}$ is a short memory process. The sample autocorrelation function (3.3.5) usually has a smaller mean square error than (3.3.6) in this case. For more details see [4].

The simulation for $k = 2$ verifies the analytic results. The autocorrelation function of ζ_k $(k = 2)$ is also derived. It can be seen that for $k = 2$ the autocorrelation function does not depend on the vector of coefficients a. Using the simulated Markov chain, the autocorrelation function of ζ_k is estimated and plotted for $\alpha = 3$. The simulation results when $k = 2$ are compared to the analytic results.

Autocorrelation function of $x_1^{(t)}$ when $k = 2$

Denote the autocorrelation function of $x_1^{(t)}$ $(k = 2)$ as $R_x(l)$, where l is the lag. So

$$R_x(l) = \frac{\mathrm{Cov}(x_1^{(t)}, x_1^{(t+l)})}{\mathrm{Var} x_1^{(t)}} = \frac{\mathrm{E}(x_1^{(t)} x_1^{(t+l)}) - \mathrm{E}(x_1^{(t)})\mathrm{E}(x_1^{(t+l)})}{\mathrm{Var}(x_1^{(t)})}.$$

$x_1^{(t)}$ is a one-dimensional, first order Markov chain so it has an autocorrelation function of the form

$$R(l) = \exp(l\lambda), \tag{3.3.7}$$

where λ is a constant and l is the lag.

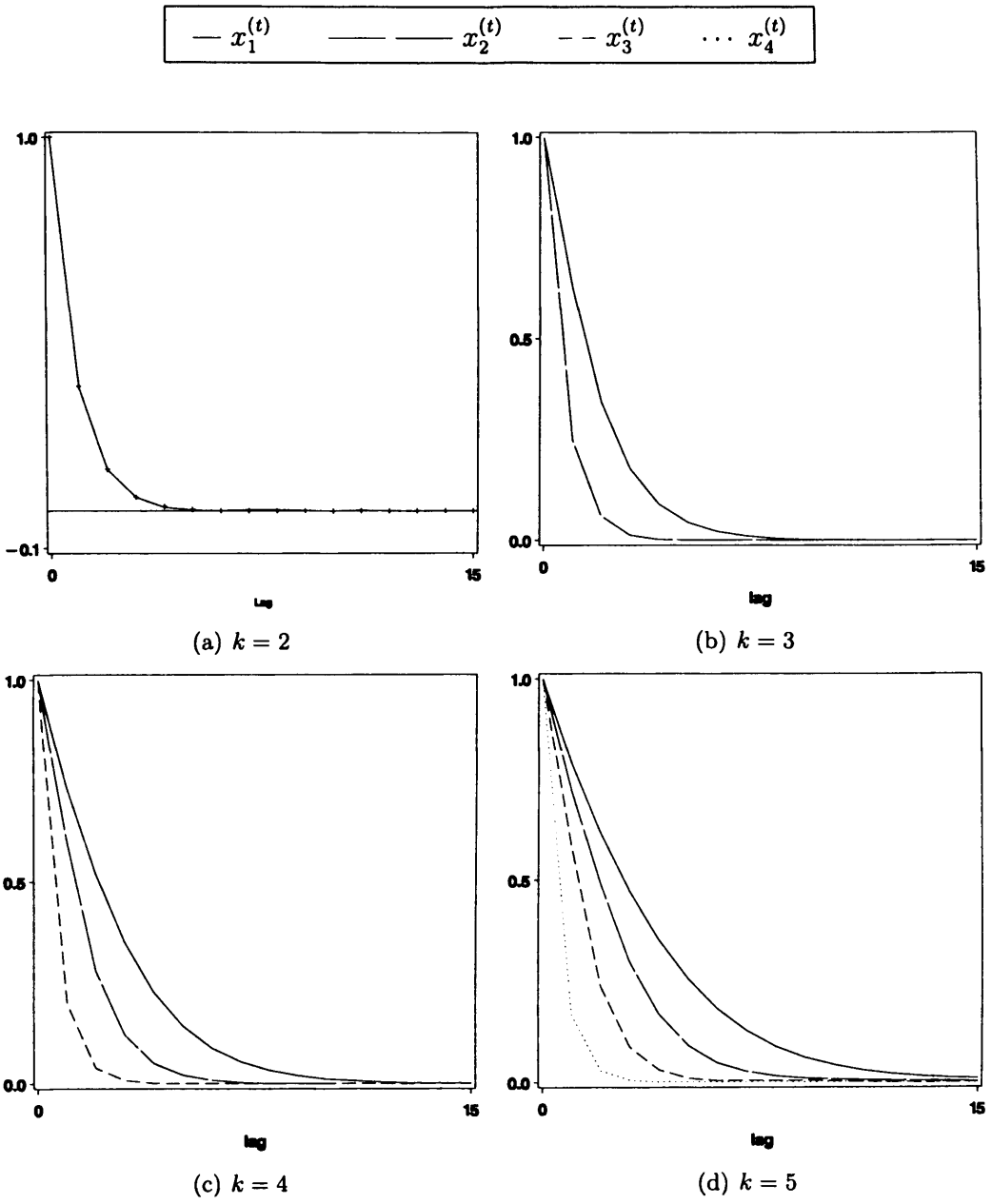
Using $p(y;x)$ from above we can derive the autocorrelation function of $x_1^{(t)}$ at lag one, and hence deduce the parameter λ from (3.3.7). Let

$$R_x(1) = \frac{\mathrm{Cov}(x_1^{(t)}, x_1^{(t+1)})}{\mathrm{Var}(x_1^{(t)})}$$

Again we use the simplified notation $x_1^{(t)} = x$ and $x_1^{(t+1)} = x'$. Using the transition density $p(x';x)$ and the density of x, $p(x) = 1$, we can say that:

$$\begin{aligned}
\mathrm{E}(xx') &= \int_0^1 \int_0^x x^2 x' dx' dx + \int_0^1 \int_x^1 x^2 x' + \frac{x^2}{x'} dx' dx = \frac{5}{18} \\
\mathrm{E}(x) &= \int_0^1 \int_0^x x^2 dx' dx + \int_0^1 \int_x^1 x^2 + \frac{x^2}{x'^2} dx' dx = \frac{1}{2} \\
\mathrm{E}(x') &= \int_0^1 \int_0^x x x' dx' dx + \int_0^1 \int_x^1 x' x + \frac{x}{x'} dx' dx = \frac{1}{2} \\
\mathrm{E}(x^2) &= \int_0^1 \int_0^x x^3 dx' dx + \int_0^1 \int_x^1 x^3 + \frac{x^3}{x'^2} dx' dx = \frac{1}{3} \\
R_x(1) &= \frac{\mathrm{E}(xx') - \mathrm{E}(x)\mathrm{E}(x')}{\mathrm{E}(x^2) - (\mathrm{E}x)^2} \\
&= \frac{1}{3}
\end{aligned}$$

As $R_x(l) = \exp(l\lambda)$; $\lambda = \log\left(\frac{1}{3}\right)$ and so $R_x(l) = \exp\left(l \log\left(\frac{1}{3}\right)\right) = 3^{-l}$. This can be verified by comparing to the plot of estimated autocorrelation for $k = 2$ in Figure 3.4.



(a) $k = 2$ (b) $k = 3$

(c) $k = 4$ (d) $k = 5$

Figure 3.4: Estimated autocorrelation function of $x_1^{(t)} \ldots x_{k-1}^{(t)}$, where k =2, 3, 4 and 5 and $t = 1\,000\,000$. In Figure (a) the exact autocorrelation function $R_x(l)$ is marked with a +.

Figure 3.4 shows that as k increases, the autocorrelation function (for any l) of $x_1^{(t)}$ increases. The autocorrelation of the largest element of the Markov chain X_t;

x_1^t when $k = 2$, x_2^t when $k = 3$, x_3^t when $k = 4$ and x_5^t when $k = 5$; are similar according to the simulation results.

Autocorrelation function of Markov Chain ζ_k when $k = 2$

It is possible to calculate the autocorrelation function analytically for ζ_2. As with $x_1^{(t)}$, ζ_2 is a one-dimensional first order Markov chain therefore it has autocorrelation function of the form (3.3.7). Now the constant λ will depend on α. We denote this parameter as λ_α and the autocorrelation function as $R_{\zeta_2}(l)$. As above we can calculate the autocorrelation function for $l \geq 0$ by finding $R_{\zeta_2}(1)$, substituting into (3.3.7) and solving for λ_α. We have that $R_{\zeta_2}(1)$ is given by

$$\begin{aligned} R_{\zeta_2}(1) &= \frac{\mathrm{Cov}([a_1(x_1^{(t)})^{1/\alpha} + a_2], [a_1(x_1^{(t+1)})^{1/\alpha} + a_2])}{\mathrm{Var}(a_1(x_1^{(t)})^{1/\alpha} + a_2)} \\ &= \frac{\mathrm{E}(x^{1/\alpha}x'^{1/\alpha}) - \mathrm{E}(x^{1/\alpha})\mathrm{E}(x'^{1/\alpha})}{\mathrm{E}(x^{2/\alpha}) - (\mathrm{E}(x^{1/\alpha}))^2} \end{aligned}$$

The expectations above can be found using the transition density $p(x, x')$. Indeed;

$$\begin{aligned} \mathrm{E}((xx')^{1/\alpha}) &= \int_0^1\int_0^x x^{\frac{1+\alpha}{\alpha}}x'^{1/\alpha}dx'dx + \int_0^1\int_x^1 x^{\frac{1+\alpha}{\alpha}}x'^{1/\alpha} + \frac{x^{\frac{1+\alpha}{\alpha}}}{x'^{\frac{1-2\alpha}{\alpha}}}dx'dx \\ \mathrm{E}(x^{1/\alpha}) &= \int_0^1\int_0^x x^{\frac{\alpha+1}{\alpha}}dx'dx + \int_0^1\int_x^1 x^{\frac{\alpha+1}{\alpha}} + \frac{x^{\frac{\alpha+1}{\alpha}}}{x'^2}dx'dx \\ \mathrm{E}(x'^{1/\alpha}) &= \int_0^1\int_0^x xx'^{1/\alpha}dx'dx + \int_0^1\int_x^1 x'^{1/\alpha}x + \frac{x}{x'^{\frac{1-2\alpha}{\alpha}}}dx'dx \\ \mathrm{E}(x^{2/\alpha}) &= \int_0^1\int_0^x x^{\frac{\alpha+2}{\alpha}}dx'dx + \int_0^1\int_x^1 x^{\frac{\alpha+2}{\alpha}} + \frac{x^{\frac{\alpha+2}{\alpha}}}{x'^2}dx'dx \end{aligned}$$

Substituting these expectations into the expression for $R_{\zeta_2}(1)$ above, we get $R_{\zeta_2}(1) = \exp(\lambda_\alpha) = \frac{\alpha}{1+2\alpha}$. So the autocorrelation function of ζ_2 is given by

$$R_{\zeta_2}(l) = \left(\frac{\alpha}{1+2\alpha}\right)^l. \tag{3.3.8}$$

Notice that this autocorrelation function is not affected by the coefficient a_1.

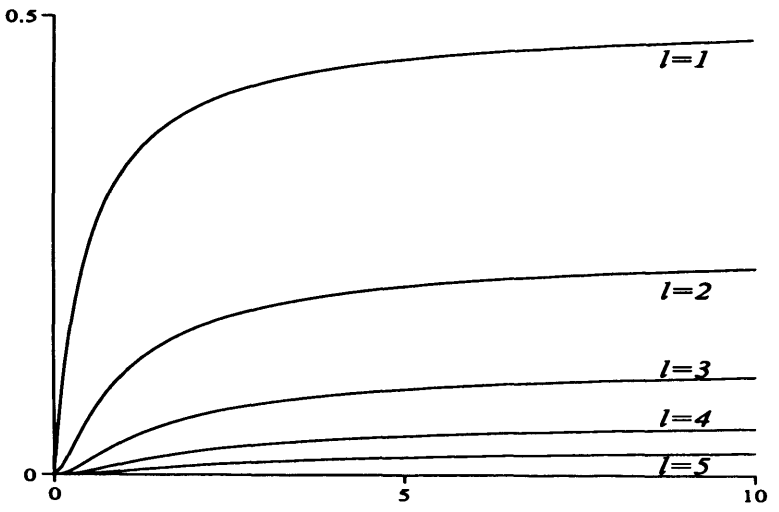
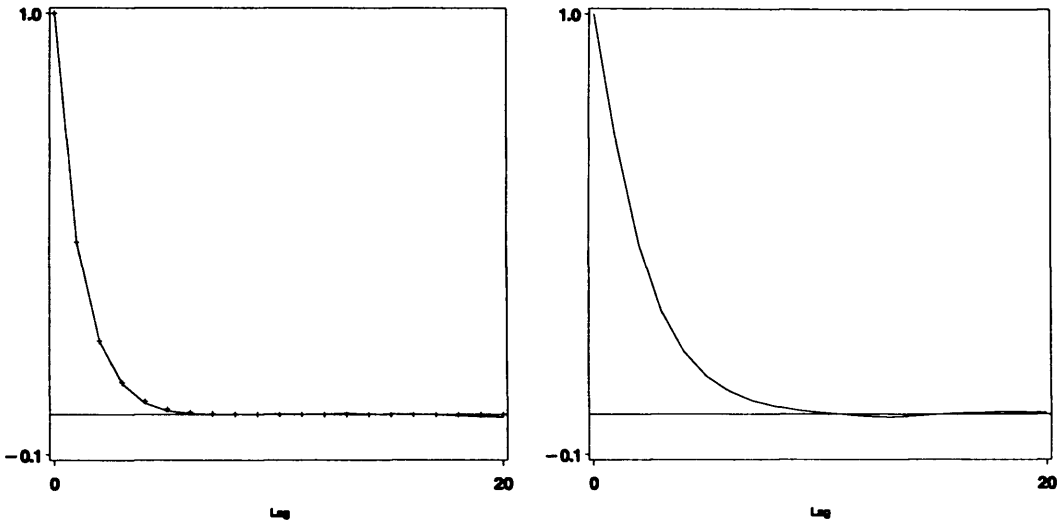


Figure 3.5: Autocorrelation function of ζ_2, (3.3.8) as a function of α. Lags $l = 1$, 2, 3, 4 and 5 are displayed as separate lines.

Figure 3.5 shows (3.3.8) plotted against α. Different values of l are plotted as separate lines. As α increases, the autocorrelation function increases for any l. Figure 3.6 shows the estimated autocorrelation functions for $k = 2$, 3, 4 and 5 with $\alpha = 3$, from simulated data. When $\alpha = 3$ and $k = 2$ the autocorrelation function of ζ_k is given by $R_{\zeta_2}(l) = (3/7)^l$. This is also plotted on Figure 3.6 with a $+$.



(a) $k = 2$ (b) $k = 3$

(c) $k = 4$ (d) $k = 5$

Figure 3.6: Estimated autocorrelation function of ζ_k from simulated data. Here k takes values 2, 3, 4 and 5, $\alpha = 3$ and $t = 1 \ldots 10\,000$. In Figure (a) the exact autocorrelation function R_{ζ_2} is marked with a $+$.

3.4 Comparison of the Estimators and Their Related Functionals on the Markov Chain

In this section we verify that the model described in Section 3.2 produces functionals with a distribution equal to the asymptotic distribution of their corresponding normalized estimator. Indeed, we must verify that for large n $\zeta_k^{\circ} \stackrel{d}{=} (m^{\circ}-m)/(y_{k,n}-m)$, $\zeta_k^{*} \stackrel{d}{=} (m^{*}-m)/(y_{k,n}-m)$, $\zeta_k^{(2)} \stackrel{d}{=} (m^{(2)}-m)/(y_{k,n}-m)$ and $\zeta_k^{(3)} \stackrel{d}{=} (m^{(3)}-m)/(y_{k,n}-m)$. This has been done in the following ways: First we derive the densities of $\zeta_k^{(2)}$ and $\zeta_k^{(3)}$ for $k = 2, 3, \ldots$ and general $\alpha > 1$. We then carry out a simulation study and compare the histograms of the normalized estimators and the functional.

3.4.1 Density Function of $\zeta_k^{(2)}$

Throughout this section we will continue to use the notation $x_1^{(t)} = x$. In order to derive the density of $\zeta_k^{(2)}$ we use the definition from above that $\zeta_k^{(2)}$ is given by $\zeta_k^{(2)} = (1+C_k)x^{1/\alpha} - C_k$, where C_k is as defined in (2.1.3). However, the derivation is valid when any coefficient $C_k^{(\Delta)} > -1$ is substituted in place of C_k. The c.d.f. of $\zeta_k^{(2)}$ is given by

$$\begin{aligned} \mathrm{P}(\zeta_k^{(2)} \leq z) &= \mathrm{P}((1+C_k)x^{1/\alpha} - C_k \leq z) \\ &= \mathrm{P}\left(x \leq \left[\frac{z+C_k}{1+C_k}\right]^{\alpha}\right) \\ &= 1 - \left(1 - \left(\frac{z+C_k}{1+C_k}\right)^{\alpha}\right)^{k-1} \quad -C_k \leq z \leq 1. \end{aligned}$$

By differentiating this expression we find that the density of $\zeta_k^{(2)}$ is given by

$$p_{\zeta_k^{(2)}}(z) = \frac{\alpha(k-1)(z+C_k)^{\alpha-1}}{(1+C_k)^{\alpha}} \left(1 - \left(\frac{z+C_k}{1+C_k}\right)^{\alpha}\right)^{k-2} \quad -C_k \leq z \leq 1. \qquad (3.4.1)$$

The density (3.4.1) can be seen to be exactly equal to the asymptotic density (2.2.19) with $\omega = \omega_4$. Notice for the case $k = 2$ we have that $C_k = \alpha/2$, by replacing

$C_k = \alpha/2$ by $C_k^{(\Delta)} = (\alpha - 1)/2$ we obtain exactly the asymptotic density (2.2.27) of $(m^* - m)/(y_{k,n} - m)$.

3.4.2 Density Function of $\zeta_k^{(3)}$

Before we derive the density of $\zeta_k^{(3)} = a_1^{(3)} x_1^{(t)} + a_i x_i^{(t)} + a_k^{(3)}$ we will simplify the notation and let $x_1^{(t)} = x$ and $x_i^{(t)} = y$. As we have the relation $a_1 + a_2 + a_3 = 1$, we can write one coefficient in term of the others. Indeed, we will write $a_1 = 1 - a_i - a_k$.

We find the density using the same technique as in Section 2.2. The one-to-one differentiable function $\Phi : \mathbb{R}^2 \to \mathbb{R}^2$ that maps (x, y) onto $\boldsymbol{\xi} = (\xi_1, \xi_2)$ is written below. During the derivation we will consider the coefficients a_i and a_k in place of $a_i^{(3)}$ and $a_k^{(3)}$. This is partly to simplify notation, but also to highlight that the density can be applied to any $\zeta_k = (1 - a_i - a_k)x + a_i y + a_k$ with coefficients a_i and a_k that satisfy $1 - a_i - a_k > 0$.

$$
\begin{aligned}
&\Phi : \\
\xi_1 &= (1 - a_i - a_k)x^{1/\alpha} + a_i y^{1/\alpha} + a_k \\
\xi_2 &= y.
\end{aligned}
$$

Its inverse is then given by,

$$
\begin{aligned}
&\Phi^{-1} : \\
x &= \left(\frac{\xi_1 - a_i \xi_2^{1/\alpha} - a_k}{1 - a_i - a_k} \right)^{\alpha} \\
y &= \xi_2
\end{aligned}
$$

where $1 - a_i - a_k \neq 0$. The elements, x and y, of the Markov chain have the same distribution as the first and i^{th} uniform order statistics from a sample of size $k - 1$. Their joint density is therefore given by,

$$
f_{U_{1,k-1}, U_{i,k-1}}(x, y) = \frac{(k-1)!}{(k-i-1)!(i-2)!} (1 - y)^{k-i-1} (y - x)^{i-2}, \quad 0 \leq x \leq y \leq 1.
$$

The Jacobian of the transformation is given by

$$\left|\frac{D\Phi^{-1}}{d\eta}\right| = \frac{\alpha}{1-a_i-a_k}\left(\frac{\eta_1 - a_i\eta_2^{1/\alpha} - a_k}{1-a_i-a_k}\right)^{\alpha-1}.$$

So using (2.2.2) the density of $\zeta_k^{(3)}$ is

$$p_{\zeta_k^{(3)}}(z) = b\int_{l_1}^{l_2}(z-a_i x^{1/\alpha}-a_k)^{\alpha-1}\left(x-\left(\frac{z-a_i x^{1/\alpha}-a_k}{1-a_i-a_k}\right)^{\alpha}\right)^{i-2}(1-x)^{k-i-1}dy$$

where $b = \frac{(k-1)!\alpha}{(1-a_2-a_3)^{\alpha}(k-i-1)!(i-2)!}$, $l_1 = \left(\frac{z-a_k}{1-a_k}\right)^{\alpha}$ and $l_2 = \begin{cases} \left(\frac{z-a_k}{a_i}\right)^{\alpha} & \text{for } a_k \le z < a_i + a_k \\ 1 & \text{for } a_i + a_k \le z \le 1. \end{cases}$

The density $p_{\zeta_k^{(3)}}(z)$ is valid where $a_k \le z \le 1$

When $k = 3$ we have that $\zeta^{(3)} = \zeta^{\circ}$ and a simpler form of the density can be found. Indeed, if we make the change of variable $u = \frac{a_2 y^{1/\alpha}}{z-a_3}$, we find that

$$p_{\zeta_3^{(3)}}(z) = c[B(l_4(z),\alpha-1,\alpha-1) - B(l_3,\alpha-1,\alpha-1)], \quad a_3 \le z \le 1 \qquad (3.4.2)$$

where

$$c = \frac{2(z-a_3)^{2\alpha-1}}{(a_2(1-a_2-a_3))^{\alpha}},$$

$$l_3 = \frac{a_2}{1-a_3},$$

$$l_4(z) = \begin{cases} 1 & \text{for } a_3 \le z < a_2 + a_3 \\ \frac{a_2}{z-a_3} & \text{for } a_2 + a_3 \le z \le 1 \end{cases}$$

and $B(z,a,b)$ is the incomplete beta function:

$$B(z,a,b) = \int_0^z u^{a-1}(1-u)^{b-1}du.$$

3.4.3 Density Function of ζ_k^*, when $k = 2$, $\alpha \ge 2$

Using the definition (3.2.7) we can see that when $k = 2$,

$$\zeta_2^* = \frac{2+\alpha-1}{2}x_1^{1/\alpha} - \frac{\alpha-1}{2}.$$

As $x_1^{(t)}$ is uniformly distributed on [0,1] when $k = 2$, we have that its density is given by $f_1(x) = 1, 0 \le x \le 1$. Therefore using the transformation $g(x) = \frac{2+\alpha-1}{2}x^{1/\alpha} - \frac{\alpha-1}{2}$

we find that the density of ζ_2^* is given by

$$f_{\zeta_2^*}(z) = \frac{2\alpha}{2+\alpha-1}\left(\frac{2z+(\alpha-1)}{2+\alpha-1}\right)^{\alpha-1} \quad \text{for } -\frac{\alpha-1}{2} \leq \zeta_2^* \leq 1. \tag{3.4.3}$$

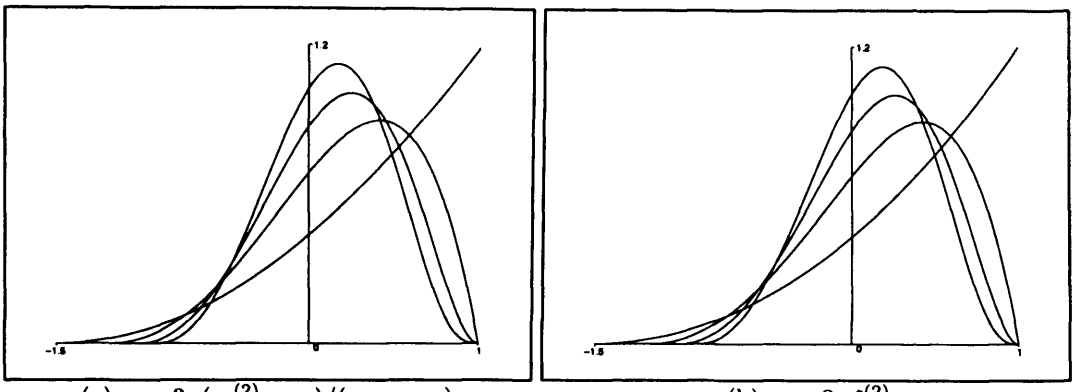
Compare this to (3.4.1) with $C_k^{(\Delta)} = \frac{\alpha-1}{2}$.

Using the above densities it was possible to calculate the moments displayed in Table 3.2. From here it can be seen that when $k = 2$ the expectation of ζ^* is 0.5 for $\alpha = 2, 5$ and 5. The variance of ζ^* increases with α. The variance of ζ^* is smaller than the variance of $\zeta^{(2)}$ for equivalent α. The mean of ζ^* is greater than that of $\zeta^{(2)}$

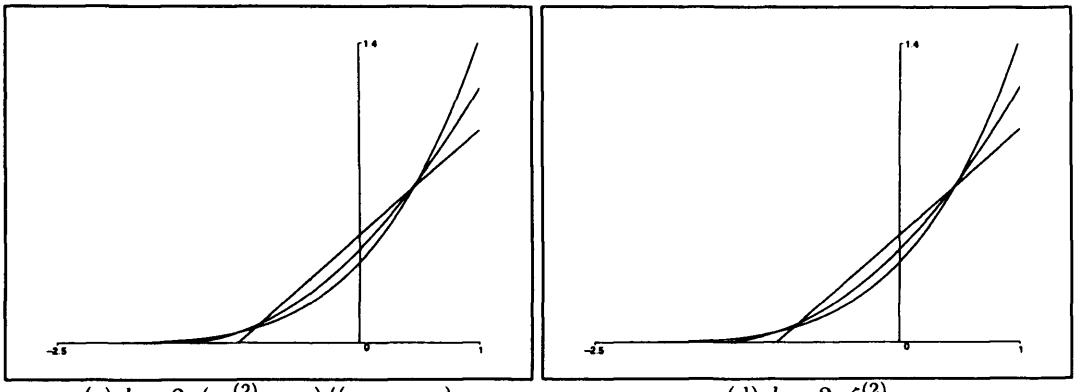
The densities of $\zeta^{(2)}$ and $(m^{(2)} - m)/(y_{k,n} - m)$ have been plotted for a variety of k and α in Figure 3.7. They verify that for n as small as 100 $\zeta^{(2)}$ is approximately equal in distribution to $(m^{(2)} - m)/(y_{k,n} - m)$.

Random variable	α	$\mathbf{E}\zeta$	$\mathbf{E}\zeta^2$	$\mathrm{Var}\zeta$
ζ_2°	2	0.333	0.333	0.222
ζ_2^*	2	0.500	0.375	0.125
ζ_2°	3	0.375	0.375	0.234
ζ_2^*	3	0.500	0.400	0.150
ζ_2°	5	0.417	0.417	0.243
ζ_2^*	5	0.500	0.429	0.179
ζ_3°	2	0.182	0.182	0.149
ζ_3°	3	0.220	0.219	0.171
ζ_3°	5	0.258	0.268	0.191

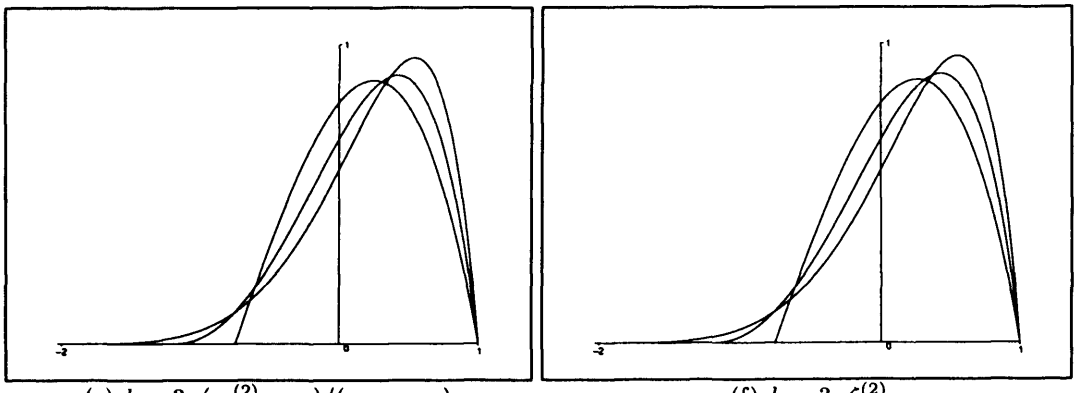
Table 3.2: Moments calculated from densities (3.4.1), (6.1.1) and (3.4.3) for $k = 2$ or 3 and $\alpha = 2$, 3 and 5.



(a) $\alpha = 3$, $(m^{(2)} - m)/(y_{k,n} - m)$ (b) $\alpha = 3$, $\zeta^{(2)}$



(c) $k = 2$, $(m^{(2)} - m)/(y_{k,n} - m)$ (d) $k = 2$, $\zeta^{(2)}$

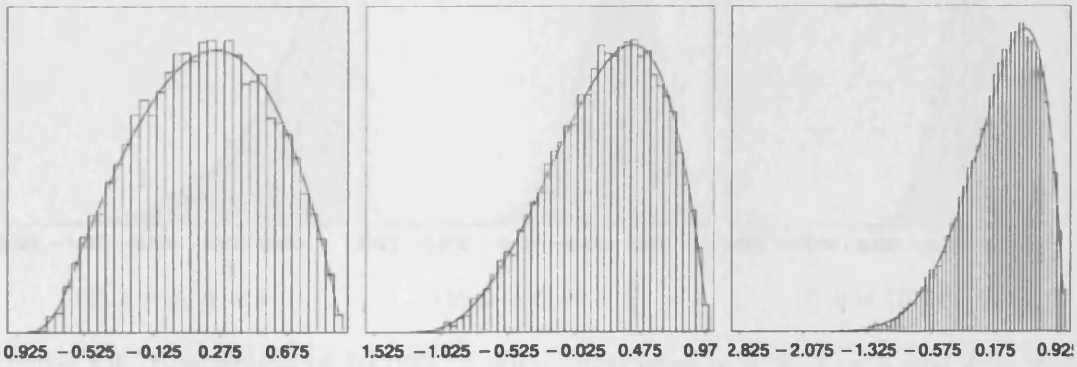


(e) $k = 3$, $(m^{(2)} - m)/(y_{k,n} - m)$ (f) $k = 3$, $\zeta^{(2)}$

Figure 3.7: Graphs on the left hand side show the asymptotic densities of the normalized estimator $(m^{(2)} - m)/(y_{k,n} - m)$. Graphs on the right show densities of $\zeta^{(2)}$. Plots (a) and (b) each show three densities with $\alpha = 3$, In each plot the different densities relate to $k = 2$, 3, 4, and 5. Plots (c) and (d) show densities with $k = 2$, plots (e) and (f) show densities with $k = 3$. There are three densities shown on each of these four plots, the three densities have parameters $\alpha = 2$, 3 and 5 respectively.

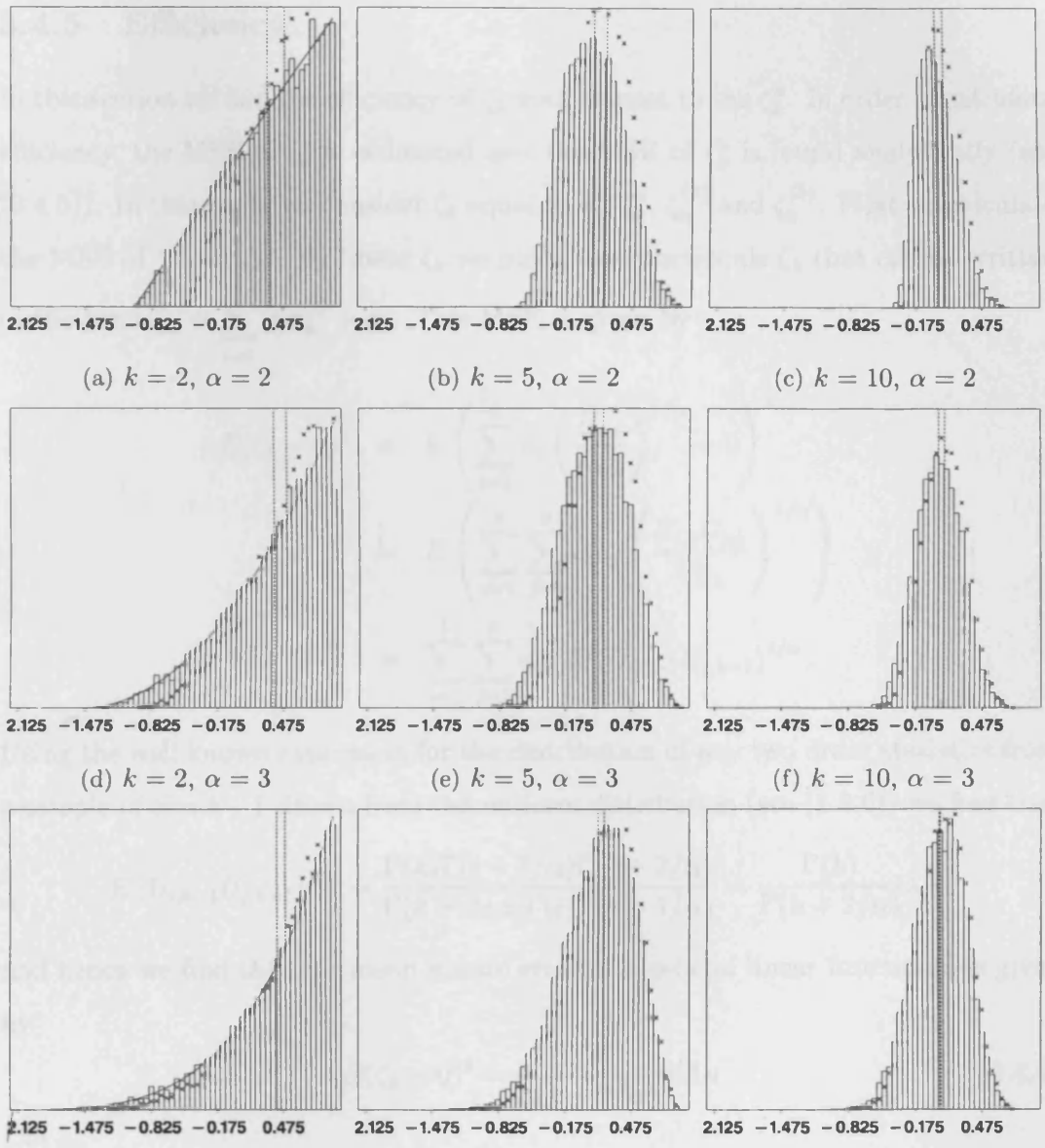
3.4.4 Histograms

This section shows the results of simulations that verify that for large n we approximately have $\zeta_k \stackrel{d}{=} (\hat{m} - m)/(y_{k,n} - m)$. 100 000 samples of size $k-1$ were drawn from a uniform distribution. The elements of each sample were sorted into ascending order and the resulting order statistics were labeled $x_1^{(t)}, \ldots, x_{k-1}^{(t)}$. For each of these samples ζ°, and ζ^* were calculated using $\alpha = 2$, 3 and 5 and $k = 2$, 3, 5 and 10. Figure 3.8 show histograms of ζ° with the density (6.1.1) plotted. It has been included to show that the derived density (6.1.1) fits the histograms of $\zeta^{(3)}$ (which is equal to ζ° when $k = 3$). Histograms of ζ° for $\alpha = 2$, 3 and 5 and $k = 2$, 5 and 10 can be found in Figure 3.9. The outline of the histogram of ζ^* is also marked, with $\times$. By comparing Figure 3.9 to Figure 2.7 we can see that ζ° appears to have a similar density to the density of $(m^\circ - m)/(y_{k,n} - m)$ when $n = 100$ and ζ^* appears to have a similar density to $(m^* - m)/(y_{k,n} - m)$ when $n = 100$. We have seen from earlier simulations that the density of $(\hat{m} - m)/(y_{k,n} - m)$ when $n = 100$ is very close to its asymptotic density. The density 3.4.1 has been plotted on the histograms related to $k = 2$ and appears to fit the histogram well.



(a) $\alpha = 2$, ζ° (b) $\alpha = 3$, ζ° (c) $\alpha = 5$, ζ°

Figure 3.8: Histograms of $\zeta^\circ = \zeta^{(3)}$. Here α =2, 3 and 5. $t = 1 \ldots 100\,000$. $k = 3$. Histograms have been plotted with density (6.1.1).



(a) $k = 2$, $\alpha = 2$ (b) $k = 5$, $\alpha = 2$ (c) $k = 10$, $\alpha = 2$

(d) $k = 2$, $\alpha = 3$ (e) $k = 5$, $\alpha = 3$ (f) $k = 10$, $\alpha = 3$

(g) $k = 2$, $\alpha = 5$ (h) $k = 5$, $\alpha = 5$ (i) $k = 10$, $\alpha = 5$

Figure 3.9: Histograms of 10 000 ζ_k° with parameter $\alpha = 2$, 3 or 5 and $k = 2$, 5 and 10 (as marked below each histogram). From the same 10 000 Markov chains ζ_k^* was calculated. The percentage frequencies of ζ_k^* (crosses) were plotted at each of the midpoints of the original histogram. The sample means of ζ_k° and ζ_k^* are plotted using vertical dashed lines. The sample mean of the ζ_k° is less than that of ζ_k^* in all of the above plots. The density of $\zeta_2^{(2)} = \zeta_2^\circ$ (3.4.1) has been plotted on the histograms relating to $k = 2$ with a solid line.

3.4.5 Efficiency

In this section we find the efficiency of ζ_k with respect to the ζ_k°. In order to calculate efficiency, the MSE of ζ_k is estimated and the MSE of ζ_k° is found analytically (see (3.4.5)). In this study we consider ζ_k equal to ζ_k°, ζ_k^*, $\zeta_k^{(2)}$ and $\zeta_k^{(3)}$. First we calculate the MSE of 'linear ζ_k'. By linear ζ_k we mean the functionals ζ_k that can be written in the form $\zeta_k = \sum_{i=1}^{k-1} a_i x_k^{(t)} + a_k$. This MSE is given by

$$\begin{aligned} E(\zeta_k - 0)^2 &= E\left(\sum_{i=1}^{k} a_i \left(\frac{U_{i,k}}{U_{k,k}}\right)^{1/\alpha} - 0\right)^2 \\ &= E\left(\sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j \left(\frac{U_{i,k}U_{j,k}}{U_{k,k}^2}\right)^{1/\alpha}\right) \\ &= \sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j E\left(U_{i,k-1}U_{j,k-1}\right)^{1/\alpha}. \end{aligned}$$

Using the well known expression for the distribution of any two order statistics from a sample of size $k-1$ drawn from the uniform distribution (see (1.3.6)) we find that

$$E\left(U_{i,k-1}U_{j,k-1}\right)^{1/\alpha} = \frac{\Gamma(k)\Gamma(i+1/\alpha)\Gamma(j+2/\alpha)}{\Gamma(k+2/\alpha)\Gamma(i)\Gamma(j+1/\alpha)} = \frac{\Gamma(k)}{\Gamma(k+2/\alpha)}\lambda_{j,i},$$

and hence we find that the mean square error of a general linear functional is given by

$$E(\zeta_k - 0)^2 = \frac{\Gamma(k)}{\Gamma(k+2/\alpha)} a'\Lambda a \tag{3.4.4}$$

and so

$$E(\zeta_k^\circ - 0)^2 = \frac{\Gamma(k)}{\Gamma(k+2/\alpha)} a^{\circ\prime}\Lambda a^\circ = \frac{\Gamma(k)}{\Gamma(k+2/\alpha)\mathbf{1}'\Lambda^{-1}\mathbf{1}}. \tag{3.4.5}$$

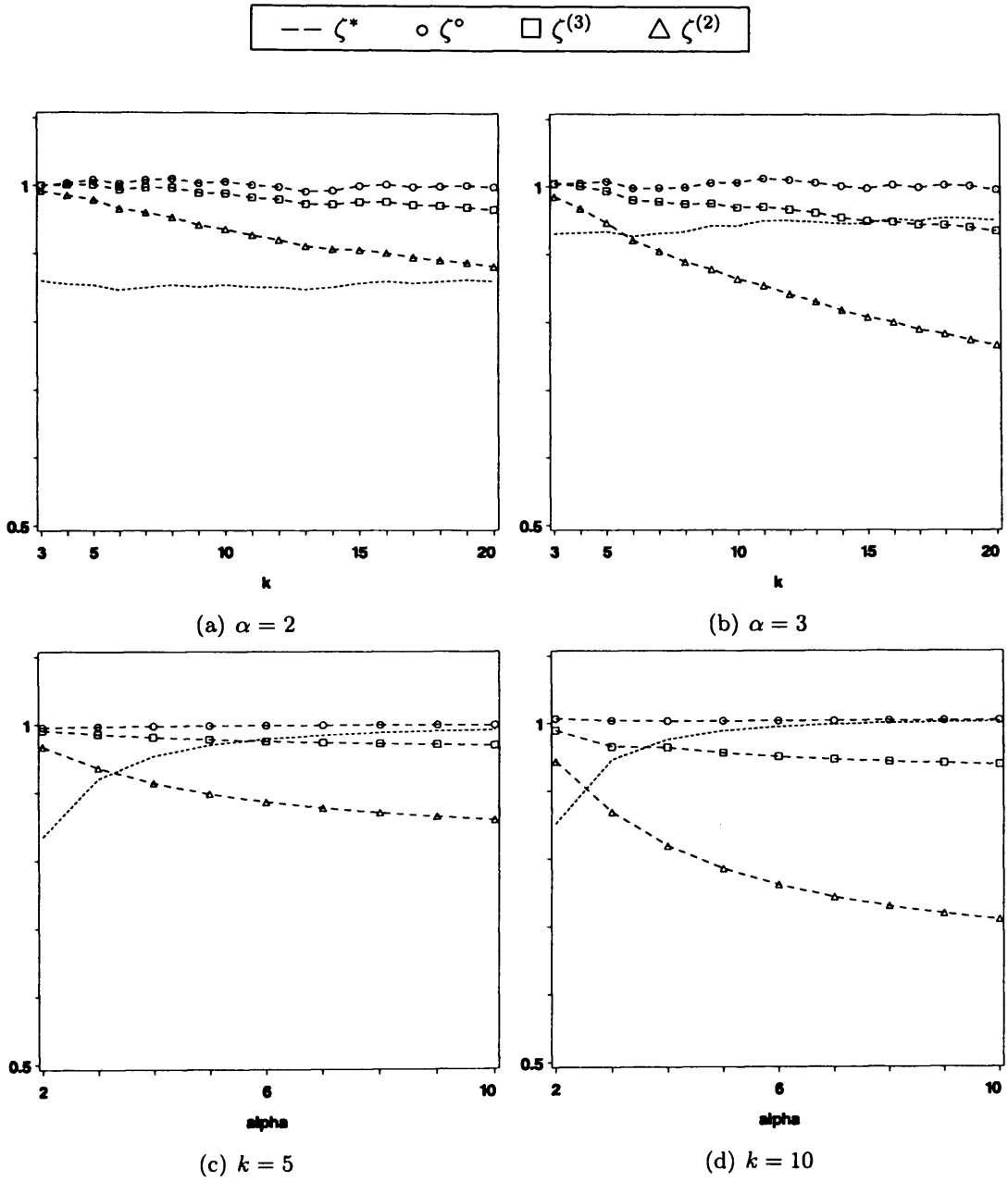
Therefore in a similar way to in Section 2.4.2 we can estimate efficiency of ζ_k with respect to ζ° by

$$\text{eff}(\zeta_k) = \frac{\Gamma(k)}{\Gamma(k+2/\alpha)\mathbf{1}'\Lambda^{-1}\mathbf{1}} \Big/ \frac{1}{R}\sum_{r=1}^{R}(\zeta_k^r - 0)^2. \tag{3.4.6}$$

Here each ζ_k^r is calculated from a separate set of $k-1$ uniform order statistics.

From (3.4.4) and (3.4.5) we can see that the efficiency of linear ζ_k is equal to the asymptotic efficiency of $\hat{m}$ for $F(x) = c_0(x-m)^\alpha + (o(x-m)^\alpha)$. When $F(x) = c_0(x-m)^\alpha$, e.g. the beta distribution with $\beta = 1$, the (non-asymptotic) efficiency of $\hat{m}$ is very close to the asymptotic efficiency of $\hat{m}$. Therefore the finite sample efficiency of $\hat{m}$ is very close to the efficiency of linear ζ_k when $F(x)$ is Beta (with $\beta = 1$) or Weibull. This is verified by comparing Figure 3.10 to Figure 2.8.

The efficiency of ζ° and ζ^* in Figure 3.10 are approximately constant in k. When $\alpha = 2$ they are approximately 1 and 0.86 respectively. The efficiency of m° and m^* when $F(x)$ is the beta c.d.f. with $\alpha = 2$, $\beta = 1$ and $n = 100$ are also 1 and 0.86 respectively (see Figure 2.11).



(a) $\alpha = 2$

(b) $\alpha = 3$

(c) $k = 5$

(d) $k = 10$

Figure 3.10: Estimated efficiency of ζ_k^*, ζ_k°, $\zeta_k^{(3)}$ and $\zeta_k^{(2)}$ with respect to (analytic) ζ_k° as defined above. In figures (a) and (b) $k = 2 \ldots 20$. In figure (c) $k = 5$ and in (d) $k = 10$. In figures (a) and (b) $\alpha = 2$ and 3 respectively. In figures (c) and (d) α takes integer values in the range $[2, 10]$.

3.5 Functional Model: Unknown Value of the tail index Study

In a similar way to Section 2.5, we now investigate how the functionals ζ°, $\zeta^{(2)}$, $\zeta^{(3)}$ and ζ^{*} react to using the wrong value of the tail index. That is to say we study the efficiency of linear functionals

$$\zeta_k(\vartheta) = \sum_{i=1}^{k} a_i(\vartheta) x_i^{(t)1/\alpha}$$

where $a_i(\vartheta)$ are coefficients as considered in Section 2.5 and $x_k^{(t)} = 1$. We also study the efficiency of the functional $\zeta^*(\vartheta)$. The functional $\zeta^*(\vartheta)$ is equal to the solution z to the following equation:

$$(\vartheta - 1) \sum_{i=1}^{k-1} \frac{1 - (x_i^{(t)})^{1/\alpha}}{(x_i^{(t)})^{1/\alpha} - z} = k.$$

When $k = 2$ the solution z is linear in $x_1^{(t)}$ and $x_2^{(t)}$.

First we derive the efficiencies analytically. We then undertake a simulation study that confirms and extends the analytic results.

3.5.1 Analytical Comparison of Estimators

It is easy to show analytical results similar to those in Section 2.5. From (3.4.5) we have the MSE of $\zeta_k^{\circ}(\alpha)$. This will be the benchmark by which the other functionals will be judged. From (3.4.4) we have that the MSE of linear $\zeta_k(\vartheta)$ is given by

$$E(\zeta_k(\vartheta) - 0)^2 = \frac{\Gamma(k)}{\Gamma(k + 2/\alpha)} (a(\vartheta))' \Lambda a(\vartheta).$$

Therefore by considering $\text{eff}(\zeta_k) = E(\zeta_k^{\circ} - 0)^2 / E(\zeta_k - 0)^2$ we can say that

$$\text{eff}(\zeta_k^{\circ}(\vartheta)) = \text{eff}(m^{\circ}(\vartheta)) = \frac{(1'\Lambda_0^{-1}1)^2}{1\Lambda^{-1}1 \cdot 1'\Lambda_0^{-1}\Lambda\Lambda_0^{-1}1}$$

and

$$\text{eff}(\zeta_k^{\bullet}(\vartheta)) = \text{eff}(\zeta_k^{\bullet}) = \text{eff}(m^{\bullet}) = \frac{1}{1\Lambda^{-1}1 \cdot \lambda_{1,1}}$$

Also, as when $k = 2$, ζ_k^* and m^* are linear functions of their respective order statistics we can say that, for $k = 2$

$$\text{eff}(\zeta_2^*) = \text{eff}(m^*(\vartheta)) = \frac{\alpha(\alpha+2)}{2\alpha^2 + 4\alpha - 2\alpha\vartheta - 2\vartheta + \vartheta^2 + 1}.$$

The efficiency $\text{eff}(\zeta_2^*(\vartheta))$ is maximized at $\vartheta = \alpha + 1$. These three efficiencies have been plotted on the graphs of efficiencies from simulation.

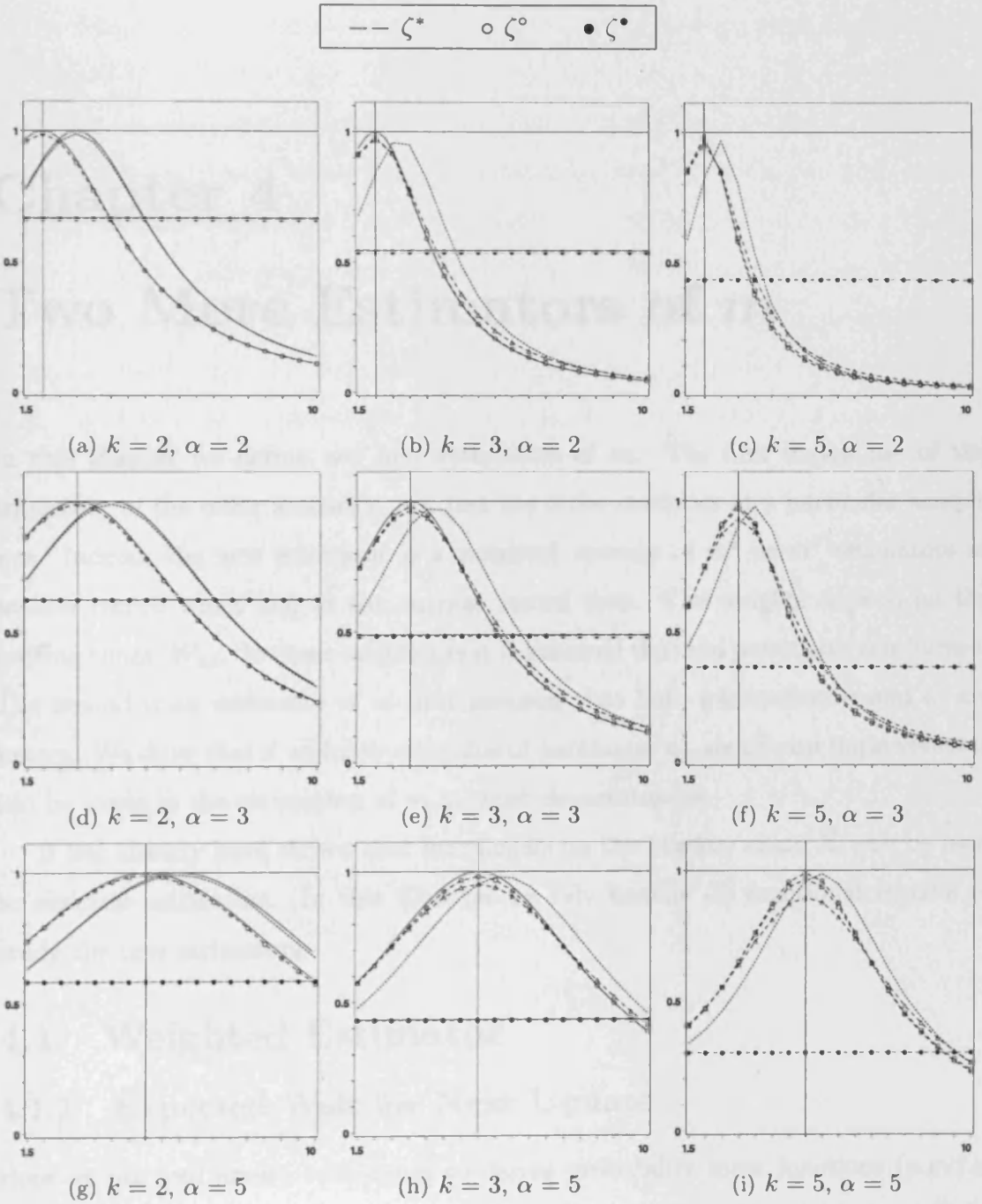
3.5.2 Simulation results

In this section we make a similar simulation study to the one in Section 2.5.2. Here 10 000 samples of size $k-1$ are drawn from a uniform distribution. The sample is sorted into ascending order, then the resulting order statistics labeled $x_1^{(t)}, \ldots, x_{k-1}^{(t)}$. These are used to create functionals $\zeta^\circ(\vartheta)$, $\zeta^*(\vartheta)$, $\zeta^{(2)}(\vartheta)$, $\zeta^{(3)}(\vartheta)$ and $\zeta^\bullet$. An estimate of the efficiency with respect to (analytic) $\zeta^\circ(\alpha)$ of these functionals is made, and has been plotted against ϑ for a range of k and α in Figure 3.11. The efficiency has been estimated by

$$\text{E}(\zeta^\circ(\alpha) - 0)^2 \Big/ \frac{1}{R}\sum_{r=1}^{10\,000} ((\zeta_k(\vartheta))^r - 0)^2 .$$

The analytical efficiencies $\text{eff}(\zeta_k^\circ)$, $\text{eff}(\zeta_k^\bullet)$ and (for $k = 2$ only) ζ_k^* are also plotted on Figure 3.11.

Figure 3.11 verifies the derived efficiencies; the derived efficiencies plotted match the simulation data well. It also shows that the value of ϑ that maximizes $\text{eff}(\zeta_k^*(\vartheta))$ is greater than α (but less than $\alpha+1$). The efficiencies of ζ_k^* (for all k and α plotted) appear to be the same as the efficiency of m^* shown in Figure 2.17.



(a) $k = 2$, $\alpha = 2$ (b) $k = 3$, $\alpha = 2$ (c) $k = 5$, $\alpha = 2$

(d) $k = 2$, $\alpha = 3$ (e) $k = 3$, $\alpha = 3$ (f) $k = 5$, $\alpha = 3$

(g) $k = 2$, $\alpha = 5$ (h) $k = 3$, $\alpha = 5$ (i) $k = 5$, $\alpha = 5$

Figure 3.11: Estimated efficiency, $\text{eff}(\zeta_k(\vartheta))$, of ζ_k^*, ζ_k° and $\zeta_k^\bullet$ as a function of ϑ. ϑ varies from 1.5 to 10, $\alpha = 2$, 3 and 5 and $k = 2$, 3 and 5. The estimated efficiency is calculated from a sample of 10 000. Also shown are the analytic efficiencies of ζ_k° and $\zeta_k^\bullet$, and where $k = 2$, ζ_k^*.

Chapter 4

Two More Estimators of m

In this chapter we define two new estimators of m. The first makes use of the trajectory of the order statistics, not just the order statistics at a particular sample size. Indeed, the new estimator is a weighted average of m° or m^* estimators at smaller record times and at the current record time. The weights depend on the waiting times, $W_{k,t}$. In these estimators it is assumed that the parameter α is known. The second is an estimator of m that assumes that both parameters α and c_0 are known. We show that if we know the value of parameter c_0, significant improvements can be made in the estimation of m in most circumstances.

It has already been shown that functionals on the Markov chain X_t can be used to simulate estimators. In this Chapter we rely heavily on functional models to study the new estimators.

4.1 Weighted Estimator

4.1.1 Expected Wait for Next Update

Here we use well known techniques to derive probability mass functions (p.m.f.s) and expectations relating to waiting times. Throughout this section we will be considering the t^{th} type 2 k^{th} records. In order to simplify the explanations, we will refer to these simply as t^{th} records, where t is equal to the record count, $N_{k,n}$. This means, for example, that if we refer to the t^{th} waiting time we mean the t^{th} type 2

k^{th} waiting time, $W_{k,t}$, and if we refer to the t^{th} record time we mean the t^{th} type 2 k^{th} record time, $T_{k,t}$.

Below we derive the expectation and variance of the waiting time for the $t+1^{\text{th}}$ record given that the t^{th} record time is known. In order to do this we first find the conditional probability mass function of the $t+1^{\text{th}}$ record time given the t^{th} record time is known. This can be derived as follows. If the t^{th} record occurs at time l, then in order for the $t+1^{\text{th}}$ record to occur at time r (where $r > l$), no records may occur at times $l+1, \ldots, r-1$, and a record must occur at time r. The probability that a record occurs at some time n is given by $\mathrm{P}(I_{k,n} = 1) = k/n$, this means that $\mathrm{P}(I_{k,n} = 0) = (n-k)/n$. This gives us

$$\begin{aligned} \mathrm{P}(T_{k,t+1} = r | T_{k,t} = l) &= \mathrm{P}(I_{k,l+1} = 0, \ldots I_{k,r-1} = 0, I_{k,r} = 1) \\ &= \frac{k(r-k-1)!l!}{r!(l-k)!}, \quad k \leq l < r \end{aligned}$$

By definition $W_{k,t+1} = T_{k,t+1} - T_{k,t+1} - 1$, this means that conditional p.m.f. of the waiting time $W_{k,t+1}$ given $T_{k,t}$ is

$$\mathrm{P}(W_{k,t+1} = w | T_{k,t} = l) = \frac{k(l+w-k)!l!}{(l+w+1)!(l-k)!}, \quad w \geq 0, k \leq l. \qquad (4.1.1)$$

Using this we can calculate the expectation of $W_{k,t+1}$, the expectation of $W_{k,t+1}^2$ and the variance of $W_{k,t+1}$. These are given by

$$\begin{aligned} \mathrm{E}(W_{k,t+1} | T_{k,t} = l) &= \frac{l-k+1}{k-1} \\ \mathrm{E}(W_{k,t+1}^2 | T_{k,t} = l) &= \frac{(2l-k+2)(l-k+1)}{(k-2)(k-1)} \\ \mathrm{Var}(W_{k,t+1} | T_{k,t} = l) &= \frac{lk(l-k+1)}{(k-2)(k-1)^2} \end{aligned}$$

This shows that, given the t^{th} record time, the waiting time of $t+1^{\text{th}}$ record is independent of record number, it depends only on k and the record time, $T_{k,t}$, of the previous record. The expected wait for the next update when $k = 1$ is infinite for all $l \geq 1$.

We now wish to investigate whether, if we wait for a new record but none come, have we reduced the amount of time that we expect to wait? That is to say, we wish to find out if the expectation of $T_{k,t+1} - n - 1$ is longer if $n = T_{k,t}$ or if $n = T_{k,t} + j$, where $j \in \mathbb{Z}_+$ and n is the current time. We assume $n < T_{k,t+1}$.

Letting $T_{k,t} = l$, the p.m.f. of $T_{k,t+1} - n - 1$ when $n = T_{k,t} = l$ is equal to (4.1.1). This means that the expectation of $T_{k,t+1} - n - 1$ is $\frac{l-k+1}{k-1}$.

The probability mass function of $T_{k,t+1} - n - 1$ where $n = T_{k,t} + j = l + j$, (assuming $n < T_{k,t+1}$) is given by the following expression

$$\begin{aligned} \mathrm{P}(T_{k,t+1} - n - 1 = w | n = l + j,\, n < T_{k,t+1}) \\ &= \mathrm{P}(I_{k,l+j+1} = 0, \ldots, I_{k,l+j+w} = 0, I_{k,l+j+w+1} = 1) \\ &= \frac{k(l+j+w-k)!(l-+j)!}{(l+j+w+1)!(l+j-k)!}. \end{aligned}$$

The expectation of the wait for record $t+1$ at time $n = l + j$ (assuming $T_{k,t} \leq l + j < T_{k,t+1}$) is given by $\frac{l+j-k+1}{k-1}$. This is larger than $\frac{l-k+1}{k-1}$. From this it can be seen that the longer you wait for a record, the longer you expect to wait for it. The relationship is linear: for every observation that you wait, you expect to wait an extra $1/(k-1)$ observations on top of your original expected wait. Indeed,

$$\mathrm{E}(T_{k,t+1} - n - 1 | n = l + j) - \mathrm{E}(T_{k,t+1} - n - 1 | n = l) = \frac{j}{k-1}.$$

This phenomenon is demonstrated through the following example. If $k = 4$ and the t^{th} type 2 4^{th} record occurs at time $n = 50$ (i.e., $T_{4,t} = 50$ for some $t \in \mathbb{Z}^+$), the expected wait for the next record, $T_{4,t+1}$, is $\frac{50-4+1}{4-1} = 15\frac{2}{3}$. If we make one more observation (so now $n = 51$) and it isn't a record, our expected wait for the next record is $\frac{51-4+1}{4-1} = 15\frac{2}{3} + \frac{1}{3} = 16$. If we then make a further 19 observations without encountering a new type 2 4^{th} record, the expected waiting time becomes $\frac{70-4+1}{4-1} = 23\frac{1}{3}$. So the expected wait for the next record has increased because of having waited longer. Figure 4.1 shows the expected wait for record $t+1$ against n, the current time.

From the discussion above we can see that the p.m.f. can be found conditionally on the current time n. We will see below that it can also be found conditionally on

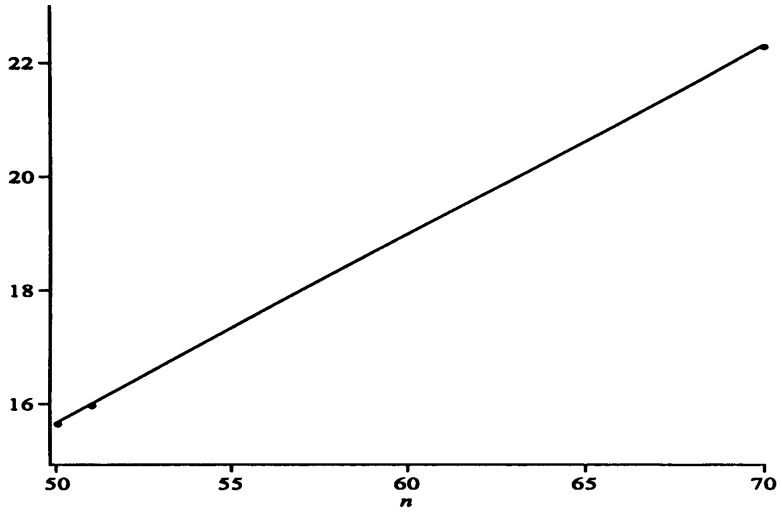


Figure 4.1: The y-axis of this figure shows $T_{4,t+1} - n + 1$, this is the expected wait until $T_{4,t+1}$ from time n. The x-axis shows the time n. We have made the assumption that $T_{4,t} \leq n < T_{4,t+1}$. Also marked on the plot with $\bullet$ are the points related to the above example: at $n = l = 50$ the expected wait is $15\frac{2}{3}$, at $n = 51$ the expected wait is 16, at $n = 70$ the expected wait is $23\frac{1}{3}$.

the value of the k^{th} order statistic at time n, provided that we know the distribution $F(x)$. At time n ($T_{k,t} \leq n < T_{k,t+1}$), the conditional p.m.f. of the wait for the next record (given $y_{k,n}$) is given by

$$\mathrm{P}[T_{k,t+1} - n - 1 = w | y_{k,n} = x] = (1 - F(x))^w F(x), \quad w \in \mathbb{N}.$$

In particular, if $n = T_{k,t}$, we have $T_{k,t+1} - n - 1 = W_{k,t+1}$ and so

$$\mathrm{P}[W_{k,t+1} = w | y_{k,n} = x] = (1 - F(x))^w F(x), \quad w \in \mathbb{N}.$$

Notice that for n such that $T_{k,t} \leq n < T_{k,t+1}$, we have that $y_{k,n}$ is constant for all n and $y_{k,n} \equiv y_{k,T_{k,t}}$.

The conditional c.d.f. of $W_{k,t+1}$ given $y_{k,n}$ ($T_{k,t} \leq n < T_{k,t+1}$) is

$$\mathrm{P}[W_{k,t+1} \leq w | y_{k,n} = x] = \sum_{i=0}^{w} (1 - F(x))^i F(x) = 1 - (1 - F(x))^{w+1}, \quad w \in \mathbb{N}. \quad (4.1.2)$$

Notice that this is independent of t.

4.1.2 A New Estimator: Weighted Average Estimator

We now consider a new estimator:

$$\hat{m}_{k,n}^{WE} = \frac{1}{T_{k,N_{k,n}} - k} \left[\sum_{j=1}^{N_{k,n}} (W_{k,j} + 1) \hat{m}_{k,T_{k,j}} \right] \tag{4.1.3}$$

where $\hat{m}_{k,n}$ is an estimator based on the k smallest order statistics from a sample of n. $\hat{m}_{k,T_{k,j}}$ is therefore the estimator at record time $n = T_{k,j}$. We will refer to the estimator $\hat{m}_{k,n}$ as the original estimator associated with the weighted estimator $\hat{m}_{k,n}^{WE}$. The notation that will be used below is as follows: if the original estimator is the optimal linear estimator, then its associated weighted estimator (WE) is denoted $m^{\circ WE}$ and if the original estimator is m^*, then the associated WE is denoted m^{*WE}.

The estimator given by (4.1.3) is a weighted average of the elements of the trajectory of estimators $\hat{m}_{k,n}$, $n = T_{k,1}, T_{k,2}, \ldots$. The weight given to each original estimator $\hat{m}_{k,T_{k,t}}$ is $\left(\frac{W_{k,t}+1}{T_{k,N_{k,n}} - k} \right)$. The numerator is one greater than the waiting time for the record. This means that, if $T_{k,t+1} = T_{k,t} + j$ then the numerator for the weight for estimator $\hat{m}_{k,T_{k,t+1}}$ is $T_{k,t+1} - T_{k,t} = j$. The denominator is the total number of random variables added since sample size $n = k$. This means that a larger weight is given to those estimator that we had to wait longer for.

The WE has been designed to create a greater correlation between consecutive members of a trajectory of estimators (as $n \to \infty$) than the original estimator. In a trajectory of original estimators as $n \to \infty$, sometimes $\hat{m}_{k,n}$ overestimates and sometimes it underestimates. By taking an average of these estimates we hope to reduce bias and produce a more predictable estimator. Larger weights are given to the estimates that we had to wait longer for, we hypothesize that if we have had to wait a long time for an estimator it may be a particularly good one. This weighting will tend to favor estimates made at larger sample sizes, which in general are better.

We wish to compare $\hat{m}_{k,n}^{WE}$ to $\hat{m}_{k,n}$ through a simulation study. Even for very large sample sizes the estimator $\hat{m}_{k,n}^{WE}$ will not have updated very often, we therefore carry out a simulation study using functionals of the Markov chain X_t (as defined

in Section 3.2). In Section 4.1.3 we define a functional on X_t that has the same distribution as $\hat{m}^{WE}_{k,T_{k,t}}$. In Table 4.1 and the scatter plots in Figures 4.2 and 4.3 we verify that the distributions of the functional and estimator are equal. We then study trajectories of this new functional and its efficiency with respect to the asymptotic functional ζ°.

4.1.3 Weighted Functional

Here we derive a functional, $\zeta^{WE}_{k,t}$, that is equal in distribution to $(\hat{m}^{WE}_{k,n} - m)/(y_{k,n} - m)$. Earlier we saw that $\zeta_{k,t}\left(\frac{U_{k,T_{k,t}}}{c_0}\right)^{1/\alpha} + m \sim \hat{m}_{k,T_{k,t}}$, therefore

$$\hat{m}^{WE}_{k,n} \sim \frac{1}{T_{k,N_{k,n}} - k}\left[\sum_{j=1}^{N_{k,n}}(W_{k,j}+1)\left(\frac{U_{k,T_{k,j}}}{c_0}\right)^{1/\alpha}\zeta_{k,j}\right] + m.$$

If we normalize by subtracting m and multiplying by $\left(\frac{c_0}{U_{k,T_{k,t}}}\right)^{1/\alpha}$, then after simplification we obtain

$$(\hat{m}^{WE}_{k,n} - m)\left(\frac{c_0}{U_{k,T_{k,t}}}\right)^{1/\alpha} \sim \frac{1}{\left(U_{k,T_{k,t}}\right)^{1/\alpha} T_{k,N_{k,n}} - k}\left[\sum_{j=1}^{N_{k,n}}(W_{k,j}+1)\left(U_{k,T_{k,j}}\right)^{1/\alpha}\zeta_{k,j}\right] \tag{4.1.4}$$

Before we define the functional $\zeta^{WE}_{k,t}$, we wish to be able to model $U_{k,T_{k,t}}$, $W_{k,t}$ and $T_{k,t}$ without having to simulate order statistics between records. We must not model $U_{k,T_{k,t}}$ independently from the functional $\zeta_{k,j}$, or the weighted functional will not be equal in distribution to its related WE. If $U_{k,\mathbf{t}}$ are the random variables that normalize the Markov chain X_t (see Section 3.2.2), we can model $U_{k,T_{k,1}}$ with $\tilde{U}_1 = U_{k,1}$. Subsequent uniform order statistics $U_{k,T_{k,t}}$ can be modeled by $\tilde{U}_t = \tilde{U}_{t-1}U_{k,\mathbf{t}}$.

Modeling $W_{k,t}$ and $T_{k,t}$

Letting

$$\tilde{W}_{k,t} + 1 = \left\lceil \frac{\log(1-y)}{\log(1-\tilde{U}_t)} \right\rceil, \quad t = 2, 3, \ldots$$

we have that $\tilde{W}_{k,t} \stackrel{d}{=} W_{k,t}$. This expression for $\tilde{W}_{k,t}$ was obtained by inverting (4.1.2). Letting $\tilde{T}_{k,1} = k$ and $\tilde{T}_{k,t} = \tilde{T}_{k,t-1} + \tilde{W}_{k,t} + 1$, $t > 1$, we have $\tilde{T}_{k,t} \stackrel{d}{=} T_{k,t}$.

Now let us define $\zeta_{k,t}^{WE}$ as follows.

$$\zeta_{k,t}^{WE} = \frac{1}{(\tilde{T}_{k,t} - k)\tilde{U}_t^{1/\alpha}} \sum_{i=1}^{t} (\tilde{W}_{k,i} + 1)\tilde{U}_i^{1/\alpha} \zeta_{k,i}. \tag{4.1.5}$$

As the R.H.S. of (4.1.5) is equal in distribution to the R.H.S. of (4.1.4), and $y_{k,n} \stackrel{d}{=} F^{-1}(U_{k,n})$, we have that $\zeta_{k,t}^{WE} \stackrel{d}{=} (\hat{m}_{k,n}^{WE} - m)/(y_{k,n} - m)$.

We say that the functional $\zeta_{k,t}$ is the original functional associated with the weighted functional $\zeta_{k,t}^{WE}$. We denote by $\zeta_{k,t}^{\circ WE}$ the weighted functional associated with the original functional $\zeta_{k,t}^{\circ}$. We denote by $\zeta_{k,t}^{*WE}$ the weighted functional associated with the original functional $\zeta_{k,t}^{*}$. We say that $\zeta_{k,t}^{\circ WE}$ is the functional related to the estimator $m_{k,T_{k,t}}^{\circ}$ and $\zeta_{k,t}^{*WE}$ is the functional related to the estimator $m_{k,T_{k,t}}^{*}$. As with earlier functionals we will drop the t and k subscripts where convenient. We will also drop k, n subscripts or n subscripts where convenient on estimators.

4.1.4 Simulation Study

In Chapter 3 it was demonstrated that if $F(x) \sim c_0(x - m)^\alpha$ as $n \to \infty$, then for large n we have $(\hat{m} - m)/(y_{k,n} - m) \stackrel{d}{=} \zeta_k$. The scatter plots in Figures 4.2 and 4.3, and moments shown in Table 4.1, verify that $(m^\circ - m)/(y_{k,n}) \stackrel{d}{=} \zeta_k^\circ$, $(m^{\circ WE} - m)/(y_{k,n} - m) \stackrel{d}{=} \zeta_k^{\circ WE}$, $(m^* - m)/(y_{k,n}) \stackrel{d}{=} \zeta_k^*$ and $(m^{*WE} - m)/(y_{k,n} - m) \stackrel{d}{=} \zeta_k^{*WE}$. The subscripts have been dropped on the functionals as the equality in distribution is true for any t.

In this simulation study we simulate 10 000 trajectories of random variables from the distribution $F(x) = c_0(x - m)^\alpha$, with $c_0 = 1$, $m = 0$ and $\alpha = 3$. Each trajectory is of length $T_{k,c+1}$. Here c is the record number such that $T_{k,c+1} < 1\,000$ and $T_{k,c+2} \geq 1\,000$. At each record number in each trajectory the order statistics are calculated, and estimates m°, m^*, $m^{\circ WE}$ and m^{*WE} are made. Similarly, 10 000 trajectories of Markov Chain X_t and $\tilde{U}_t$, $1 \leq t \leq C$ are made. Here C is such that $\tilde{T}_{k,C+1} < 10^{13}$ and $\tilde{T}_{k,C+1} \geq 10^{13}$. For each t the functionals ζ°, ζ^*, $\zeta^{\circ WE}$ and ζ^{*WE} are calculated. These functionals use the parameter $\alpha = 3$. We also define the

Estimator	Mean				Std Dev			
k	2	3	4	5	2	3	4	5
$\frac{m^\circ - m}{y_{k,n} - m}$	0.261	0.167	0.122	0.101	0.511	0.423	0.364	0.321
$\frac{m^* - m}{y_{k,n} - m}$	0.409	0.291	0.228	0.193	0.409	0.360	0.320	0.288
$\frac{m^{\circ WE} - m}{y_{k,n} - m}$	0.499	0.288	0.195	0.149	0.447	0.391	0.342	0.302
$\frac{m^{*WE} - m}{y_{k,n} - m}$	0.692	0.466	0.350	0.285	0.371	0.330	0.296	0.268
ζ°	0.268	0.171	0.122	0.097	0.510	0.423	0.361	0.324
ζ^*	0.414	0.294	0.227	0.189	0.408	0.361	0.318	0.291
$\zeta^{\circ WE}$	0.493	0.286	0.200	0.154	0.457	0.384	0.335	0.303
ζ^{*WE}	0.686	0.464	0.355	0.289	0.373	0.325	0.291	0.268

Table 4.1: Sample mean and standard deviation of normalized estimators m°, m^*, $m^{\circ WE}$ and m^{*WE} and their related functionals. Mean and standard deviation are calculated over 10 000 samples of estimators at $n = T_{k,c}$ and functionals at $\tilde{T}_{k,\tilde{c}}$, where c and $\tilde{c}$ are as defined above.

record number $\tilde{c}$ such that $\tilde{T}_{k,\tilde{c}+1} < 1\,000$ and $\tilde{T}_{k,\tilde{c}+2} \geq 1\,000$.

Table 4.1 shows the means and standard deviations for various k of the normalized estimators and their related functionals. The means and standard deviations of estimators are based on 10 000 runs of the estimators $\hat{m}_{k,T_{k,c}}$ and $\hat{m}^{WE}_{k,T_{k,c}}$ and the functionals $\zeta_{k,\tilde{c}}$ and $\zeta^{WE}_{k,\tilde{c}}$. The means of the weighted estimators and weighted functionals are larger than the means of their associated original estimators and functionals. The variances of the weighted estimators and weighted functionals are less than the variance of their associated original estimators and functionals. The sample means and variances of the estimators and functionals whose distribution we have shown to be equal are similar, for example the mean of $m^{\circ WE}$ is similar to that of $\zeta^{\circ WE}$ for all k.

The scatter plots in Figures 4.2 and 4.3 show the estimators and functionals at record time c (or $\tilde{c}$), plotted against the same estimator or functional at time $c+1$ or $(\tilde{c}+1)$. The scatter plots demonstrate that the weighted functionals defined in

(4.1.5) can be used to model a trajectory of weighted estimators at record times. By considering the shape made by the points of the scatter plots and their Pearson's correlation coefficient, it can be seen that the WE has greater correlation between consecutive elements of its trajectory than the original estimator.

The plots in Figure 4.4 show a single trajectory (of length C) of $\zeta^{\circ}_{k,t}$ and $\zeta^{*}_{k,t}$ where $k = 4$ and 5. Also plotted are the associated weighted functionals $\zeta^{\circ WE}_{k,t}$ and $\zeta^{*WE}_{k,t}$ (for the same values of k) and the random variable $\tilde{U}_t^{1/\alpha}$ simulated at the same time as $\zeta^{\circ WE}_{k,t}$ and $\zeta^{*WE}_{k,t}$. In all of these plots $\alpha = 3$. From these plots it can be seen that the weighted functionals have a less erratic trajectory than their associated original functionals; in each trajectory, the most extreme data points of the original functionals are more extreme than the most extreme data points of their associated weighted functionals; in the trajectory of the weighted functional, the occurrence of a very large observation next to a very small observation is rarer than in the trajectory of the original functional.

Figure 4.5 shows the efficiency (3.4.6) of ζ°, $\zeta^{\circ WE}$, ζ^{*} and ζ^{*WE} plotted against t. The efficiency is calculated from 10 000 runs. At $t = 1$ the weighted functional is equal to its associated original functional and so their efficiencies are equal. For low t, as t increases the efficiency of the weighted functional decreases. The efficiency of the weighted functionals is fairly constant for larger t. The efficiency of the original functional varies about a constant value, (approximately 1 for ζ° and 0.93 for ζ^{*}). The efficiency of the WEs is lower than that of their related original functional. This difference in efficiency reduces as k increases. The trajectory of efficiency of the original estimator is much less smooth than the trajectory of efficiency of the WE. It can be seen from Figure 4.5 that (although it improves with k) the WE is not as efficient as the original estimator.

We now attempt to explain why the WE fails and make some suggestions as to what could be done to improve it. Figure 4.6 (a) shows $\tilde{U}_t^{1/\alpha}\zeta^{\circ}_{k,t}$ plotted against the record number (t) for 100 trajectories from $t = 1, \ldots, 100$. We refer to $\tilde{U}_t^{1/\alpha}\zeta^{\circ}_{k,t}$ as

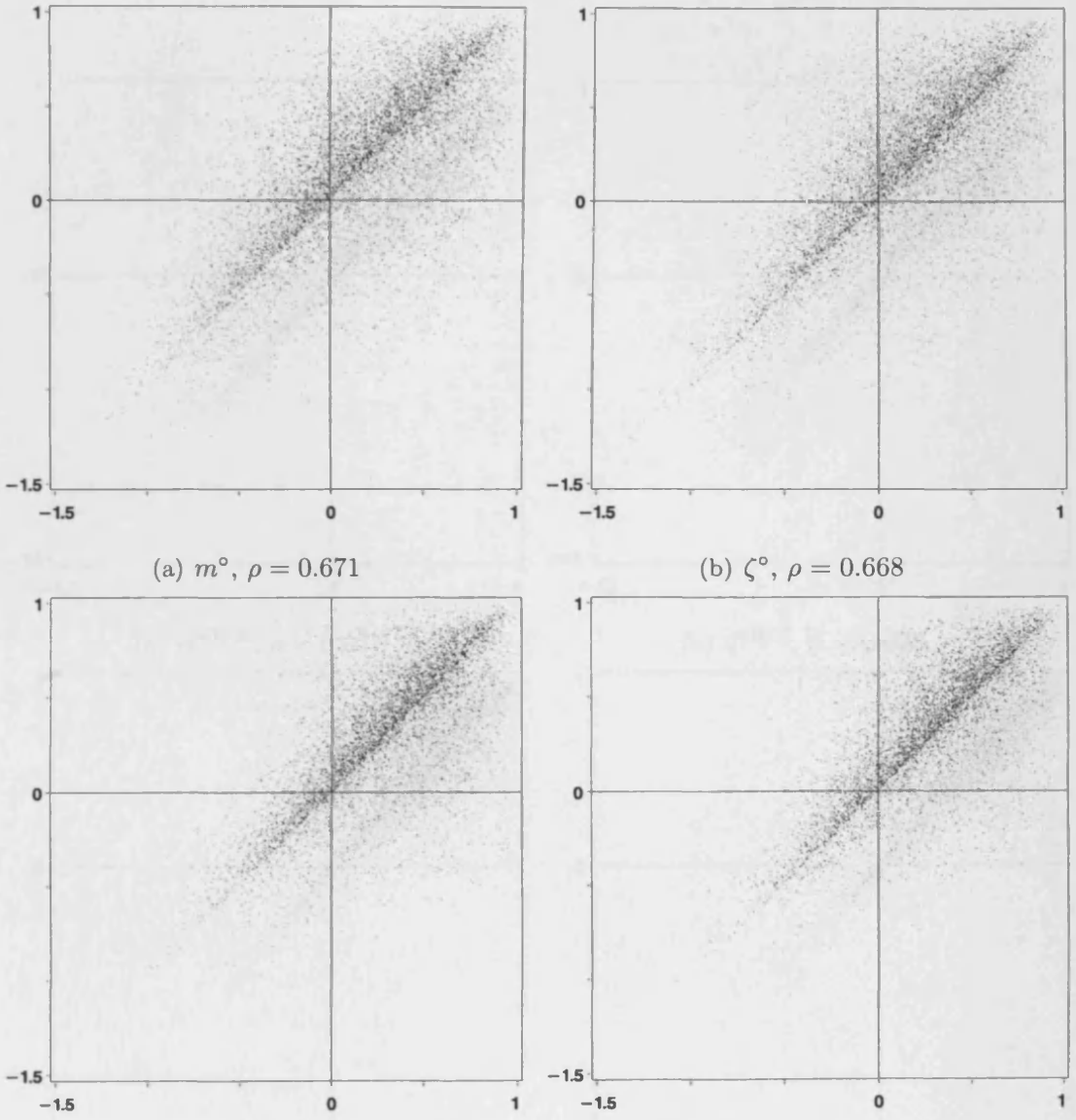
an *un-normalized functional.* We have that $\tilde{U}_t^{1/\alpha}\zeta^\circ_{k,t} \stackrel{d}{=} m^\circ - m$ at $n = \tilde{T}_{k,t}$. (b) and (c) show the un-normalized functionals from the same samples as (a), but this time plotted against $\log(\tilde{T}_{k,t})$ and $\log(\tilde{W}_{k,t})$ respectively. In (a), (b) and (c) $k = 5$ and $\alpha = 3$. These three plots show that the quality of the un-normalized functional (and hence the estimator m°) increases with record number, sample size, and waiting time. As $\zeta^\circ_{k,t}$ (and $\zeta^*_{k,t}$) are stationary in t, the improvement of the un-normalized functional can be explained by considering the expectation of $\tilde{U}_t$: $\mathrm{E}(\tilde{U}_t) = \left(\frac{k}{k+1}\right)^t$. Clearly this decreases to zero as $t \to \infty$, and so the un-normalized functional will also decrease to zero as $t \to \infty$. Record number, sample size and waiting time are highly dependent variables. Indeed, as $\tilde{T}_{k,t} \stackrel{d}{=} T_{k,t}$, from (3.1.4) we have that for large t the expectation of $\tilde{T}_{k,t}$ is approximately $\exp(t/k)$ for large t, and the expectation of $\tilde{W}_{k,t}$ for large t is approximately $\exp(t/k) - \exp((t-1)/k)$. The quality of the functional (and hence the related estimator) improves very quickly as record number (or waiting time or sample size) increases. Our WE fails as it gives too large a weight to estimators at small sample sizes.

Figure 4.6 (d) shows the un-normalized functional $\tilde{U}_t{}^{1/\alpha}\zeta^\circ_{k,t}$ plotted against $\log(\tilde{W}_{k,t})$. Here $t = \tilde{c}$ (where $\tilde{c}$ is as defined above), $k = 5$ and $\alpha = 3$. The data from 10 000 separate runs has been plotted. This plot shows no correlation between wait and estimator quality. By considering (d) with (a), (b) and (c) we can conclude that estimators improve with waiting time only because of the increase in sample size. The initial hypothesis that there would be higher correlation between waiting time and estimator quality than between record number and estimator quality or record time and estimator quality appears to be false.

An improvement on the idea of a WE could perhaps be made by weighting later estimators more heavily (see (4.1.6)), or estimating using an average of the latest two original estimates (see (4.1.7)).

$$\hat{m}_{k,n}^{WE1} = \frac{1}{\sum_{j=1}^{N_{k,n}} \exp(W_{k,j}+1)} \left[\sum_{j=1}^{N_{k,n}} (\exp(W_{k,j}+1)) \hat{m}_{k,T_{k,j}} \right] \tag{4.1.6}$$

$$\hat{m}_{k,T_{k,t}}^{WE2} = \frac{\hat{m}_{k,t-1} + \hat{m}_{k,t}}{2}. \tag{4.1.7}$$



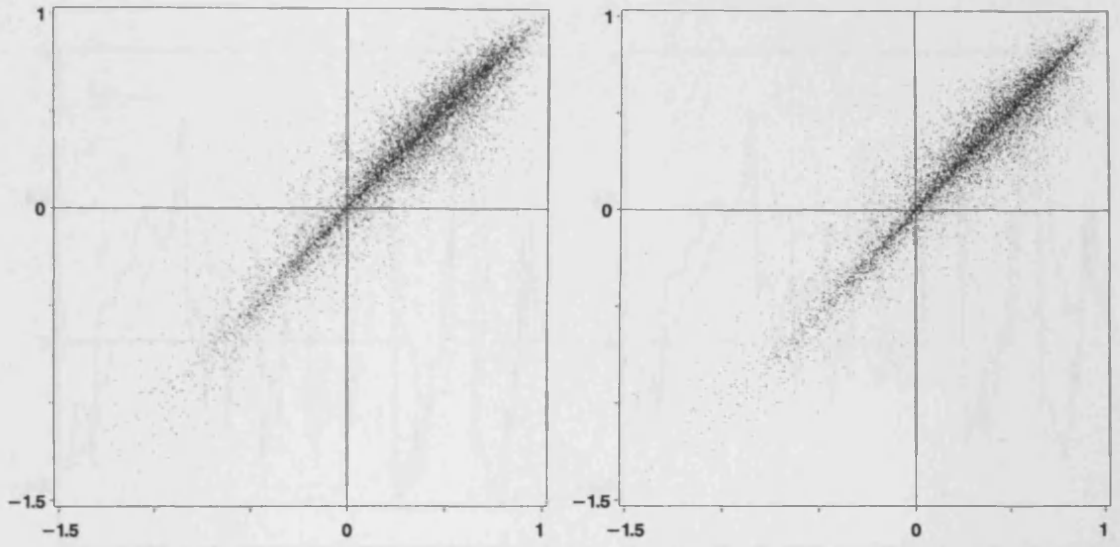
(a) m°, $\rho = 0.671$

(b) ζ°, $\rho = 0.668$

(c) m^*, $\rho = 0.662$

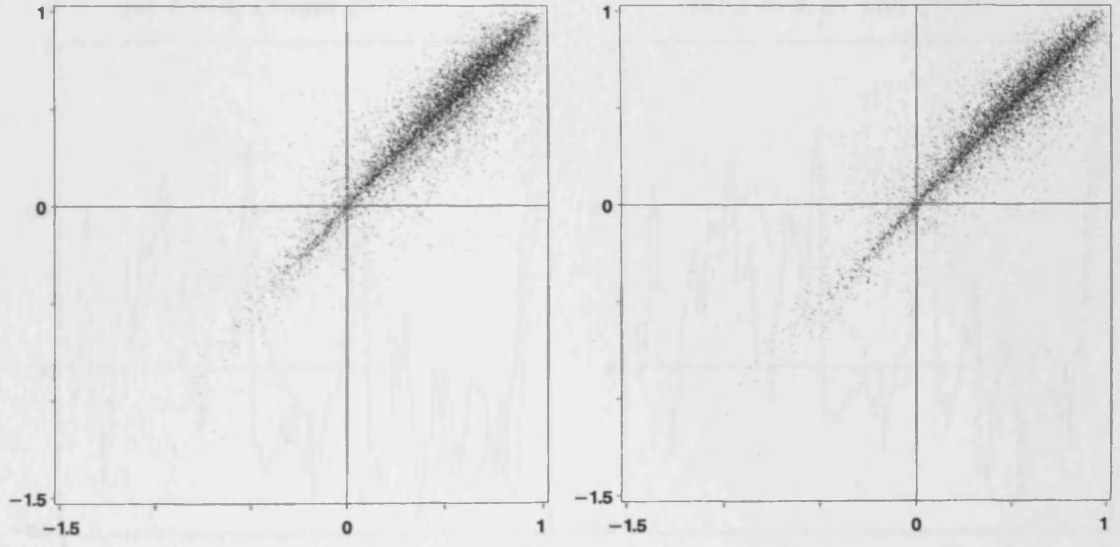
(d) ζ^*, $\rho = 0.659$

Figure 4.2: Scatter graphs of estimators or functionals (as indicated below plots) at record time c (or $\tilde{c}$), plotted against the same estimator or functional at record time $c+1$ (or $\tilde{c}+1$). Here c and $\tilde{c}$ are defined as above. Also shown below each plot is the Pearson's correlation coefficient, ρ.



(a) $m^{\circ WE}$, $\rho = 0.927$

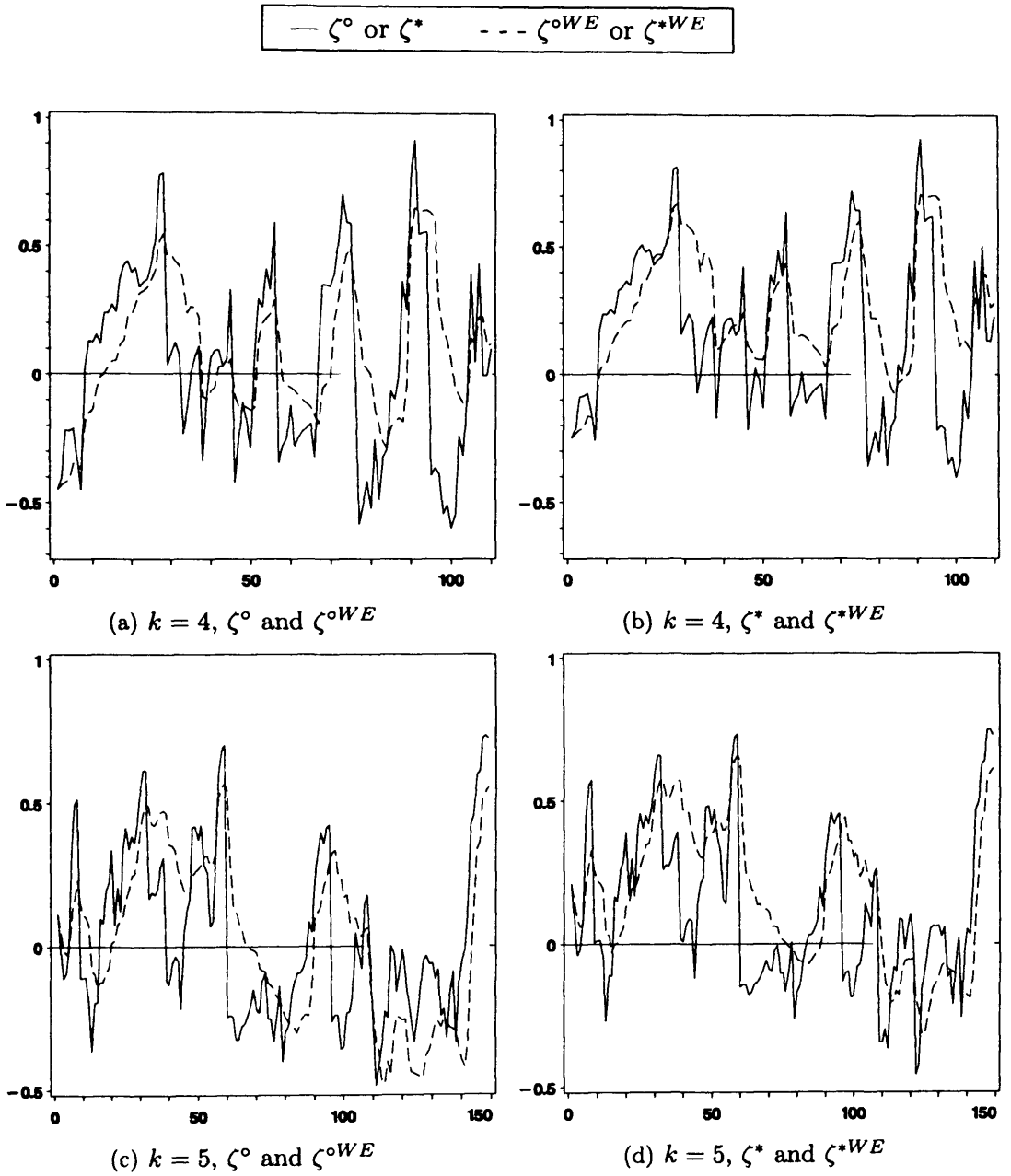
(b) $\zeta^{\circ WE}$, $\rho = 0.923$



(c) m^{*WE}, $\rho = 0.917$

(d) ζ^{*WE}, $\rho = 0.913$

Figure 4.3: Scatter graphs of weighted estimators or weighted functionals (as indicated below plots) at record time c (or $\tilde{c}$), plotted against the same estimator or functional at record time $c+1$ (or $\tilde{c}+1$). Here c and $\tilde{c}$ are defined as above. Also shown below each plot is the Pearson's correlation coefficient, ρ.



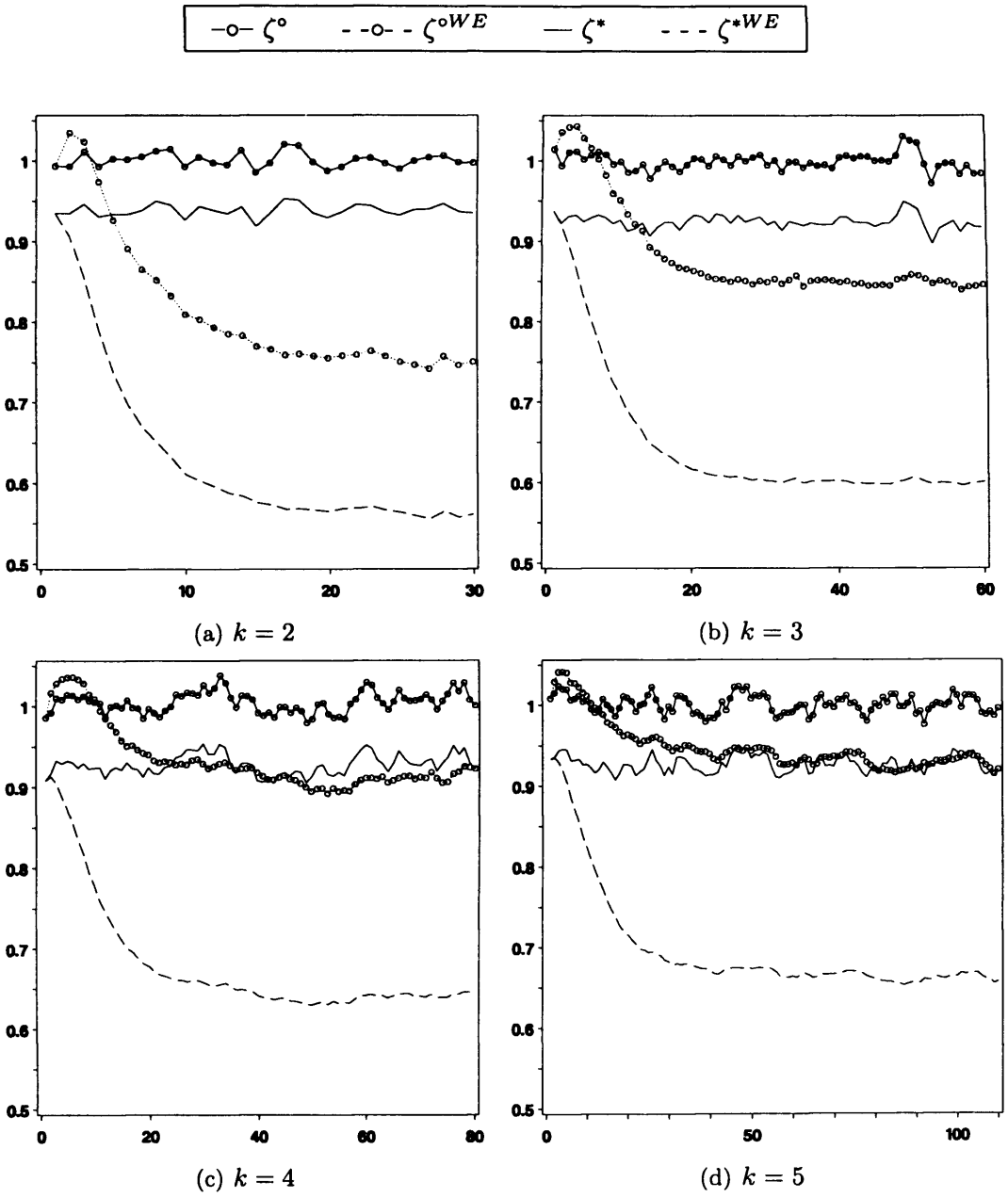
(a) $k = 4$, ζ° and $\zeta^{\circ WE}$

(b) $k = 4$, ζ^* and ζ^{*WE}

(c) $k = 5$, ζ° and $\zeta^{\circ WE}$

(d) $k = 5$, ζ^* and ζ^{*WE}

Figure 4.4: Trajectories of $\zeta^\circ_{k,t}$, $\zeta^*_{k,t}$, $\zeta^{\circ WE}_{k,t}$ and $\zeta^*_{k,t}$. Also marked is $\tilde{U}_t^{1/\alpha}$. In the simulation of these trajectories we used $\alpha = 3$ and $k = 4$ or 5 as marked



(a) $k = 2$ (b) $k = 3$

(c) $k = 4$ (d) $k = 5$

Figure 4.5: Efficiency as defined in (3.4.6) of $\zeta^{\circ}_{k,t}$, $\zeta^{\circ WE}_{k,t}$, $\zeta^{*}_{k,t}$ and $\zeta^{*WE}_{k,t}$ plotted against t. Efficiency is calculated from 10 000 trajectories. The value of k used is marked below each plot. In each plot $\alpha = 3$.

4.2 Estimating m, c_0 known

Consider the problem of estimating the endpoint of a distribution that satisfies (1.2.5), where α and c_0 are known. This section shows that a very simple estimator of m, where c_0 and α are known, is often far more efficient than the optimal linear estimator of m (where only knowledge of α is used). In Section 4.2.1 we define this simple estimator. We then give some analytic results concerning its quality. In Section 4.2.2 we define a functional on the Markov chain X_t that is equal in distribution to our simple estimator (defined in Section 4.2.1) normalized. In Section 4.2.3 we carry out a simulation study to asses its performance on finite samples using the functional defined in Section 4.2.2.

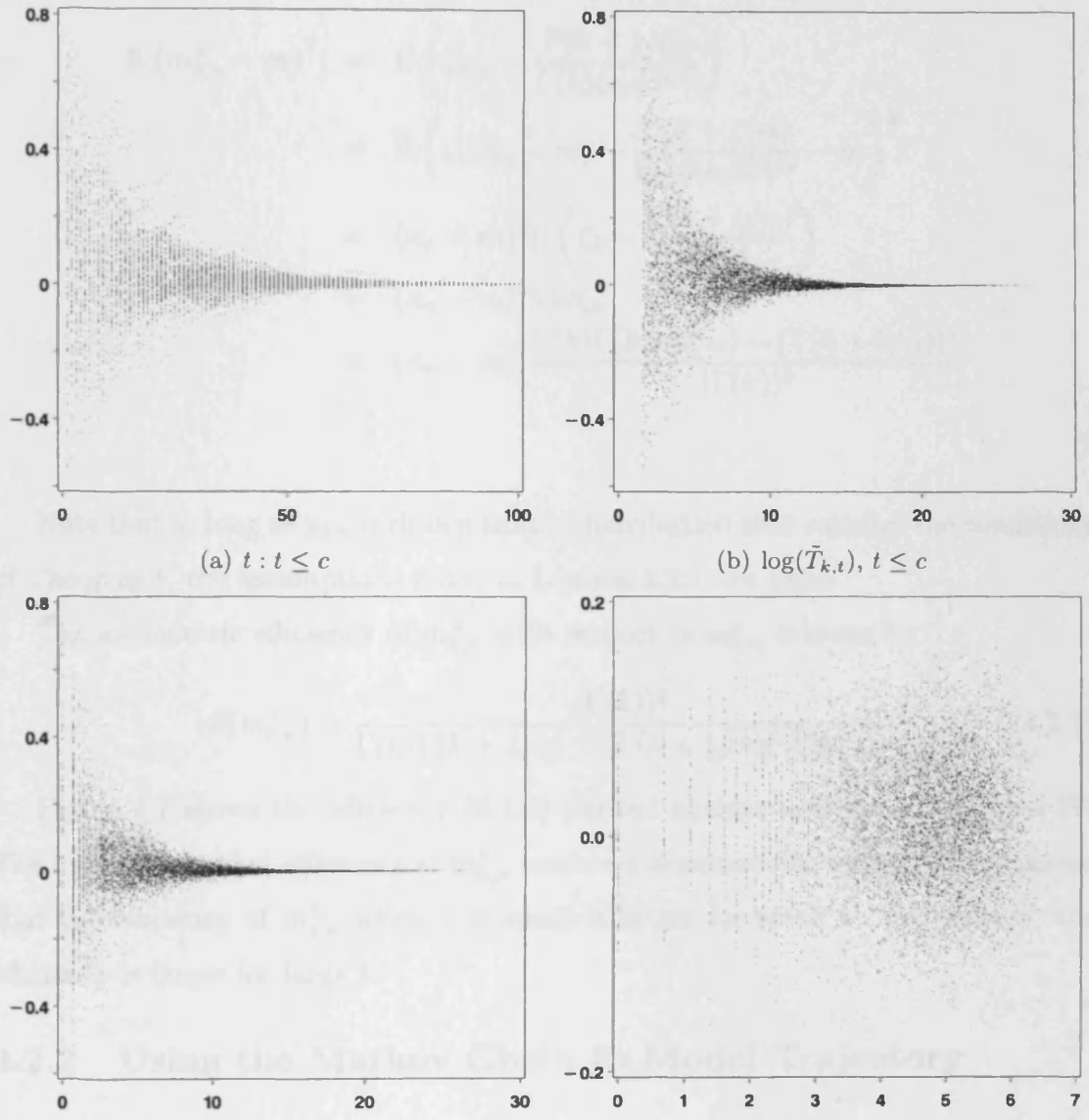
4.2.1 Defining estimator and analysis

Let us define an estimator as,

$$m_{k,n}^{\times} = y_{k,n} - \frac{\Gamma(k+1/\alpha)}{\Gamma(k)(c_0 n)^{1/\alpha}}. \tag{4.2.1}$$

If $F(x)$ can be approximated by (1.2.5), we have that for large n, $\mathrm{E}(y_{k,n}) \sim \frac{\Gamma(k+1/\alpha)}{\Gamma(k)(c_0 n)^{1/\alpha}} + m$ (Proposition 2.2 of [37]). This means that asymptotically the estimator $m_{k,n}^{\times}$ has zero bias. Its variance can also be easily calculated.

Lemma 4.2.1. *Let $\xi_k = \frac{y_{k,n}-m}{\kappa_n - m}$ and assume that asymptotically ξ_k has c.d.f. (1.2.7) and $\kappa_n - m = (c_0 n)^{-1/\alpha}$. Then $m_{k,n}^{\times}$ is a consistent estimator as $k \longrightarrow \infty$ with variance* $\mathrm{Var}(m_{k,n}^{\times}) = (\kappa_n - m)^2 \frac{\Gamma(k)\Gamma(k+2/\alpha) - (\Gamma(k+1/\alpha))^2}{(\Gamma(k))^2}$.



(a) $t : t \leq c$

(b) $\log(\tilde{T}_{k,t})$, $t \leq c$

(c) $\log(\tilde{W}_{k,t})$

(d) $\log(\tilde{W}_{k,t})$, $t = c$

Figure 4.6: (a) shows $\tilde{U}_t^{\,1/\alpha}\zeta^{\circ}_{k,t}$ plotted against the record number (t) for 100 trajectories from $t = 1, \ldots, 100$. (b) and (c) show the same un-normalized functionals, $\tilde{U}_t^{\,1/\alpha}\zeta^{\circ}_{k,t}$, plotted against $\log(\tilde{T}_{k,t})$ and $\log(\tilde{W}_{k,t})$ respectively. (d) shows the un-normalized functional $\tilde{U}_t^{\,1/\alpha}\zeta^{\circ}_{k,\tilde{c}}$ plotted against $\log(\tilde{W}_{k,\tilde{c}})$. Here $\tilde{c}$ is defined as above. The data from 100 separate runs has been plotted. In all of the plots $k = 5$ and $\alpha = 3$.

Proof. It is easy to show that $E\xi_k = \frac{\Gamma(k+1/\alpha)}{\Gamma(k)}$ and $E\xi_k^2 = \frac{\Gamma(k+2/\alpha)}{\Gamma(k)}$. Therefore

$$\begin{aligned}
\mathrm{E}\left(m_{k,n}^{\times} - m\right)^2 &= \mathrm{E}\left(y_{k,n} - \frac{\Gamma(k+1/\alpha)}{\Gamma(k)(c_0 n)^{1/\alpha}}\right) \\
&= \mathrm{E}\left(\zeta_k(\kappa_n - m) - \frac{\Gamma(k+1/\alpha)}{\Gamma(k)(c_0 n)^{1/\alpha}} - m\right)^2 \\
&= (\kappa_n - m)^2 \mathrm{E}\left(\zeta_k - \frac{\Gamma(k+1/\alpha)}{\Gamma(k)}\right)^2 \\
&= (\kappa_n - m)^2 \mathrm{Var}\zeta_k \\
&= (\kappa_n - m)^2 \frac{\Gamma(k)\Gamma(k+2/\alpha) - (\Gamma(k+1/\alpha))^2}{(\Gamma(k))^2}
\end{aligned}$$

□

Note that as long as $y_{k,n}$ is drawn from a distribution that satisfies the conditions of Theorem 1, the assumptions made in Lemma 4.2.1 are valid.

The asymptotic efficiency of $m_{k,n}^{\times}$ with respect to $m_{k,n}^{\circ}$ is given by

$$\mathrm{eff}(m_{k,n}^{\times}) = \frac{(\Gamma(k))^2}{\Gamma(k)\Gamma(k+2/\alpha) - (\Gamma(k+1/\alpha))^2 \mathbf{1}'\Lambda^{-1}\mathbf{1}}. \tag{4.2.2}$$

Figure 4.7 shows the efficiency (6.1.4) plotted against α for $k = 5$, 10 and 20. The figure shows that efficiency of $m_{k,n}^{\times}$ increases dramatically with α. It also shows that the efficiency of $m_{k,n}^{\times}$ when α is small is larger for small k. For large α, the efficiency is larger for large k.

4.2.2 Using the Markov Chain to Model Trajectory

In Section 4.2.3 we carry out a simulation study to asses the performance of $m_{k,n}^{\times}$ against $m_{k,n}^{\circ}$. First we define a functional on the Markov chain X_i (as defined in Section 3.2.2) that is equal in distribution to $(m_{k,n}^{\times} - m)/(y_{k,n} - m)$, where $F(x)$ can be any continuous distribution, denoted $F_g(x)$.

The Markov chain X_i can be used to model order statistics at records from any continuous distribution with inverse $F^{-1}(x)$. Indeed, in Section 4.1 we showed that we can model the k smallest order statistics from a uniform sample at record times:

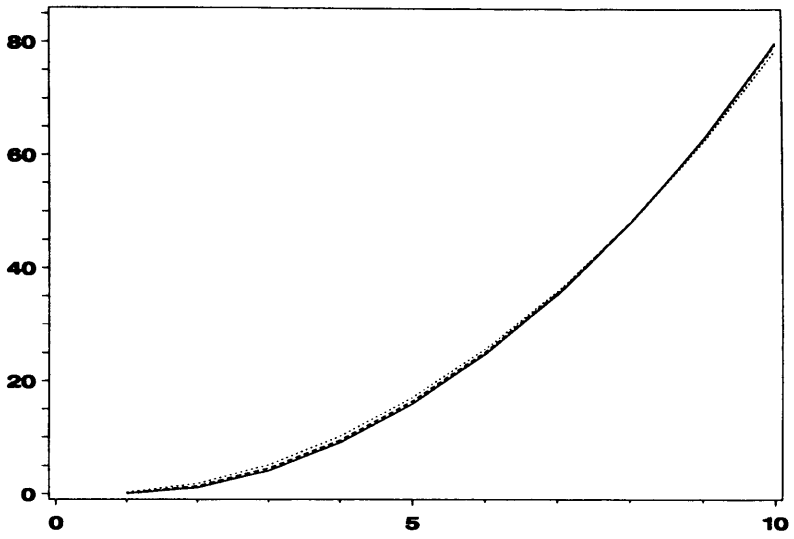


Figure 4.7: Asymptotic efficiency (6.1.4) of $m^{\times}_{k,n}$ with respect to $m^{\circ}_{k,n}$ against α. Here $k = 5$ (dotted line), $k = 10$ (dashed line) and $k = 20$ (solid line). α ranges from 1 to 10.

$U_{i,T_{k,t}} \stackrel{d}{=} \tilde{U}_t x_i^{(t)}$, $i = 1, \ldots, k$ where $x_k^{(t)} \equiv 1$. We also showed that $T_{k,t} \stackrel{d}{=} \tilde{T}_{k,t}$. From Representation 3.2.1 we have that for a continuous c.d.f. $F(x)$ with inverse $F^{-1}(x)$, we have $F^{-1}(U_{k,n}) \stackrel{d}{=} y_{k,n}$. We can therefore easily simulate random variables equal in distribution to order statistics at record times from a general distribution $F(x)$. These random variables are equal to $F^{-1}(\tilde{U}_t x_i^{(t)})$, $i = 1, \ldots, k$. We can model the record times using $\tilde{T}_{k,t}$.

Let us define

$$\zeta^{\times}_{k,t} = 1 - \frac{\Gamma(k + 1/\alpha)}{\Gamma(k)(c_0 \tilde{T}_{k,t})^{1/\alpha}(F_g^{-1}(\tilde{U}_t) - m)},$$

where $F_g^{-1}(x)$ is the inverse of some general continuous distribution $F_g(x)$. We have that $\zeta^{\times}_{k,t} \stackrel{d}{=} (m^{\times}_{k,T_{k,t}} - m)/(y_{k,T_{k,n}} - m)$ where $m^{\times}_{k,n}$ estimates the endpoint of some continuous distribution $F_g(x)$ and $y_{k,n}$ is the k^{th} order statistic from a sample of size n drawn from the distribution $F_g(x)$. Note that when $m = 0$, $\zeta^{\times}_{k,t} F_g^{-1}(\tilde{U}_t) \stackrel{d}{=} m^{\times}$.

4.2.3 Simulation Study

Here we carry out a simulation study (using the functional $\zeta^{\times}_{k,t}$) into how $m^{\times}_{k,n}$ performs when calculating m from finite samples drawn from various distributions. The distributions we consider are, the beta distribution (with c.d.f.s $F_\beta(x;3,0.5)$, $F_\beta(x;3,1)$ and $F_\beta(x;3,3)$), the Weibull distribution (with $\alpha = 3$), the Uniform [0,1] distribution and the Gamma distribution (with $\alpha = 3$). The Gamma distribution is given by $F_\gamma(x;\alpha) = \frac{\gamma(\alpha,x)}{\Gamma(\alpha)}$, $x \geq 0$ where with $\gamma(\alpha,x) = \int_0^x y^{k-1}\exp(-y)dy$.

10 000 trajectories of $\zeta^{\times}_{k,t}$, $t = 1,\ldots,C$ were made, where $F_g(x)$ was equal to each of the six distributions described above and C is as defined in Section 4.1.4. Figure 4.8 (a) and (b) show the trajectories of $F_g^{-1}(\tilde{U}_t)\zeta^{\times}_{k,t} + m$ and $F_g^{-1}(\tilde{U}_t)\zeta^{\circ}_{k,t} + m$ from a single run of the simulations described above where $F_g(x)$ is $F_\beta(x;3,1)$. As $F_g(x)$ is $F_\beta(x;3,1)$, we have $m = 0$. Trajectories are plotted for $k = 4$ and $k = 5$. Plots (c) and (d) show $\zeta^{\times}_{k,t}$ and $\zeta^{\circ}_{k,t}$ from the same trajectories as (a) and (b) respectively. It can be seen from this figure that $\zeta^{\times}_{k,t}$ outperforms $\zeta^{\circ}_{k,t}$ in the trajectories shown.

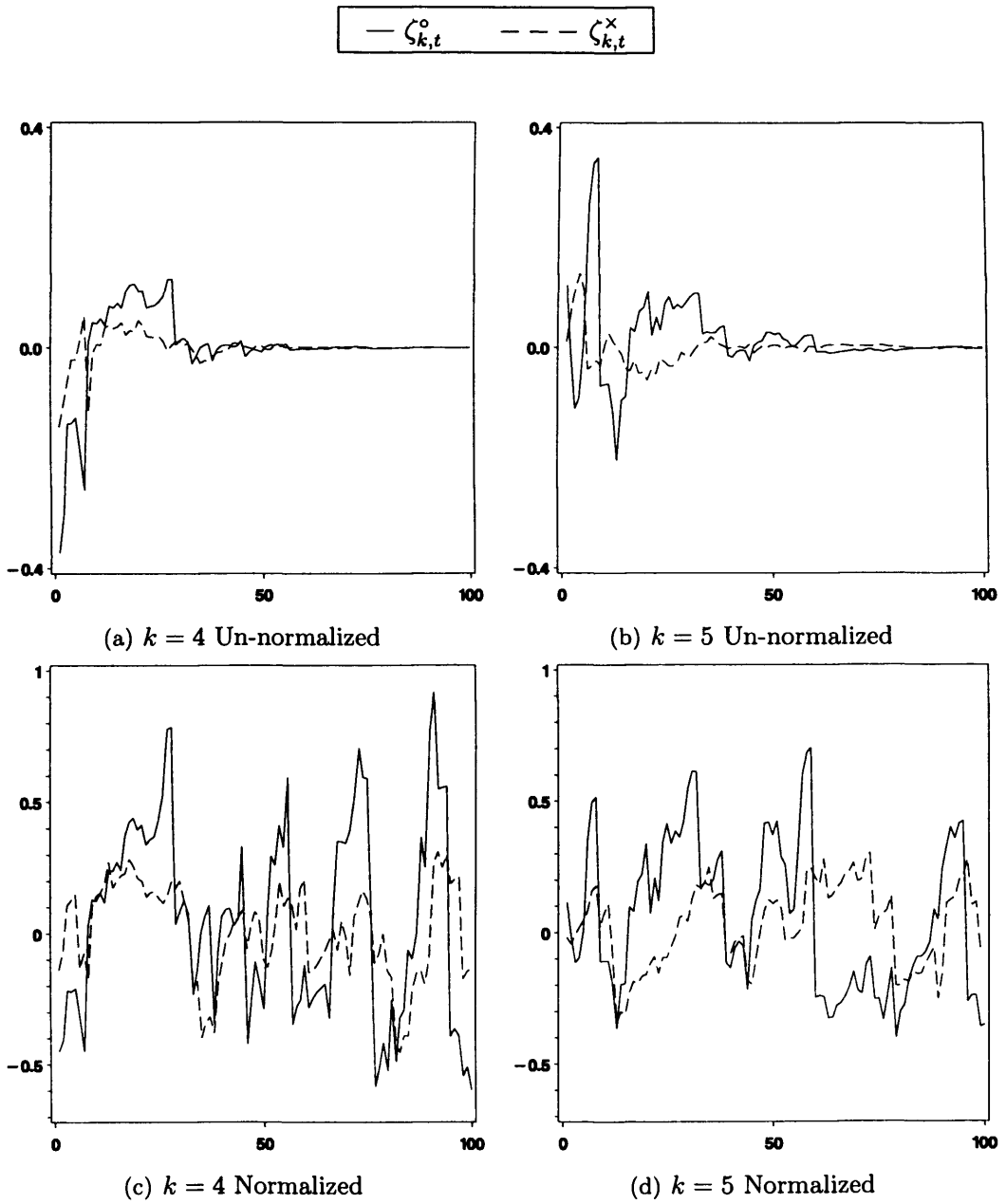
In order to calculate $\zeta^{\times}_{k,n}$, each of these distributions must be approximated by a function of the form $F(x) = c_0(x-m)^\alpha$, so that α and c_0 can be calculated. For the beta distribution $c_0 = \frac{1}{B(\alpha,\beta)\alpha}$ and the tail index is equal to α. This means that $F_\beta(x;3,0.5)$ is approximated by $F(x) = 0.3125x^3$ and $F_\beta(x;3,3)$ is approximated by $F(x) = 10x^3$. When $\beta = 1$ $F_g(x)$ is already in the form of the approximation, indeed, $F_\beta(x;3,1) = x^3$. The beta distributions and their approximations have been plotted in Figure 4.9 (b), (d) and (f). For the Weibull distribution, $c_0 = 1$ and the tail index is equal to α. This means that the Weibull distribution when $\alpha = 3$ is approximated by $F(x) = x^3$. For the uniform distribution, $c_0 = 1$ and the tail index is equal to 1. The uniform distribution exactly equal to its approximation, $F(x) = x$. For the gamma distribution, $c_0 = \frac{1}{\alpha\Gamma(\alpha)}$ and the tail index is equal to α. This means that when $\alpha = 3$ we approximate the gamma distribution by $F(x) = 0.6x^\alpha$. The Weibull distribution, uniform distribution, gamma distribution

and their approximations have been plotted in Figure 4.10 (b), (d) and (f). All of the above approximations are valid for $m \leq x \leq c_0^{-1/\alpha} + m$, and in each case $m = 0$.

(a), (c) and (e) of Figures 4.9 and 4.10 show the estimated efficiency as defined in (3.4.6) of $\zeta_{k,t}^{\times}$ with respect to ζ_k° (asymptotic) and the estimated efficiency of $\zeta_{k,t}^{\circ}$ from simulation with respect to ζ_k° asymptotically. The analytic efficiency (6.1.4) is also plotted (grey line). These efficiencies are plotted against k. (b), (d) and (f) of Figures 4.9 and 4.10 show the distribution $F_g(x)$ (solid line) and its approximation $F(x) = c_0(x - m)^{\alpha}$ (dashed line).

In Figure 4.9 (a) and (b), $F_g(x) = F_\beta(x; 3, 0.5)$. The approximation $F(x)$ is smaller than the distribution $F_\beta(x; 3, 0.5)$ for all $0 < x < 1$. For small k the estimated efficiency of $\zeta_{k,t}^{\times}$ is greater than the analytic efficiency, for large k it is smaller then the analytic efficiency. In Figure 4.9 (c) and (d), $F_g(x) = F_\beta(x; 3, 1)$, this means that the approximation is exactly equal to $F_g(x)$. The estimated efficiency is therefore very close to the analytic efficiency. In Figure 4.9 (e) and (f), $F_g(x) = F_\beta(x; 3, 3)$. (f) shows that the approximation is larger than $F_g(x)$ for $0 < x \leq 10^{-1/3}$. The estimated efficiency is less than the analytic efficiency for all k plotted. As k increases both the estimated efficiency and the analytic efficiency decrease, the estimated efficiency decreases faster. In Figure 4.10 (a) and (b) $F_g(x)$ is Weibull with $\alpha = 3$. The approximation $F(x)$ is larger than the Weibull distribution for all $0 < x < 1$. The Weibull distribution is close to the approximation for $0 \leq x \leq 0.5$. The estimated efficiency of $\zeta_{k,t}^{\times}$ in this case is only slightly lower than the analytic efficiency. In Figure 4.10 (c) and (d) $F_g(x)$ is uniform. This means that the approximation is exactly equal to $F(x)$ for all x. The estimated efficiency of $\zeta_{k,t}^{\times}$ is less than one (and so less than the efficiency of $\zeta_{k,t}^{\circ}$). It decreases with k. The estimated efficiency is very close to the analytic efficiency (6.1.4). In Figure 4.10 (e) and (f), $F_g(x) = F_\gamma(x)$. (f) shows that the approximation is greater than $F_\gamma(x)$ for all $0 < x \leq 6^{1/3}$. The estimated efficiency is lower than the analytic efficiency and decreases more quickly than the analytic efficiency as k increases. From both Figures 4.9 and 4.10 we can

say that generally if the approximation $F(x)$ is close to the distribution $F_g(x)$, the estimated efficiency is close to the analytic efficiency. From these figures we can also see that the estimated efficiency of $\zeta^{\circ}_{k,t}$ is close to its analytic efficiency, 1.



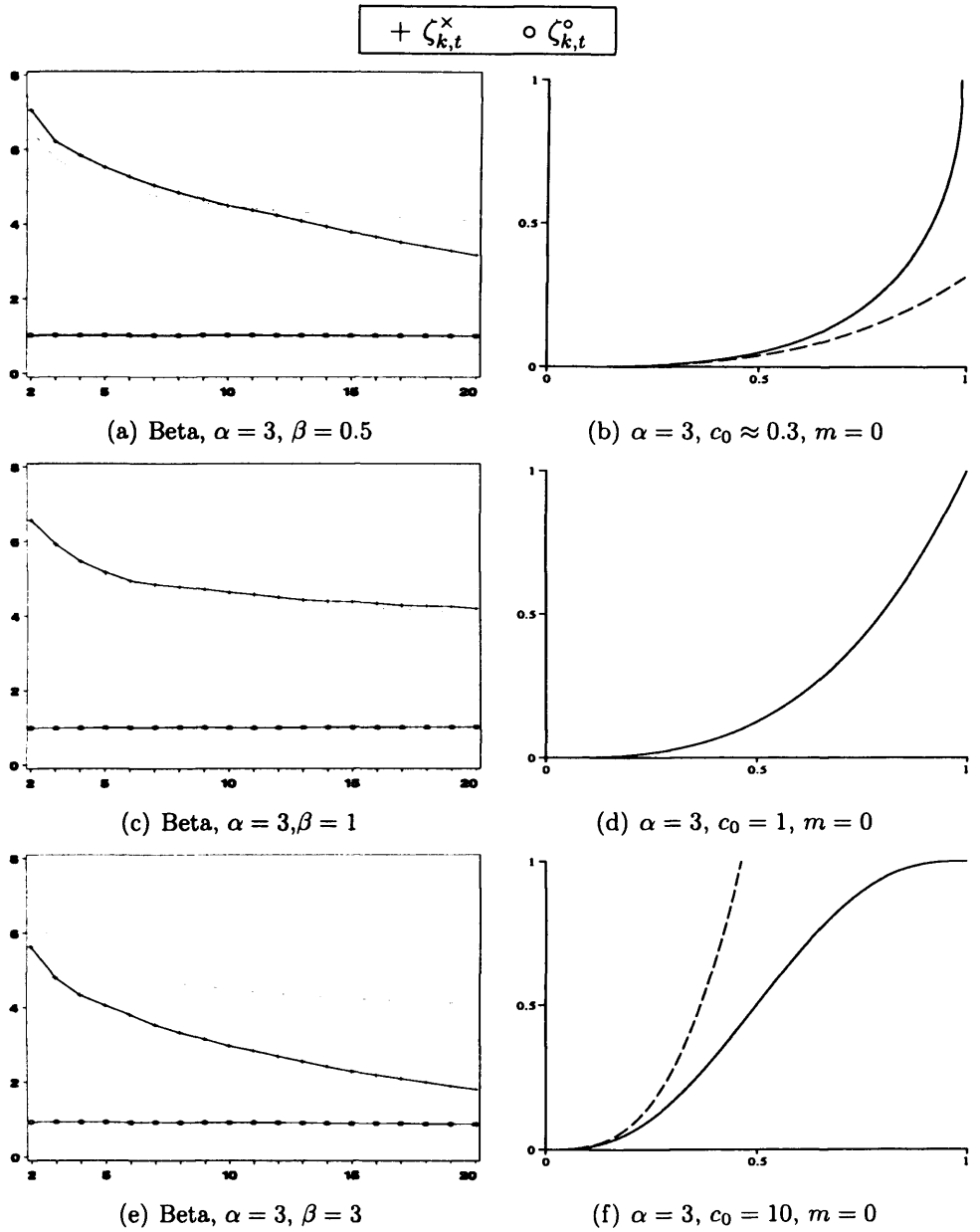
(a) $k = 4$ Un-normalized

(b) $k = 5$ Un-normalized

(c) $k = 4$ Normalized

(d) $k = 5$ Normalized

Figure 4.8: (a) and (b) show trajectories of $\zeta^{\circ}_{k,t} F_g^{-1}(\tilde{U}_t) + m$ and $\zeta^{\times}_{k,t} F_g^{-1}(\tilde{U}_t) + m$ where $k = 4$ and 5 respectively. (c) and (d) show trajectories of $\zeta^{\circ}_{k,t}$ and $\zeta^{\times}_{k,t}$ where $k = 4$ and 5 respectively. In all four plots we have $F_g(x) = F_\beta(x; 3, 1)$.



(a) Beta, $\alpha = 3$, $\beta = 0.5$

(b) $\alpha = 3$, $c_0 \approx 0.3$, $m = 0$

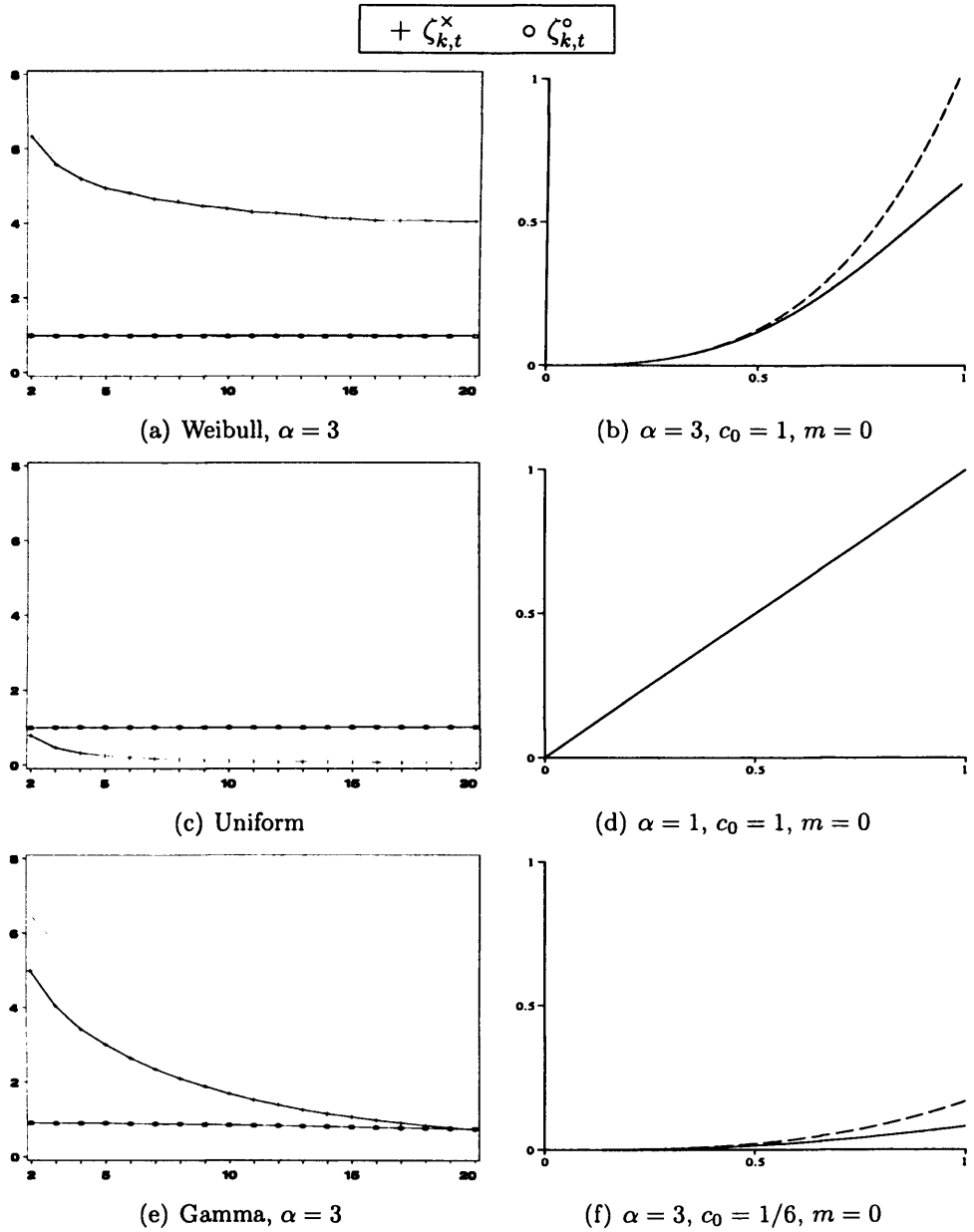
(c) Beta, $\alpha = 3$,$\beta = 1$

(d) $\alpha = 3$, $c_0 = 1$, $m = 0$

(e) Beta, $\alpha = 3$, $\beta = 3$

(f) $\alpha = 3$, $c_0 = 10$, $m = 0$

Figure 4.9: Graphs (a), (c), and (e) show the efficiency with respect to ζ° (asymptotically) of $\zeta^{\times}_{k,t}$ and $\zeta^{\circ}_{k,t}$ (from simulation) plotted against k. Also shown (grey line) is the asymptotic efficiency (6.1.4) of $m^{\times}_{k,n}$ with respect to $m^{\circ}_{k,n}$. In all plots $F_g(x)$ is beta. In (a) $\alpha = 3$ and $\beta = 0.5$, in (c) $\alpha = 3$ and $\beta = 1$, and in (e) $\alpha = 3$ and $\beta = 3$. These distributions (solid line) and their approximations (dashed line) are shown in (b), (d) and (f).



(a) Weibull, $\alpha = 3$

(b) $\alpha = 3$, $c_0 = 1$, $m = 0$

(c) Uniform

(d) $\alpha = 1$, $c_0 = 1$, $m = 0$

(e) Gamma, $\alpha = 3$

(f) $\alpha = 3$, $c_0 = 1/6$, $m = 0$

Figure 4.10: Graphs (a), (c), and (e) show the efficiency with respect to ζ° (asymptotically) of $\zeta^{\times}_{k,t}$ and $\zeta^{\circ}_{k,t}$ (from simulation) plotted against k. Also shown (grey line) is the asymptotic efficiency (6.1.4) of $m^{\times}_{k,n}$ with respect to $m^{\circ}_{k,n}$. In (a) $F_g(x)$ is Weibull, in (c) $F_g(x)$ is uniform, and in (e) $F_g(x)$ is gamma. These distributions (solid line) and their approximations (dashed line) are shown in (b), (d) and (f).

Chapter 5

Meteorological Applications

In this chapter we consider applications of record theory to real world (meteorological) situations.

First, in Section 5.1 we discuss the reporting of records in the media and show that the occurrences of records on their own cannot be used to draw conclusions about whether, or by how much, the climate is changing.

In Section 5.2 we consider mean monthly sea level data from six locations in the Netherlands - Delfzijl, Harlingen, Den Helder, Ijmuiden, Maassluis and Hellevoetsluis. We calculate the expected number of records and estimate probability mass functions for the number of records that we would expect to see in such a time series if it were stationary.

In Section 5.3 we introduce the Singular Spectrum Analysis algorithm and the program CaterpillarSSA. The aim of this section and Section 5.4 is to use SSA to split the Harlingen time series into two components; one, a forecastable time series consisting of trend and regular oscillation (the reconstruction); the other, a stationary time series of the remaining noise (the residual). We then can use the standard techniques already discussed in this thesis to make estimates of the upper and lower endpoints of this residual. We use CaterpillarSSA to forecast the reconstruction. Finally we add the forecast to the estimates of m and M to obtain forecasts of the endpoints of the distribution of monthly sea level. In Section 5.3

we investigate the effects on the number of records observed of different methods (within the CaterpillarSSA framework) of removing trend and noise. In Section 5.4 we select some separation methods from Section 5.3 and forecast the upper and lower endpoints of the mean monthly sea level for each month of three different years.

5.1 Reported Records: How Likely Are They?

5.1.1 Expected Number of Reported Records

The media reports record temperatures and, due to worries about climate change, imply that the occurrence of records means that the climate is non-stationary. Below we discuss the probability of observing at least one record in one of the final elements of a collection of stationary time series.

The probability of a record occurring after a particular number of (i.i.d.) observations is straightforward to calculate. Consider a time series of length n. As defined in Section 3.1.1, a new random variable is a 1^{st} record if it is the smallest member of the time series so far. If the time series is made of continuous independent identically distributed random variables, the probability of a new observation being the smallest so far is the same as the probability of it being the second, third, fourth, or n^{th} smallest. Therefore the probability of the n^{th} member of a time series being a record is $\frac{1}{n}$. Notice that the first observation therefore is always a record.

Consider a time series 100 elements long. The probability of the final element of this time series being a record value is $\frac{1}{100}$. An example of a reported meteorological record with this probability is: if there exist 100 years of total monthly precipitation data, there is a 1/100 chance that, say, this August was the wettest on record. This of course assumes that the time series of total precipitation in August is made up of continuous independent identically distributed random variables. As 1/100 is small, 'Wettest August on Record' could be considered a news-worthy event. However, if we consider all of the different ways a record could occur, it is no longer surprising

that every year we get at least one record event reported.

If we consider 365 time series (one for each day of the year), where each time series is 100 elements (years) long, stationary, and independent of the other time series, then the probability that any of the 365 times series has a record as its final entry is

$$1 - \text{P(No records in the final element of any of the time series)} = 1 - \left(\frac{99}{100}\right)^{365} = 0.974.$$

The above probability could be interpreted as the probability that in any given year, at least one day is the, say, hottest for 100 years. However, an unlikely assumption has been made: that the 365 time series are independent of each other. This would require that, for example, the temperature on 04/10/2000 is independent of the temperature on 05/10/2000. It has also been assumed that the distribution of the temperature for each day of the year is stationary and that each day is independent of the same day in all of the other 99 years considered (i.e. 04/10/1901 is independent of 04/10/1902, 04/10/1903, ..., 04/10/2000), this is a reasonable assumption. Although in order to apply the above probability to meteorological records an unlikely assumption must be made, the example illustrates that the likelihood of a record occurring greatly depends on the number of time series being considered.

Events that are more likely to come from independent time series and be more interesting to the public are; record monthly averages, record monthly totals and records for a particular city. When you consider this includes particularly cold, wet, dry or windy months, or years, or Wimbledon tournaments..., the chance of a 'news-worthy' record occurring in any particular year is again close to one. For example, the probability that in a year, at least one of the months, in at least one of five locations, is found to be the most extreme (hottest, coldest, wettest, driest, windiest) for 100 years is $1 - (\frac{99}{100})^{5\times12\times5} = 0.95$. In order to calculate this probability we are considering $5 \times 12 \times 5 = 300$ time series. Again we must make the

assumptions that the time series are made up of i.i.d. elements and that each time series is independent of all of the others. This second assumption of independence between time series relies on independence between climatic elements (e.g. between high and low temperature - clearly false), between locations (likely if locations are not too close together) and between months (likely). This dependence need not be a problem, we can make assumptions as to the level of dependence between some climatic events, and assume independence between others. The events of a record high and a record low (temperature or precipitation) are mutually exclusive. We can say that the probability of a record (either high or low) mean monthly temperature occurring in at least one of 12 time series of temperature that are 100 entries (years) long is

$$1 - \text{P(No record high or low temperature occurs in any of the 12 time series)}$$
$$= 1 - \left(\frac{98}{100}\right)^{12} = 0.22.$$

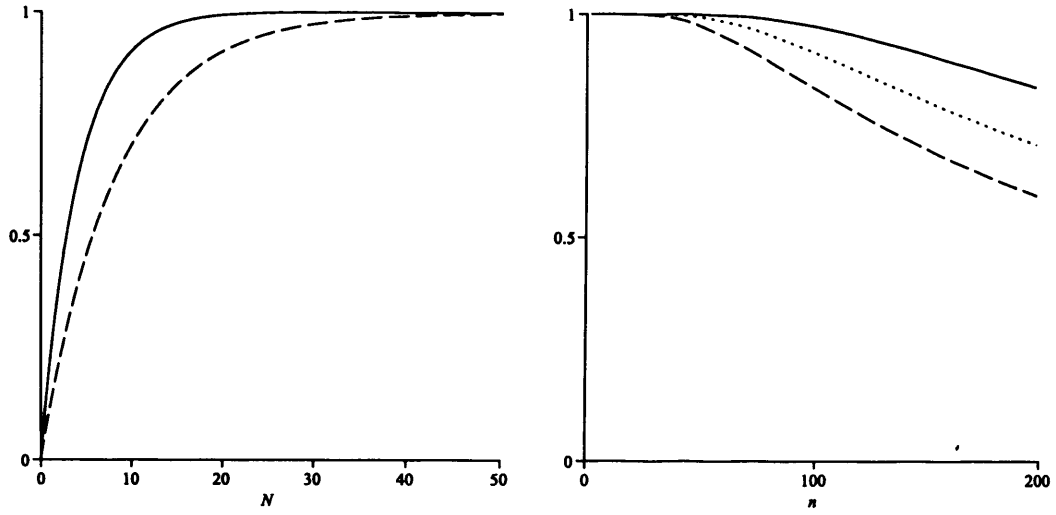
Although this is a fairly small probability, if more independent events are considered, the probability increases dramatically (see Figure 5.1).

Although most meteorological data sets go back much further, reports are often made when an event occurs that is the most extreme for some 'round' number of years, say 10 or 20 years.

We can now estimate the following probability:

$$\begin{aligned} \text{P} \ (\ & \text{At least one of the following occurs in the current year:} \\ & \text{Highest or lowest recorded monthly temperature for } n \text{ years;} \\ & \text{Highest or lowest total monthly precipitation for } n \text{ years;} \\ & \text{Strongest winds recorded for } n \text{ years)} \\ & = 1 - \left(\frac{(n-2)^{12\times5\times2}(n-1)^5}{n^{12\times5\times2+5}}\right) \end{aligned}$$

This probability is plotted as a function of n in Figure 5.1. It is likely to be an underestimate because of the many events that could not be included in the



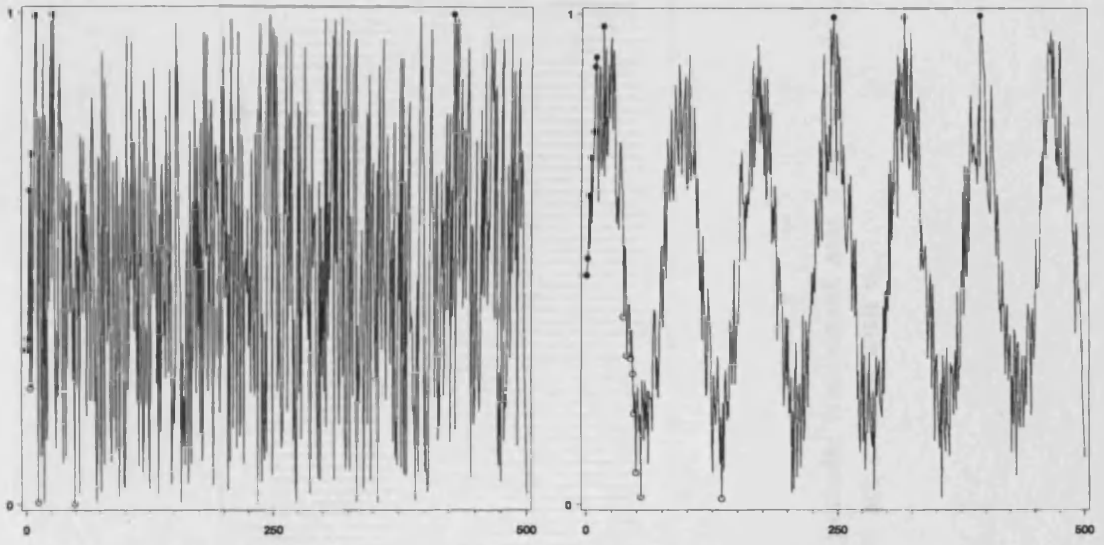
(a) Number of time series, N

(b) Length of time series, n

Figure 5.1: (a) shows a plot of $1-\left(\frac{98}{100}\right)^{12\times N}$ (solid line) and $1-\left(\frac{99}{100}\right)^{12\times N}$ (dashed line), $N=1,\ldots,50$. This demonstrates how increasing the number of independent time series increases the probability of observing a record event in the current year. (b) shows a plot of $1-\left(\frac{n-2}{n}\right)^{12\times 15}$ (solid line), $1-\left(\frac{n-1}{n}\right)^{12\times 15}$ (dashed line), and $1-\left(\frac{(n-2)^{12\times 5\times 2}(n-1)^5}{n^{12\times 5\times 2+5}}\right)$ (dotted line) $n=2,\ldots,200$. This demonstrates how the length of time series affects the probability of observing a record event in the current year. The dotted line represents an estimate of the actual probability that the current year will have at least one record (for time series of length n).

estimate due to dependence. These include; sea level, hours of sunlight, mean monthly temperature/precipitation, statistics for particular special day or period (e.g. coldest/warmest Christmas, most days of a sporting event rained off), highest annual/seasonal) mean temperature/precipitation or most amount of rainfall ever recorded to fall in one hour.

always increase the number of observed records. For example a time series described by the equation $p'_n = (r_n + 10\sin(n\pi/2) + 10)/11$ is plotted in Figure 5.3. It has produced only 4 maximal record values when 6.79 were expected. The number of records observed depends on the amplitude of the periodicity compared to the noise, the period of the cycle, and the phase of the periodicity at $n = 1$. A periodicity with a large amplitude will tend to create records in the time series only at extreme values of the periodicity, this can reduce the total number of records. The plot of t_n demonstrates that an increasing trend increases the number of maximal record values (17 observed) and decreases the number of minimal record values (3 observed). A decreasing trend would have the opposite effect. The plot showing a time series of noise with increasing variance demonstrates that this causes an increase in the number of maximal and minimal records (observed are 34 and 36 respectively).



(a) Stationary

(b) Periodicity

(c) Trend

(d) Increasing variance

Figure 5.2: Plots show four time series with record high values marked with • and record low values marked with ○. (a) shows s_n, a stationary time series of pure noise, (b) shows p_n, a time series with a periodicity and noise, (c) shows t_n, a time series with an increasing trend and noise, (d) shows v_n, a time series with noise whose variance is increasing with time.

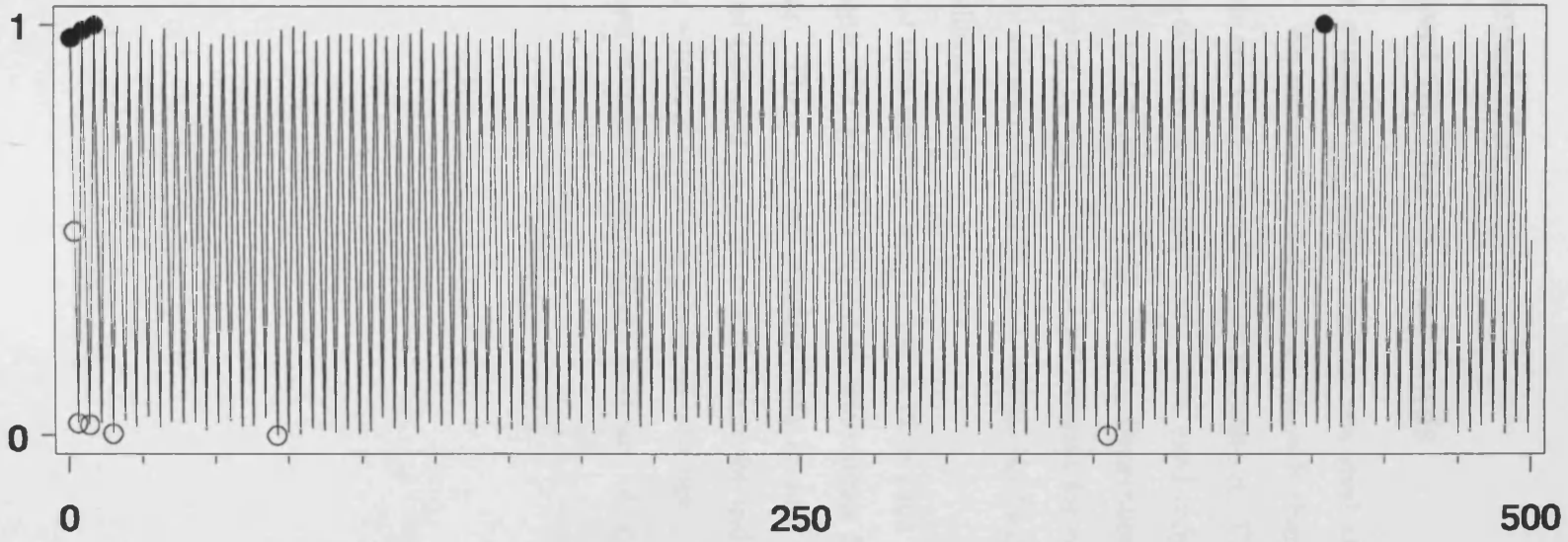


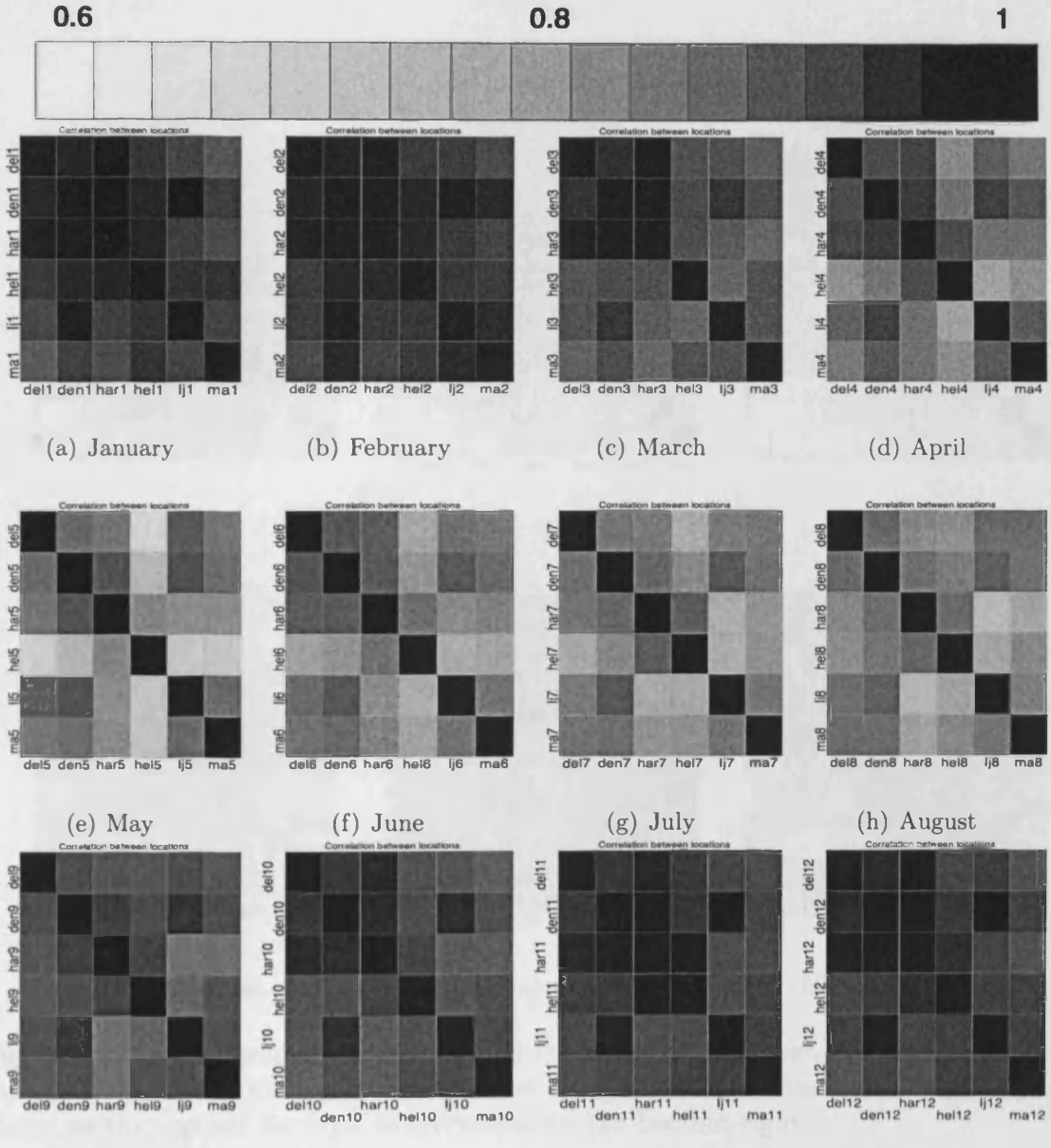
Figure 5.3: Plot of time series p'_n. This is a time series with a large periodic component and a small random noise component. Record high values are marked with • and record low values are marked with ∘.

5.2 Sea level data

5.2.1 Sea level in The Netherlands

In this section we consider the mean monthly sea level at six measuring stations in the Netherlands. We assume that for every month there is some maximum (and minimum) possible sea level and we try to estimate it. The data was obtained from http://www.gloss-sealevel.org/data/. We have used only data from the 122 years from January 1885 to December 2006. Before December 1885 measurements were rounded to the nearest 5, and so are not convenient for considering the occurrences of records. The locations considered are Delfzijl, Harlingen, Den Helder, Ijmuiden, Maassluis and Hellevoetsluis.

The six original time series of length $122 \times 12 = 1464$ were split into $6 \times 12 = 72$ time series of length 122, giving a separate time series for each location and each month of the year. There are no missing data in any of these time series. The different months of the year can be considered to be independent (see Figure 5.5), however, there is evidence that the locations are not. The Pearson's correlation coefficients between locations (see Figure 5.4) are all greater than 0.7 (most are greater than 0.9).

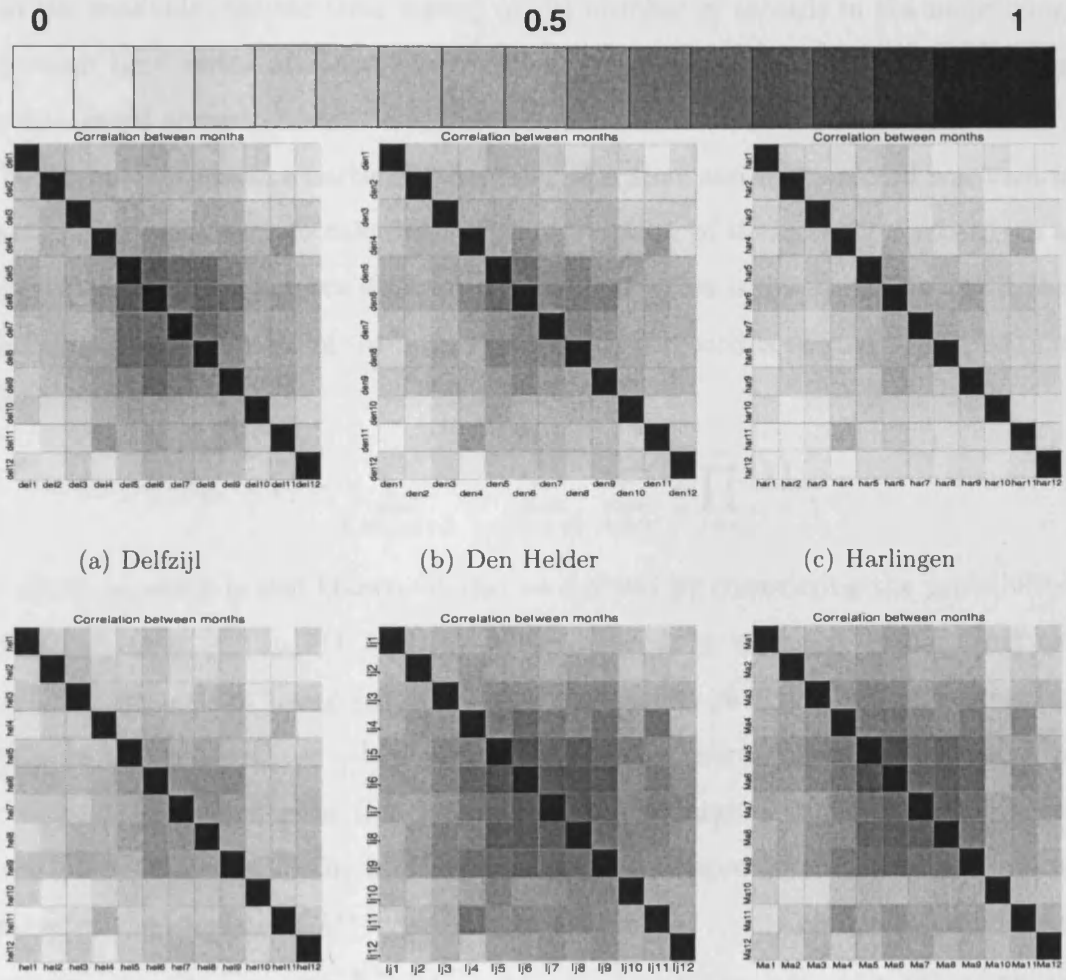


(a) January (b) February (c) March (d) April

(e) May (f) June (g) July (h) August

(i) September (j) October (k) November (l) December

Figure 5.4: Correlograms show Pearson's correlation coefficient. Each plot shows the coefficient for a different month and all six location. The order of the variables in the plot are (from top left to bottom right) Delfzijl, Den Helder, Harlingen, Hellevoetsluis, Ijmuiden and Maassluis.



(a) Delfzijl (b) Den Helder (c) Harlingen

(d) Hellevoetsluis (e) Ijmuiden (f) Maassluis

Figure 5.5: Correlograms show Pearson's correlation coefficient. Each plot shows the coefficient for a different location and all 12 months of the year. January is shown as the top left through to December in the bottom right.

5.2.2 Occurrences of Records

In this section we make Monte Carlo estimates of the expected number of records for a stationary continuous time series of 122 elements. We compare this to estimates (from the available discrete time series) of the number of records in the underlying continuous time series of mean monthly sea level. This is done for each of the locations listed above.

The probability that a particular member of a time series is a record was shown in Section 5.1 to be easy to calculate. The probability of there being x records in a time series of length n is more difficult to calculate. This is due to the fact that the probability of any member of the time series being a record is related to its position in the time series.

$$\mathrm{P}(N_{i,n} = x) = \sum_{i_x=i_{x-1}+1}^{n} \cdots \sum_{i_3=i_2+1}^{n-x+3} \sum_{i_2=2}^{n-x+2} \frac{1}{n} \prod_{j=2}^{x} \frac{1}{i_j - 1}$$

The above equation is well known. It can be derived by considering the probability of observations $y_{i_1}, \ldots, y_{i_x}$ $(1 = i_1 < i_2 < \ldots < i_x \leq n)$ being records and all other observations not being records. This probability is given by the expression $\frac{1}{1}\frac{i_1}{i_1+1}\frac{i_1+1}{i_1+2}\cdots\frac{i_2-2}{i_2-1}\frac{1}{i_2}\frac{i_2}{i_2+1}\cdots = \frac{1}{i_2-1}\cdots\frac{1}{i_x-1}\frac{1}{n}$. Summing over all possible indices, $1 = i_1 < i_2 < \ldots < i_x \leq n$, gives the desired result. For large n and x this expression becomes very computationally intensive. Below is an approximate probability mass function of the number of k^{th} type 1 records (k=1, ..., 10) and an approximate probability mass function for $\sum_{i=1}^{k} N^*_{i,122}$. These were obtained through Monte Carlo simulations: 10 000 time series of length 122 were simulated. For each time series (indexed i) the number of k^{th} $(k = 1, \ldots, 10)$ records were recorded. For simplicity in this section the number of k^{th} type 1 records at $n = 122$ for time series i has been denoted N^i_k. $\mathrm{P}(N_{k,122} = x)$ can then be estimated by $\sum_{i=1}^{10\,000} \frac{I_x(N^i_k)}{10000}$. Here $I_x(a)$ is an indicator function such that $I_x(a) = 1$ if $a = x$ and $I(a)_x = 0$ otherwise. The

expected number of k^{th} type 1 records in a time series 122 elements long is given by:

$$E(N^*_{k,n}) = \sum_{i=k}^{122} \frac{1}{i}. \tag{5.2.1}$$

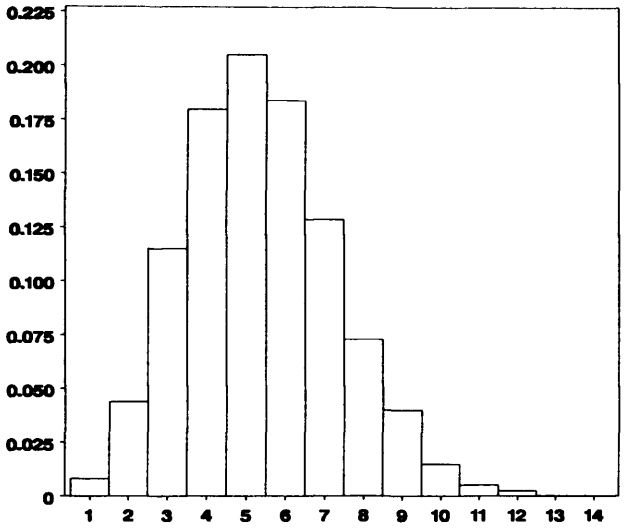


Figure 5.6: Estimated probability mass function of the number of 1^{st} records occurring in a time series of length 122.

The sea level data obtained for this study are discrete. They come from rounding (to the nearest integer) the true continuous sea level. It is the occurrence of records in this true, continuous data that we would like to estimate. We do this by adding a small amount of noise to each of the variables in the discrete time series. Denote this new time series by $y'_i = y_i + \varsigma$, where y_i, $i = 1, 2, \ldots, n$ is the original, discrete time series and ς is an independent uniform [0,0.1] random variable. The expected number of records in the time series y'_i is the same as the expected number of records in the underlying continuous time series. The new time series will not have any weak records. Denote the number of minimal type 1 k^{th} records in the time series y'_i, $i = 1, \ldots, n$ as $N'^*_{k,n}$ and the number of maximal type 1 k^{th} records in this time series as $N'^{+*}_{k,n}$.

By considering Table 5.1 we can see that the expected number of type 1 1^{st} records in a stationary time series of length 122 is 5.4. Let us classify observing

	$P(N^*_{k,122} \leq x)$									
$x \setminus k$	1	2	3	4	5	6	7	8	9	10
0	0	0.01	0.02	0.02	0.03	0.04	0.05	0.06	0.06	0.07
1	0.27	0.05	0.09	0.12	0.15	0.18	0.20	0.22	0.25	0.01
2	0.53	0.16	0.24	0.30	0.35	0.39	0.43	0.47	0.50	0.05
3	0.75	0.34	0.45	0.52	0.58	0.63	0.66	0.69	0.72	0.16
4	0.89	0.55	0.65	0.72	0.77	0.80	0.83	0.85	0.87	0.35
5	0.96	0.73	0.81	0.86	0.89	0.91	0.93	0.94	0.95	0.55
6	0.99	0.86	0.91	0.94	0.96	0.97	0.97	0.98	0.98	0.73
7	1.00	0.94	0.96	0.98	0.98	0.99	0.99	0.99	1.00	0.86
8	1.00	0.98	0.99	0.99	0.99	1.00	1.00	1.00	1.00	0.94
9	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98
10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
11	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$EN^*_{k,122}$	5.4	4.4	3.9	3.6	3.3	3.1	2.9	2.8	2.7	2.6

Table 5.1: Estimated cumulative probability function of the number of k^{th} type 1 records in a time series of length 122 ($N^*_{k,122}$). Also shown is the expectation of $N^*_{k,122}$. This was calculated using (5.2.1) and verified using the Monte Carlo Simulation.

4 or fewer type 1 1st records in a time series of length 122 as observing a lower than expected number of records. Let us also classify observing 7 or greater type 1 1st records in a time series of length 122 as observing too many records. We can now say that the number of type 1 1st record highs was found to be higher than expected in 69 of the 72 time series in Table 5.3. The number of record lows was lower than expected in 47 of the 72 time series in Table 5.3. This indicates that the time series are not stationary (see Section 5.1). We can assign a probability to each month and location of observing the calculated number of minimal and maximal records. However, we cannot calculate the probability of making the $12 \times 6 = 72$ observations, as the locations (and to some extent months) are not independent.

In order to draw conclusions about whether more (or fewer) records are being observed than would be expected of stationary time series we must consider a collection of independent time series. There is low correlation between months for the number of observed records. This is especially true of the location Harlingen, therefore for Harlingen, each month can be considered to be an independent time series. Another

x	$P(\sum_{k=1}^{10} N^*_{k,122} = x)$	$P(\sum_{k=1}^{10} N^*_{k,122} \leq x)$
25	0.01	0.01
26	0.01	0.02
27	0.01	0.03
28	0.03	0.06
29	0.04	0.10
30	0.05	0.15
31	0.07	0.22
32	0.09	0.31
33	0.09	0.40
34	0.10	0.50
35	0.10	0.60
36	0.09	0.69
37	0.09	0.78
38	0.06	0.84
39	0.05	0.89
40	0.04	0.93
41	0.03	0.96
42	0.01	0.97
43	0.02	0.99
44	0.01	0.99
45	0.00	0.99
46	0.00	1.00
$E\sum_{k=1}^{10} N^*_{k,122}$	34.56	

Table 5.2: Estimated cumulative probability function of the sum of the number of k^{th},$k = 1, \ldots, 10$, type 1 records in a time series of length 122. Also shown is the expectation of $\sum_{k=1}^{10} N^*_{k,122}$. This was calculated using (5.2.1) and verified using the Monte Carlo Simulation.

source of independent data can be found if we consider not just the 1^{st} type 1 record values (of the minimum and the maximum) but also the 2^{nd}, $3^{\text{rd}}, \ldots, k^{\text{th}}$ type 1 record values. For a time series that is stationary with i.i.d.r.v.s, the number of type 1 i^{th} records is independent of the number of type 1 j^{th} records for $i \neq j$. In order to draw conclusions about whether the monthly sea level data in Harlingen forms stationary time series we investigate whether the values of $\sum_{k=1}^{10} N^*_{k,122}$ and $\sum_{k=1}^{10} N^{+*}_{k,122}$ for each month of the Harlingen time series are close to the expected

	Minimal 1st records						Maximal 1st records					
Month	Del	Den	Har	Hel	Ij	Ma	Del	Den	Har	Hel	Ij	Ma
Jan	7	6	6	5	7	3	7	9	7	**10**	10	10
Feb	6	6	6	5	5	4	7	8	7	7	8	7
Mar	5	6	4	3	6	2	9	11	9	11	10	9
Apr	3	3	3	1	5	3	9	12	11	**7**	10	8
May	7	4	3	4	8	2	6	8	**5**	9	7	8
Jun	3	3	5	1	3	1	**14**	**9**	10	10	12	**14**
Jul	2	1	2	1	2	1	11	10	9	10	10	8
Aug	4	5	3	1	6	1	7	**14**	9	**11**	10	16
Sep	6	4	4	5	6	4	14	**7**	8	9	7	8
Oct	3	6	3	5	7	3	10	7	8	10	7	**6**
Nov	2	2	2	3	3	3	13	10	10	**13**	10	12
Dec	4	4	3	4	4	3	9	8	**9**	7	11	6

Table 5.3: $N'^{*}_{1,n}$ and $N'^{+*}_{1,n}$. Record numbers written in bold font indicate where weak records occurred in the original discrete time series.

value of 34.56. We also show that we can use CaterpillarSSA to create time series from the Harlingen time series that have this property. For simplicity of notation we will denote $\sum_{k=1}^{10} N^{*}_{k,n}$ as $\sum N_{10,n}$, $\sum_{k=1}^{10} N^{+*}_{k,n}$ as $\sum N^{+}_{10,n}$, $\sum_{k=1}^{10} N'^{*}_{k,n}$ as $\sum N'_{10,n}$ and $\sum_{k=1}^{10} N'^{+*}_{k,n}$ as $\sum N'^{+}_{10,n}$.

5.3 Using CaterpillarSSA on Sea level Data

5.3.1 Introduction to CaterpillarSSA

In this section we use singular spectrum analysis (SSA) to decompose the original Harlingen time series into a slowly varying trend and oscillatory components. By taking these elements away from the raw time series we obtain a stationary time series. Extremes can be predicted for this stationary time series in the usual way, after which seasonal variation and trend can be forecasted and added to the prediction of the extreme.

SSA is a model-free method of time series analysis. Thorough descriptions of the theory behind SSA as well as examples can easily be found in texts such as [17]. The program 'CaterpillarSSA' (version 3.1) was used to perform this analysis (see

http://www.gistatgroup.com/cat/ for more information). This program performs the SSA algorithm automatically allowing the user to simply select a number of parameters for the decomposition and reconstruction.

Decomposition

Decomposition requires the user to select just one parameter, the lag (L) with which to create a trajectory matrix. If the time series to be analyzed is denoted $Y = (y_1, \ldots, y_n)$, the trajectory matrix X is the $L \times K$ matrix (where $K = n - L + 1$) consisting of the L-dimensional vectors $X_i = (y_i, \ldots, y_{i+L-1})$. Discussion on determining the correct value of L can be found in literature. [21] advises that L should be large enough but no larger than $n/2$.

Reconstruction

When the selection of L has been made 'CaterpillarSSA' performs singular value decomposition. In doing this d elementary matrices, X_i, are created where $d = rankX$ and $X = X_1 + \ldots + X_d$. Each X_i is given by $X_i = \sqrt{\lambda_i} U_i V_i'$ where λ_i is the i^{th} eigenvalue of the trajectory matrix X and U_i and V_i are the left and right i^{th} eigenvectors of the trajectory matrix respectively. The collection $(\sqrt{\lambda_i}, U_i, V_i)$ is called the i^{th} eigentriple (marked as ET on plots below) of the trajectory matrix. It is these eigentriples that must be selected to reconstruct X. If eigentriples $i_1, \ldots, i_m$ are selected, we create a matrix $X_I = X_{i_1} + \ldots + X_{i_m}$. If correctly separated, some of the eigentriples will reconstruct the trend present in the original time series and others (often pairs of eigentriples) will reconstruct periodicities. The final stage performed by 'CaterpillarSSA' is to average diagonally across each of $X_{i_1}, \ldots, X_{i_m}$ to create $\tilde{X}_{i_1}, \ldots, \tilde{X}_{i_m}$. These can be summed to create $\tilde{X}$. $\tilde{X}$ is the trajectory matrix of the reconstructed series. The trajectory matrix of the residual time series is given by $X - \tilde{X}$. Our aim is to discover if it is possible to reconstruct the time series so that the reconstruction can be used to forecast trend and periodicities into the future. We also require estimates of the extremes of the residual time series to

be accurately made.

5.3.2 Analysis

The periodogram feature of 'CaterpillarSSA' can help us to identify the periods that are present in the original time series. These can be identified by the location of sharp spikes on the periodogram. Table 5.4 shows the main periods present in the periodograms of the original time series. They are listed in descending order of height of peak. Therefore the number listed first is likely to correspond to the strongest period for each month.

Month	Peaks of periodogram (descending order of height)
January	61, 8, 6, 9, 5, 3, 2, 24
February	30, 2, 8, 13, 3
March	6, 11, 3, 4, 8, 24
April	11, 17, 30, 2, 24, 4
May	10, 20, 40, 15
June	17, 2, 11
July	6, 4, 5, 15
August	12, 20, 40, 2, 6
September	2, 17, 4, 30
October	3, 14, 6, 24, 8
November	30, 2, 4, 15
December	6, 9, 5, 40, 3

Table 5.4: Main peaks found in periodograms of each month of the Harlingen time series. The highest peaks correspond to the most prominent periods (in years) in the time series.

Figure 5.8 shows time series for sea level at Harlingen for each month of the year. Figure 5.9 shows $\sum N'_{10,n}$ and $\sum N'^{+}_{10,n}$ against the expected value $\mathrm{E}\sum N_{10,n} = \sum_{j=1}^{10}\sum_{i=j}^{n} 1/i$, $n = 1, \ldots, 122$, for each month of the year. Also shown on Figure 5.9 is a $> 90\%$ confidence interval for the number of records. This confidence interval was calculated using Monte Carlo simulation. 10 000 stationary time series of length

122 were simulated (by drawing random variables from a normal distribution). At each sample size the c.d.f. of $\sum N_{10,n}$ was estimated. The confidence interval was then taken to be the values s that satisfy $x \leq s \leq y$ where x and y are given by $x = \max_t \left(\mathrm{P}\left(\sum N_{10,n} \leq t \right) < 0.05 \right)$ and $y = \min_t \left(\mathrm{P}\left(\sum N_{10,n} \geq t \right) < 0.05 \right)$. For the 1st type 1 record this confidence can be read off Table 5.2 as $28 \leq s \leq 40$. The confidence interval cannot be exactly 90% as $\sum N'_{10,n}$ and $\sum N'^{+}_{10,n}$ are discrete. Note that there is no independence between $\mathrm{P}\left(\sum N_{10,i} \right)$, $10 < i \leq n$, this means that care must be taken when drawing conclusions from how often the observed number of records falls within the confidence interval. Also plotted on these graphs are the number of records observed in residual time series when just the trend has been removed (dashed line) and when the trend and the main cycle had been removed (dotted line). It can be seen that in most cases the removal of these elements from the original time series is not enough to bring all of the resulting residuals within the confidence intervals. However, it is clear that the removal of the trend goes some way to making the time series stationary and removal of the period is also important. By stationary, we mean in the sense that the distribution of record number is as expected for a series where the elements are i.i.d. random variables.

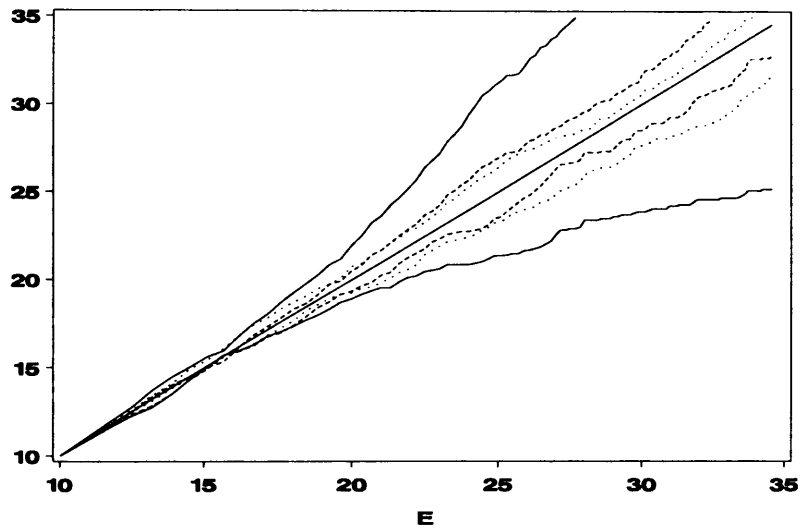


Figure 5.7: Mean number of records over the 12 months plotted against expected number of records $E(\sum N_{10,n})$ for a stationary time series. The solid lines relate to $\sum N'_{10,n}$ and $\sum N'^{+}_{10,n}$ from the original time series, the dashed line to $\sum N'_{10,n}$ and $\sum N'^{+}_{10,n}$ from the residuals of the time series with the trend removed and the dotted line to $\sum N'_{10,n}$ and $\sum N'^{+}_{10,n}$ from the residual left from removing the trend and the main periodicity.

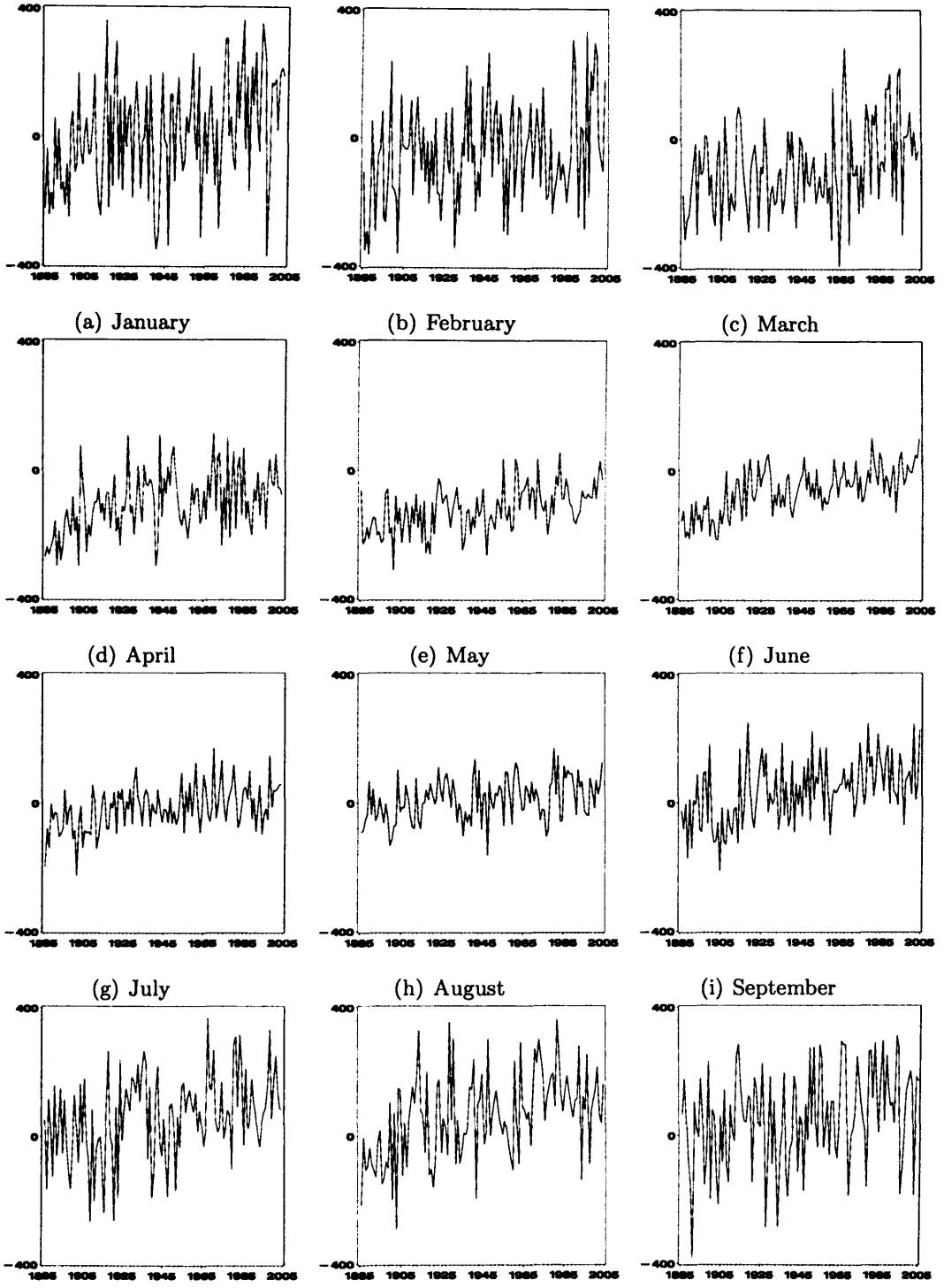
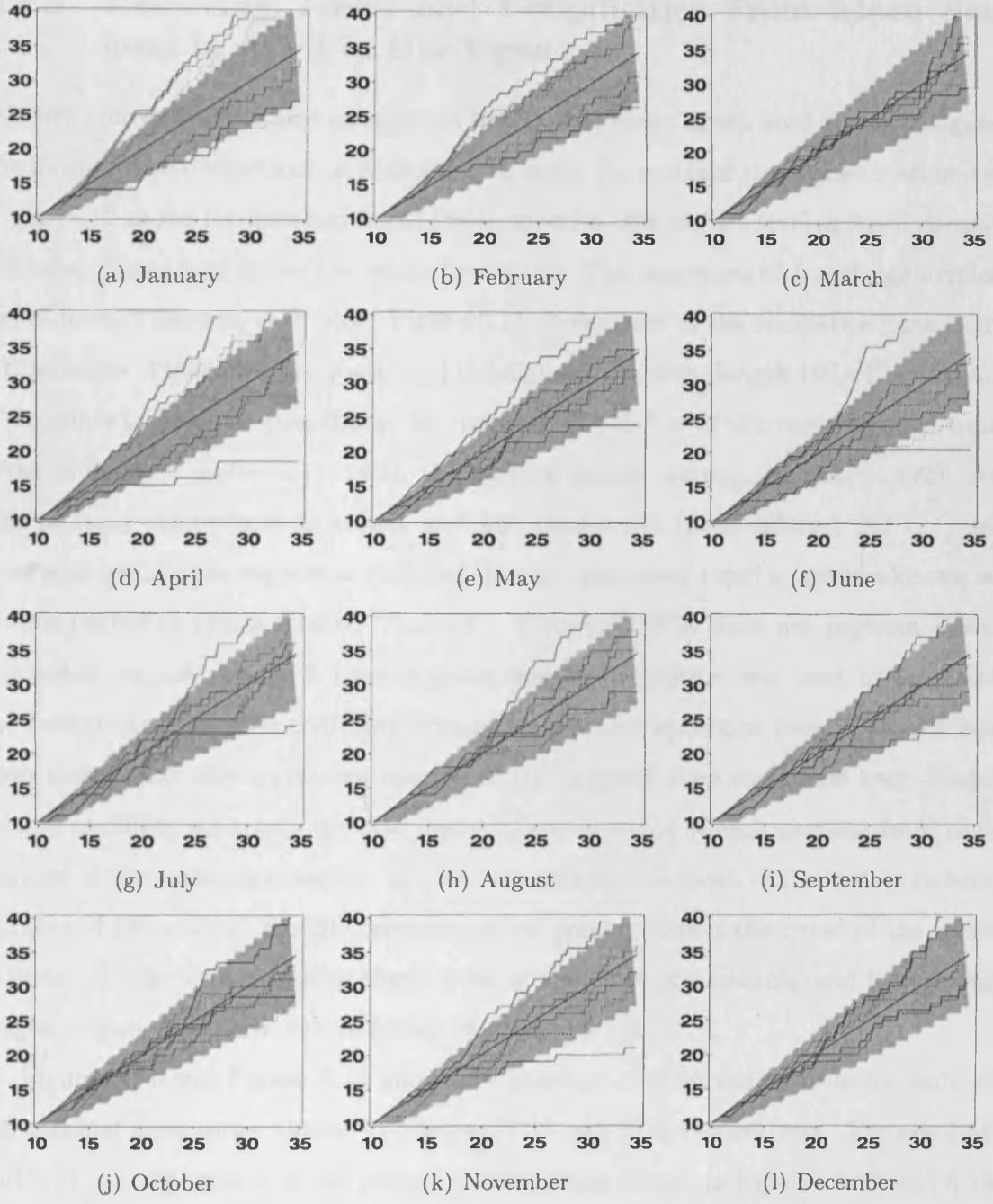


Figure 5.8: Time series of sea level at Harlingen for each month of the year from 1885 to 2000.



(a) January (b) February (c) March

(d) April (e) May (f) June

(g) July (h) August (i) September

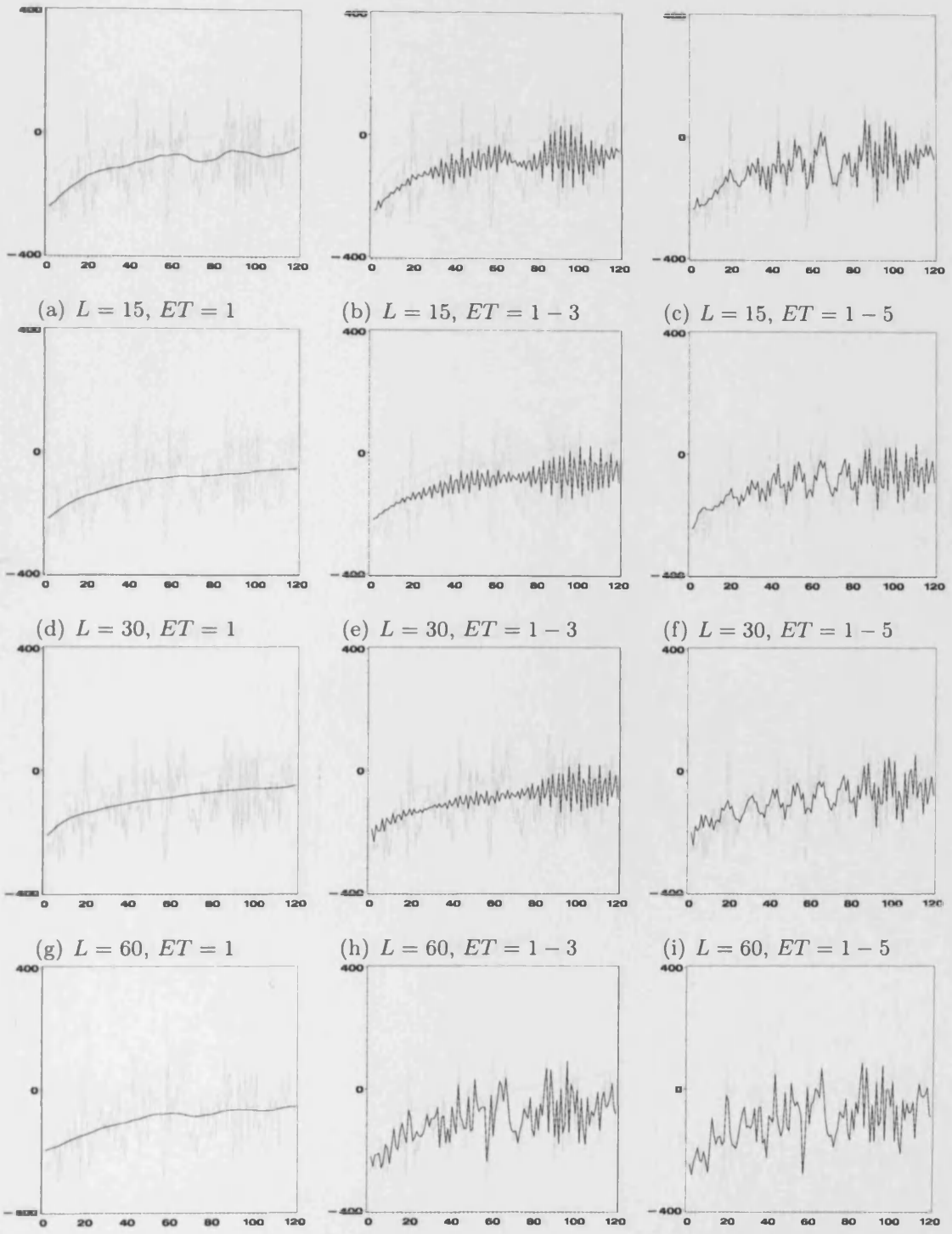
(j) October (k) November (l) December

Figure 5.9: $\sum N'_{10,n}$ and $\sum N'^{+}_{10,n}$ plotted against expected number of records $E(\sum N_{10,n})$ for the original time series, the residual time series left after removing the trend (dashed line) and the residual time series left after removing the trend and the main periodicity (dotted line). The $> 90\%$ confidence interval (as described above) for the number of records $\sum N_{10,n}$ is shaded in grey.

5.3.3 Removing Trend and Periodicities From Mean Sea level in April in Harlingen

We now concentrate on just one month of the time series of sea level and investigate the most effective selections of L and ET to make the residual time series stationary. Figure 5.10 shows reconstructions of the time series of mean sea level in April (length 122) and Figure 5.12 shows the related residuals. The selections of L and eigentriples are indicated beneath each plot. Figure 5.11 shows part of the reconstructions from a time series of mean monthly sea level throughout the year (length $122 \times 12 = 1464$). The points plotted are just those that relate to April (i.e. if the reconstructed time series is denoted $y_i, i = 1, \ldots 1464$, the plotted points are $y_{12i-8}, i = 1, \ldots 122$). As well as using eigentriples to reconstruct the time series (plots labeled 'ET=..') we have also used linear regression (labeled 'Linear regression') and a method known as double centering (plots labeled 'Double'). CaterpillarSSA does not perform linear regression reconstruction, a least-squares regression routine was used to calculate the best-fit line. Double centering acts after the decomposition stage: After L has been chosen and the trajectory matrix of the original time series has been made, double centering subtracts the row and column averages of each element from each element of the trajectory matrix. It then adds the overall mean of the matrix to each element of the matrix. Double centering gives good results if the trend of the series is linear. Under this centering there is no access to approximating and forecasting stages. Figure 5.13 show the residuals related to Figure 5.11.

Figure 5.14 and Figure 5.15 show the number of observed records for each of the residual time series shown in Figures 5.12 and 5.13 respectively. Figures 5.16 and 5.17 are histograms of the points in time series shown in Figures 5.12 and 5.13 respectively.



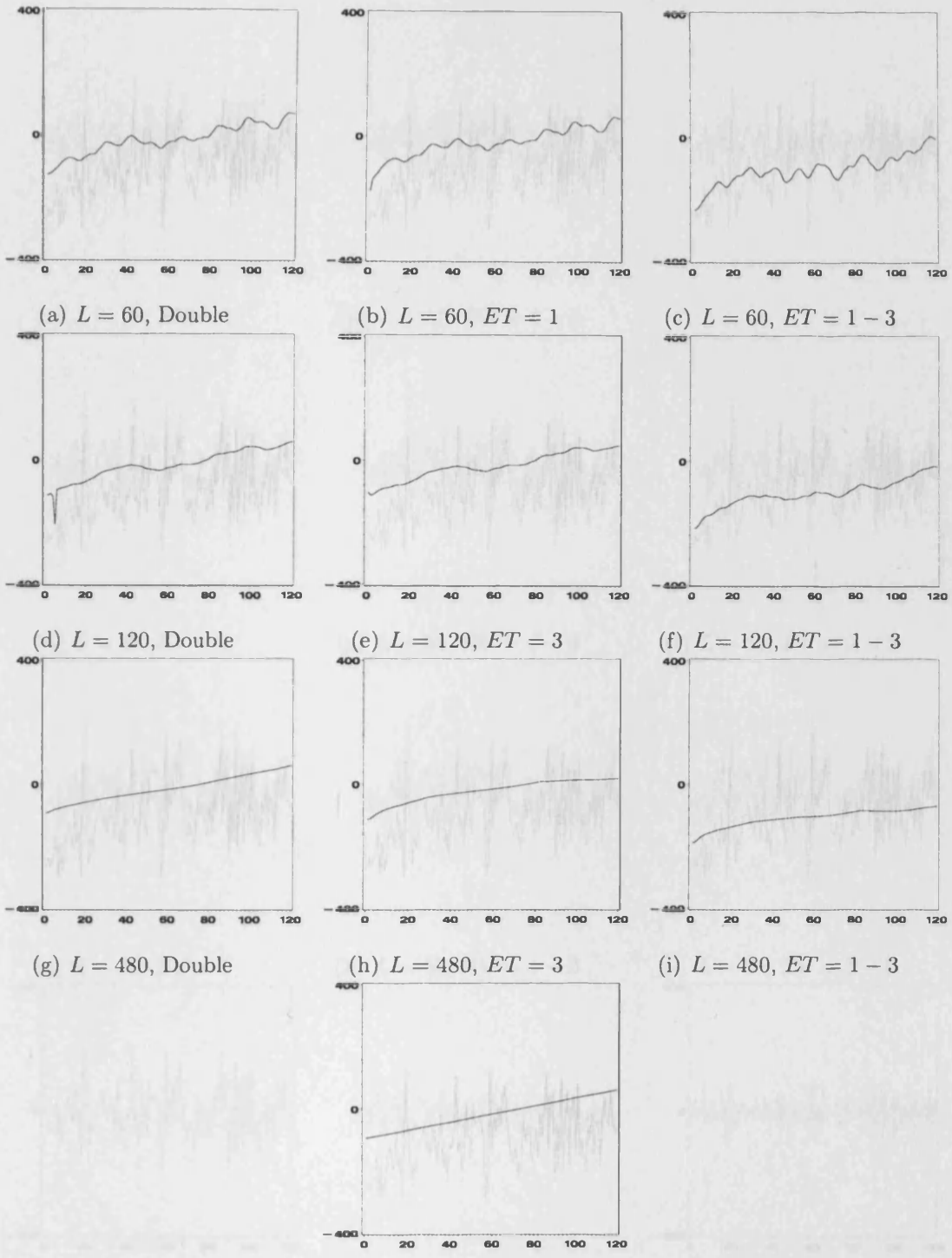
(a) $L = 15$, $ET = 1$ (b) $L = 15$, $ET = 1 - 3$ (c) $L = 15$, $ET = 1 - 5$

(d) $L = 30$, $ET = 1$ (e) $L = 30$, $ET = 1 - 3$ (f) $L = 30$, $ET = 1 - 5$

(g) $L = 60$, $ET = 1$ (h) $L = 60$, $ET = 1 - 3$ (i) $L = 60$, $ET = 1 - 5$

(j) $L = 20$ Double (k) $L = 15$ $ET = 1 - 7$ (l) $L = 15$ $ET = 1 - 9$

Figure 5.10: Raw time series of sea level at Harlingen in April (grey line) and the reconstruction of the main periods and trend achieved as labeled (black line). Using just April data.



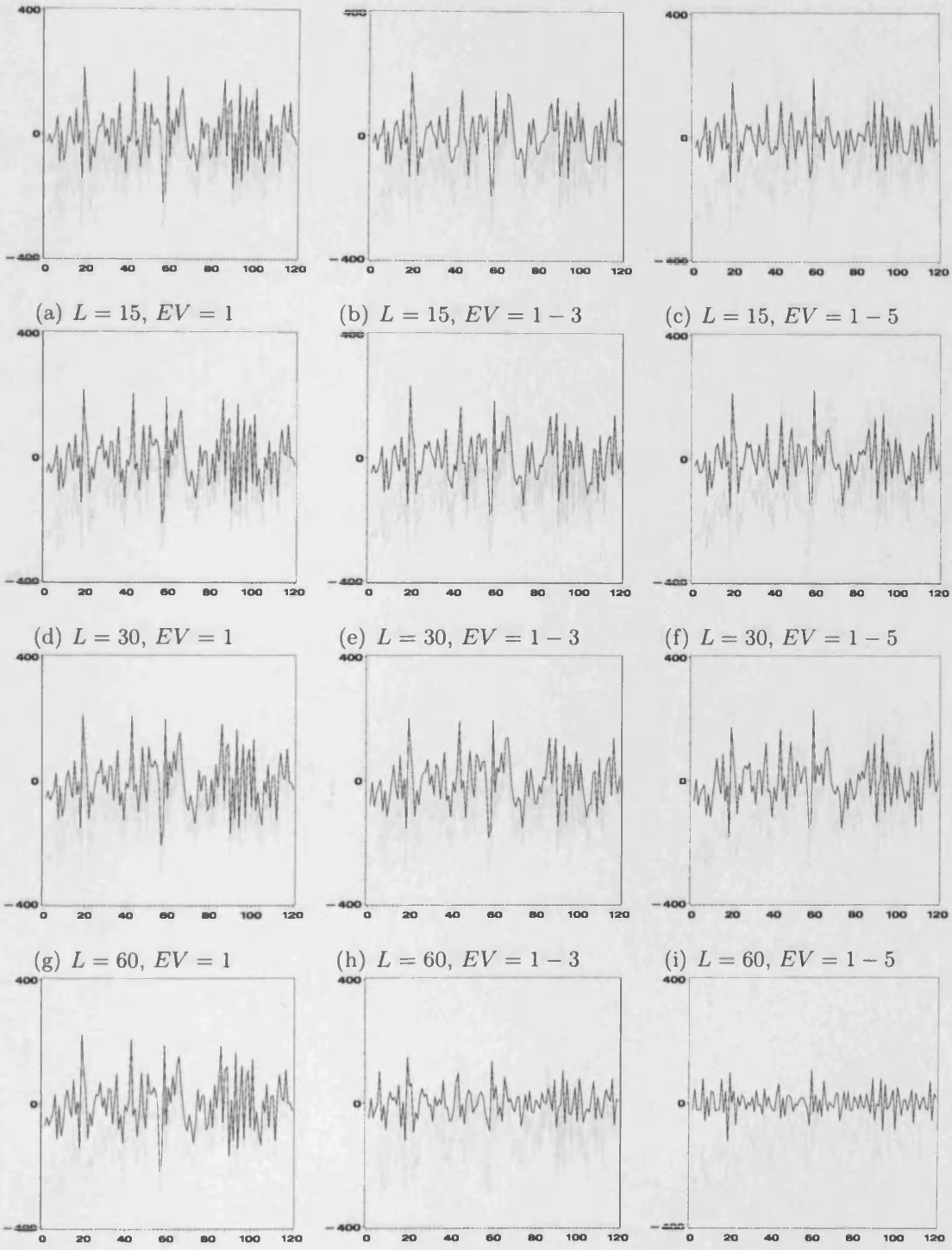
(a) $L = 60$, Double (b) $L = 60$, $ET = 1$ (c) $L = 60$, $ET = 1 - 3$

(d) $L = 120$, Double (e) $L = 120$, $ET = 3$ (f) $L = 120$, $ET = 1 - 3$

(g) $L = 480$, Double (h) $L = 480$, $ET = 3$ (i) $L = 480$, $ET = 1 - 3$

(j) Linear regression

Figure 5.11: Raw time series of sea level at Harlingen in April (grey line) and the reconstruction achieved as labeled (black line). Using all Harlingen data then considering just April.



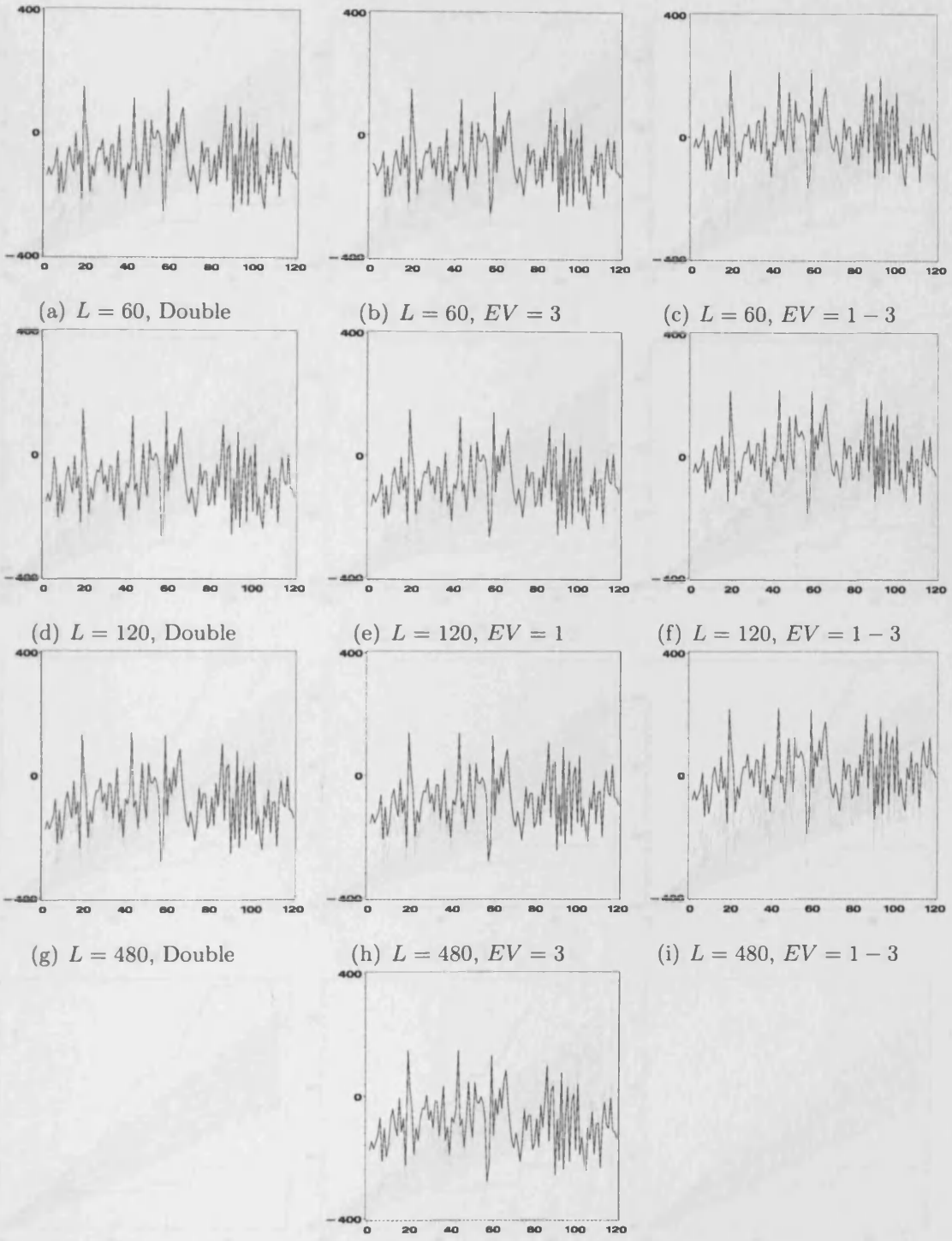
(a) $L = 15$, $EV = 1$ (b) $L = 15$, $EV = 1 - 3$ (c) $L = 15$, $EV = 1 - 5$

(d) $L = 30$, $EV = 1$ (e) $L = 30$, $EV = 1 - 3$ (f) $L = 30$, $EV = 1 - 5$

(g) $L = 60$, $EV = 1$ (h) $L = 60$, $EV = 1 - 3$ (i) $L = 60$, $EV = 1 - 5$

(j) $L = 20$ Double (k) $L = 15$ $EV = 1 - 7$ (l) $L = 15$ $EV = 1 - 9$

Figure 5.12: Original time series of sea level at Harlingen in April (grey line) and residual created by the extraction of reconstructed series from the original time series (black line). Using just April data.



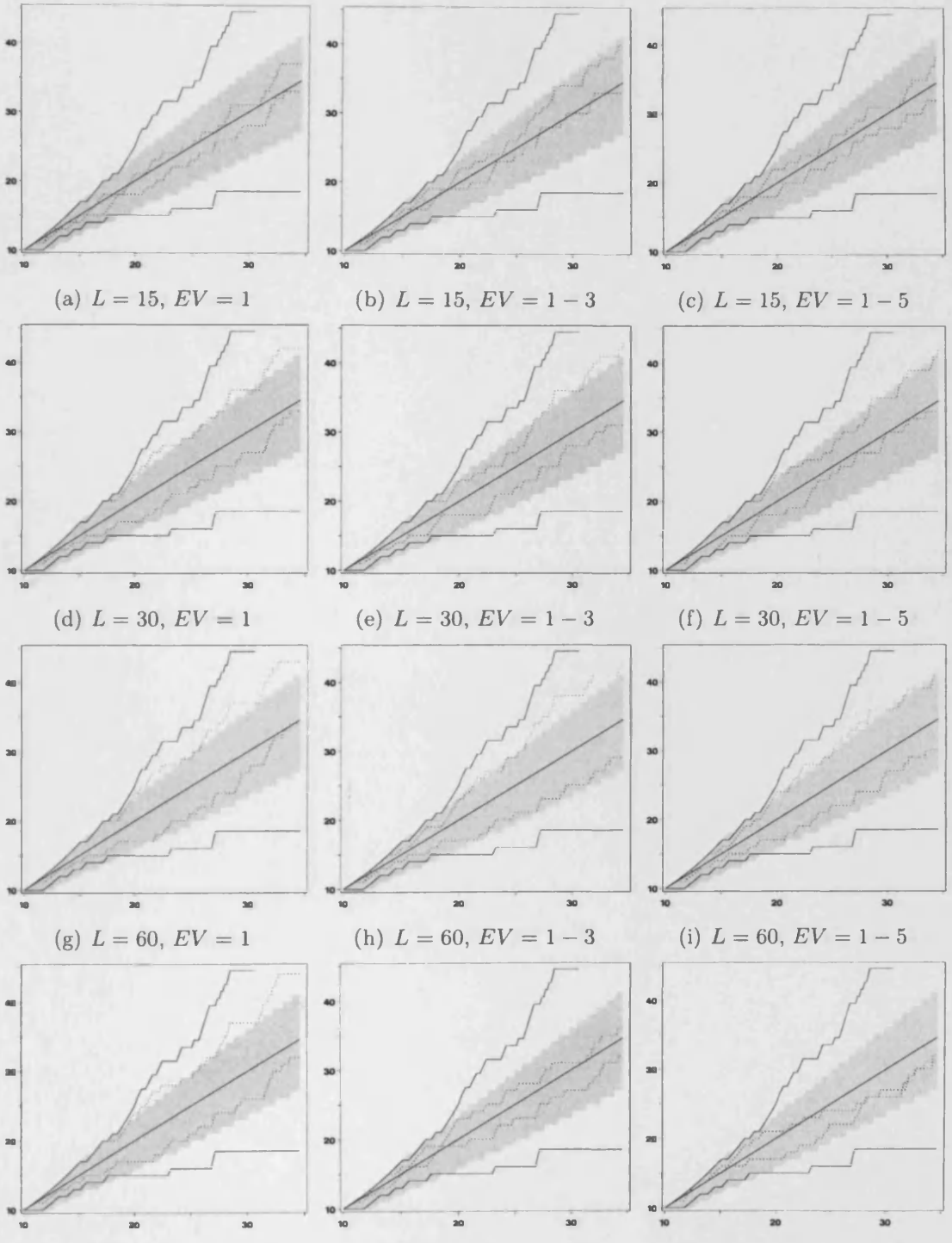
(a) $L = 60$, Double (b) $L = 60$, $EV = 3$ (c) $L = 60$, $EV = 1 - 3$

(d) $L = 120$, Double (e) $L = 120$, $EV = 1$ (f) $L = 120$, $EV = 1 - 3$

(g) $L = 480$, Double (h) $L = 480$, $EV = 3$ (i) $L = 480$, $EV = 1 - 3$

(j) Linear regression

Figure 5.13: Original time series of sea level at Harlingen in April (grey line) and residual created by the extraction of reconstructed series from the original time series (black line). Using all Harlingen data then considering just April.



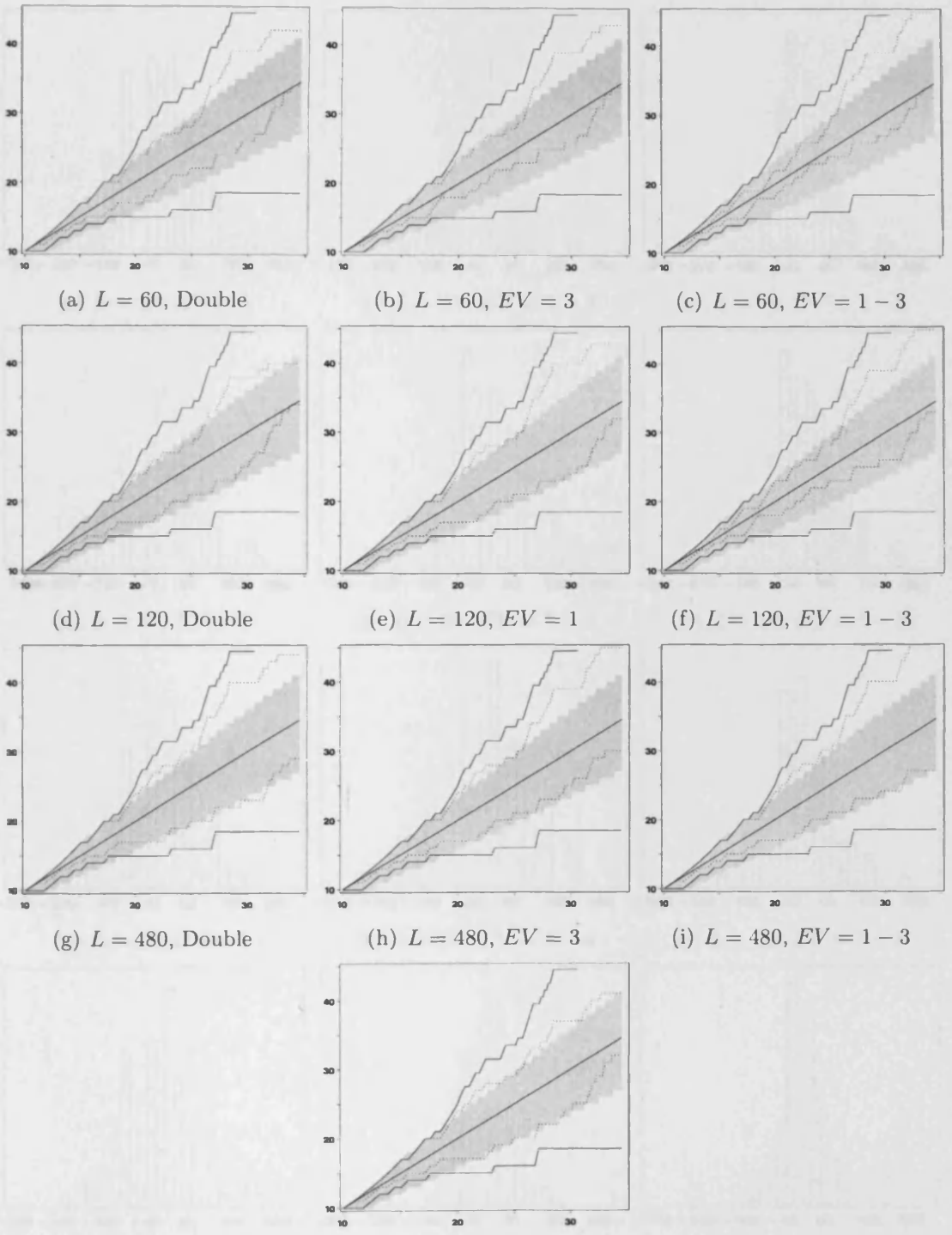
(a) $L = 15$, $EV = 1$ (b) $L = 15$, $EV = 1 - 3$ (c) $L = 15$, $EV = 1 - 5$

(d) $L = 30$, $EV = 1$ (e) $L = 30$, $EV = 1 - 3$ (f) $L = 30$, $EV = 1 - 5$

(g) $L = 60$, $EV = 1$ (h) $L = 60$, $EV = 1 - 3$ (i) $L = 60$, $EV = 1 - 5$

(j) $L = 20$ Double (k) $L = 15$ $EV = 1 - 7$ (l) $L = 15$ $EV = 1 - 9$

Figure 5.14: $\sum N'_{10,n}$ and $\sum N'^{+}_{10,n}$ (for original time series and residual) plotted against expected number of records $E(\sum N_{10,n})$. Using just April data. The '$> 90\%$' confidence interval for the number of records $\sum N_{10,n}$ is shaded in grey.



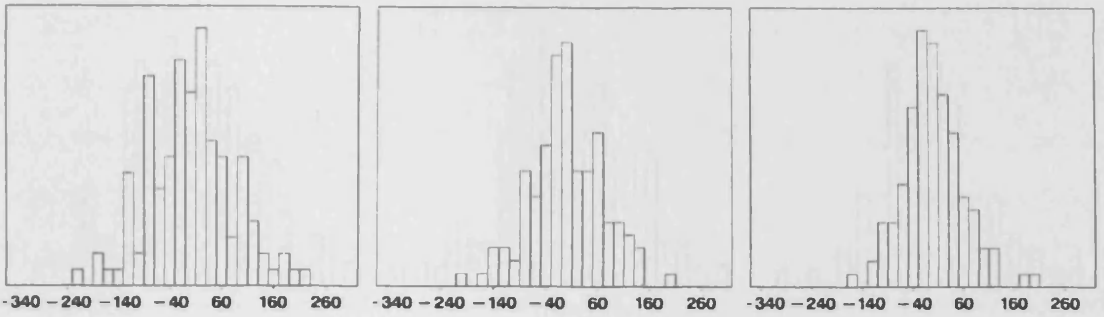
(a) $L = 60$, Double (b) $L = 60$, $EV = 3$ (c) $L = 60$, $EV = 1 - 3$

(d) $L = 120$, Double (e) $L = 120$, $EV = 1$ (f) $L = 120$, $EV = 1 - 3$

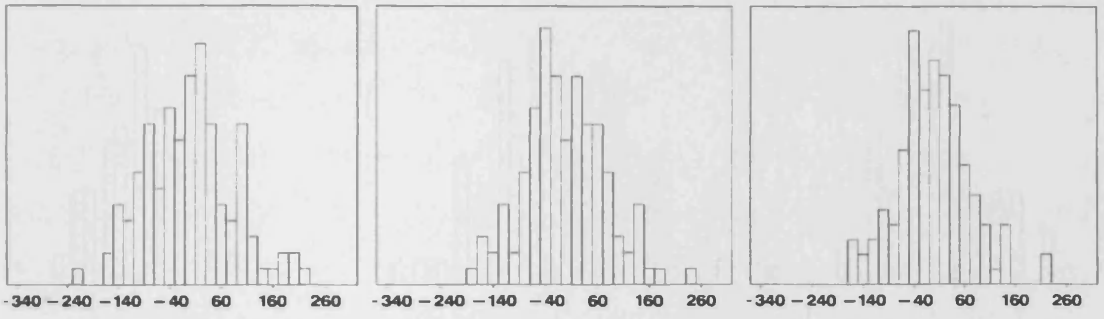
(g) $L = 480$, Double (h) $L = 480$, $EV = 3$ (i) $L = 480$, $EV = 1 - 3$

(j) Linear regression

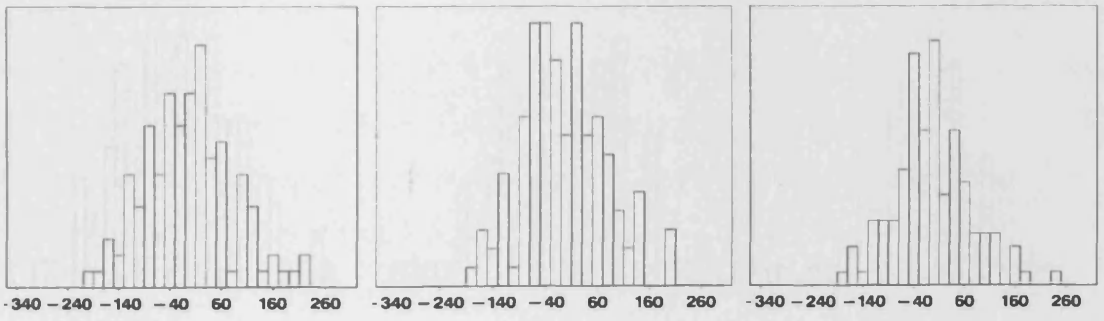
Figure 5.15: $\sum N'_{10,n}$ and $\sum N'^{+}_{10,n}$ (for original time series and residual) plotted against expected number of records $E(\sum N_{10,n})$. Using all data then considering just April. The '> 90%' confidence interval for the number of records $\sum N_{10,n}$ is shaded in grey.



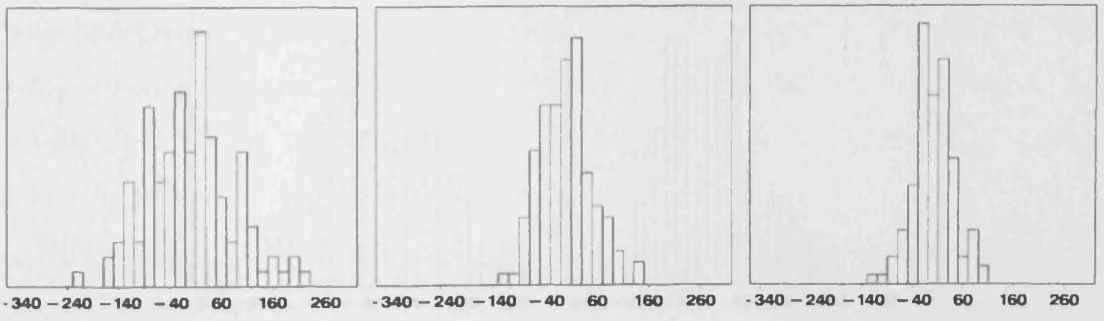
(a) $L = 15$, $EV = 1$ (b) $L = 15$, $EV = 1 - 3$ (c) $L = 15$, $EV = 1 - 5$



(d) $L = 30$, $EV = 1$ (e) $L = 30$, $EV = 1 - 3$ (f) $L = 30$, $EV = 1 - 5$

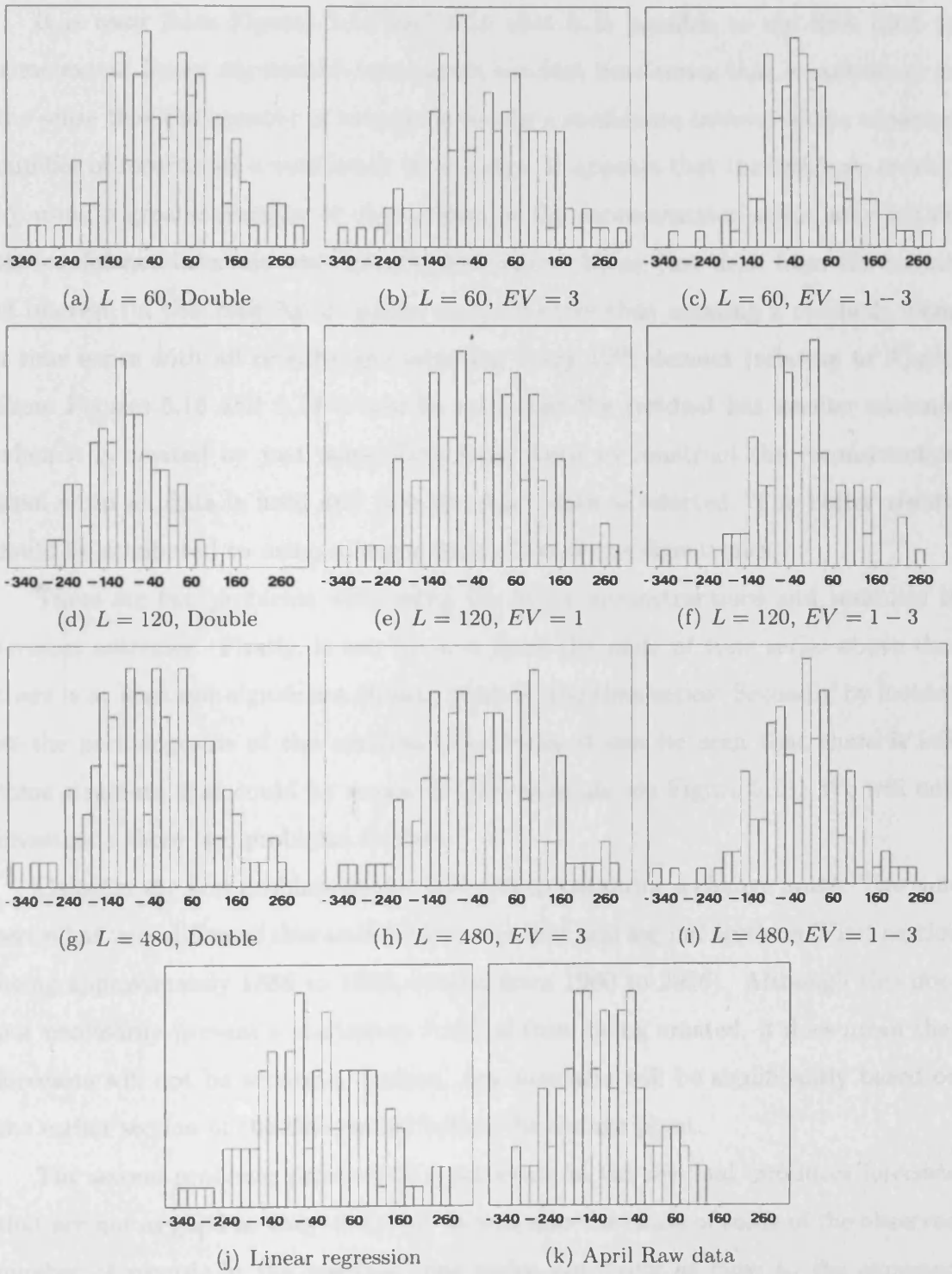


(g) $L = 60$, $EV = 1$ (h) $L = 60$, $EV = 1 - 3$ (i) $L = 60$, $EV = 1 - 5$



(j) $L = 20$ Double (k) $L = 15$ $EV = 1 - 7$ (l) $L = 15$ $EV = 1 - 9$

Figure 5.16: Histograms of residual time series created by removing trend and/or cycles. Using just April data.



(a) $L = 60$, Double (b) $L = 60$, $EV = 3$ (c) $L = 60$, $EV = 1 - 3$

(d) $L = 120$, Double (e) $L = 120$, $EV = 1$ (f) $L = 120$, $EV = 1 - 3$

(g) $L = 480$, Double (h) $L = 480$, $EV = 3$ (i) $L = 480$, $EV = 1 - 3$

(j) Linear regression (k) April Raw data

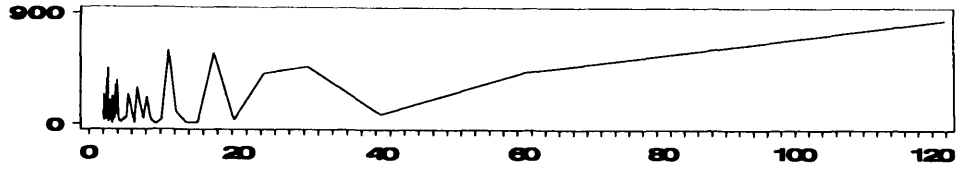
Figure 5.17: Histograms of residual time series created by removing trend and/or cycles. Using all Harlingen data then considering just April.

It is clear from Figures 5.14 and 5.15 that it is possible to use SSA (and to some extent linear regression) to create a residual time series that is stationary in the sense that the number of records is within a confidence interval of the expected number of records for a stationary time series. It appears that the residuals created by using a greater number of eigentriples in the reconstruction stage, stay within the confidence intervals with greater frequency. Using just data from the month of interest (in this case April) seems more effective than creating a residuals using a time series with all months and selecting every 12^{th} element (relating to April). From Figures 5.16 and 5.17 it can be seen that the residual has smaller variance when it is created by just using data from April to construct the reconstruction than when all data is used and just the April data is selected. The better results could be attributed to using a larger proportion of the eigentriples.

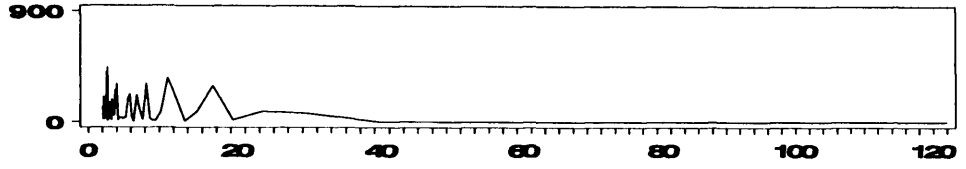
There are two problems with using the above reconstructions and residuals to forecast extremes. Firstly, it can be seen from the plots of time series above that there is at least one significant change point in the time series. Secondly by looking at the periodograms of the residual time series it can be seen that there is still some structure that could be removed. For example see Figure 5.18. We will now investigate these two problems further.

Consider the first problem of the time series containing a change point. The time series has very different characteristics in the first and second sections (First section being approximately 1885 to 1960, second from 1960 to 2006). Although this does not necessarily prevent a stationary residual from being created, it does mean that forecasts will not be accurate. Indeed, any forecasts will be significantly based on the earlier section of the time series before the change point.

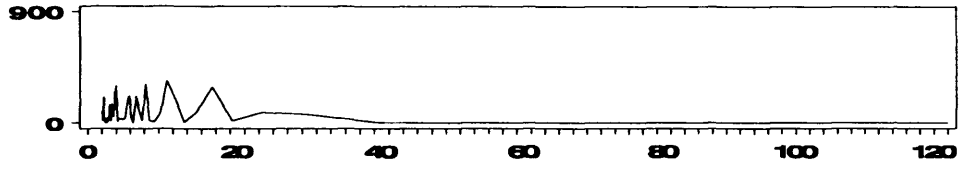
The second problem; presence of a structure in the residual, produces forecasts that are not as good as they could be. It was also the cause of some of the observed number of records in the residual time series not being as close to the expected values as others. The problem is easily overcome by using more eigenvalues in the



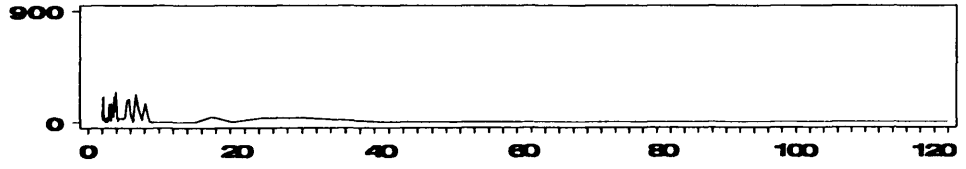
(a) Raw data



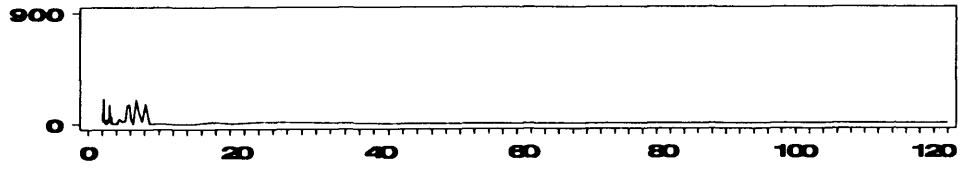
(b) $ET = 1$



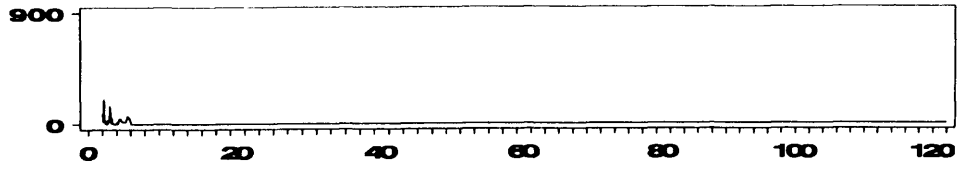
(c) $ET = 1 - 3$



(d) $ET = 1 - 5$



(e) $ET = 1 - 7$



(f) $ET = 1 - 9$

Figure 5.18: Figure (a) shows the periodogram of the raw time series of sea level at Harlingen in April. Plots (b)-(f) show the periodograms of the residuals created by the extraction of reconstructed series from the raw time series. The eigentriples used in the reconstruction are marked below the periodograms.

reconstruction of the time series. Doing this will inevitably introduce more noise into the reconstruction, leading to a poor forecast. These factors must be weighed up against each other.

5.3.4 Change Point

The plots in Figure 5.19 show reconstruction from either single, pairs or three eigentriples of the time series of mean monthly sea level from January 1885 to December 2007. They show clearly that both the amplitude of periodicities and the gradient of trend are not regular. This indicates the presence of a change point.

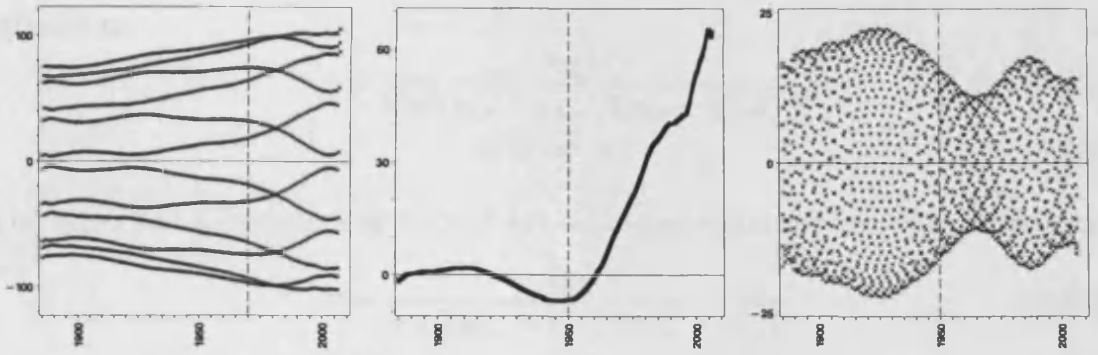
We can only use CaterpillarSSA to forecast accurately if the time series has no change point. We must therefore discard any data from the years before the latest change point occurred. However, if we continue to only consider the data relating to April there is not much data to work with. Below we consider the mean monthly sea level in Harlingen for all months of the year from three different sections.

5.4 Estimating Tail Index and Endpoint

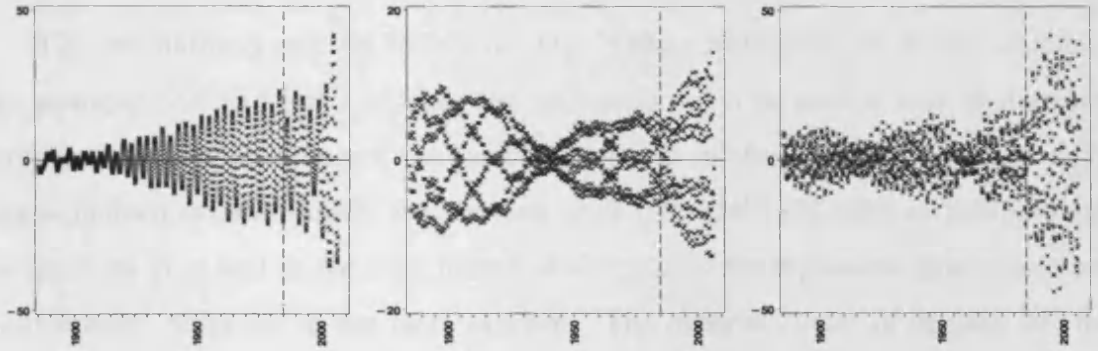
5.4.1 Methodology

In this section we use SSA to remove the trend and periods from sections of the original time series of mean monthly sea level in Harlingen (of all months). The first section considered is the fairly stable period from 1885 to 1904. We then consider the period from 1973 to 1992. The reconstructions in Figure 5.19 would suggest that this section contains a change point. Finally we consider the period of most up-to-date data: 1987-2006. Note that the period say, 1973-1992, describes all dates from 1st January 1973 to 31st December 1992 inclusive, so each time series has a length of 240 observations.

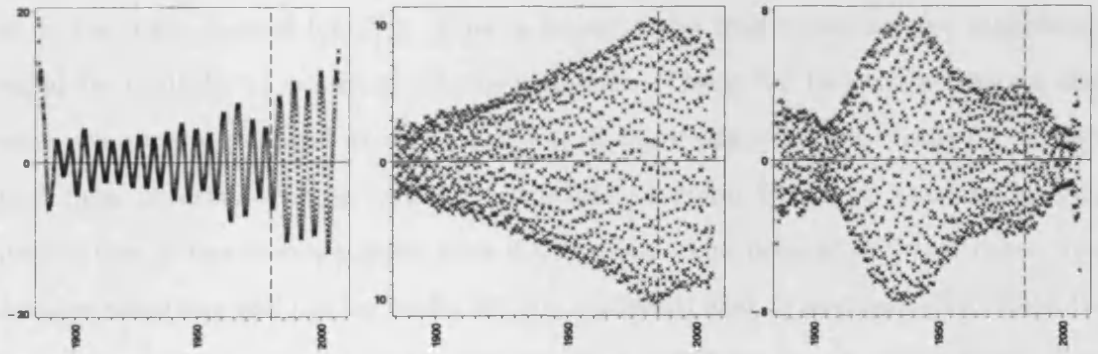
The next step is to estimate the lower and upper tail indices, α and α^u respectively and the lower and upper endpoints and m and M using the following



(a) L=660 ET=1 & 2 (b) L=660 ET=4 (c) L=660 ET=7 & 8



(d) L=660 ET=13 & 14 (e) L=660 ET=20 (f) L=660 ET=23 & 24 & 25



(g) L=660 ET=29 (h) L=660 ET=38 (i) L=660 ET=41

Figure 5.19: Reconstruction time series that demonstrate irregular periodicities and trends. The original time series of mean monthly sea level at Harlingen was decomposed with lag $L = 660$. Eigentriples or groups of eigentriples were then used to reconstruct the time series. The reconstructions that demonstrate irregular patterns are plotted above. The most recent apparent change point has been marked using a vertical dotted line.

estimators:

$$\hat{\alpha} = \frac{\log(k)}{\log((y_{k,n} - y_{3,n})/(y_{3,n} - y_{2,n}))} \tag{5.4.1}$$

$$\hat{m} = m^{\circ}(\hat{\alpha}). \tag{5.4.2}$$

The upper tail parameters $\hat{M}$ and α^u are estimated using the following estimators:

$$\hat{\alpha}^u = \frac{\log(k)}{\log((y_{k,n} - y_{3,n})/(y_{3,n} - y_{2,n}))} \tag{5.4.3}$$

$$\hat{M} = M^{\circ}(\hat{\alpha}^u) = a_1(\hat{\alpha}^u)y_{n,n} + \ldots + a_k(\hat{\alpha}^u)y_{n-k+1,n}. \tag{5.4.4}$$

The estimator $\hat{\alpha}$ can be found in [11]. Other estimators of α are available, for example [22] and [19]. Indeed, the estimation of α is itself a very challenging problem, however, it was not the primary concern in this study. $m^{\circ}(\vartheta)$ and $a_i(\vartheta)$ are as defined earlier. Finally for sections 1885-1904 and 1973-1992 we compare our estimations of m and M for each month of the year to the minimum (and maximum respectively) observed in the next 14 years. The observed value of m (and M) for each month should not be viewed as the 'true value', but rather as an upper bound for m (or lower bound for M). This is because the true minimum (or maximum) would be unlikely to occur in 14 observations. There are two main reasons that more years were not used as a comparison; if more post-forecast observations were used then observed values would be unlikely to come from the same underlying distribution as the forecast (note that if there is a trend present any year other than the forecasted one will not be useful for comparison); also, it was desired to make the second forecast a relatively up-to-date one, this would not be possible if many years were required after the forecast for comparative purposes. In the analysis described above a value of $k = 10$ was used. Figure 5.24 shows the results of an investigation into the effect of changing k on the estimates of tail index and endpoint.

5.4.2 Results

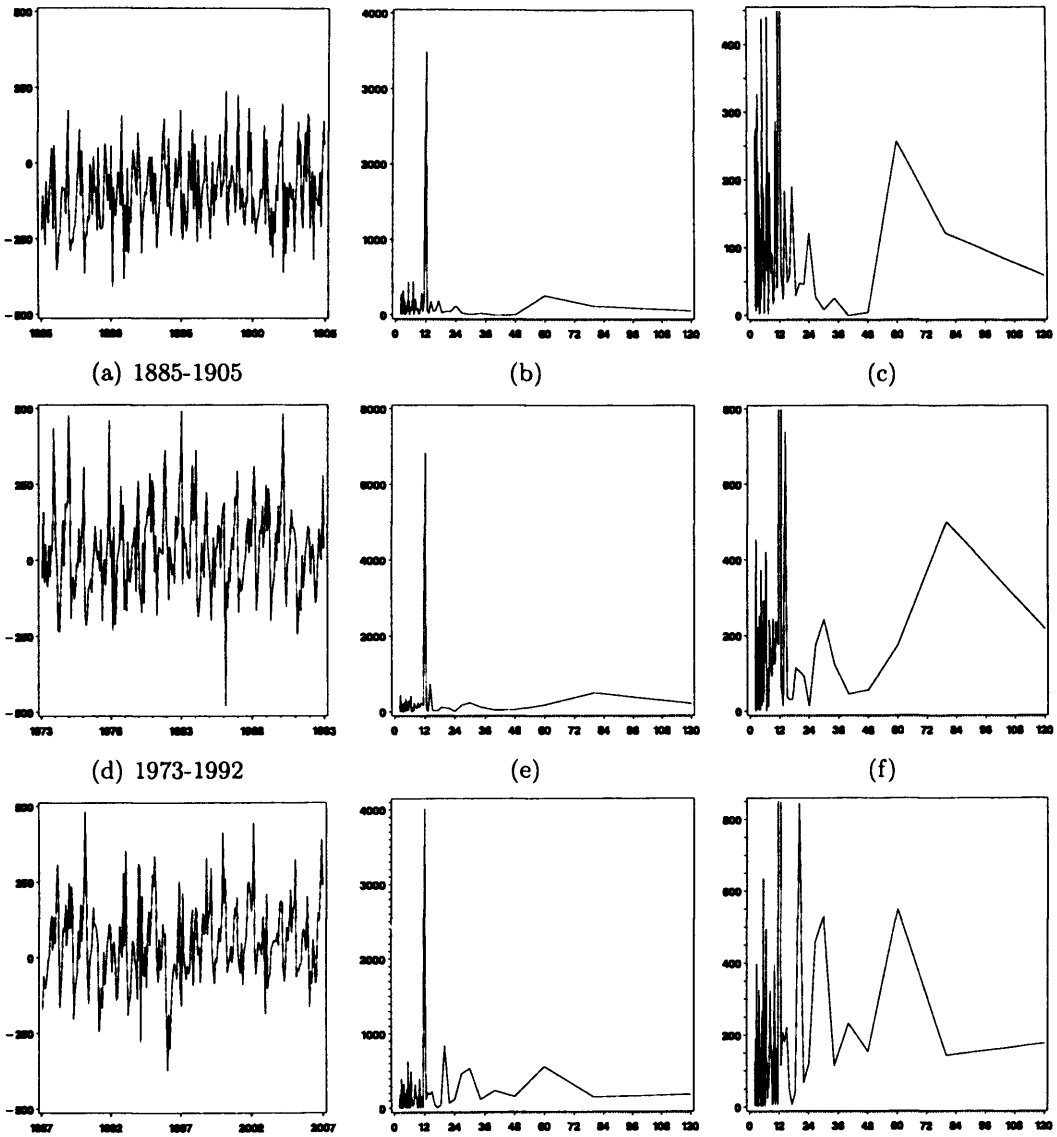
(a), (d) and (g) in Figure 5.20 show the time series of sea level in Harlingen over the three 20 year periods; 1885-1904, 1973-1992 and 1987-2006. The remaining plots in

these Figures are the periodogram of the time series in (a), (d) and (g). (b), (e) and (h) in each of these Figures shows that the time series all include a strong signal with period 12. Plots (c), (f) and (i) show the periodograms with the y-axis cut off above the second largest spike of the periodogram, this is in order to see the other periods that are present in the time series more clearly.

(a) and (b) from Figures 5.21, 5.22 and 5.23 show estimates of α and m using (5.4.1) and (5.4.2) respectively with $k = 10$. In the plot, each point represents an estimation from the residual of a time series where the associated reconstruction used eigentriples $1 - j$ (where j is the value on the x-axis of the plots). Estimates of α and m where $\hat{\alpha} < 0$ or $\hat{\alpha} > 6$ have not been included. Shown in (c) of Figures 5.21, 5.22 and 5.23 are forecasts of the minimum mean monthly sea level in Harlingen and (for Figures 5.21 or 5.22 only) the minimum observed mean monthly sea level for the next 14 years. (c) shows a selection of forecasts, relating to some or all of the following sets of eigentriples: 1, 1-3, 1-10, 1-30 and 1-60. Forecasts are not shown for sets of eigentriples that produce estimates $\hat{\alpha} < 0$ or $\hat{\alpha} > 6$. In Figures 5.21, 5.22 and 5.23, plots (d) (e) and (f) show similar estimates and forecasts, for the same periods, but for the upper tail.

From Figure 5.20 we can see the following. The section of the time series from January 1885 to December 1904 has a small trend and fairly regular periodicities, mainly period 12. The section of the time series from January 1973 to December 1992 has a small trend and periodicities that are less regular than the section from 1885 to 1904. This section of the time series has a greater variance and mean than the first section. This data was chosen as it is relatively up-to-date whilst still having a reasonable post-forecast data set for comparison. The section of the time series 1987 to 2006 was considered as it is the most up-to-date 20 year period with available data. The section of the time series from January 1987 to December 2006 has a small trend and periodicities that are less regular than the section from 1885 to 1904. Again the main period is 12. The mean and variance of this section are

similar to that of the section from 1973 to 1992.



(a) 1885-1905 (b) (c)

(d) 1973-1992 (e) (f)

(g) 1987-2006 (h) (i)

Figure 5.20: (a), (d) and (g) show the time series of mean monthly sea level in Harlingen during the three 20 year time periods as labeled. (b), (e) and (h) show periodograms of each time series. (c) shows the same periodogram but with the y-axis cut off above the second largest spike, so that spikes from periods other than 12 can be seen more clearly.

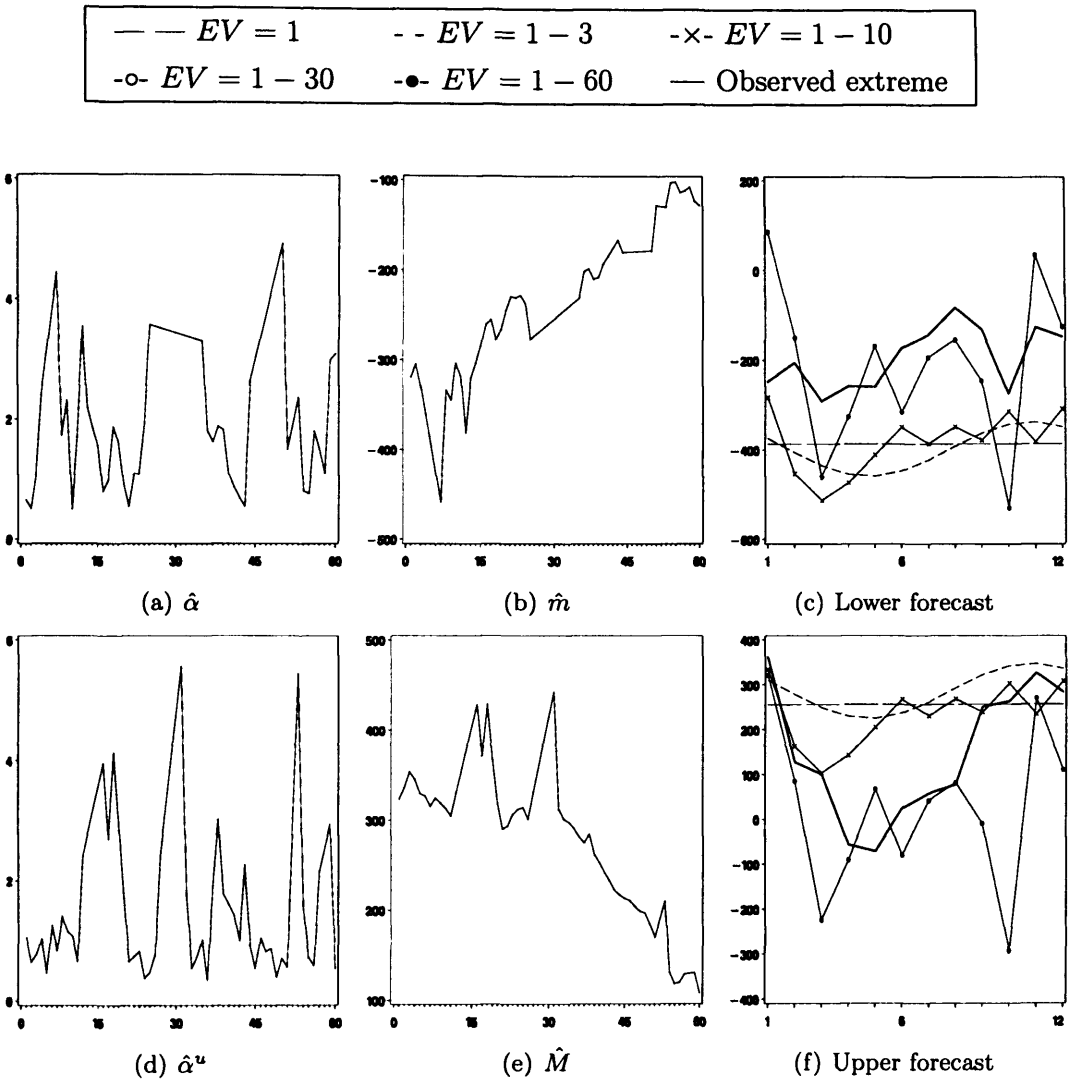


Figure 5.21: Figure (a) shows $\hat{\alpha}$ against the number of eigentriples, j, used in the corresponding reconstruction. Figure (b) shows $\hat{m}$ against j. The estimation of m uses the corresponding $\hat{\alpha}$ from (a). Figure (c) shows a forecast of the minimum mean monthly sea level for each month of the year 1905. The forecasts use the reconstruction of 1885-1904 as a base. Only the reconstructions related to 1, 3, 10, 30 and 60 eigentriples are shown. Also shown in (c) is the minimum mean monthly sea level observed for the period 1905-1918 for each month. Graphs (d), (e) and (f) show similar information but for the upper tail.

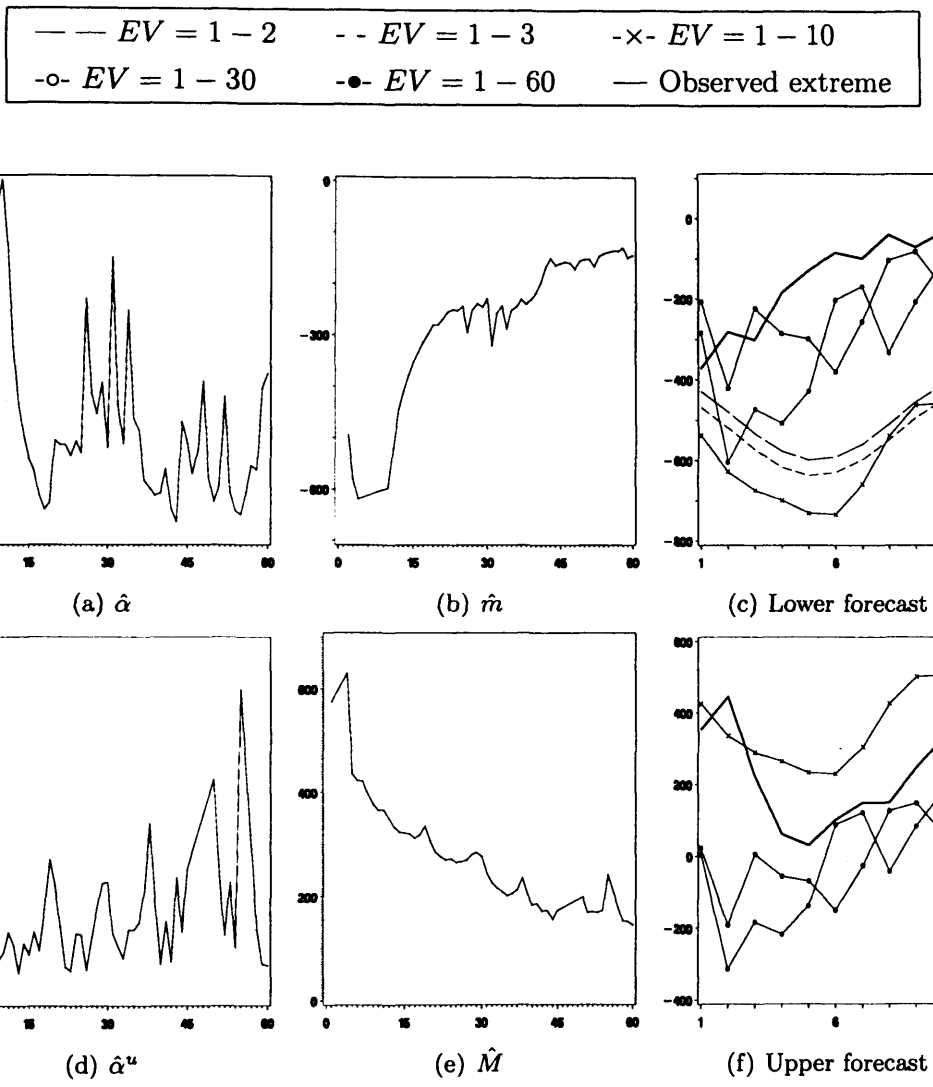


Figure 5.22: Figure (a) shows $\hat{\alpha}$ against the number of eigentriples, j, used in the corresponding reconstruction. Figure (b) shows $\hat{m}$ against j. The estimation of m uses the corresponding $\hat{\alpha}$ from (a). Figure (c) shows a forecast of the mean monthly sea level for each month of the year 1993. The forecasts use the reconstruction of 1973-1992 as a base. Also shown is the minimum mean monthly sea level observed for the period 1993-2006 for each month. (d), (e) and (f) show the similar information but for the upper tail.

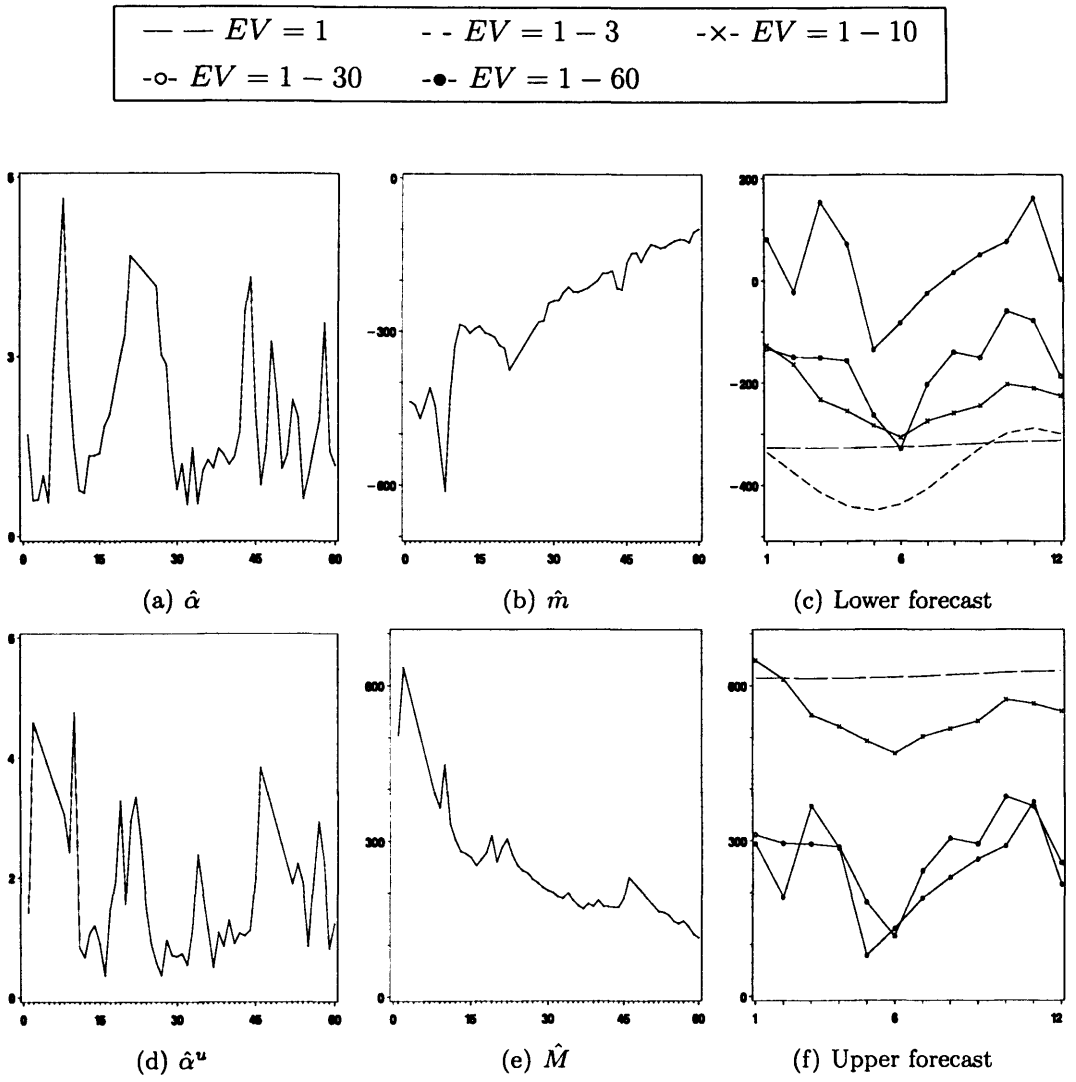


Figure 5.23: Figure (a) shows $\hat{\alpha}$ against the number of eigentriples, j, used in the corresponding reconstruction. Figure (b) shows $\hat{m}$ against j. The estimation of m uses the corresponding $\hat{\alpha}$ from (a). Figure (c) shows a forecast of the mean monthly sea level for each month of the year 2007. The forecasts use the reconstruction of 1987-2006 as a base. No data is available after 2006. (d), (e) and (f) show similar information but for the upper tail.

Plots (c) and (f) of Figures 5.21, 5.22 and 5.23 show us that the final forecasts vary a great deal depending on the set of eigentriples chosen. In general, it appears that the more eigentriples used in a lower forecast, the larger the forecast becomes. That is, if few eigentriples are used, it is less likely that an observation will occur that is less than the forecast. Where more eigentriples have been used, the observed minimum (plotted with the plain solid line) is frequently lower than the lower forecast. Similarly for the upper forecast: when few eigentriples were used the observed maximum sea level was generally lower than the upper forecast. When lots of eigentriples were used the observed maximum sea level was often higher than the upper forecast. This is an obvious consequence of the method. Using too many eigentriples should be avoided.

The upper and lower forecasts (these are equal to the reconstruction forecast plus M and reconstruction forecast plus m respectively) for any particular set of eigentriples, are based on the same reconstruction forecast, and so the difference between them is constant and equal to $M - m$. This can be seen by looking at plots (c) and (f) of Figures 5.21, 5.22 and 5.23.

In Figures 5.21 (c) and (f) it can be seen that the reconstruction made using $ET = 1$ relates to the trend of the time series as the forecast is close to a straight line. Similarly in Figure 5.23 (c) and (f) it can be seen that the reconstruction using $ET = 1$ relates to the trend of the time series. In Figure 5.22 (c) it appears that the reconstruction using $ET = 3$ relates to the trend of the time series as the forecasts for eigentriples $ET = 1$ and $ET = 1 - 3$ are close to parallel.

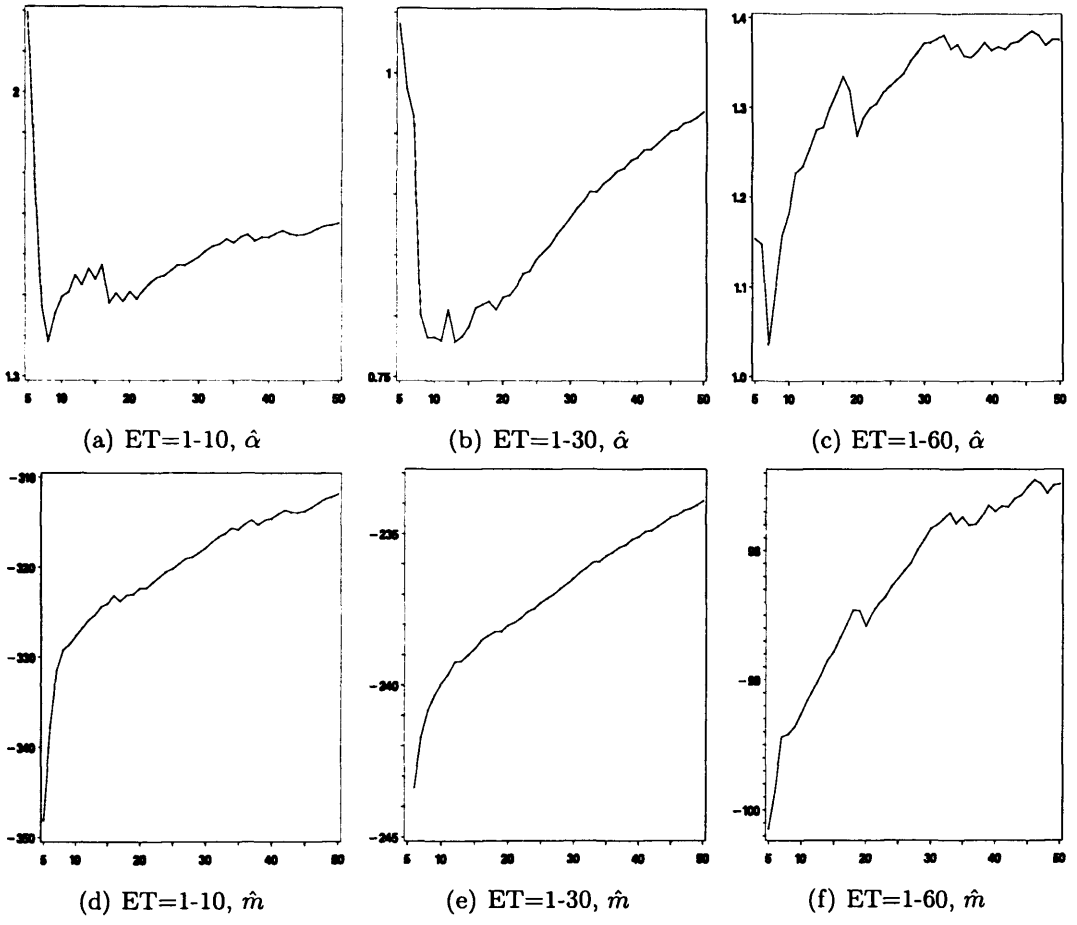
Estimates, $\hat{\alpha}$ and $\hat{\alpha}^u$ vary with the eigentriples used seemingly randomly. Estimates of α greater than 6 or less than 0 were ignored as it is unlikely that the true value of α would be outside these bounds (in fact m° cannot be applied if $\alpha < 0$). When m (or M) was estimated using one of these unlikely estimates of α, $\hat{m}$ and $\hat{M}$ were also unlikely (much smaller or larger than any observed value). Each of the three sections produced similar numbers of poor estimates of α and α^u. The poor

Eigentriples	$\hat{\alpha}^u$	$\hat{M}$
1-13	8.9	587.8
1-14	14.2	847.3
1-15	8.9	642.1
1-28	10.7	686.6
1-29	14.3	833.7
1-30	64.3	2911.0
1-52	10.0	308.1
1-58	8.3	184.1

Table 5.5: Table shows discounted estimates of $\hat{\alpha}^u$ from the time series of 1885-1904. Also shown in the table are the eigenvalues related to the residual and the estimate of M.

estimates of α^u and their corresponding eigentriples and $\hat{M}$ are shown in Table 5.5 for the time period 1885-1904.

The estimates of α and m in Figures 5.21, 5.22 and 5.23 are all made with $k = 10$. Figure 5.24 shows estimates of α and m from the residuals created after the extraction of various reconstructions from the time series of sea level in Harlingen from 1987-2006. It can be seen that altering k changes the estimates of both α and m. In general it appears that as k increases so does the tail index and the estimate of m. For any set of eigentriples shown, the difference between the largest estimate of α and the smallest is of a similar order of magnitude (< 1). The difference between the largest estimate of m and the smallest estimate of m decreases as the number of eigentriples included in the reconstruction increases. This is likely to be because including more eigentriples in the reconstruction phase results in general in a residual with a lower amplitude.



(a) ET=1-10, $\hat{\alpha}$ (b) ET=1-30, $\hat{\alpha}$ (c) ET=1-60, $\hat{\alpha}$

(d) ET=1-10, $\hat{m}$ (e) ET=1-30, $\hat{m}$ (f) ET=1-60, $\hat{m}$

Figure 5.24: Figures (a)-(c) shows $\hat{\alpha}$ against k. Figures (d)-(f) shows $\hat{m}$ against k. The sample used for estimating α and m are the residuals created by extracting reconstructions from the time series of sea level in Harlingen in the period 1987-2006. The reconstructions use eigentriples as indicated below the graphs.

Chapter 6

Conclusion

6.1 Summary

In Chapter 2 we compared two well known estimators of m and some simple estimators of m. The well known estimators were the maximum likelihood estimator, and the optimal linear estimator. The density and efficiency of the estimators were the main factors investigated. We first did this under the assumption that α was known. A second comparison of the estimators was undertaken where the value of the tail index α was unknown. Here the efficiency of the estimators was compared. In these comparisons the sample size, n was fixed.

In Chapter 3 we introduced the idea of the sample $y_1, \ldots, y_n$ as a time series. This is often a natural way to consider the sample, as can be seen in Chapter 5. Looking at the sample in this way allowed us to consider record observations. Section 3.1 was dedicated to introducing terminology, notation and some basic results related to records. No new material was present here. In Section 3.2 we introduced a way of modeling order statistics and normalized estimators at record times. This model was verified, then used in the remainder of the chapter as part of similar simulation studies as Chapter 2.

In Chapter 4 we defined two estimators of m. In Section 4.1 we introduced the so called weighted estimator (WE). Before defining and investigating this estimator, in Section 4.1.1 we demonstrated some interesting properties of the conditional ex-

pectation of waiting times (defined in Section 3.1). In Section 4.1.2 we defined the WE. The WE is a weighted average of the estimate from the current sample size and estimates at smaller sample sizes. This estimator was investigated through a simulation study using an extension to the functional model developed in Chapter 3. It was found that this estimator had smaller efficiency than the optimal linear estimator, some analysis was undertaken to explain this and suggestions for improvements were made. The second estimator introduced in this chapter assumed that the value of parameter c_0 was known (the value of α was also assumed to be known). It was shown analytically and through simulation that this extra knowledge produces dramatic improvements in efficiency for some distributions, $F(x)$.

In Chapter 5 we considered real-life applications of the theory from Chapters 2 and 3. We began by applying the record theory discussed in Section 3.1 to the number of record events that are reported in the media. In Section 5.2 we considered data of average monthly sea level in six observation stations in the Netherlands. We showed how the number of observed records can help us to identify that the time series were non-stationary. We used the CaterpillarSSA algorithm to spilt sections of the time series of mean monthly sea level in Harlingen in April into a forcastable reconstruction and a stationary residual time series. The upper and lower tail indices were estimated and the optimal linear estimator was used to predict m and M of the residual. CaterpillarSSA was used to forecast the next 12 months of mean monthly sea level for each section of data. The forecast was added to the estimates of m and M.

6.1.1 The Estimators

- The Maximum Likelihood Estimator, m^*.

  The maximum likelihood estimator m^* is the smallest solution z to the following equation:

$$(\alpha - 1)\sum_{j=1}^{k-1}\frac{(y_{k,n} - y_{j,n})}{(y_{j,n} - z)} = k$$

  where $\alpha \geq 2$.

- The Optimal Linear Estimator, m°.

  The optimal linear estimator is equal to $m^\circ = \sum_{i=1}^{k} a_i^\circ y_{i,n}$ where a° is equal to

$$a^\circ = \arg\min_{a:a'\,\mathbf{1}=1} a'\Lambda a = \frac{\Lambda^{-1}\,\mathbf{1}}{\mathbf{1}'\Lambda^{-1}\,\mathbf{1}}.$$

  Here $\Lambda = \|\lambda_{i,j}\|_{i,j=1}^{k}$ is a symmetric $k \times k$ matrix and $\lambda_{i,j}$ is defined for $i \geq j$ by (1.3.4).

- Three Non-zero Coefficients Estimator, $m^{(3)}$.

  The linear estimator $m^{(3)}$ is defined to be equal to

$$a_1^{(3)} y_{1,n} + (1 - a_1^{(3)} - a_k^{(3)}) y_{i,n} + a_k^{(3)} y_{k,n}$$

  where $a_1^{(3)}$, $a_k^{(3)}$ and i are chosen to minimize the MSE of $m^{(3)}$.

- Two Non-zero Coefficients Estimator, $m^{(2)}$.

  The linear estimator $m^{(2)}$ is defined to be

$$m_k^{(2)} = (1 + C_k) y_{1,n} - C_k y_{k,n}, \quad C_k = \frac{\lambda_{k,1} - \lambda_{1,1}}{\lambda_{1,1} - 2\lambda_{k,1} + \lambda_{k,k}}$$

  where $\lambda_{i,j}$ are as defined in (1.3.4).

- Minimum Order Statistic, $m^\bullet$.

  The minimum order statistic is the variable $y_{1,n}$.

- Weighted Estimator, $\hat{m}^{WE}$.

  The weighted estimator (or WE), $\hat{m}^{WE} = \hat{m}^{WE}_{k,n}$ is equal to

$$\hat{m}^{WE}_{k,n} = \frac{1}{T_{k,N_{k,n}} - k} \left[\sum_{j=1}^{N_{k,n}} (W_{k,j} + 1) \hat{m}_{k,T_{k,j}} \right]$$

- Estimator with c_0 known, $m^{\times}$.

  The estimator $m^{\times} = m^{\times}_{k,n}$ assumes that α and c_0 are known. It is equal to

$$m^{\times}_{k,n} = y_{k,n} - \frac{\Gamma(k + 1/\alpha)}{\Gamma(k)(c_0 n)^{1/\alpha}}.$$

6.1.2 Estimators when $k = 2$

- Maximum Likelihood Estimator, m^*_2.

  When $k = 2$, the maximum likelihood estimator is equal to

$$m^*_2 = y_{1,n} - \frac{\alpha - 1}{2}(y_{2,n} - y_{1,n}).$$

  As $y_{2,n} \geq y_{1,n}$ and $\alpha \geq 2$ we have that $m^*_2 \leq y_{1,n}$.

- Optimal linear Estimator, m°_2.

  When $k = 2$, the optimal linear estimator is defined to be

$$m^{\circ}_2 = (1 + C_2) y_{1,n} - C_2 y_{2,n}, \quad C_2 = \frac{\alpha}{2}.$$

6.1.3 Summary of Densities

In calculating the densities below we assumed that the order statistics are from a sample drawn from the Weibull distribution with parameter α. Here $\omega_3 = y_{1,n} - m$ and $\omega_4 = y_{k,n} - m$.

- The density of the normalized estimators, $(m^{(2)} - m)/(y_{1,n} - m)$ for all integer $k \geq 2$, is given by

$$p^{(2)}_3(z) = \frac{n! \alpha C_k^{\alpha}}{(n-k)!(1 + C_k - z)^{\alpha+1}} \sum_{r=0}^{k-2} \frac{(-1)^r}{(k-2-r)! r! \left[n - (k-r-1) \left(1 - \left(\frac{C_k}{1 + C_k - z} \right)^{\alpha} \right) \right]^2}$$

$$-\infty \leq z \leq 1.$$

- Asymptotically as $n \to \infty$ we have,

$$\lim_{n\to\infty} p_3^{(2)}(z) = \frac{\alpha C_k^{\alpha}}{(1+C_k-z)^{\alpha+1}}, -\infty \le z \le 1.$$

- The density of the normalized estimators, $(m^{(2)}-m)/(y_{k,n}-m)$ for all integer $k \ge 2$, is given by

$$p_4^{(2)}(z) = \frac{n!\alpha(z+C_k)^{\alpha-1}}{(n-k)!(1+C_k)^{\alpha}} \sum_{r=0}^{k-2} \frac{(-1)^r}{(k-2-r)!r!\left[n-(k-r-1)\left(1-\left(\frac{z+C_k}{1+C_k}\right)^{\alpha}\right)\right]^2}$$
$$-C_k \le z \le 1.$$

- Asymptotically as $n \to \infty$ we have,

$$\lim_{n\to\infty} p_4^{(2)}(z) = \frac{\alpha(k-1)(z+C_k)^{\alpha-1}}{(1+C_k)^{\alpha}} \left(1-\left(\frac{z+C_k}{1+C_k}\right)^{\alpha}\right)^{k-2}, -C_k \le z \le 1$$

This is equal to the density of the related functional $\zeta_k^{(2)}$.

- The density of the functional $\zeta_k^{(3)}$ (which is related to the estimator $m^{(3)}$) is given by

$$p_{\zeta_3^{(3)}}(z) = c[B(l_4(z), \alpha-1, \alpha-1) - B(l_3, \alpha-1, \alpha-1)], \quad a_3 \le z \le 1 \quad (6.1.1)$$

where

$$c = \frac{2(z-a_3)^{2\alpha-1}}{(a_2(1-a_2-a_3))^{\alpha}},$$

$$l_3 = \frac{a_2}{1-a_3},$$

$$l_4(z) = \begin{cases} 1 & \text{for } a_3 \le z < a_2+a_3 \\ \frac{a_2}{z-a_3} & \text{for } a_2+a_3 \le z \le 1 \end{cases}$$

and $B(z,a,b)$ is the incomplete beta function:

$$B(z,a,b) = \int_0^z u^{a-1}(1-u)^{b-1}du.$$

z	$p(z)$	$\lim_{n\to\infty} p(z)$	Support
$(m^{(2)}-m)/\omega_3$	$\frac{n(n-1)\alpha C_k^\alpha}{(1+C_2-z)^{\alpha+1}}\left(n-1+\left(\frac{C_2}{1+C_2-z}\right)^\alpha\right)^{-2}$	$\frac{\alpha C_2^\alpha}{(1+C_2-z)^{\alpha+1}}$	$-\infty < z \leq 1$
$(m^\circ-m)/\omega_4$	$\frac{n(n-1)\alpha(z+C_2)^{\alpha-1}}{(1+C_2)^\alpha}\left(n-1+\left(\frac{z+C_2}{1+C_2}\right)^\alpha\right)^{-2}$	$\frac{\alpha(z+C_2)^{\alpha-1}}{(1+C_2)^\alpha}$	$-C_2 < z \leq 1$
$(m^*-m)/\omega_3$	$\frac{2n(n-1)\alpha(\alpha-1)^\alpha}{(\alpha+1-2z)^{\alpha+1}}\left(n-1+\left(\frac{\alpha-1}{\alpha+1-2z}\right)^\alpha\right)^{-2}$	$\frac{2\alpha(\alpha-1)^\alpha}{(\alpha+1-2z)^{\alpha+1}}$	$-\infty < z < 1$
$(m^*-m)/\omega_4$	$\frac{2n(n-1)\alpha(2z+\alpha-1)^{\alpha-1}}{(\alpha+1)^\alpha}\left(n-1+\left(\frac{2z+\alpha-1}{\alpha+1}\right)^\alpha\right)^{-2}$	$\frac{2\alpha(2z+\alpha-1)^{\alpha-1}}{(\alpha+1)^\alpha}$	$-\frac{\alpha-1}{2} < z < 1$
ζ°	$\frac{\alpha(z+C_2)^{\alpha-1}}{1+C_2}$	-	$-C_2 < z < 1$
ζ^*	$\frac{2\alpha(2z+\alpha-1)^{\alpha-1}}{(2+\alpha-1)^\alpha}$	-	$-\frac{\alpha-1}{2} < z < 1$

Table 6.1: Densities of estimators and functionals when $k = 2$.

6.1.4 Efficiency

- The asymptotic efficiency when $k = 2$ of estimator $m^\circ(\vartheta)$ with respect to $m^\circ(\alpha)$ is given by the following.

$$\text{eff}(m^\circ(\vartheta)) = \frac{\alpha+2}{\alpha+2+\alpha(1-\frac{\vartheta}{\alpha})}.$$

  The analytic efficiency of $m^\circ(\alpha)$ is $\text{eff}(m^\circ(\alpha)) = 1$.

- The asymptotic efficiency when $k = 2$ of estimator $m^*(\vartheta)$ with respect to $m^\circ(\alpha)$ is given by the following.

$$\text{eff}(m^*(\vartheta)) = \frac{\alpha(\alpha+2)}{2\alpha^2+4\alpha-2\alpha\vartheta-2\vartheta+\vartheta^2+1}. \tag{6.1.2}$$

- The efficiency (6.1.2) is maximized at $\vartheta = \alpha+1$, where it is equal to 1.

- The asymptotic efficiency when $k = 2$ of $m^*(\alpha)$ with respect to $m^\circ(\alpha)$ is given by

$$\text{eff}(m^*(\vartheta)) = \frac{\alpha(\alpha+2)}{\alpha(\alpha+2)+1}.$$

- The asymptotic efficiency of $m^{\bullet}$ when $k = 2$ is given by

$$\text{eff}(m^{\bullet}) = \frac{\alpha + 2}{2(\alpha + 1)}.$$

- The asymptotic efficiency of $m^{(2)}(\vartheta)$ with respect to $m^{\circ}(\alpha)$ for any integer k such that $k \geq 2$ is given by

$$\text{eff}(m^{(2)})(\vartheta) = \frac{\mathbf{1}'_{(2)}\Lambda^{-1}_{(2)}\mathbf{1}_{(2)}}{\mathbf{1}'\Lambda^{-1}\mathbf{1}}.$$

Here $\mathbf{1}_{(2)} = (1, 1)'$ and

$$\Lambda_{(2)} = \begin{pmatrix} \lambda_{11} & \lambda_{k1} \\ \lambda_{k1} & \lambda_{kk} \end{pmatrix}. \tag{6.1.3}$$

- The asymptotic efficiency of $m^{\times}_{k,n}$ with respect to $m^{\circ}_{k,n}$ is given by

$$\text{eff}(m^{\times}_{k,n}) = \frac{(\Gamma(k))^2}{\Gamma(k)\Gamma(k + 2/\alpha) - (\Gamma(k + 1/\alpha))^2 \mathbf{1}'\Lambda^{-1}\mathbf{1}}. \tag{6.1.4}$$

6.1.5 Transition Densities

- The transition density of $x^{(t)}_1$ when $k = 2$ is given by

$$p(y; x) = \begin{cases} x, & 0 \leq y \leq x \\ x + \frac{x}{y^2}, & x \leq y \leq 1 \end{cases}$$

- The transition density of ζ_k for $k = 2$ is given by

$$p_{\zeta}(z; \zeta) = \begin{cases} \frac{\alpha(z - a_2)^{\alpha - 1}(\zeta - a_2)^{\alpha}}{a_1^{2\alpha}}, & a_2 \leq z \leq \zeta \\ \frac{\alpha(z - a_2)^{\alpha - 1}(\zeta - a_2)^{\alpha}}{a_1^{2\alpha}} + \frac{\alpha(\zeta - a_2)^{\alpha}}{(z - a_2)^{\alpha + 1}}, & \zeta < z \leq 1 \end{cases}$$

6.1.6 Autocorrelation Function

- The autocorrelation function of the 1-dimensional Markov chain X_1 (where $k = 2$) is $R_x(l) = \frac{1}{3^l}$, where l is lag.

- The autocorrelation function of any linear functional when $k = 2$, ζ_2, is given by $R_{\zeta}(l) = \left(\frac{\alpha}{1 + 2\alpha}\right)^l$.

6.2 Conclusions

In this final section we collect the main results of this thesis. We make concluding remarks regarding the strengths and weaknesses of the work carried out. We also reflect on the implications of the results and any further work that could be carried out.

6.2.1 Densities and Histograms

Investigations into the distributions of the estimators, such as deriving densities analytically and plotting histograms, was useful as a first step. It enabled basic properties of the estimators to be found, such as their mean and variance for finite samples. Deriving the densities also acted as verification of results from simulation studies. Further work could involve the derivation of densities of estimators when $F(x)$ is not the c.d.f. of the Weibull distribution.

Results

- If $F_a(x;\alpha) = c_0(x-m)^\alpha$ then for $\alpha > 0$ and x in the neighborhood of m
  - when $\beta = 1$, $F_\beta(x;\alpha,1) = F_a(x;\alpha)$;
  - when $\beta > 1$, $F_\beta(x;\alpha,1) < F_a(x;\alpha)$;
  - when $\beta < 1$, $F_\beta(x;\alpha,1) > F_a(x;\alpha)$;
  - $\Psi_\alpha(x) < F_a(x;\alpha)$.
- Normalized linear estimators, $\hat{m}_{k,n}$, converge quickly to their asymptotic distribution when $F(x)$ is Weibull as $n \to \infty$.
- The order statistics from a sample drawn from the Weibull distribution converge quickly to their asymptotic distribution as $n \to \infty$.
- The order statistics from a sample drawn from $F_\beta(x;\alpha,1)$ converge quickly to their asymptotic distribution as $n \to \infty$ for $\alpha > 0$.

- The derived densities from Sections 2.2.4 and 2.2.6 were found to be a good fit to simulated data.
- Estimators normalized in different ways had very different distributions.
- The different estimators had similar distribution:
  - When $\alpha = 1$, $m^{\circ}_{k,n} = m^{(2)}_{k,n} = m^{(3)}_{k,n}$.
  - The skewness of the distribution of each estimator is similar.
  - The mean and variance of the distribution of each estimator is different.
  - The $(m^* - m)/(y_{k,n} - m)$ has a smaller variance than $(m^\circ - m)/(y_{k,n} - m)$.
  - $(m^\circ - m)/(y_{k,n} - m)$ has a smaller variance than $(m^* - m)/(y_{k,n} - m)$.
- As k increases the mean and variance of $(m^\circ - m)/(y_{k,n} - m)$ and $(m^* - m)/(y_{k,n} - m)$ both decrease.
- As α increases the mean of $(m^* - m)/(y_{k,n} - m)$ remains fairly constant.
- As α increases the mean of $(m^\circ - m)/(y_{k,n} - m)$ increases.
- As α increases the variance of estimators increases.

6.2.2 Efficiency, Mean Square Error and Bias

The results regarding the efficiency of the estimators showed that the MLE is not as efficient as the optimal linear estimator for estimating the endpoint of many of the distributions tested. The Weibull distribution is a famous case of where the MLE is poor due to the discontinuity of the likelihood function at m. See for example [6]. However, it is a standard distribution and therefore poor behavior is a significant problem.

Results

- The MSE of the estimator $m^{(2)} = a_1^{(2)} y_{1,n} + a_k^{(2)} y_{k,n}$ is greater than the estimator $m_{k,n}^{\Delta_i} = a_i^{\Delta_i} y_{i,n} + a_k^{\Delta_i} y_{k,n}$, where $i = 1, 2, \ldots, k-1$ (see Figure 2.1) for some α and k.

- For $F(x)$ Weibull:

  - For small α and for small k the MLE has smaller efficiency and larger bias than m°.

  - The efficiency and bias of $m^{(2)}$ and $m^{(3)}$ are better than the MLE for some small α and small k.

- For $F(x) = F_\beta(x; \alpha, 1)$:

  - The efficiency of the optimal linear estimator is very close to 1.

  - For small α and for small k the optimal linear estimator outperforms the MLE.

- When $F(x)$ is equal to (1.1.2):

  - When n is equal to 1 000 and k is large, the efficiency of all estimators is poor.

  - When n is equal to 10 000 the efficiency of the optimal linear estimator is approximately constant (and larger than 1) for all k.

  - The efficiency of $m^{(3)}$ and $m^{(2)}$ decreases with k.

  - The efficiency of $m^{(3)}$ and $m^{(2)}$ remain larger than the MLE even for large k.

- The results from this section were verified using the functional model.

6.2.3 Wrong α

The investigation into using the wrong value of tail index was useful and interesting. It is important to know how the estimators react to using the wrong value of tail index. The main results from this section were that the MLE achieved maximum efficiency when the wrong value of tail index, ϑ^*, was used and that $\alpha \leq \vartheta^* \leq \alpha + 1$. This result could be used to improve the maximum likelihood estimator. Further work could be to find the exact value of ϑ^* for different values of k. This may not be possible, in which case upper and lower bounds could perhaps be found. It would also be useful to find the reason behind this interesting behavior.

Results

- For estimators m°, $m^{(2)}$, $m^{(3)}$ and m^*, $\text{eff}(\hat{m}(\vartheta)) > \text{eff}(m^\bullet(\vartheta))$ for a wide range of ϑ.
- The range of ϑ for which the estimators were more efficient than $y_{1,n}$ was larger if k was small.
- The range of ϑ for which the estimators were more efficient than $y_{1,n}$ was larger if α was large.
- For the MLE, the value of ϑ that made the estimator most efficient was ϑ^* where $\alpha < \vartheta^* \leq \alpha + 1$.
- When $k = 2$, $\vartheta^* = \alpha + 1$.
- These results were verified using the functional model.

6.2.4 Waiting times

In this thesis a literature review concerning waiting times was conducted. This has since been extended into a paper (submitted) on stopping rules for k-adaptive search algorithms. Here the expected waiting time was combined with the expectation of

the ratio $\frac{y_{k,T_{k,t+1}}-y_{k,T_{1,t+1}}}{y_{k,T_{k,t}}-y_{k,T_{1,t}}}$ (related to the quality of linear estimators) to produce a stopping rule.

Results

- The expectation of $W_{1,t}$, is infinite when $t > 1$.
- The expectation of $W_{k,t}$, is finite when $k > 1$.
- The expectation of $T_{k,t+1} - n$ (where $T_{k,t} \leq n < T_{k,t+1}$) increases linearly with n.
- The conditional distribution of $W_{k,t}$ given $y_{k,T_{k,t}}$ can be used to model waiting times.

6.2.5 Simulating order statistics at record times

Defining the Markov chain equal in distribution to (normalized) order statistics at records was extremely important to Chapters 3 and 4. It enabled record values to be modeled at sample sizes that otherwise would be been way beyond the capability of the available computing power.

Results

The simulation of order statistics at record values featured in Chapters 3 and 4. It was used to verify many of the results listed in Sections 6.2.1, 6.2.2, and 6.2.3, and played an integral part in determining results listed in Sections 6.2.6 and 6.2.7

6.2.6 The Weighted Estimator

The weighted estimator that we defined and investigated was unsuccessful as it failed to improve on the MSE of the existing linear estimators. However, it is likely that further work in this area could produce an estimator with smaller MSE than existing estimators. In the work presented here, suggestions of alternative weighted

estimators were made based on heuristic arguments. Further work could involve deriving an estimator by considering the expected improvement in a linear estimator as sample size increases.

Results

- The WE has a smoother trajectory than m°.
- The efficiency and bias of the WE are worse than m°.
- Estimator quality and $W_{k,c}$ are not correlated at record number c, where c is such that $T_{k,c+1} < 1\,000$ and $T_{k,c+2} \geq 1\,000$.
- Estimators improve with sample size, record number and waiting time.

6.2.7 The estimator $m^{\times}_{k,n}$

The work conducted concerning $m^{\times}_{k,n}$ showed that for a variety of c.d.f.s, when the parameter c_0 is known even a very simple estimator can be more efficient than the optimal linear estimator.

Results

- When $\alpha \geq 2$ the asymptotic analytic efficiency of the estimator $m^{\times}_{k,n}$ with respect to $m^{\circ}_{k,n}$ is much greater than 1.
- When $\alpha = 1$ the asymptotic analytic efficiency of the estimator $m^{\times}_{k,n}$ with respect to $m^{\circ}_{k,n}$ is much less than 1.
- As k increases the asymptotic analytic efficiency of $m^{\times}_{k,n}$ with respect to $m^{\circ}_{k,n}$ decreases.
- As α increases the asymptotic analytic efficiency of $m^{\times}_{k,n}$ with respect to $m^{\circ}_{k,n}$ increases.

- When $F(x)$ is either the Weibull or the beta distribution and $\alpha = 3$, the estimated efficiency (from simulation) of $m_{k,n}^{\times}$ is much greater than $m_{k,n}^{\circ}$.
- When $F(x)$ is the gamma distribution the estimated efficiency of $m_{k,n}^{\times}$ is greater than $m_{k,n}^{\circ}$ for small k.
- When $F(x)$ is the uniform distribution, the estimated efficiency (from simulation) of $m_{k,n}^{\times}$ is much less than $m_{k,n}^{\circ}$.
- The estimated efficiency (from simulations) differed greatly from the analytic asymptotic efficiency when the distribution $F(x)$ was very different from $F_a(x) = c_0(x - m)^{\alpha}$.

6.2.8 Meteorological Applications

Section 5.1 shows that, with care, the number of observed (maximal and minimal) records in a time series could be used to hypothesize that the time series is stationary. However, it also discussed that it is possible to observe the expected number of maximal and minimal records in a time series with periodicities.

Results

- If the elements of climate (e.g. average rainfall) are assumed to form a stationary time series, the expected number of 'news-worthy' records is high.
- Trend, periodicities, increased variation are factors that affect the number of records observed in a time series.
- In data of mean monthly sea level from six different observation stations in The Netherlands,
  - there is high correlation between locations;
  - correlation between months of the year at any particular location is fairly low.

- Consider the estimated number of records in the $6 \times 12 = 72$ continuous time series of mean monthly sea level at six measuring stations in The Netherlands.
  - The expected number of type 1 1^{st} records in each time series (if time series were stationary) is 5.4.
  - The number of record highs was higher than expected in 69 of the 72 time series.
  - The number of record lows was lower than expected in 47 of the 72 time series.

6.2.9 Using CaterpillarSSA

CaterpillarSSA is an excellent tool for time series analysis as it allows the expert user to consider the physical mechanisms behind the behavior of the time series. In this thesis greater consideration should have been given to the choice of eigentriples in order to draw robust conclusions about the maximum and minimum sea level. The estimation of the tail index is a difficult problem that also needs greater consideration.

It is possible that change points in the time series of sea level could be better dealt with by using climate models to explain the changes in behavior. The time series could be split into two reconstructions and the residual noise. The first reconstruction should be explainable and forcastable using climate models. The second reconstruction should be forecastable using SSA (i.e. not contain change points). The endpoints of the distribution of the residual could then be estimated using extreme value estimators.

Results

- Removing the trend and periodicities from the time series of mean monthly sea level in Harlingen (using CaterpillarSSA), and then selecting just the data

relating to April was not as effective as just using data from April at the reconstruction stage.

- Using more eigentriples in the reconstruction produced residuals with smaller variances and whose record counts were closer to those expected in a stationary time series.
- The residual when just data from April is used to create the reconstruction have smaller variance than the residuals when all data is used to create the reconstruction and then just the April data is selected (see Figures 5.16-5.17).
- Some structure had been left in the residuals (shown in Figures 5.12-5.13) after removing the reconstructions.
- There is at least one change point in the time series of mean monthly sea level in Harlingen in April from 1885 to 2006.
- The final upper and lower forecasts (reconstructed forecast plus M and reconstructed forecast plus m respectively) varied a great deal depending on the number of eigentriples used.
- The value of the parameter k affected the estimation of α, α^u, m and M a great deal.

Bibliography

[1] M Ahsanullah, *Record values*, The exponential distribution (N. Balakrishnan and Asit P. Basu, eds.), Gordon and Breach Publishers, Amsterdam, 1995, Theory, methods and applications, pp. 279–296.

[2] J. Beirlant, Y. Goegebeur, J. Teugels, and J. Segers, *Statistics of extremes*, John Wiley & Sons Ltd., Chichester, 2004.

[3] R. E. Benestad, *Record-values, nonstationarity tests and extreme value distributions*, Global and Planetary Change **44** (2004), no. 12, 11–26.

[4] Jan Beran, *Statistics for long-memory processes*, Monographs on Statistics and Applied Probability, vol. 61, Chapman and Hall, New York, 1994.

[5] J. Bunge and C. M. Goldie, *Record sequences and their applications*, Stochastic processes: theory and methods, Handbook of Statist., vol. 19, North-Holland, Amsterdam, 2001, pp. 277–308.

[6] R. C. H. Cheng and L. Traylor, *Non-regular maximum likelihood problems*, J. Roy. Statist. Soc. Ser. B **57** (1995), no. 1, 3–44, With discussion and a reply by the authors.

[7] P. Cooke, *Statistical inference for bounds of random variables*, Biometrika **66** (1979), no. 2, 367–374.

[8] ———, *Optimal linear estimation of bounds of random variables*, Biometrika **67** (1980), no. 1, 257–258.

[9] S. Csörgő and D. M. Mason, *Simple estimators of the endpoint of a distribution*, Extreme value theory (Oberwolfach, 1987), Lecture Notes in Statist., vol. 51, Springer, New York, 1989, pp. 132–147.

[10] H. A. David and H. N. Nagaraja, *Order statistics*, third ed., Wiley Series in Probability and Statistics, Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, 2003.

[11] L. de Haan, *Estimation of the minimum of a function using order statistics*, J. Amer. Statist. Assoc. **76** (1981), no. 374, 467–469.

[12] M. J. Dixon and J. A. Tawn, *The effect of non-stationarity on extreme sea-level estimation*, J. Roy. Statist. Soc. Ser. C. **48** (1999), no. 2, 135–151.

[13] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modelling extremal events for insurance and finance*, Springer-Verlag, Berlin, 2003.

[14] F. G. Foster and A. Stuart, *Distribution-free tests in time-series based on the breaking of records*, J. Roy. Statist. Soc. Ser. B. **16** (1954), 1–13; discussion 13–22.

[15] J. Galambos, *The asymptotic theory of extreme order statistics*, second ed., Robert E. Krieger Publishing Co. Inc., Melbourne, FL, 1987.

[16] Ned Glick, *Breaking records and breaking boards*, Amer. Math. Monthly **85** (1978), no. 1, 2–26.

[17] N. Golyandina, V. Nekrutkin, and A. Zhigljavsky, *Analysis of time series structure*, Monographs on Statistics and Applied Probability, vol. 90, Chapman & Hall/CRC, Boca Raton, FL, 2001, SSA and related techniques.

[18] E. J. Gumbel, *Statistics of extremes*, Columbia University Press, New York, 1958.

[19] P. Hall, *On estimating the endpoint of a distribution*, Ann. Statist. **10** (1982), no. 2, 556–568.

[20] E. Hamilton, V. Savani, and A Zhigljavsky, *Estimating the minimal value of a function in global random search: Comparison of estimation procedures*, Models and Algorithms for Global Optimization, Springer Optimization and Its Applucations, vol. 4.

[21] H. Hassani, *Singular spectrum analysis: Methodology and comparison*, Journal of Data Science **5** (2007), no. 2, 239–257.

[22] B. M. Hill, *A simple general approach to inference about the tail of a distribution*, Ann. Statist. **3** (1975), no. 5, 1163–1174.

[23] Z. Ignatov, *Ein von der Variationsreihe erzeugter Poissonscher Punktprozeß*, Annuaire Univ. Sofia Fac. Math. Méc. **71** (1976/77), no. 2, 79–94 (1986).

[24] ———, *Point processes generated by order statistics and their applications*, Point processes and queuing problems (Colloq., Keszthely, 1978), Colloq. Math. Soc. János Bolyai, vol. 24, North-Holland, Amsterdam, 1981, pp. 109–116.

[25] S. Kotz and S. Nadarajah, *Extreme value distributions*, Imperial College Press, London, 2000.

[26] H. N. Nagaraja, *Record values and related statistics. A review*, Comm. Statist. Theory Methods **17** (1988), no. 7, 2223–2238.

[27] V. B. Nevzorov, *Records: mathematical theory*, Translations of Mathematical Monographs, vol. 194, American Mathematical Society, Providence, RI, 2001, Translated from the Russian manuscript by D. M. Chibisov.

[28] J. Pintér, *Global optimization in action*, Nonconvex Optimization and its Applications, vol. 6, Kluwer Academic Publishers, Dordrecht, 1996, Continuous and Lipschitz optimization: algorithms, implementations and applications.