Parameter Estimation and Prediction Interval Construction for Location–Scale Models with Nuclear Applications

By

Xingli Wei

A Dissertation

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree of

Doctor of Philosophy

McMaster University

@Copyright by Xingli Wei, July, 2014

Ph.D Math(Stats) McMaster University

Hamilton, Ontario

TITLE: Parameter Estimation and Prediction Interval Construction for Location–Scale Models with Nuclear Applications

AUTHOR: Xingli Wei

SUPERVISOR: Dr. Fred M. Hoppe

Contents

List of Figures **vi**

Acknowledgement **viii**

Abstract **ix**

1 Introduction **1**
1.1 Motivation 1
1.2 Literature Review and Preliminaries 4
1.3 Thesis Outline and Contributions 11

2 A Simple Method for Estimating Parameters of Location-Scale Families **13**
2.1 Introduction 13
2.2 Estimation Based on the Sample Mean and Sample Standard Deviation 16
2.2.1 Equivariance Principle 16
2.2.2 Unbiasedness Principle 18
2.2.3 Minimum Risk Principle 23
2.3 Estimation Based on the Sample Mean and Gini's Mean Difference . 31
2.4 An Adaptive Unbiased Estimator for Location-Scale Families 33

3 A Mixed Method for Estimating Location, Scale and Shape Parameters of a 3-Parameter Distribution **41**
3.1 Mixed Method 41
3.2 Example 1: Generalized Extreme Value (GEV) Distribution 43
3.2.1 Introduction 43
3.2.2 Methods of Estimation for the Generalized Extreme Value Distribution 46
3.2.3 Simulation Results and Conclusions 52
3.3 Example 2: Power Gumbel Distribution 56
3.3.1 Maximum Likelihood Method 56
3.3.2 Mixed Method 58
3.3.3 Simulation Results and Conclusions 61

4 Bayesian Estimation for the Gumbel Minimum Distribution **69**
4.1 Bayesian Method 69
4.2 Full Bayesian Estimation for Gumbel Minimum Distribution 72
4.2.1 An Approximation Theorem for an Exponential Sum of Gumbel Random Variables 72
4.2.2 Priors for the Gumbel Minimum Distribution 77

4.2.3 Bayesian Estimators Based on Jeffrey's Non-informative Prior 80
4.2.4 Bayesian Estimators Based on an Asymptotically Conjugate Prior 83
4.3 Bayesian-Maximum Likelihood Estimation for the Gumbel Minimum Distribution . . . 86
4.3.1 Estimation for the Gumbel Minimum Distribution Using a Non-informative Prior . . . 86
4.3.2 Estimation for the Gumbel Maximum Distribution Using a Noninformative Prior . . . 89
4.4 Another Example: Estimation for the Normal Distribution . . . 93

5 Frequentist Prediction Interval for a Location–Scale Family 97
5.1 Introduction . . . 97
5.2 Construction of a Conditional Prediction Interval . . . 98
5.3 Construction of an Unconditional Prediction Interval . . . 102
5.4 Construction of Prediction Intervals Using a Monte Carlo Method . . . 104

6 Bayesian Prediction Intervals for the Gumbel Minimum Distribution 107
6.1 Construction of a Bayesian Prediction Interval Using a Noninformative Prior on a when b is Known . . . 108
6.1.1 Prediction Interval for Y_{max} . . . 108
6.1.2 Prediction Intervals for $Y_{(j)}$ . . . 111
6.2 Construction of a Bayesian Prediction Interval Using a Conjugate Prior for a with Known b . . . 114
6.2.1 Prediction Interval for Y_{max} . . . 116
6.2.2 Prediction Interval for $Y_{(j)}$ . . . 119
6.3 Constructing Bayesian Prediction Intervals Using Gibbs Sampling . . . 121
6.3.1 Case 1: a Unknown and b Known . . . 124
6.3.2 Case 2: Both a and b Unknown . . . 128

7 Comparative Study of Parameter Estimation and Prediction Intervals for the Gumbel Distribution 135
7.1 Introduction . . . 135
7.2 Methods of Parameter Estimation for the Gumbel Distribution . . . 137
7.3 Simulation Results and Conclusions . . . 148
7.4 Methods of Prediction Interval Construction for the Gumbel Minimum Distribution . . . 153
7.4.1 The Approximation for the Distribution . . . 153
7.4.2 Conditional and Unconditional Probability Method . . . 156
7.4.3 Monte Carlo Estimation of the Distribution Percentiles and Method 1 . . . 158
7.4.4 Markov Chain Monte Carlo Estimation of the Distribution Percentiles and Method 2 . . . 158

7.5 Numerical Examples 159

8 Resampling and Prediction Interval Construction 173

8.1 Introduction 173
8.2 Bootstrap Resampling 174
8.3 A New Resampling Method 176
8.4 Another Look at the Bootstrap Method 179
8.5 Smoothed Resampling 180
8.6 A Simple Version of the Generalized Bootstrap 183
8.7 Numerical Examples 188
8.8 Wilks' Formula for Symmetric Distributions in the GRS Method . . 205
8.8.1 Symmetrization 207
8.8.2 Comparison with the Generalized Bootstrap 212

A Useful Formulas 219

B Bibliography 221

List of Figures

1.1 Pickering A Feeder Pipes 2

3.1 Location estimator of GEV with $a = 12, b = 5, \kappa = 0.25, 10000$ trials . 53
3.2 Scale estimator of GEV with $a = 12, b = 5, \kappa = 0.25, 10000$ trials . . . 53
3.3 Shape estimator of GEV with $a = 12, b = 5, \kappa = 0.25, 10000$ trials . . 54
3.4 Location estimator of GEV with $a = 12, b = 5, \kappa = -0.25, 10000$ trials 54
3.5 Scale estimator of GEV with $a = 12, b = 5, \kappa = -0.25, 10000$ trials . . 55
3.6 Shape estimator of GEV with $a = 12, b = 5, \kappa = -0.25, 10000$ trials . 55
3.7 Distribution function and density of power Gumbel 60
3.8 Location estimator of power Gumbel with $a = 5, b = 12, r = 5$ 67
3.9 Scale estimator of power Gumbel with $a = 5, b = 12, r = 5$ 67
3.10 Shape estimator of power Gumbel with $a = 5, b = 12, r = 5$ 67
3.11 Location estimator of power Gumbel with $a = 1, b = 5, r = 5$ 68
3.12 Scale estimator of power Gumbel with $a = 12, b = 5, r = 5$ 68
3.13 Shape estimator of power Gumbel with $a = 12, b = 5, r = 5$ 68

7.1 Bias of location estimator of Gumbel 151
7.2 MSE of location estimator of Gumbel 151
7.3 Bias of scale estimator of Gumbel 152
7.4 MSE of scale estimator of Gumbel 152
7.5 Difference between two quantiles for $n = 23, m = 8, j = 1$ with 2000000 simulations 160
7.6 Difference between two quantiles for $n = 23, m = 8, j = 3$ with 2000000 simulations 161
7.7 Difference between two quantiles for $n = 23, m = 8, j = 8$ with 2000000 simulations 161
7.8 Predictive distribution of Gumbel 167

8.1 Scatter plot of standard normal distribution 210
8.2 Scatter plot of standard Cauchy distribution 210
8.3 Scatter plot of Laplace distribution 210
8.4 Scatter plot of Student distribution with df=5 211
8.5 Scatter plot of Student distribution with df=10 211
8.6 Scatter plot of logistic distribution 211
8.7 Scatter plot of uniform distribution 212
8.8 Estimated Probability of $P(W > \chi_\gamma)$ with $P(X_1 \leq \chi_\gamma) = 0.95$ and $X_1 \sim Uniform(-1, 1)$ 214
8.9 Estimated Probability of $P(W > \chi_\gamma)$ with $P(X_1 \leq \chi_\gamma) = 0.95$ and $X_1 \sim$ *logistic distribution with* $a = 0$ $b = 5$ 214

8.10 Estimated Probability of $P(W > \chi_\gamma)$ with $P(X_1 \leq \chi_\gamma) = 0.95$ and $X_1 \sim Norm(0,1)$ 215
8.11 Estimated Probability of $P(W > \chi_\gamma)$ with $P(X_1 \leq \chi_\gamma) = 0.95$ and $X_1 \sim Laplace(0,1)$ 216
8.12 Estimated Probability of $P(W > \chi_\gamma)$ with $P(X_1 \leq \chi_\gamma) = 0.95$ and $X_1 \sim t - distribution\ with\ df = 1$ 216
8.13 Estimated Probability of $P(W > \chi_\gamma)$ with $P(X_1 \leq \chi_\gamma) = 0.95$ and $X_1 \sim t - distribution\ with\ df = 5$ 217
8.14 Estimated Probability of $P(W > \chi_\gamma)$ with $P(X_1 \leq \chi_\gamma) = 0.95$ and $X_1 \sim t - distribution\ with\ df = 10$ 217

Acknowledgement

This thesis would not be finished in time and shape without the help and support of my supervisor Dr. Fred M. Hoppe. My greatest thanks go to him for his supporting, guiding, and working very closely with me for the last few years. I feel tremendously lucky to have had the opportunity to work with him. I am very grateful for his patience, encouragement, criticism, persistence, and friendly relationship with me.

A very special thanks goes to the other members of my committee: Dr. Alexander Rosa and Dr Paul Sermer. They were always available for my questions, and their suggestions and comments helped me a lot to stay on my path and to do the right thing. I would like to thank them for their comments, suggestions and review that helped to produce this document.

Finally, I would like to thank my family, Hua Zheng, and Cynthia Wei, for their love and support. This dissertation is dedicated to them.

Abstract

This thesis presents simple efficient algorithms to estimate distribution parameters and to construct prediction intervals for location–scale families. Specifically, we study two scenarios: one is a frequentist method for a general location–scale family and then extend to a 3–parameter distribution, another is a Bayesian method for the Gumbel distribution. At the end of the thesis, a generalized bootstrap resampling scheme is proposed to construct prediction intervals for data with an unknown distribution.

Our estimator construction begins with the equivariance principle, and then makes use of unbiasedness principle. These two estimates have closed form and are functions of the sample mean, sample standard deviation, sample size, as well as the mean and variance of a corresponding standard distribution. Next, we extend the previous result to estimate a 3-parameter distribution which we call a mixed method. A central idea of the mixed method is to estimate the location and scale parameters as functions of the shape parameter.

The sample mean is a popular estimator for the population mean. The mean squared error (MSE) of the sample mean is often large, however, when the sample size is small or the scale parameter is greater than the location parameter. To reduce the MSE of our location estimator, we introduce an adaptive estimator. We will illustrate this by the example of the power Gumbel distribution.

The frequentist approach is often criticized as failing to take into account the uncertainty of an unknown parameter, whereas a Bayesian approach incorporates such uncertainty. The present Bayesian analysis for the Gumbel data is achieved numerically as it is hard to obtain an explicit form. We tackle the problem by providing an approximation to the exponential sum of Gumbel random variables.

Next, we provide two efficient methods to construct prediction intervals. The first one is a Monte Carlo method for a general location-scale family, based on our previous parameter estimation. Another is the Gibbs sampler, a special case of Markov Chain Monte Carlo. We derive the predictive distribution by making use of an approximation to the exponential sum of Gumbel random variables .

Finally, we present a new generalized bootstrap and show that Efron's bootstrap re-sampling is a special case of the new re-sampling scheme. Our result overcomes the issue of the bootstrap of its "inability to draw samples outside the range of the original dataset." We give an applications for constructing prediction intervals, and simulation shows that generalized bootstrap is better than that of the bootstrap when the sample size is small. The last contribution in this thesis is an improved GRS method used in nuclear engineering for construction of non-parametric tolerance intervals for percentiles of an unknown distribution. Our result shows that the required sample size can be reduced by a factor of almost two when the distribution is symmetric. The confidence level is computed for a number of distributions and then compared with the results of applying the generalized bootstrap. We find that the generalized bootstrap approximates the confidence level very well.

Chapter 1

Introduction

1.1 Motivation

CANDU, which stands for Canadian Deuterium Uranium reactor, is a registered trademark of Atomic Energy of Canada Limited (AECL). The feeder piping system in a CANDU reactor is a main component of the Primary Heat Transport System (PHTS) that carries the primary coolant between the reactor and the steam generators. This system is composed of an outlet section and inlet section and Figure 1.1 shows some Pickering Nuclear Generating Station (PNGS) A feeder pipes.

The feeder piping system is exposed to a corrosive environment, and thus it is part of the most sensitive structural elements of the power plant. The flow accelerated corrosion (FAC) process is influenced by flow turbulence, temperature, chemical environment, and the geometry of the pipe section. Because of this, different pipes experience different rates of corrosion. It has been observed that the outlet feeders, the first pipe weld in the reactor outlet end fitting, may encounter considerable wall thinning occurring on the inside surface of the feeder piping. Gradual wall thinning can cause the pipes to leak, and may lead to a hazardous situation. Hence, it is crucial to inspect the wall thinning of piping systems.

In the middle of the 1990s, wall thinning was postulated as a degradation mechanism. Critical sections of the pipes in PHTS are periodically inspected to measure their wall thickness using ultrasonic probes. Such inspections are also intended

to identify pipes that are near the end of life, so that their replacement in the near future can be planned.

At the end of 2004 Ontario Power Generation (OPG) discovered that the walls of some pipes at Pickering NGS A Unit 1 were becoming perilously thin. As a result, Unit 4 was shut down in order to perform an inspection on April 2, 2005.

The outlet feeder have been categorized by OPG into 4 groups based on similarity of shape, diameter, and required thickness. Figure 1.1 shows some Pickering A feeder pipes.

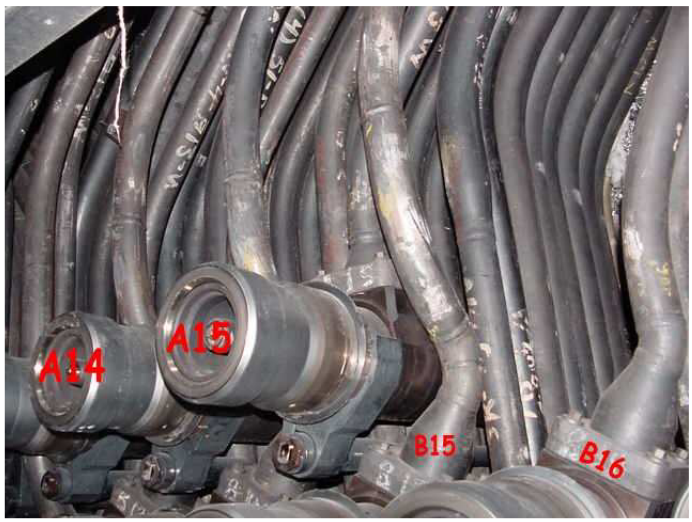


Figure 1.1: Pickering A Feeder Pipes

Ultrasonic examination were performed at the pipe side of the grayloc weld where the more severe thinning was expected to be located based on previous experience. Two kinds of instruments were used to scan over 360^0 circumference of the pipes. The minimum wall thicknesses were recorded. The following table shows the number of feeders and number of inspected feeders within each group for feeders undergoing planned inspection.

Table 1.1: Grouping of Outlet Feeder Pipe Categories

Group	Combined Feeder Categories	No. of Outlet Pipes	No. of Inspected Pipes
1	13, 15, 16, 16A, 18	36	23
2	5, 7, 17, 17B, 19, 436, HUB	52	16
3	10	200	16
4	8, 9, 10A	46	14

OPG's objective was to determine what sample size for inspection would be acceptable in order to obtain sufficient confidence that the minimum wall thickness of pipes was not less than a prescribed minimum acceptable thickness. It was agreed that the appropriate statistical tool was a one–sided prediction interval for future minimum observations. If the lower limit of the prediction interval was found to be greater than the required minimum thickness, then no additional sampling of the uninspected feeder pipes would be needed.

Hoppe (unpublished) pointed out that feeder pipes thickness obey the normal distribution which belongs to the exponential family. Thus the minimum values of wall thicknesses of feeder pipes may reasonably be expected to follow a Gumbel minimum distribution. Hoppe constructed a prediction interval using the method of Hahn [50, 51] which is designed for normal data.

Hoppe and Fang [58] proposed a hybrid method to construct Bayesian prediction intervals for the minimum of any specified number of future measurements from a Gumbel distribution based on previous observations. In their approach, they combined Bayesian ideas with maximum likelihood, by considering a noninformative prior on the location and then estimating the scale by maximum–likelihood method.

In this thesis, we will give more detailed discussion of construction prediction intervals, including parametric methods (frequentist and Bayesian approaches) and non-parametric methods.

1.2 Literature Review and Preliminaries

This section provides a review of the relevant literature as well as some statistical methods and techniques that will be needed in subsequent chapters.

Extreme value distributions arise as limiting distributions, as the sample size increases, for maxima or minima (called extreme values) of a sample of independent, identically distributed random variables. The general cumulative distribution function (CDF)is given by

$$F(x) = \exp\left\{-\left[1-\kappa\left(\frac{x-a}{b}\right)\right]^{\frac{1}{\kappa}}\right\}, \quad -\infty < a < \infty,\ b > 0, 1-\kappa\left(\frac{x-a}{b}\right) > 0.$$

The extreme value distributions were first derived by Fisher and Tippett in 1928. The case $\kappa \to 0$ is often referred to as the Gumbel Distribution, named after the German mathematician and engineer Emil Gumbel [48], because of his extensive study of this distribution.

Extreme Value Theory (EVT) is the theory of modelling and measuring "extreme" events which occur with very small probability. This theory has been shown to be useful in various real world applications, including hydrology, meteorology, and engineering for modeling extreme events: see, for example, Gumbel [48] for modeling annual maximum flood flows, Sutcliffe [107] for estimation extreme storm rainfalls, Tomlison [111] for estimating frequencies of annual maximum storm rainfalls in New Zealand, Davenport [22], Alexandersson [3] and Bergstrom [10] for estimating annual maximum wind speed, Longin [78] for modeling extreme stock market price movements, Persson and Ryden [91], Martucci [83] for a model of significant wave height, and Hoppe and Fang [58] for estimating minimum feeder pipe thicknesses.

Estimation and prediction arise quite naturally in many real life situations. There are two major "schools" of statistical estimation. The frequentist school views

a parameter $\underline{\theta} = (\theta_1, \cdots, \theta_k)$ as a fixed but unknown parameter. In particular, $\underline{\theta}$ is not considered to be random. The Bayesian school models $\underline{\theta}$ itself as a random variable. One advantage of Bayesian over frequentist methods is that Bayesians use information that they might have about $\underline{\theta}$ through the use of a prior distribution.

All estimation procedures are based on a random sample $\underline{X} = (X_1, X_2, \cdots, X_n)$ from a population. In this thesis we will generally assume that the population distribution is described by a density function $f(x|\underline{\theta})$ where $\underline{\theta}$ is unknown. We call the set of possible values for the parameter $\underline{\theta}$ the parameter space Θ, assumed to be a subset of $\mathbb{R}^k$. The parameter estimator is a *statistic*, some function of the data, which is written as

$$\hat{\underline{\theta}} = \hat{\theta}(X_1, X_2, \cdots, X_n)$$

whose domain includes the sample space $\mathcal{X}$ of $\underline{X}$. More generally for certain functions $d(\underline{\theta})$ of the parameters, any statistic $T = T(\underline{X})$ that approximates $d(\underline{\theta})$ is called an estimator of $d(\underline{\theta})$.

There are various suggestions for measures of "closeness" of an estimator $\hat{\underline{\theta}}$ to the true parameter $\underline{\theta}$. One measure is bias, which describes the distance between the estimator's mean and the parameter $\underline{\theta}$, and is defined as, for $k = 1$,

$$Bias(\hat{\underline{\theta}}) = E(\hat{\underline{\theta}}) - \underline{\theta} \tag{1.1}$$

If it is 0, the estimator $\hat{\underline{\theta}}$ is said to be unbiased. In general, any increasing function of the absolute distance $|\hat{\underline{\theta}} - \underline{\theta}|$ would serve to measure the goodness of an estimator. For example, the mean absolute error (MAE), $E(|\hat{\underline{\theta}} - \underline{\theta}|)$, is a reasonable alternative.

There are, however, more important performance characteristics for an estimator than merely unbiasedness. The mean squared error (MSE) is one of the most

important ones and is defined by

$$MSE(\hat{\underline{\theta}}) = E[(\hat{\underline{\theta}} - \underline{\theta})^2] \tag{1.2}$$

This is also called the risk function of an estimator, with $(\hat{\underline{\theta}} - \underline{\theta})^2$ called the quadratic loss function. The MSE measures the average squared difference between the estimator $\hat{\underline{\theta}}$ and the parameter $\underline{\theta}$. Estimators that have small MSE are considered good because its expected distance from $\underline{\theta}$ is small.

It is easy to check that the mean square error can be decomposed into a variance term and a bias term:

$$MSE(\hat{\underline{\theta}}) = Var(\hat{\underline{\theta}}) + [E(\hat{\underline{\theta}}) - \underline{\theta}]^2$$

This implies that both quantities are important and need to be as small as possible to achieve good estimation performance.

The MSE has two components, one measures the variability of the estimator (precision) and the other measures its bias (accuracy). Thus, MSE has at least two advantages over other distance measures: first, it is analytically tractable and, secondly, it has this intuitive interpretation.

Simulation is a numerical technique for conducting experiments on the computer. Usually, when we refer to "simulations" in statistics, we mean "Monte Carlo simulations". A typical Monte Carlo simulation involves the following steps:

- Generate K independent data sets: $\{T_1, \cdots, T_K\}$, under the conditions of interest
- Compute the numerical value of statistic T for each data set
- If K is large enough, then $E(T) \approx \frac{1}{K}\sum_{i=1}^{K} T_i$.

The strong law of large numbers states that $\frac{1}{K}\sum_{i=1}^{K} T_i \to E(T)$ as $K \to \infty$. Thus,

the mean bias error (MBE) and MSE can be calculated by

$$\begin{aligned} MBE &= \frac{1}{K}\sum_{i=1}^{K}\left(\hat{\theta}_i - \theta\right) \\ MSE &= \frac{1}{K}\sum_{i=1}^{K}\left(\hat{\theta}_i - \theta\right)^2 \end{aligned}$$

In practice, we sometimes calculate the mean absolute error (MAE):

$$MAE = \frac{1}{K}\sum_{i=1}^{K}\left|\hat{\theta}_i - \theta\right|$$

Another performance measure for estimators $\hat{\underline{\theta}}$ is the probability that the estimator's error is within an acceptable range. Let $\{\hat{\underline{\theta}}_n\}$ be a family of estimators based on increasing sample sizes n. The family is consistent if, for any $\varepsilon > 0$,

$$P\left[|\hat{\underline{\theta}}_n - \underline{\theta}| \geq \varepsilon\right] \longrightarrow 0 \;\; as \; n \longrightarrow 0 \tag{1.3}$$

Informally, we say that $\hat{\underline{\theta}}_n$ is consistent if $\hat{\underline{\theta}}_n \longrightarrow \underline{\theta}$ in probability as $n \longrightarrow 0$.

A prediction interval (PI) represents the uncertainty in prediction and offers a range that the target is highly likely to lie within. Compared to point predictions, PIs provide more information about possible future observations and are practically more useful.

Let $\underline{X} = (X_1, X_2, \cdots, X_n)$ represent n observations from a population with a distribution function $F(x;\underline{\theta}) := P_{\underline{\theta}}(X \leq x)$ depending on p parameters $\underline{\theta}$, and let $\underline{Y} = (Y_1, Y_2, \cdots, Y_m)$ denote m future observations of some random variable having a distribution determined by $\underline{\theta}$. The usual case of interest is where each Y_i has the same distribution as X_i but this is not necessary.

Prediction of a future observation Y or a function $T(\underline{Y})$, based on previously observed data $\underline{X}$, is a fundamental problem in statistics. Broadly speaking, there are two approaches to the problem: frequentist and Bayesian. A frequentist predictive

distribution for Y, based on data $\underline{X}$, has the form

$$F\left(y|\underline{x};\underline{\theta}\right) = P_{\underline{\theta}}\left(Y \leq y|\underline{X} = \underline{x}\right) \tag{1.4}$$

A Bayesian predictive distribution for Y, given by Aitchison and Dunsmore (1975), has the form

$$F_B\left(y|\underline{x}\right) = \int F\left(y|\underline{x};\underline{\theta}\right)\pi\left(\underline{\theta}|\underline{x}\right)d\underline{\theta} \tag{1.5}$$

$F\left(y|\underline{x};\underline{\theta}\right)$ is the prediction we would make for future data if $\underline{\theta}$ were known exactly, but since it is not, we average this quantity over $\pi\left(\underline{\theta}|\underline{x}\right)$, the posterior density of $\underline{\theta}$.

If $\underline{X}$ and Y are not independent, we need to find their joint distribution. Below, we assume that $\underline{X}$ and Y are independent except for their common dependence on $\underline{\theta}$. In this case, the predictive function will be reduced to

$$F\left(y|\underline{x};\underline{\theta}\right) = F\left(y;\underline{\theta}\right) = P_{\underline{\theta}}\left(Y \leq y\right)$$

To obtain an estimator of $F\left(y|\underline{x};\underline{\theta}\right)$, say $\hat{F}\left(y|\underline{x};\underline{\theta}\right)$, a simple method is to substitute an estimator $\hat{\underline{\theta}}$ of $\underline{\theta}$ into $F(y|\underline{x};\underline{\theta})$, that is

$$\hat{F}\left(y|\underline{x};\underline{\theta}\right) = F(y|\underline{x};\hat{\underline{\theta}})$$

It is usually suggested to use a pivotal method to find $F\left(y|\underline{x};\underline{\theta}\right)$.

A prediction region for Y with prescribed probability γ is a random set $D_\gamma(\underline{X})$ such that, exactly or (approximately, if no exact region is available)

$$P_{\underline{\theta}}\left\{Y \in D_\gamma(\underline{X})|\underline{X}\right\} = \gamma \tag{1.6}$$

for all $\underline{\theta}$, where $\gamma \in (0,1)$ is fixed. If $\underline{X}$ and Y are independent, the dependence on $\underline{X}$ can be suppressed, and then the above equation can be rewritten as

$$P_{\underline{\theta}}\left\{Y \in D_\gamma(\underline{X})\right\} = \gamma$$

Here the probability refers to the joint distribution of $(Y, \underline{X})$. Without loss of generality, if we consider a one-sided upper γ-prediction interval of the form $(-\infty, U_\gamma(\underline{X}))$, then the above equation can be rewritten as

$$P_{\underline{\theta}}\{Y \leq U_\gamma(\underline{X}\} = \gamma$$

and then $U_\gamma(\underline{X})$ is called an upper γ-prediction limit for Y.

We hope to look for an exact solution to above equation, but this usually relies on the existence of a suitable pivotal quantity. Thus, in general, we look for an approximate solution.

Next, we first introduce the pivotal method. Lawless and Fredette [73] describe a general pivotal method by employing a pivotal quantity

$$W = q(\underline{X}, Y)$$

If W is monotone, and if there exists a point w_γ satisfying $F_W(w_\gamma) = \gamma$, we have

$$P(W \leq w_\gamma) = \gamma = P_{\underline{\theta}}\{Y \leq U_\gamma(\underline{X})\}$$

for some $U_\gamma(\underline{X})$.

The major difficulty is to obtain the value w_γ, since it is hard to find the exact distribution of W. Lawless [72] attacks the problem for a location-scale family by conditioning on ancillary statistics, but his approach still requires a complicated numerical calculation. In general, there are three methods to find w_γ:

1. Find an approximation for the distribution of W.

2. Find the exact distribution of W.

3. Estimate the percentiles of the distribution of W by Monte Carlo simulation.

Patel [90] provides an extensive survey of the literature on the pivotal method, which includes the normal, lognormal, binomial, exponential, Weibull (including

Gumbel maximum), gamma, inverse Gaussian, Pareto, negative binomial and Poisson distributions. Hahn & Meeker [52] explain in detail how to construct prediction intervals for the normal, binomial, and Poisson distributions.

For applications involving the Weibull and Gumbel distribution, several authors have derived prediction bounds for order statistics from a future sample. These include Mann and Saunders [79], Mann [81], Antle and Rademaker [5], Lawless [72], Mann, Schafer and Singpurwalla [80], Engelhardt and Bain [30, 31, 32], Fertig, Mayer and Mann [36], Kushary [70], Hsieh [59], Escobar and Meeker [34], Nelson [85], Nordman and Meeker [86], and Krishnamoorthy et al . [69].

When a pivotal quantity is not available, a simple method to obtain an estimated prediction limit $\hat{U}_\gamma(\underline{X})$ is to use an approximation

$$\gamma = P_{\underline{\theta}}\left\{Y \leq U_\gamma(\underline{X})\right\} \approx P_{\hat{\underline{\theta}}}\left\{Y \leq \hat{U}_\gamma(\underline{X})\right\}$$

where $\hat{\underline{\theta}}$ is an estimator for $\underline{\theta}$ based on $\underline{X}$. It is well known that (Barndorff-Nielsen and Cox, 1996)

$$P_{\hat{\underline{\theta}}}\left\{Y \leq \hat{U}_\gamma(\underline{X})\right\} = \gamma + O\left(\frac{1}{n}\right) \tag{1.7}$$

There are a number of ways proposed in the literature to improve the coverage probability. For example, Beran [9] calibrates prediction limits by using a simulation-based method; Barndorff-Nielsen and Cox [8], and Vidoni [113, 114], and Ueki and Fueda [112] obtain prediction limits with coverage error of order $O\left(n^{-\frac{3}{2}}\right)$ by correcting the quantile of the estimated predictive distribution; Hall, Peng, and Tajvidi [53] suggest a bootstrap procedure to increase coverage accuracy of prediction intervals; Lawless and Fredette [73] calibrate prediction limits based on asymptotic pivotals; Fonseca, Giuummole, and Vidoni [39] improve the coverage probability by defining a well-calibrated predictive distribution.

Bayesian prediction problems have been discussed by Dunsmore [28], Aitchison and Dunsmore [1], and Geisser [41]. Smith [104] points out that a frequentist method is usually better than Bayesian method, but the latter can perform in the presence of some accurate prior. Ren, Sun and Dey [99] verify this for a normal distribution. They show that Bayesian prediction intervals are better under suitable priors.

1.3 Thesis Outline and Contributions

This thesis is concerned with the estimation of parameters and construction of prediction intervals (PIs) for future observations. We first consider parametric methods. Our study proposes novel methods for parameter estimation and prediction interval construction from the Bayesian and frequentist points of view. Next, we present a new nonparametric bootstrap method for the construction of prediction intervals. In the remainder of the thesis, we compare the performance of estimators and prediction intervals for the Gumbel minimum distribution using extensive computer simulation. Note that similar results hold, mutatis mutandis, for the Gumbel maximum distribution.

This thesis is organized into eight chapters, and the main contributions are listed below in order of appearance:

- Three new algorithms to estimate location and scale parameters for location-scale model, including a new adaptive sample mean, presented in Chapter 2.
- A new mixed method developed for estimating a three–parameter distribution, detailed in Chapter 3.
- The new full Bayesian estimators and Bayesian–maximum likelihood estimators for the Gumbel distribution, given in Chapter 4.

- The frequentist prediction intervals constructed by two existing methods and by a Monte Carlo method, all based on new estimators in Chapter 2, detailed in Chapter 5.

- The Gibbs sampling procedure for construction of Bayesian prediction intervals for the Gumbel minimum distribution, based on new estimators in Chapter 4, described in Chapter 6.

- The comparative study of parameter estimation and prediction intervals for the Gumbel minimum distribution, detailed in Chapter 7.

- A new generalized bootstrap and its application for estimating tolerance limit and constructing prediction intervals for an unknown distribution, presented in Chapter 8.

Chapter 2

A Simple Method for Estimating Parameters of Location-Scale Families

Parameter estimation is one of the most important topics in statistical inference. In this chapter, a statement of parameter estimation for a location-scale model including the Gumbel distribution is specified.

2.1 Introduction

The location-scale family is an important class of probability distributions sharing a certain form. The family includes the normal, Cauchy, extreme value, and Student-t distributions. Additionally, some distributions such as the lognormal and Weibull can be transformed into location-scale models by applying a logarithmic transformation.

Definition 2.1. Let $g(x)$ be any probability density function (pdf) and let a and $b > 0$ be given constants. Then the function

$$f(x|a,b) = \frac{1}{b} g\left(\frac{x-a}{b}\right) \tag{2.1}$$

is called the pdf of a location–scale family.

Definition 2.2. Let $g(x)$ be any pdf. Then for any $-\infty < a < \infty, b > 0$, the family of pdfs $\frac{1}{b} g\left(\frac{x-a}{b}\right)$, indexed by the parameters (a, b), is called a location–scale family with standard pdf $g(x)$; a is called location parameter and b is the scale parameter.

The effect of introducing both the location and scale parameters is to stretch $(b > 1)$ or contract $(b < 1)$ the graph of $g(x)$ with the scale parameter and then to

shift the graph by the location parameter, so that any value of x that was greater than 0 is now greater than a.

It is easily proved that if the distribution of the random variable X is a member of the location-scale family, then X can be expressed as

$$X = a + bZ$$

where Z has the standard distribution with density function $g(z)$.

Theorem 2.1. *(Casella [12]) Let Z be a random variable with pdf $g(z)$. Suppose $E(Z)$ and $Var(Z)$ exist. If X is a random variable with pdf $\frac{1}{b}g\left(\frac{x-a}{b}\right)$, then*

$$E(X) = a + bE(Z) \quad \textit{and} \quad Var(X) = b^2 Var(Z)$$

In particular, if $E(Z) = 0$ and $Var(Z) = 1$, then $E(X) = a$ and $Var(X) = b^2$.

Pitman [93] first carried out a systematic study of estimation of location and scale parameters through the use of fiducial functions. He showed that for a random sample $\{X_1, \cdots, X_n\}$ from a density $f(x|a,b) = \frac{1}{b}g\left(\frac{x-a}{b}\right)$ the fiducial function for the estimation of a is

$$\phi_a(x_1, \cdots, x_n) = \frac{\int_0^\infty \prod_{i=1}^n \frac{1}{b^{n+1}} g\left(\frac{x_i-a}{b}\right) db}{\int_0^\infty \int_{-\infty}^\infty \prod_{i=1}^n \frac{1}{b^{n+1}} g\left(\frac{x_i-a}{b}\right) dadb}$$

and the fiducial function for the estimation of b is

$$\phi_b(x_1, \cdots, x_n) = \frac{\int_{-\infty}^\infty \prod_{i=1}^n \frac{1}{b^{n+1}} g\left(\frac{x_i-a}{b}\right) da}{\int_0^\infty \int_{-\infty}^\infty \prod_{i=1}^n \frac{1}{b^{n+1}} g\left(\frac{x_i-a}{b}\right) dadb}$$

The mean, median, etc. of the fiducial distribution of ϕ_a and ϕ_b are estimators of a and b, respectively.

Chen [14] provides a simple method for estimating location and scale parameters for a location-scale family such that standard density $g(0) = 0.5$. He considers

the pivotal

$$\xi = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{b^2}$$

Let ξ_α be the αth quantile of ξ . Then

$$P\left(\xi_{\frac{\alpha}{2}} < \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{b^2} < \xi_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

Thus, a $(1-\alpha)100\%$ confidence interval for b is

$$\left(\sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\xi_{1-\frac{\alpha}{2}}}}, \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\xi_{\frac{\alpha}{2}}}}\right)$$

When α approaches 1 (that is, the confidence approaches 0), the above confidence interval contracts to a point estimator of b,

$$\hat{b} = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\xi_{\frac{\alpha}{2}}}}$$

To estimate a, Chen defines a pivotal

$$\zeta = \frac{\sum_{i=k+1}^{n-k} \frac{X_{(i)}}{n-2k} - a}{b}, \quad 0 \leq p \leq 0.5$$

where $X_{(i)}$ denoe order statistics $k = \lfloor (n-1)p \rfloor$ is the integer part of $(n-1)p$. Chen gave a range for p, not a specific value. As before, let ζ_α be the αth quantile of ζ . Then

$$P\left(\zeta_{\frac{\alpha}{2}} < \frac{\sum_{i=k+1}^{n-k} \frac{X_{(i)}}{n-k} - a}{b} < \zeta_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

It implies that

$$P\left(\frac{\sum_{i=k+1}^{n-k} X_{(i)}}{n-2k} - b\zeta_{\frac{1-\alpha}{2}} < a < \frac{\sum_{i=k+1}^{n-k} X_{(i)}}{n-2k} + b\zeta_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

Let α approach 1. Then, the fact that $\zeta_{0.5} = 0$ would imply that a point estimator of a is

$$\hat{a} = \frac{\sum_{i=k+1}^{n-k} X_{(i)}}{n-2k}$$

which is in fact the $(100p)\%$ trimmed sample mean.

In the next section, we derive a simple estimator for location a and scale b in case the first two moments exist. The method proposed for the location-scale distribution family is adjusted according to different optimization criteria

2.2 Estimation Based on the Sample Mean and Sample Standard Deviation

As we pointed out in Chapter 1, an estimator of parameter is a function of the data. Thus, a parameter estimation problem is usually formulated as an optimization one. Because of different optimization criteria, a given problem may be solved in different ways. The purpose of this section is to construct location and scale parameters obeying appropriate criteria simultaneously.

2.2.1 Equivariance Principle

In general, statistical inference or estimation based on data should not be affected by a transformations on the data or by reordering of the data, that is, the statistical procedure should be invariant.

Definition 2.3. Let $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ be a family of distributions. Let $\mathcal{G} = \{g\}$ be a class of transformations of the sample space, i.e. $g : \mathcal{X} \longrightarrow \mathcal{X}$, that is a group under composition. If

$$\begin{aligned} \theta(g(x)) &= g(\theta(x)), \quad x \in \mathbb{R}^k \\ \theta(x+c) &= c + \theta(x), \quad x \in \mathbb{R}^k, \ c \in \mathbb{R}^k \end{aligned}$$

then such an estimator is called equivariant with respect to the transformation g.

In particular, if $\{X_1, \cdots, X_n\}$ come from a location-scale family, and if $\hat{a}$ and $\hat{b}$ are estimators of the parameters a and b, then for any real number α and non-negative

number β

1. $\hat{a}$ is location–invariant if

$$\hat{a}(\alpha+\beta X_1,\cdots,\alpha+\beta X_n)=\alpha+\beta\hat{a}(X_1,\cdots,X_n)$$

2. $\hat{b}$ is dispersion–invariant if

$$\hat{b}(\alpha+\beta X_1,\cdots,\alpha+\beta X_n)=\beta\hat{b}(X_1,\cdots,X_n)$$

and we say that $\hat{a},\hat{b}$ are equivariant estimators of $a,\ b$.

Let $\bar{X}_n=\frac{1}{n}\sum_{i=1}^n X_i$ and $S_n=\sqrt{\frac{1}{n-1}\sum_{i=1}^n\left(X_i-\bar{X}_n\right)^2}$, it is well known that $\bar{X}_n$ and S_n are location and dispersion invariant respectively.

Proposition 2.2. *(Oliveira [87])*
The following are general equivariant location and dispersion estimators based on sample mean and standard deviation:

$$\begin{aligned}\hat{a}(X_1,\cdots,X_n) &= \bar{X}_n+t_n(0,1)S_n\\ \hat{b}(X_1,\cdots,X_n) &= k_n(0,1)S_n\end{aligned}$$

Proof. Let $\hat{a}(X_1,\cdots,X_n)=t_n(\bar{X}_n,S_n)$ be the required location statistic. We then have

$$\begin{aligned}\hat{a}(\alpha+\beta X_1,\cdots,\alpha+\beta X_n) &= t_n\left[\bar{X}_n(\alpha+\beta X_1,\cdots,\alpha+\beta X_n),S_n(\alpha+\beta X_1,\cdots,\alpha+\beta X_n)\right]\\ &= t_n\left(\alpha+\beta\bar{X}_n,\beta S_n\right)\end{aligned}$$

On the other hand,

$$\hat{a}(\alpha+\beta X_1,\cdots,\alpha+\beta X_n) = \alpha+\beta\hat{a}(X_1,\cdots,X_n)=\alpha+\beta t_n(\bar{X}_n,S_n)$$

so that t_n has the form

$$t_n\left(\alpha+\beta\bar{X}_n,\beta S_n\right)=\alpha+\beta t_n(\bar{X}_n,S_n).$$

Setting $\bar{X}_n = 0$ and $S_n = 1$, we obtain

$$t_n(\alpha, \beta) = \alpha + t_n(0,1)\beta$$

Therefore, the general location statistic based on the sample mean and standard deviation is

$$\hat{a}(X_1, \cdots, X_n) = \bar{X}_n + t_n(0,1)S_n$$

Let $\hat{b}(X_1, \cdots, X_n) = k_n(\bar{X}_n, S_n)$ be the required dispersion statistic. We then have

$$\begin{aligned}\hat{b}(\alpha + \beta X_1, \cdots, \alpha + \beta X_n) &= k_n\left[\bar{X}_n(\alpha + \beta X_1, \cdots, \alpha + \beta X_n), S_n(\alpha + \beta X_1, \cdots, \alpha + \beta X_n)\right] \\ &= k_n\left(\alpha + \beta\bar{X}_n, \beta S_n\right)\end{aligned}$$

On the other hand,

$$\hat{b}(\alpha + \beta X_1, \cdots, \alpha + \beta X_n) = \beta\hat{a}(X_1, \cdots, X_n) = \beta k_n(\bar{X}_n, S_n)$$

so that k_n has the form

$$k_n\left(\alpha + \beta\bar{X}_n, \beta S_n\right) = \beta k_n(\bar{X}_n, S_n)$$

Setting $\bar{X}_n = 0$ and $S_n = 1$, we obtain

$$k_n(\alpha, \beta) = k_n(0,1)\beta$$

Therefore, the general dispersion statistic based on the sample mean and standard deviation is

$$\hat{b}(X_1, \cdots, X_n) = k_n(0,1)S_n$$

2.2.2 Unbiasedness Principle

In statistics, the bias of an estimator is the difference between this estimator's expected value and the true value of the parameter being estimated. Expectations are

an "average" taken over all possible values.

Definition 2.4. For observations $\underline{X} = (X_1,\ X_2, \cdots,\ X_n)$ based on a distribution with parameter θ, and for $\hat{d}(\underline{X})$ an estimator for $d(\theta)$, the bias is the mean of the difference $\hat{d}(\underline{X}) - d(\theta)$, That is,

$$Bias(\theta) = E\left[\hat{d}(\underline{X})\right] - d(\theta)$$

If $Bias(\theta) = 0$ for all values of the parameter, then $\hat{d}(\underline{X})$ is called an unbiased estimator of $d(\theta)$. Any estimator that is not unbiased is called biased.

Definition 2.5. We say $f(x) = O\left[p(x)\right]$ as $x \to q$ if there exists a constant M such that $|f(x)| \leq M|p(x)|$ in some punctured neighborhood of q, that is for $x \in (q-\delta, q+\delta) \setminus \{q\}$ for some value of δ.

We say $f(x) = o\left[p(x)\right]$ as $x \to q$ if $\lim_{x\to q} \frac{f(x)}{p(x)} = 0$. This implies that there exists a punctured neighborhood of q on which $p(x)$ does not vanish.

Lemma 2.3. *(Cramer [21], P.353)*

$$\begin{aligned} E(S_n) &= \sqrt{\frac{n}{n-1}}\,\sigma_X + O\left(\frac{1}{n}\right) \\ Var(S_n) &= \frac{m_4 - \sigma_X^4}{4(n-1)\sigma_X^2} + O\left(\frac{1}{n^2}\right) \end{aligned}$$

where $m_4 = E\left[(X - E(X)\right]^4$ *is the* 4th *central moment.*

Lemma 2.4. *(Weba [118]) Suppose* $E\left(|X|^{8+\alpha}\right) < \infty$ *for some* $\alpha > 0$ *and* $m_r = E\left[(X - E(X)\right]^r$ *is the* r^{th} *central moment. Then*

$$\begin{aligned} E(S_n) &= \sigma_X + \frac{b_1}{n} + \frac{b_2}{n^2} + o\left(\frac{1}{n^2}\right) \\ Var(S_n) &= \frac{c_1}{n} + \frac{c_2}{n^2} + o\left(\frac{1}{n^2}\right) \end{aligned}$$

where

$$
\begin{aligned}
b_1 &= -\frac{m_4 - m_2^2}{8m_2^{3/2}}, \\
b_2 &= -\frac{1}{4}m_2^{1/2} + \frac{m_6 - 3m_2m_4 - 6m_3^2 + 2m_2^3}{16m_2^{5/2}} - 15\frac{(m_4 - m_2^2)^2}{128m_2^{7/2}} \\
c_1 &= \frac{m_4 - m_2^2}{4m_2} \\
c_2 &= \frac{m_2}{2} - \frac{m_6 - 3m_2m_4 - 6m_3^2 + 2m_2^3}{8m_2^2} + 7\frac{(m_4 - m_2^2)^2}{32m_2^3}
\end{aligned}
$$

Following the method of Oliveira, we construct unbiased consistent equivariant estimators of location and scale parameters for a location-scale family based on the sample mean and standard deviation.

Theorem 2.5. *Let $\{X_1, \cdots, X_n\}$ come from a location-scale family with density given by $f(x) = \frac{1}{b}g\left(\frac{x-a}{b}\right)$ and let $Z = \frac{X-a}{b}$ with pdf $g(z)$. Suppose that $E(Z) = \mu_Z$, $Var(Z) = \sigma_Z^2$, $S_{Z,n}$ and $S_{X,n}$ are the sample standard deviations of Z, X, respectively. Then*

$$\hat{a} = \bar{X}_n - \frac{\mu_Z}{E(S_{Z,n})}S_{X,n} \tag{2.2}$$

$$\hat{b} = \frac{1}{E(S_{Z,n})}S_{X,n} \tag{2.3}$$

are unbiased equivariant estimators for location a and scale b.

There are many practical cases where it is impossible to compute the exact value of $E(S_{Z,n})$. Often we can estimate the expected value by Monte Carlo simulation. Improved estimates of such expected values can be had by further increasing the number of simulation repetitions.

Proof. As the unbiasedness principle implies

$$E(\hat{a}) = E\left[\bar{X}_n + t_n(0,1)S_{X,n}\right] = (a + b\mu_Z) + t_n(0,1)E(S_{X,n}) = a$$

This requires

$$b\mu + t_n(0,1)E(S_{X,n}) = 0$$

Then we have

$$t_n(0,1) = -\frac{b\mu_Z}{E(S_{X,n})} = -\frac{\mu_Z}{E(S_{Z,n})}$$

Similarly,

$$E(\hat{b}) \quad = \quad E\left[k_n(0,1)S_{X,n}\right] = k_n(0,1)E(S_{X,n}) = b$$

This requires

$$k_n(0,1) = \frac{b}{E(S_{X,n})} = \frac{1}{E(S_{Z,n})}$$

Corollary 2.6. *Let $\{X_1, \cdots, X_n\}$ come from a location-scale family with density given by $f(x) = \frac{1}{b}g\left(\frac{x-a}{b}\right)$ and let $Z = \frac{X-a}{b}$ with pdf $g(z)$. Suppose that $E(Z) = \mu$, $Var(Z) = \sigma^2$ and $S_{X,n}$ is sample standard deviations. Then*

$$\hat{a} \quad = \quad \bar{X}_n - \left(\sqrt{\frac{n-1}{n}}\frac{\mu}{\sigma}\right) S_{X,n} \tag{2.4}$$

$$\hat{b} \quad = \quad \left(\sqrt{\frac{n-1}{n}}\frac{1}{\sigma}\right) S_{X,n} \tag{2.5}$$

are approximately unbiased, equivariant and consistent estimators for location a and scale b.

Proof. It is well know that $\bar{X}_n \longrightarrow E(X) = a + b\mu,\ S_n \longrightarrow \sigma_X = b\sigma$ as $n \longrightarrow \infty$. Obviously,

$$\hat{a} \longrightarrow a + b\mu - \frac{\mu}{\sigma}b\sigma = a \quad as\ n \longrightarrow \infty$$

$$\hat{b} \longrightarrow \frac{1}{\sigma}b\sigma = b \quad as\ n \longrightarrow \infty$$

Corollary 2.7. *Let $\{X_1, \cdots, X_n\}$ come from a location-scale family with density*

given by $f(x) = \frac{1}{b}g\left(\frac{x-a}{b}\right)$ *and let* $Z = \frac{X-a}{b}$ *with pdf* $g(z)$*. Suppose that* $E(Z) = \mu$*,* $Var(Z) = \sigma^2$*,* $\mu_r = E\left[(Z - E(Z)\right]^r$ *is the* r^{th} *central moment, and* S_n *is the sample standard deviation. If* $E\left(|Z|^{8+\alpha}\right) < \infty$ *for some* $\alpha > 0$*, then*

$$\hat{a} = \bar{X}_n - \frac{128n^2\mu\sigma^7}{128n^2\sigma^8 - 16n\sigma^5(\mu_4 - \sigma^4) - 32\sigma^8 + 8\sigma^2(\mu_6 - 3\sigma^2\mu_4 - 6\mu_3^2 + 2\sigma^6) - 15(\mu_4 - \sigma^4)^2} S_n \tag{2.6}$$

$$\hat{b} = \frac{128n^2\sigma^7}{128n^2\sigma^8 - 16n\sigma^5(\mu_4 - \sigma^4) - 32\sigma^8 + 8\sigma^2(\mu_6 - 3\sigma^2\mu_4 - 6\mu_3^2 + 2\sigma^6) - 15(\mu_4 - \sigma^4)^2} S_n \tag{2.7}$$

are approximately unbiased, equivariant and consistent estimators for location a *and scale* b*.*

Proof. Let $m_r = E\left[(X - E(X)\right]^r$ be the rth central moment. Suppose that $\mu_r < \infty$, then

$$\begin{aligned} m_r &= \int_{-\infty}^{\infty} (x - E(X))^r f_X(x)dx = \int_{-\infty}^{\infty} (x - a - b\mu)^r f_X(x)dx \\ &= \int_{-\infty}^{\infty} (x - a - b\mu)^r \frac{1}{b} g\left(\frac{x-a}{b}\right) dx = \int_{-\infty}^{\infty} (bz - b\mu)^r g(z)\, dz \\ &= b^r \mu_r \end{aligned}$$

Notice that $m_2 = \sigma_X^2 = b^2\sigma^2$, we have

$$\begin{aligned} b_1' &= -\frac{\mu_4 - \sigma^4}{8\sigma^2} b, \\ b_2' &= \left[-\frac{1}{4}\sigma + \frac{\mu_6 - 3\sigma^2\mu_4 - 6\mu_3^2 + 2\sigma^6}{16\sigma^5} - 15\frac{(\mu_4 - \sigma^4)^2}{128\sigma^7}\right] b \\ c_1' &= \frac{\mu_4 - \sigma^4}{4\sigma^2} b^2 \\ c_2' &= \left[\frac{\sigma^2}{2} - \frac{\mu_6 - 3\sigma^2\mu_4 - 6\mu_3^2 + 2\sigma^6}{8\sigma^4} + 7\frac{(\mu_4 - \sigma^4)^2}{32\sigma^6}\right] b^2 \end{aligned}$$

From Lemma 2.4, we have

$$E(S_n) = \left(\sigma - \frac{\mu_4 - \sigma^4}{8n\sigma^2} + \frac{-\frac{1}{4}\sigma + \frac{\mu_6 - 3\sigma^2\mu_4 - 6\mu_3^2 + 2\sigma^6}{16\sigma^5} - 15\frac{(\mu_4-\sigma^4)^2}{128\sigma^7}}{n^2}\right) b$$

$$= \frac{128n^2\sigma^8 - 16n\sigma^5(\mu_4 - \sigma^4) - 32\sigma^8 + 8\sigma^2(\mu_6 - 3\sigma^2\mu_4 - 6\mu_3^2 + 2\sigma^6) - 15(\mu_4 - \sigma^4)^2}{128n^2\sigma^7}\, b$$

therefore

$$t_n(0,1) = -\frac{b\mu}{E(S_n)}$$

$$= -\frac{128n^2\mu\sigma^7}{128n^2\sigma^8 - 16n\sigma^5(\mu_4 - \sigma^4) - 32\sigma^8 + 8\sigma^2(\mu_6 - 3\sigma^2\mu_4 - 6\mu_3^2 + 2\sigma^6) - 15(\mu_4 - \sigma^4)^2}$$

and

$$k_n(0,1) = \frac{128n^2\sigma^7}{128n^2\sigma^8 - 16n\sigma^5(\mu_4 - \sigma^4) - 32\sigma^8 + 8\sigma^2(\mu_6 - 3\sigma^2\mu_4 - 6\mu_3^2 + 2\sigma^6) - 15(\mu_4 - \sigma^4)^2}$$

So we obtain

$$\hat{a} = \bar{X}_n$$

$$- \frac{128n^2\mu\sigma^7}{128n^2\sigma^8 - 16n\sigma^5(\mu_4 - \sigma^4) - 32\sigma^8 + 8\sigma^2(\mu_6 - 3\sigma^2\mu_4 - 6\mu_3^2 + 2\sigma^6) - 15(\mu_4 - \sigma^4)^2} S_n$$

$$\hat{b} = \frac{128n^2\sigma^7}{128n^2\sigma^8 - 16n\sigma^5(\mu_4 - \sigma^4) - 32\sigma^8 + 8\sigma^2(\mu_6 - 3\sigma^2\mu_4 - 6\mu_3^2 + 2\sigma^6) - 15(\mu_4 - \sigma^4)^2} S_n$$

It is well know that $\bar{X}_n \longrightarrow E(X) = a + b\mu,\ S_n \longrightarrow \sigma_X = b\sigma$ as $n \longrightarrow \infty$. Obviously,

$$\hat{a} \longrightarrow a + b\mu - \frac{\mu}{\sigma} \times b\sigma = a \quad as\ n \longrightarrow \infty$$

$$\hat{b} \longrightarrow \frac{1}{\sigma} \times b\sigma = b \quad as\ n \longrightarrow \infty$$

2.2.3 Minimum Risk Principle

Let $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ be a family of distributions. Suppose that we are interested in estimating $d(\theta)$ for a fixed function $d : \Theta \longrightarrow \mathbb{R}$. Let $\hat{d}(\underline{X})$ be an estimator of $d(\theta)$,

and consider the function

$$L\left[\theta, \hat{d}(\underline{X})\right] = \left[\hat{d}(\underline{X}) - d(\theta)\right]^2 \tag{2.8}$$

where the "loss" may be considered as a distance between $\hat{d}(\underline{X})$ and $d(\theta)$. Then the risk of the estimator of $\hat{d}(\underline{X})$ is

$$R\left[\theta, \hat{d}(\underline{X})\right] = EL\left[\theta, \hat{d}(\underline{X})\right] = \int L\left[\theta, \hat{d}(\underline{x})\right] P_\theta(d\underline{x}) \tag{2.9}$$

Roughly speaking, an estimator $\hat{d}_1(\underline{X})$ is said to be as good as an estimator $\hat{d}_2(\underline{X})$ if

$$R\left[\theta, \hat{d}_1(\underline{X})\right] \leq R\left[\theta, \hat{d}_2(\underline{X})\right], \quad for\ all\ \theta \in \Theta$$

and $\hat{d}_1(\underline{X})$ is said to be better than $\hat{d}_2(\underline{X})$ if

$$\begin{aligned} R\left[\theta, \hat{d}_1(\underline{X})\right] &\leq R\left[\theta, \hat{d}_2(\underline{X})\right], \quad for\ all\ \theta \in \Theta \\ R\left[\theta, \hat{d}_1(\underline{X})\right] &< R\left[\theta, \hat{d}_2(\underline{X})\right], \quad for\ some\ \theta \in \Theta \end{aligned}$$

Suppose $\underline{X} = (X_1, X_2, \cdots, X_n)$ is a random sample from distribution with density $f(x; \theta_1, \theta_2)$, depending on two parameters θ_1 and θ_2. Let $\dot{\theta}_1(\underline{X})$ and $\dot{\theta}_2(\underline{X})$ be estimators of θ_1 and θ_2, respectively. Later in this section, we consider following the loss functions

$$\begin{aligned} L_1 &= (\theta_2 - \dot{\theta}_2)^2 K(\theta_1, \theta_2) \\ L_2 &= (\theta_1 - \dot{\theta}_1)^2 K(\theta_1, \theta_2) \\ L_3 &= (\theta_1 + \theta_2 - \xi)^2 K(\theta_1, \theta_2) \\ L_4 &= (\theta_1 - \dot{\theta}_1)^2 (\theta_2 - \dot{\theta}_2)^2 K(\theta_1, \theta_2) \end{aligned}$$

where ξ is a linear combination of $\dot{\theta}_1$ and $\dot{\theta}_2$ and $K(\theta_1, \theta_2)$ is any positive function.

Theorem 2.8. *(Part 1 of the Theorem is due to Goodman [43], Part 2 of the Theorem is due to Mann [79]*

Let (θ_1, θ_2) be two parameters of a distribution and $K(\theta_1, \theta_2) = 1$. Assume that there exist unbiased estimators $(\hat{\theta}_1, \hat{\theta}_2)$ with

$$Cov\begin{pmatrix}\hat{\theta}_1\\ \hat{\theta}_2\end{pmatrix} = \theta_2^2 \begin{pmatrix} A & B \\ B & C \end{pmatrix}$$

where A, B, C are constants that do not depend on (θ_1 or θ_2). The following two statements are true.

1. *The minimum risk estimator $\dot{\theta}_2$, under loss L_1, among all estimators $\hat{\theta}_2(k_2) = k_2\hat{\theta}_2$ depending on an arbitrary parameter k_2 is*

   $$\dot{\theta}_2 = \frac{\hat{\theta}_2}{1+C}$$

   The risk is independent of θ_1 and has the value

   $$R_1\left(\theta_2, \dot{\theta}_2\right) = \frac{C}{1+C}\theta_2^2$$

2. *The minimum risk estimator $\dot{\theta}_1$, under loss L_2, among all estimators $\hat{\theta}_1(k_1) = \hat{\theta}_1 + k_2\hat{\theta}_2$ depending on an arbitrary parameter k_1 is*

   $$\dot{\theta}_1 = \hat{\theta}_1 - \frac{B}{1+C}\hat{\theta}_2$$

   The risk is independent of θ_1 and has the value

   $$R_2\left(\theta_1, \dot{\theta}_1\right) = \left[A - \frac{B^2}{1+C}\right]\theta_2^2$$

Theorem 2.9. *(Mann [79]*

If the conditions of Theorem 2.8 are true, then among all estimators that are linear combinations of the unbiased estimators, $\hat{\xi}(k_1, k_2) = k_1\theta_1 + k_2\theta_2$, the estimators that produce risk under L_3 with $K(\theta_1, \theta_2) = \frac{1}{\theta_2^2}$ that is invariant with respect to θ_1 and θ_2, and is minimum are in the class

$$\theta_1(k_2) = \theta_1 + k_2\theta_2 \quad and \quad \theta_2(k_2) = k_2\theta_2.$$

We can find k_1 and k_2 that result in minimum risk and risk independent of θ_1 and θ_2 using

$$R_3\left[\theta_1+\theta_2,\hat{\xi}(k_1,k_2)\right]=\frac{Var(k_1\theta_1+k_2\theta_2)+\left[(1-k_1)\theta_1+(1-k_2)\theta_2\right]^2}{\theta_2^2}$$

Let $\hat{a}$ and $\hat{b}$ be unbiased estimator for location a and scale b such that

$$\hat{a}=\bar{X}_n+t_nS_n \quad and\ \hat{b}=k_nS_n$$

where $t_n,\ k_n$ are given in Theorem 2.6. Since $\bar{X}=a+b\bar{Z},\ S_X=bS_Z$, we now have

$$\begin{aligned}
&Var(\bar{X})=\frac{1}{n}b^2Var(Z)\\
&Var(S_X)=b^2Var(S_Z)\\
&Cov(\bar{X},S_X)=Cov(a+b\bar{Z},bS_Z)=b^2Cov(\bar{Z},S_Z)\\
&Var(\bar{X}_n+t_nS_n)=Var(\bar{X})+t_n^2Var(S_X)+2t_nCov(\bar{X},S_X)\\
&\qquad\qquad=b^2\left[\frac{1}{n}Var(Z)+t_n^2Var(S_Z)+2t_nCov(\bar{Z},S_Z)\right]\\
&Cov(\bar{X}_n+t_nS_n,k_nS_n)=k_nCov(\bar{X}_n,S_X)+k_nt_nVar(S_X^2)=b^2k_n\left[Cov(\bar{Z},S_Z)+t_nVar(S_Z^2)\right]
\end{aligned}$$

Therefore,

$$Cov\begin{pmatrix}\hat{a}\\ \hat{b}\end{pmatrix}=b^2\begin{pmatrix}\frac{1}{n}Var(Z)+t_n^2Var(S_Z)+2t_nCov(\bar{Z},S_Z) & k_n\left[Cov(\bar{Z},S_Z)+t_nVar(S_Z^2)\right]\\ k_n\left[Cov(\bar{Z},S_Z)+t_nVar(S_Z^2)\right] & k_n^2Var(S_Z)\end{pmatrix}$$

Now using results of Lemma 2.3, we have

$$\begin{aligned}
C &= k_n^2Var(S_Z)=k_n^2\frac{\mu_4-\sigma^4}{4(n-1)\sigma^2}\\
&= \frac{k_n^2\sigma^2}{4(n-1)}\left(\frac{\mu_4}{\sigma^4}-1\right)=\frac{k_n^2\sigma^2}{4(n-1)}(\beta_2-1)\\
&= \frac{k_n^2\sigma^2}{4(n-1)}(\gamma_2+2)
\end{aligned}$$

where $\gamma_2=\beta_2-3$ is the kurtosis coefficient of $Z=\frac{X-a}{b}$.

To find A and B, we need to calculate $Cov(\bar{Z},S_Z)$. Letting $H(x,y)=x\sqrt{y}$,

we expand $H(x,y)$ using a Taylor series expansion at the point $(\mu,1)$ with $E(Z)=\mu$, that is

$$\begin{aligned}
H(x,y) &= H(\mu,1) + (x-\mu)\left.\frac{\partial H}{\partial x}\right|_{\substack{x=\mu\\y=1}} + (y-1)\left.\frac{\partial H}{\partial y}\right|_{\substack{x=\mu\\y=1}} + \frac{1}{2}(x-\mu)^2\left.\frac{\partial^2 H}{\partial x^2}\right|_{\substack{x=\mu\\y=1}} \\
&+ \frac{1}{2}(y-1)^2\left.\frac{\partial^2 H}{\partial y^2}\right|_{\substack{x=\mu\\y=1}} + (x-\mu)(y-1)\left.\frac{\partial^2 H}{\partial x\partial y}\right|_{\substack{x=\mu\\y=1}} \\
&= \mu + (x-\mu)\times 1 + (y-1)\times\frac{\mu}{2} + \frac{1}{2}(x-\mu)^2\times 0 + \frac{1}{2}(y-1)^2\times(-\frac{\mu}{4}) \\
&+ (x-\mu)(y-1)\times\frac{1}{2} \\
&= -\frac{\mu}{8} + \frac{x}{2} + \frac{3\mu}{4}y - \frac{\mu}{8}y^2 + \frac{(x-\mu)y}{2}
\end{aligned}$$

Now let $\bar{Z}=x, S_Z^2=y$, we have

$$\begin{aligned}
H(\bar{Z},S_Z^2) &= \bar{Z}S_Z = \bar{Z}\sqrt{S_Z^2} \\
&= -\frac{\mu}{8} + \frac{\bar{Z}}{2} + \frac{3\mu}{4}S_Z^2 - \frac{\mu}{8}S_Z^4 + \frac{(\bar{Z}-\mu)S_Z^2}{2}
\end{aligned}$$

It is well known that

$$\begin{aligned}
Cov(\bar{Z},S_Z^2) &= E\left[(\bar{Z}-\mu)S_Z^2\right] = \frac{\mu_3}{n} \\
Var(S_Z^2) &= \frac{\mu_4}{n} - \frac{n-3}{n(n-1)}\sigma^4 \\
E(S_Z^4) &= Var(S_Z^2) + [E(S_Z^2)]^2 = \frac{\mu_4}{n} + \frac{n^2-2n+3}{n(n-1)}\sigma^4
\end{aligned}$$

Therefore, if $\{Z_1,\cdots,Z_n\}$ come from a distribution such that $E(Z)=\mu$, $Var(Z)=\sigma^2$, $\mu_r = E\left[(Z-E(Z)\right]^r$ is the rth central moment, $\bar{Z}$ is sample mean and S_n is sample variance. Then

$$\begin{aligned}
Cov(\bar{Z},S_Z) &= E(\bar{Z}S_Z) - E(\bar{Z})E(S_Z) = E\left(\bar{Z}\sqrt{S_Z^2}\right) - E(\bar{Z})E(S_Z) \\
&= E\left(-\frac{\mu}{8} + \frac{\bar{Z}}{2} + \frac{3\mu}{4}S_Z^2 - \frac{\mu}{8}S_Z^4 + \frac{(\bar{Z}-\mu)S_Z^2}{2}\right) - \sqrt{\frac{n}{n-1}}\mu\sigma \\
&= \frac{3}{8}\mu + \frac{3}{4}\mu\sigma^2 - \sqrt{\frac{n}{n-1}}\mu\sigma - \frac{\mu}{8}\left(\frac{\mu_4}{n} + \frac{n^2-2n+3}{n(n-1)}\sigma^4\right) + \frac{\mu_3}{2n}
\end{aligned}$$

so then

$$B = k_n \left[Cov(\bar{Z}, S_Z) + t_n Var(S_Z^2)\right]$$
$$= k_n \left[\frac{3}{8}\mu + \frac{3}{4}\mu\sigma^2 - \sqrt{\frac{n}{n-1}}\mu\sigma - \frac{\mu}{8}\left(\frac{\mu_4}{n} + \frac{n^2-2n+3}{n(n-1)}\sigma^4\right) + \frac{\mu_3}{2n} + t_n\left(\frac{\mu_4}{n} - \frac{n-3}{n(n-1)}\sigma^4\right)\right]$$

and

$$A = \frac{1}{n}Var(Z) + t_n^2 Var(S_Z) + 2t_n Cov(\bar{Z}, S_Z)$$
$$= \frac{1}{n}\sigma^2 + t_n^2\frac{\mu_4 - \sigma^4}{4(n-1)\sigma^2} + 2t_n\left[\frac{3}{8}\mu + \frac{3}{4}\mu\sigma^2 - \sqrt{\frac{n}{n-1}}\mu\sigma - \frac{\mu}{8}\left(\frac{\mu_4}{n} + \frac{n^2-2n+3}{n(n-1)}\sigma^4\right) + \frac{\mu_3}{2n}\right]$$

Corollary 2.10. *If $\{X_1, \cdots, X_n\}$ come from a location-scale family with density given by $f(x) = \frac{1}{b}g\left(\frac{x-a}{b}\right)$ and let $Z = \frac{X-a}{b}$ with pdf $g(z)$. Suppose that $E(Z) = \mu$, $Var(Z) = \sigma^2$, $\mu_r = E\left[(Z - E(Z)\right]^r$ is the rth central moment, $r = 2, 3, 4$, γ_2 is the kurtosis coefficient of Z, and S_n is the sample variance of $\{X_1, \cdots, X_n\}$. Suppose that $\hat{a} = \bar{X}_n + t_n S_n$ and $\hat{b} = k_n S_n$ are unbiased estimators, then*

$$Cov\binom{\hat{a}}{\hat{b}} = b^2 \begin{pmatrix} A & B \\ B & C \end{pmatrix}$$

where A, B, C are constants that do not depend on a or b with

$$A = \frac{1}{n}\sigma^2 + t_n^2\frac{\mu_4 - \sigma^4}{4(n-1)\sigma^2} + 2t_n\left[\frac{3}{8}\mu + \frac{3}{4}\mu\sigma^2 - \sqrt{\frac{n}{n-1}}\mu\sigma - \frac{\mu}{8}\left(\frac{\mu_4}{n} + \frac{n^2-2n+3}{n(n-1)}\sigma^4\right) + \frac{\mu_3}{2n}\right]$$
$$B = k_n\left[\frac{3}{8}\mu + \frac{3}{4}\mu\sigma^2 - \sqrt{\frac{n}{n-1}}\mu\sigma - \frac{\mu}{8}\left(\frac{\mu_4}{n} + \frac{n^2-2n+3}{n(n-1)}\sigma^4\right) + \frac{\mu_3}{2n} + t_n\left(\frac{\mu_4}{n} - \frac{n-3}{n(n-1)}\sigma^4\right)\right]$$
$$C = \frac{k_n^2\sigma^2}{4(n-1)}(\gamma_2 + 2)$$

and the following two statements are true:

1. *The minimum risk estimator $\dot{b}$, under loss $L_1 = (b - \dot{b})^2$, among all estimators $\hat{b}(k_1) = k_1\hat{b}$ depending on an arbitrary parameter k_1 is*

$$\dot{b} = \frac{\hat{b}}{1+C} = \frac{k_n}{1+C}S_n$$

The risk is independent of a and has the value

$$R_1\left(b,\dot{b}\right)=\frac{C}{1+C}b^2$$

2. *The minimum risk estimator $\dot{a}$, under loss $L_2=(a-\dot{a})^2$, among all estimators $\hat{a}(k_2)=\hat{a}+k_2\hat{b}$ depending on an arbitrary parameter k_2 is*

$$\dot{a}=\hat{a}-\frac{B}{1+C}\hat{b}=\bar{X}+\left(t_n-\frac{B}{1+C}k_n\right)S_n$$

The risk is independent of a and has the value

$$R_2\left(a,\dot{a}\right)=\left[A-\frac{B^2}{1+C}\right]b^2$$

Now we give an example using a Gumbel maximum distribution. The cdf of the standard Gumbel random variable is given by

$$F(z)=e^{-e^{-z}},\quad -\infty<z<\infty$$

and the pdf by

$$f(z)=e^{-z}e^{-e^{-z}}$$

Let γ denote Euler constant. It is known that

$$E(Z)=\mu=\gamma\ \ and\ \ \ Var(Z)=\sigma^2=\frac{\pi^2}{6}$$

Let $\mu_k=E\left[Z-E(Z)\right]^k$ be the kth central moment. Clearly $\mu_1=0$. For $k\geq 2$, we have

$$\begin{aligned}\mu_k &= \int_{-\infty}^{\infty}(z-\mu)^k f(z)dz=\int_{-\infty}^{\infty}(z-\gamma)^k e^{-z}e^{-e^{-z}}dz\\ &= \sum_{j=0}^{k}\binom{k}{j}(-\gamma)^{k-j}\int_{-\infty}^{\infty}z^j e^{-z}e^{-e^{-z}}dz\end{aligned}$$

Setting $t = e^{-z}$, then

$$\mu_k = \sum_{j=0}^{k} (-1)^j \binom{k}{j} (-\gamma)^{k-j} \int_0^\infty (\ln t)^j e^{-t} dt$$

According to Proposition A.2 in the Appendix

$$\int_0^\infty x^{v-1} (\ln x)^n e^{-ux} dx = \frac{\partial^n}{\partial v^n} \left[u^{-v} \Gamma(v) \right], \quad v \geq 1, u > 0, n = 0, 1, 2, \cdots,$$

In our case, we take $v = 1, u = 1$, then

$$\int_0^\infty (\ln t)^j e^{-t} dt = \Gamma^{(j)}(1)$$

and therefore

$$\mu_k = \sum_{j=0}^{k} (-1)^j \binom{k}{j} (-\gamma)^{k-j} \Gamma^{(j)}(1), \quad k \geq 2$$

Proposition 2.11. *(Choi [19]*

$$\Gamma^{(n+1)}(1) = -\gamma \Gamma^{(n)}(1) + n! \sum_{k=1}^{n} \frac{(-1)^{k+1}}{(n-k)!} \zeta(k+1) \Gamma^{(n-k)}(1), \quad n \geq 0,$$

where $\zeta(p)$ is the Riemann zeta function

$$\zeta(p) = \sum_{k=1}^{n} \frac{1}{k^p}, \quad p > 1$$

Let $\hat{a}$ and $\hat{b}$ be unbiased estimators for location a and scale b such that

$$\hat{a} = \bar{X}_n + t_n S_n \quad \textit{and} \quad \hat{b} = k_n S_n$$

where $t_n,\ k_n$ are given in Theorem 2.6. Then we have

$$\begin{aligned}
Var(\hat{a}) &= Var(\bar{X}_n + t_n S_n) = Var(\bar{X}) + t_n^2 Var(S_n) + 2t_n Cov(\bar{X}, S_n) \\
&= b^2 \left[\frac{1}{n} Var(Z) + t_n^2 Var(S_Z) + 2t_n Cov(\bar{Z}, S_Z) \right] \\
&= b^2 \left\{ \frac{1}{n} \sigma^2 + t_n^2 \frac{\mu_4 - \sigma^4}{4(n-1)\sigma^2} \right\} \\
&\quad + 2t_n \left[\frac{3}{8} \mu + \frac{3}{4} \mu \sigma^2 - \sqrt{\frac{n}{n-1}} \mu \sigma - \frac{\mu}{8} \left(\frac{\mu_4}{n} + \frac{n^2 - 2n + 3}{n(n-1)} \sigma^4 \right) + \frac{\mu_3}{2n} \right]
\end{aligned}$$

and

$$\begin{aligned} Var(\hat{b}) &= Var(k_n S_n) = k_n^2 Var(S_n) = k_n^2 b^2 Var(S_Z) = k_n^2 b^2 \frac{\mu_4 - \sigma^4}{4(n-1)\sigma^2} \\ &= b^2 \frac{k_n^2 \sigma^2}{4(n-1)} \left(\frac{\mu_4}{\sigma^4} - 1 \right) = b^2 \frac{k_n^2 \sigma^2}{4(n-1)} (\gamma_2 + 2) \end{aligned}$$

where $\gamma_2 = \beta_2 - 3$ is kurtosis coefficient of $Z = \frac{X-a}{b}$.

It is well known that

$$\begin{aligned} \zeta(2) &= \frac{\pi^2}{6} \approx 1.6449, \quad \zeta(4) = \frac{\pi^4}{90} \approx 1.0823, \quad \zeta(6) = \frac{\pi^8}{9450} \approx 1.0173, \\ \zeta(3) &\approx 1.20205 \quad , \zeta(5) \approx 1.03692, \quad \zeta(7) \approx 1.00834, \zeta(8) = \frac{\pi^4}{9450} \approx 1.00407, \\ \zeta(9) &\approx 1.002008 \end{aligned}$$

It is easy to calculate μ_k and then obtain $Var(\hat{a})$.

2.3 Estimation Based on the Sample Mean and Gini's Mean Difference

In some cases, the second moment may not exist. We then consider using Gini's mean difference to substitute for the sample standard deviations and then derive new estimators for the location–scale model.

A symmetric function on variables $\underline{x} = (x_1, x_2, \cdots, x_n)$ is a function that is unchanged by any permutation of its variables. For example, $h(\underline{x}) = kthmax(\underline{x})$ is symmetric, where $kthmax(\cdot)$ is the kth largest value among $\{x_1, x_2, \cdots, x_n\}$.

Definition 2.6. Let $X_i \overset{\text{iid}}{\sim} F, i = 1, 2, \cdots, n$, $h(X_1, X_2, \cdots, X_r)$ be a symmetric function and $E|h(X_1, X_2, \cdots, X_r)| < \infty$. Let $\theta(F) = E[h(X_1, X_2, \cdots, X_r)]$, a U-statistic U_n with kernel h of order r is defined as

$$U_n = \frac{1}{C_n^r} \sum_C h(X_1, X_2, \cdots, X_r)$$

where math notation $\sum_C$ ranges over all subsets of size r chosen from $\{1, 2, \cdots, n\}$.

For scale estimation, we can use Gini's estimator given by

$$G_n = \frac{1}{C_n^2} \sum_{i<j} |x_j - x_i|$$

with the kernel $h = |x_j - x_i|$ of order 2. Obviously, Gini's mean difference is a special case of a U-statistic.

Following a similar method as Theorem 2.5, we can obtain another estimator for the location–scale model.

Theorem 2.12. *If $\{X_1, \cdots, X_n\}$ is a random sample from a location-scale family with density given by $f(x) = \frac{1}{b} g\left(\frac{x-a}{b}\right)$ and let $Z = \frac{X-a}{b}$ with pdf $g(z)$. Suppose that $E(Z) = \mu_Z$, and $G_{X,n}$ is Gini's mean difference. Then*

$$\hat{a} = \bar{X}_n - \frac{\mu_Z}{E(G_{Z,n})} G_{X,n} \tag{2.10}$$

$$\hat{b} = \frac{1}{E(G_{Z,n})} G_{X,n} \tag{2.11}$$

are unbiased equivariant estimators for location a and scale b.

It has been showed that $E[U_n] = \theta(F)$, that is, U-statistics are unbiased. The mean of Gini's mean difference has been derived in the literature as follows:

$$E(G_{X,n}) = 4 \int xF(x)dF(x) - 2 \int xdF(x)$$

To obtain consistency of a scale estimator, one has to multiply $G_{X,n}$ with a consistent factor K suggested by the literature such as

$$\hat{b} = KG_{X,n}$$

But it is not easy to find K for some distributions.

Proof. Let $\hat{b} = t_b G_{X,n}$, then

$$\begin{aligned} E(\hat{b}) &= E\left(t_b G_{X,n}\right) = t_b E\left(\frac{1}{C_n^2}\sum_{i<j}|X_j - X_i|\right) \\ &= t_b b E\left(\frac{1}{C_n^2}\sum_{i<j}|Z_j - Z_i|\right) \\ &= t_b b E\left(G_{Z,n}\right) \end{aligned}$$

This gives

$$t_b = \frac{1}{E\left(G_{Z,n}\right)}$$

Let $\hat{a} = \bar{X} + t_a G_{X,n}$, then

$$\begin{aligned} E(\hat{b}) &= E\left(\bar{X} + t_a G_{X,n}\right) = a + b\mu_Z + t_a E\left(\frac{1}{C_n^2}\sum_{i<j}|X_j - X_i|\right) \\ &= a + b\mu_Z + t_a b E\left(\frac{1}{C_n^2}\sum_{i<j}|Z_j - Z_i|\right) \\ &= a + b\mu_Z + t_a b E\left(G_{Z,n}\right) \end{aligned}$$

We then have

$$t_a = -\frac{\mu_Z}{E\left(G_{Z,n}\right)}$$

2.4 An Adaptive Unbiased Estimator for Location-Scale Families

Suppose that $\underline{X}$ is a random sample of size n from a location-scale model with location parameter a and scale parameter b. Let N_1 denote the number of sample points less than c, N_2 denote the number of sample point great than c, and n_1, n_2 denote observed values of N_1, N_2 respectively. Obviously, $N_1 + N_2 = n$, and the expectation of N_1 is

$$E(N_1) = nP\left(X \leq c\right)$$

If $N_1 > E(N_1)$, then it is quite possible that $\bar{X}$ is smaller than μ_X, the population mean, although $E(\bar{X}) = \mu_X$. Thus $\bar{X}$ is not good estimator for the population mean μ_X when n is small. However, it is well known that $\bar{X}$ is a good estimator for μ_X when n is large. Notice that

$$\frac{N_1}{N_2} \longrightarrow \frac{P(X \leq c)}{P(X > c)}, \quad as\ n \to \infty$$

Definition 2.7. Let $\{X_1, \cdots, X_n\}$ be a random sample from a population. We separate the sample into two groups $\underline{T}$ and $\underline{S}$ such that

$$\begin{aligned} \underline{T} &= \{X_i | X_i \leq c, i = 1, 2, \cdots, n\} \\ \underline{H} &= \{X_i | X_i > c, i = 1, 2, \cdots, n\} \end{aligned}$$

where N_1 and N_2 denote the number of points in the sets $\underline{T}$ and $\underline{H}$, respectively. Then, $N_1 + N_2 = n$, and the cluster mean is given by

$$\bar{X}^* = P\left(X_1 \leq c\right)\left(\frac{1}{N_1}\sum_{i=1}^{N_1} T_i\right) + \left[1 - P\left(X_1 \leq c\right)\right]\left(\frac{1}{N_2}\sum_{i=1}^{N_2} H_i\right) \tag{2.12}$$

For a location–scale model $Z = \frac{X_1 - a}{b}$, and we can take c to be the location parameter, and then $P\left(X_1 \leq c\right) = (Z \leq 0)$. We have developed an algorithm to estimate n_1 for the power Gumbel distribution, and details will be given in next chapter. However, it is still difficult to estimate n_1 even if we know the exact value of $P(X \leq c)$. To circumvent this problem, we can first choose c, then estimate the probability $P(X \leq c)$. Now, the problem turns into estimating the success probability of a binomial distribution.

Let N_1 be the number of successes, that is, $\#\{i : X_i \leq c\}$, in a random sample of size n. A classical frequentist estimator of $P(X \leq c)$ in the binomial experiment is the sample proportion

$$\hat{P}(X \leq c) = \frac{N_1}{n}$$

This estimator is both the MLE and the uniformly minimum variance unbiased estimator (UMVUE) of $P(X \leq c)$. If we substitute this estimator into equation (2.12), it then follows that

$$\bar{X}^* = \bar{X}$$

Wilson [120] introduced a slight modification of the above estimator by adding two successes and two failures. Thus his point estimator is

$$\hat{P}(X \leq c) = \frac{N_1 + 2}{n + 4}$$

If we consider a uniform prior $p \sim U(0, 1)$, a Bayesian estimator for the proportion is

$$\hat{P}(X \leq c) = \frac{N_1 + 1}{n + 2}$$

which is equivalent to the Laplace estimator.

The conjugate prior of the binomial distribution is the Beta distribution. If the prior is $Beta(s, t)$, then the posterior distribution of p is $Beta(N_1 + s, n - N_1 + t)$. Bayesian estimator for proportion is

$$\hat{P}(X \leq c) = \frac{N_1 + s}{n + s + t} = \left(\frac{n}{n + s + t}\right)\left(\frac{N_1}{n}\right) + \left(\frac{s + t}{n + s + t}\right)\left(\frac{s}{s + t}\right)$$

which is weighted average of MLE and prior mean.

Compared with traditional sample mean $\bar{X}$, the mean $\bar{X}^*$ is adaptive according to the value of N_1. We give properties of $\bar{X}^*$ in the following Theorem.

Theorem 2.13. *The mean of $\bar{X}^*$ has following properties:*

1. $E\left(\bar{X}^*\right) = E(X_1)$
2. *If $N_1 = nP(X_1 \leq c)$ or $X_1 = X_2 = \cdots = X_n$, then $\bar{X}^* = \bar{X}$*
3. *If $N_1 > nP(X_1 \leq c)$, then $\bar{X}^* > \bar{X}$*

4. If $N_1 < nP(X_1 \leq c)$, then $\bar{X}^* < \bar{X}$

Proof. 1.

$$\begin{aligned} E\left(\bar{X}^*\right) &= E\left[P\left(X_1 \leq c\right)\left(\frac{1}{N_1}\sum_{i=1}^{N_1} T_i\right) + \left[1 - P\left(X_1 \leq c\right)\right]\left(\frac{1}{n_2}\sum_{i=1}^{N_2} H_i\right)\right] \\ &= P\left(X_1 \leq c\right) E\left(\frac{1}{N_1}\sum_{i=1}^{N_1} T_i\right) + \left[1 - P\left(X_1 \leq c\right)\right] E\left(\frac{1}{N_2}\sum_{i=1}^{N_2} H_i\right) \end{aligned}$$

Let I denote an indicator function and write $T_i = X_i I_{\{X_i \leq c\}}$ and $H_i = X_i I_{\{X_i > c\}}$. Then

$$E\left(\frac{1}{N_1}\sum_{i=1}^{N_1} T_i\right) = E\left(\frac{1}{N_1}\sum_{i=1}^{n} X_i I_{\{X_i \leq c\}}\right)$$

where N_1 is a binomial random variable with parameters n and $p = P(X \leq c)$. For $N_1 = 0$ we define

$$\frac{1}{N_1}\sum_{i=1}^{n} X_i I_{\{X_i \leq c\}} = 0$$

It then follow that

$$E\left(\frac{1}{N_1}\sum_{i=1}^{n} X_i I_{\{X_i \leq c\}}\right) = \sum_{k=1}^{n}\left\{E\left[\frac{1}{k}\sum_{i=1}^{n} X_i I_{\{X_i \leq c\}} \mid N_1 = k\right] P(N_1 = k)\right\}$$

Notice that the conditional distribution of $\{T_i\}$ given $N_1 = k$ is

$$\begin{aligned} & P\left(T_1 \leq t_1, \cdots, T_n \leq t_n | N_1 = k\right) = \frac{P\left(T_1 \leq t_1, \cdots, T_n \leq t_n, N_1 = k\right)}{P(N_1 = k)} \\ = & \sum_{1 \leq i_1 \leq \cdots \leq i_k \leq n} \frac{P\left(T_1 \leq t_1, \cdots, T_n \leq t_n, X_{i_1} \leq c, \cdots, X_{i_k} \leq c\right)}{P(N_1 = k)} \\ = & \frac{1}{P(N_1 = k)} \sum_{1 \leq i_1 \leq \cdots \leq i_k \leq n} \left\{\left[\prod_{d=1}^{k} P\left(X_{i_d} I_{\{X_{i_d} \leq c\}} \leq t_{i_d}\right)\right]\left[\prod_{j \neq i_1, \cdots, i_k} P\left(X_j I_{\{X_j > c\}} \leq t_j\right)\right]\right\} \end{aligned}$$

Since $\{T_i\}$ are still i.i.d, we only need to consider the marginal distribution $P\left(T_i \leq t_i | N_1 = k\right)$, that is

$$P\left(T_1 \leq \infty, \cdots, T_{i-1} \leq \infty, T_i \leq t_i, T_{i+1} \leq \infty, \cdots, T_n \leq \infty | N_1 = k\right)$$

and then the conditional pdf of T_i is

$$f(t|N_1 = k) = \begin{cases} \frac{\binom{n-1}{k-1}p^{k-1}(1-p)^{n-k}}{P(N_1=k)} \frac{f(t)}{P(X \leq c)} & t \leq c \\ 0 & \text{else} \end{cases}$$

Therefore

$$\begin{aligned} E\left(\frac{1}{N_1}\sum_{i=1}^{N_1} T_i\right) &= E\left(\frac{1}{N_1}\sum_{i=1}^{n} X_i I_{\{X_i \leq c\}}\right) \\ &= \sum_{k=1}^{n} E(T_1|N_1 = k)P(N_1 = k) \\ &= \frac{1}{P(X \leq c)}\int_{-\infty}^{c} tf(t)\sum_{k=1}^{n}\binom{n-1}{k-1}p^{k-1}(1-p)^{n-k}dt \\ &= \frac{1}{P(X \leq c)}\int_{-\infty}^{c} tf(t)dt \\ &= \frac{E\left(XI_{\{X \leq c\}}\right)}{P(X \leq c)} \end{aligned}$$

similar, we have

$$f(h|N_2 = k) = \begin{cases} \frac{\binom{n-1}{k-1}p^{k-1}(1-p)^{n-k}}{P(N_2=k)} \frac{f(h)}{1-P(X \leq c)} & h > c \\ 0 & else \end{cases}$$

and

$$E\left(\frac{1}{N_2}\sum_{i=1}^{N_2} H_i\right) = \frac{E\left(XI_{\{X > c\}}\right)}{1 - P(X \leq c)}$$

It then follows that

$$\begin{aligned} E\left(\bar{X}^*\right) &= P(X \leq c)\frac{E\left(XI_{\{X \leq c\}}\right)}{P(X \leq c)} + [1 - P(X \leq c)]\frac{E\left(XI_{\{X > c\}}\right)}{1 - P(X \leq c)} \\ &= E(X) \end{aligned}$$

2.

$$\begin{aligned}
\bar{X}^* - \bar{X} &= \left[P\left(X_1 \le c\right)\left(\frac{1}{N_1}\sum_{i=1}^{N_1} T_i\right) + \left[1 - P\left(X_1 \le c\right)\right]\left(\frac{1}{N_2}\sum_{i=1}^{N_2} H_i\right)\right] - \frac{1}{n}\sum_{i=1}^{n} X_i \\
&= \left(\frac{P(X_1 \le c)}{N_1} - \frac{1}{n}\right)\sum_{i=1}^{N_1} T_i + \left(\frac{1 - P\left(X_1 \le c\right)}{N_2} - \frac{1}{n}\right)\sum_{i=1}^{N_2} H_i \\
&= \frac{nP(X_1 \le c) - N_1}{n}\left(\frac{1}{N_1}\sum_{i=1}^{N_1} T_i\right) + \frac{n\left[1 - P\left(X_1 \le c\right)\right] - N_2}{n}\left(\frac{1}{N_2}\sum_{i=1}^{N_2} H_i\right) \\
&= \left[P(X_1 \le c) - \frac{N_1}{n}\right]\left(\frac{1}{N_1}\sum_{i=1}^{N_1} T_i - \frac{1}{N_2}\sum_{i=1}^{N_2} H_i\right)
\end{aligned}$$

By definition, we have

$$T_i \le x_{(N_1)}, \quad H_i > x_{(N_1)}$$

Since $\frac{1}{N_1}\sum_{i=1}^{N_1} T_i \le x_{(N_1)}$ and $\frac{1}{N_2}\sum_{i=1}^{N_2} H_i > x_{(N_1)}$, then

$$\frac{1}{N_1}\sum_{i=1}^{N_1} T_i < \frac{1}{N_2}\sum_{i=1}^{N_2} H_i$$

Therefore, if $N_1 < nP(X_1 \le c)$, then $\bar{X}^* < \bar{X}$, and if $N_1 > nP(X_1 > c)$, then $\bar{X}^* > \bar{X}$. If $N_1 = nP(X_1 \le c)$ or $X_1 = X_2 = \cdots = X_n$, then $\bar{X}^* = \bar{X}$.

□

Next, we will construct adaptive unbiased estimators for a location-scale family.

Theorem 2.14. *Let $\{X_1, \cdots, X_n\}$ be a random sample from a location-scale family with density given by $f(x) = \frac{1}{b}g\left(\frac{x-a}{b}\right)$ with $E(Z_i) = \mu_Z$ and $X_i = a + bZ_i$. Let S_{X,N_1} and S_{X,N_2} denote the sample standard deviations of T and H, respectively. Then unbiased estimators for location a and scale b are given by*

$$\hat{a} = \bar{X}^* - \frac{2\mu_Z}{E\left(S_{Z,N_1} + S_{Z,N_2}\right)} S^*_{X,n} \tag{2.13}$$

$$\hat{b} = \frac{1}{E\left(S'_{Z,n}\right)} S'_{X,n} \tag{2.14}$$

where we take $c = a$ in the definition of $\bar{X}^*$, and

$$\begin{aligned} S^*_{X,n} &= \frac{1}{2}\left(S_{X,N_1} + S_{X,N_2}\right) \\ S'_{X,n} &= \left[\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}^*\right)^2\right]^{\frac{1}{2}} \end{aligned}$$

and $E\left(S'_{Z,n}\right)$ can be obtained by Monte Carlo method.

Proof. According to the definition of $\bar{X}^*$, $T_i = a + bZ_i$ for $i \in S_T$ where S_T is a subset of $S = \{1, 2, \cdots, n\}$, and write such Z_i as Z_i^T. Similar, we denote $H_j = a + bZ_j$ for $j \in S_H$ where $S_T \bigcup S_H = S$ and write such Z_j as Z_i^H.

$$\begin{aligned} S^2_{X,N_1} &= \frac{1}{N_1 - 1}\sum_{i=1}^{N_1}\left(T_i - \bar{T}\right)^2 = \frac{1}{N_1 - 1}\sum_{i=1}^{N_1}\left[(a + bZ_i^T) - (a + b\bar{Z}^T)\right]^2 \\ &= \frac{b^2}{N_1 - 1}\sum_{i=1}^{N_1}\left(Z_i^T - \bar{Z}^T\right)^2 = b^2 S^2_{Z,N_1} \end{aligned}$$

Similar, we have $S^2_{X,N_2} = b_2 S^2_{Z,N^2}$. Then

$$E\left(S^*_{X,n}\right) = \frac{1}{2}E\left(S_{X,N_1} + S_{X,N_2}\right) = \frac{b}{2}E\left(S_{Z,N_1} + S_{Z,N_2}\right)$$

Obviously, we get

$$\begin{aligned} E\left(\hat{a}\right) &= E\left[\bar{X}^* - \frac{2\mu_Z}{E\left(S_{Z,n_1} + S_{Z,n_2}\right)}S^*_{X,n}\right] \\ &= (a + b\mu_Z) - b\mu_Z \\ &= a \end{aligned}$$

We notice that

$$\begin{aligned}
\bar{X}^* &= P(X_1 \le a)\left(\frac{1}{N_1}\sum_{i=1}^{N_1} T_i\right) + [1 - P(X_1 \le a)]\left(\frac{1}{N_2}\sum_{i=1}^{N_2} H_i\right) \\
&= P(X_1 \le a)\left[\frac{1}{N_1}\sum_{i=1}^{N_1}(a + bZ_i)\right] + [1 - P(X_1 \le a)]\left[\frac{1}{N_2}\sum_{i=1}^{N_2}(a + bZ_j)\right] \\
&= a + b\left\{P(X_1 \le a)\left(\frac{1}{N_1}\sum_{i=1}^{N_1} Z_i^T\right) + [1 - P(X_1 \le a)]\left(\frac{1}{N_2}\sum_{i=1}^{N_2} Z_j^H\right)\right\} \\
&= a + b\bar{Z}^*
\end{aligned}$$

where

$$\bar{Z}^* = P(Z_1 \le 0)\left(\frac{1}{N_1}\sum_{i=1}^{N_1} T_i^Z\right) + [1 - P(Z_1 \le 0)]\left(\frac{1}{N_2}\sum_{i=1}^{N_2} H_i^Z\right)$$

with

$$\begin{aligned}
\underline{T}^Z &= \{Z_i | Z_i \le 0, i = 1, 2, \cdots, n\} \\
\underline{H}^Z &= \{Z_i | Z_i > 0, i = 1, 2, \cdots, n\}
\end{aligned}$$

Therefore,

$$\begin{aligned}
S'^2_{X,n} &= \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}^*\right)^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left[(a + bX_i) - (a + b\bar{Z}^*)\right]^2 \\
&= \frac{b^2}{n-1}\sum_{i=1}^{n}\left(Z_i - \bar{Z}^*\right)^2
\end{aligned}$$

This gives $S'_{X,n} = bS'_{Z,n}$. It then follows that

$$E(\hat{b}) = E\left[\frac{1}{E\left(S'_{Z,n}\right)} S'_{X,n}\right] = b$$

□

Based on this result, we propose a mixed method to estimate the power Gumbel distribution in the next chapter.

Chapter 3

A Mixed Method for Estimating Location, Scale and Shape Parameters of a 3-Parameter Distribution

In this chapter, we propose a new mixed method to estimate location, scale and shape parameters for a 3-parameter distribution. In this method, we estimate the shape parameter by maximizing the log–likelihood function or the spacing product function, which are functions of shape parameter after taking location and scale parameters as functions of the shape parameter, respectively.

Motivation came from work by Sermer and Hoppe (personal communication) in fitting bundle powers in a CANDU reactor. At any time (or at any reactor state) there is a collection of bundle powers, say $\{B_i : i = 1, 2, \cdots, 13 \times 480\}$ (Note that (13 = number of bundles in a channel and 480 = total number of reactor channels). Let $B_{\max}$ be the maximum bundle power over the whole reactor core. It was found that $B_{\max}$ could be approximated by a maximum Gumbel distribution in which the second exponent was replaced by a polynomial of fixed degree with unknown coefficients. In our work we consider a monomial with an unknown degree.

3.1 Mixed Method

The maximum product of spacing estimation (MPS) was introduced independently by Cheng and Amin [15], and Ranneby [98]. Suppose we have a random sample $\{X_1, \cdots, X_n\}$ from a continuous distribution with distribution function $F_\theta,\ \theta \in \Theta$.

The MPS estimator $\hat{\theta}$ of θ maximizes the product of spacings,

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \prod_{i-1}^{n+1} [F_\theta(X_i) - F_\theta(X_{i-1})] \tag{3.1}$$

where $F_\theta(X_0) = 0$, $F_\theta(X_{n+1}) = 1$. Since the product of spacings in the above equation is always bounded, the MPS method can generate asymptotically optimal estimates even when MLE breaks down due to unbounded likelihood functions (Cheng and Traylor [16]).

When $F_\theta(x)$ has a density $f_\theta(x)$, the logarithm of $\prod_{i-1}^{n+1} [F_\theta(X_i) - F_\theta(X_{i-1})]$ can be approximated by

$$\sum \ln f_\theta(X_{(i)})[X_{(i)} - X_{(i-1)}] = \sum \ln f_\theta(X_{(i)}) + \sum \ln[X_{(i)} - X_{(i-1)}] \tag{3.2}$$

Since $\sum \ln[X_{(i)} - X_{(i-1)}]$ is a constant not depending on θ, maximizing the product of spacings in $\prod_{i-1}^{n+1} \left[F_\theta(X_{(i)}) - F_\theta(X_{(i-1)})\right]$ is asymptotically equivalent to maximizing $\sum \ln f_\theta(X_{(i)})$, i.e., the log likelihood.

The generalized spacings estimator (GSE) of θ is defined to be the argument θ which minimizes

$$M(\theta) = \sum_{i=1}^{n+1} h\left[(n+1)D_i(\theta)\right]$$

where

$$D_i(\theta) = F_\theta(X_{(i)}) - F_\theta(X_{(i-1)}), \quad i = 1, 2, \cdots, n$$

and $h : (0, \infty) \longrightarrow \mathbb{R}$ is a strictly convex function. Some standard choices of $h(\cdot)$ that have been used in the context of goodness-of-fit testing, are: $h(x) = -\ln x,\ x \ln x,\ x^2, -\sqrt{x},\ \frac{1}{x}$ and $|x-1|$. When using $h(x) = -\ln x$, MPS can be obtained by minimizing

$$M(\theta) = -\sum_{i=1}^{n+1} \ln\left[F_\theta(X_{(i)}) - F_\theta(X_{(i-1)})\right] \tag{3.3}$$

or using $h(x) = x \ln x$, we have

$$M(\theta) = \sum_{i=1}^{n+1} \{(n+1)D_i(\theta) \log [(n+1)D_i(\theta)]\} \tag{3.4}$$

Let a, b and θ denote location, scale and shape parameter respectively. Given a sample $\{X_1, X_2, \cdots, X_n\}$, the new mixed method is given as follow:

Step 1 Write estimators of a and b as function of θ using one of Theorems 2.5, 2.12 and 2.14;

Step 2 Estimate the shape parameter by maximizing the log likelihood function or the spacing product function.

Step 3 Substitute $\hat{\theta}$ for θ into the Theorem that was used in Step 1 to obtain $\hat{a}$ and $\hat{b}$.

3.2 Example 1: Generalized Extreme Value (GEV) Distribution

The generalized extreme-value (GEV) distribution, introduced by Jenkinson [61], has been used to model a wide variety of natural extremes in many applications, including floods, rainfall, wind speeds, wave height, and other situations involving extremes.

3.2.1 Introduction

The GEV distribution has cumulative distribution function

$$F(x|a,b,\kappa) = \begin{cases} \exp\left\{-\left[1-\kappa\left(\frac{x-a}{b}\right)\right]^{1/\kappa}\right\} & \text{if } \kappa \neq 0 \\ \exp\left[-\exp\left(-\frac{x-a}{b}\right)\right] & \text{in the limit as } \kappa \to 0 \end{cases}$$

and for $\kappa \neq 0$, its density function is

$$f(x|a,b,\kappa) = \frac{1}{b}\left[1-\kappa\left(\frac{x-a}{b}\right)\right]^{1/\kappa-1} \exp\left\{-\left[1-\kappa\left(\frac{x-a}{b}\right)\right]^{1/\kappa}\right\}$$

where $\kappa(x-a) < b$ with $-\infty < a < \infty$, $0 < b < \infty$, and $-\infty < \kappa < \infty$. Here a, b and κ are the location, scale, and shape parameters, respectively. The value of κ

dictates the tail behavior of CDF, and thus we refer to κ as the shape parameter. For $\kappa < 0$, $\kappa > 0$ and $\kappa \to 0$, the GEV distribution function reduces to Frechet, Weibull, and Gumbel distributions, respectively. In general, The case $\kappa = 0$ is interpreted as the limit $\kappa \to 0$. Let $Z = \frac{X-a}{b}$, the class of Generalized Extreme Value Distributions can be defined as having standard distributions below.

1. Gumbel $\kappa = 0$

$$F(x) = e^{-e^{-z}}$$

2. Frechet $\kappa = -\frac{1}{\xi} < 0$

$$F(x) = e^{-\left(1-\frac{z}{\xi}\right)^{\xi}}$$

3. Weibull $\kappa = \frac{1}{\xi} > 0$

$$F(x) = e^{-\left(1+\frac{z}{\xi}\right)^{-\xi}}$$

The mean and variance of the GEV are given by:

$$\begin{aligned} E(X) &= a + \frac{b}{\kappa}\left[1 - \Gamma(1+\kappa)\right], \quad \kappa > -1 \\ Var(X) &= \frac{b^2}{\kappa^2}\left[\Gamma(1+2\kappa) - \Gamma^2(1+\kappa)\right] \quad \kappa > -\frac{1}{2} \end{aligned}$$

For integer t, the tth moment exists only if $\kappa > -\frac{1}{t}$. In the literature, generally κ is constrained to lie in $(-0.5, 0.5)$ to satisfy finite variance and maximum likelihood regularity conditions.

Let $\{X_{(1)}, X_{(2)}, \cdots, X_{(n)}\}$ denote the order statistics of a random sample, $\{X_1, X_2, \cdots, X_n\}$, from a continuous population with *cdf* $F_X(x)$ and *pdf* $f_X(x)$. Then the *pdf* of $X_{(j)}$ is

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x)\left[F_X(x)\right]^{j-1}\left[1 - F_X(x)\right]^{n-j}$$

So

$$
\begin{aligned}
E\left(X_{(j)}\right) &= \int_{-\infty}^{\infty} x f_{X_{(j)}}(x) dx \\
&= \frac{n!}{(j-1)!(n-j)!} \int_{-\infty}^{\infty} x \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j} dF_X \\
&= \frac{n!}{(j-1)!(n-j)!} \int_0^1 x(F) F^{j-1} (1-F)^{n-j} dF \\
&= j \binom{n}{j} \int_0^1 x(F) F^{j-1} \sum_{r=1}^{n-j} \binom{n-j}{r} (-1)^r F^{n-j-r} dF \\
&= j \binom{n}{j} \sum_{r=1}^{n-j} \left[(-1)^r \binom{n-j}{r} \int_0^1 x(F) F^{n-1-r} dF \right]
\end{aligned}
$$

where $F = F_X(x)$ and $x(F)$ is inverse of F. For the GEV distribution,

$$
x(F) = \begin{cases} a + \frac{b}{\kappa}\left[1 - (-\ln F)^{\kappa}\right] & \text{if } \kappa \neq 0 \\ a - b \ln(-\ln F) & \text{in the limit as } \kappa \to 0 \end{cases}
$$

For $\kappa \neq 0$, set $t = -\ln F$, then

$$
\begin{aligned}
\int_0^1 x(F) F^{n-1-r} dF &= \int_0^1 \left\{ a + \frac{b}{\kappa}\left[1 - (-\ln F)^{\kappa}\right] \right\} F^{n-1-r} dF \\
&= \left(a + \frac{b}{\kappa}\right) \int_0^1 F^{n-1-r} dF - \frac{b}{\kappa} \int_0^1 (-\ln F)^{\kappa} F^{n-1-r} dF \\
&= \frac{a\kappa + b}{(n-r)\kappa} - \frac{b}{\kappa} \int_0^{\infty} t^{\kappa} e^{-(n-r)t} dt \\
&= \frac{a\kappa + b}{(n-r)\kappa} - \frac{b}{\kappa} \frac{\Gamma(1+\kappa)}{(n-r)^{\kappa+1}}
\end{aligned}
$$

Thus

$$
E\left(X_{(j)}\right) = \frac{j}{\kappa} \binom{n}{j} \sum_{r=1}^{n-j} \left[(-1)^r \binom{n-j}{r} \left(\frac{a\kappa + b}{(n-r)} - b \frac{\Gamma(1+\kappa)}{(n-r)^{\kappa+1}} \right) \right] \tag{3.5}
$$

Let $G_{Z,n} = \frac{1}{C_n^2}\sum_{i<j}|Z_j - Z_i|$ denote Gini's mean difference of $Z = \frac{X-a}{b}$. We next derive the mean of Gini's mean difference $E(G_{Z,n})$.

$$\begin{aligned} E(G_{Z,n}) &= 4\int zF(z)dF(z) - 2\int zdF(z) \\ &= \frac{1}{\kappa}\Gamma(1+\kappa)\left(2 - \frac{1}{2^{\kappa-1}}\right) \end{aligned} \tag{3.6}$$

3.2.2 Methods of Estimation for the Generalized Extreme Value Distribution

Common methods used for estimating the GEV parameters are the method of maximum likelihood and the method of L-moments. Hosking et al. [56] and Hosking [57] show that small-sample maximum likelihood parameter (MLE) estimators are very unstable and recommend probability weighted moment (PWM) estimators, which are equivalent to L moment estimators. Morrison and Smith [84] proposed two methods for estimating the GEV parameters that combine both maximum likelihood and L-moment methods. Wong and Li [121] shown that the MPS performs more stably than MLE in a simulation study when sample sizes are small. However, PWM estimators have small mean absolute error (MAE) than MPS estimators.

In this section, a new mixed method parameter estimators for the GEV distribution is introduced based on a combination of the MPS and our previous results. The new estimation procedures are motivated by non-regular problems in estimating the shape parameter κ using maximum likelihood, and by the fact that we cannot impose the constraint $\kappa(x-a) < b$ when using the L-moment method. The mixed estimators possess more attractive properties than that of L-moment estimators when the sample size is small.

1. Moment Estimators

The moment estimators of the parameters of the GEV distribution are given by

$$\begin{aligned}
\hat{a} &= \bar{x} - \frac{\hat{b}}{\hat{\kappa}}\left[1 - \Gamma(1+\hat{\kappa})\right] \\
\hat{b} &= \frac{|\hat{\kappa}|}{\sqrt{\Gamma(1+2\hat{\kappa}) - \Gamma^2(1+\hat{\kappa})}} S_{X,n} \\
\hat{\tau} &= sign(\hat{\kappa}) \frac{-\Gamma(1+3\hat{\kappa}) + 3\Gamma(1+\hat{\kappa})\Gamma(1+2\hat{\kappa}) - 2\Gamma^3(1+3\hat{\kappa})}{\left[\Gamma(1+2\hat{\kappa}) - \Gamma^2(1+\hat{\kappa})\right] 3/2}
\end{aligned}$$

where $\hat{\tau}$ is the sample skewness. There is no explicit solution for κ.

2. L-moment Estimators

The L moment estimators for the GEV distribution (Hosking et al., [56]) are

$$\begin{aligned}
\hat{a} &= \hat{\lambda}_1 - \frac{\hat{b}}{\hat{\kappa}}\left[1 - \Gamma(1+\hat{\kappa})\right] \\
\hat{b} &= \frac{\hat{\lambda}_2 \hat{\kappa}}{(1 - 2^{-\hat{\kappa}})\Gamma(1+\hat{\kappa})} \\
\hat{\kappa} &= 7.8590c + 2.9554c^2 \\
c &= \frac{2}{3+\hat{\tau}} - \frac{\ln 2}{\ln 3}
\end{aligned}$$

where $\hat{\lambda}_1$, $\hat{\lambda}_3$ and $\hat{\lambda}_3$ are L-moment estimators and $\hat{\tau} = \frac{\hat{\lambda}_3}{\hat{\lambda}_2}$ is L-skewness.

Let X be a real-valued random variable with CDF $F(x)$, and let $X_{1:n}, \cdots, X_{n:n}$ be the order statistics of a random sample size of n from the distribution of X. Define the L-moments of X to be the quantiles

$$\lambda_k = \frac{1}{k}\sum_{j=0}^{k-1}(-1)^j \binom{k-1}{j} E\left(X_{k-j:k}\right), \quad k = 1, 2, \cdots$$

The first few L-moments are

$$\begin{aligned}
\lambda_1 &= E(X_{1:1}) = E(X) \\
\lambda_2 &= \frac{1}{2}E\left(X_{2:2} - X_{1:2}\right) \\
\lambda_3 &= \frac{1}{3}E\left(X_{3:3} - 2X_{2:3} + X_{1:3}\right)
\end{aligned}$$

Therefore, the L-moment ratio of X is defined by

$$\tau_k = \frac{\lambda_k}{\lambda_2}, \quad k = 3, 4, \cdots$$

Greenwood et al.[45] define PWMs (probability weighting moments) as

$$\beta_k = \int_0^1 x(F) F^k dF$$

and Hosking [57] has shown that

$$\begin{aligned} \lambda_1 &= \beta_0 \\ \lambda_2 &= 2\beta_1 - \beta_0 \\ \lambda_3 &= 6\beta_2 - 5\beta_1 + \beta_0 \end{aligned}$$

Landwehr et al. [71] have show that the unbiased estimator of β_k is given by

$$\hat{\beta}_k = \frac{1}{n} \sum_{i=1}^{n} \frac{(i-1)(i-2)\cdots(i-k)}{(n-1)(n-2)\cdots(n-k)} x_{i:n}$$

Wang [117] points out that the above procedure is not particularly efficient and gives the following results:

$$\begin{aligned} \hat{\lambda}_1 &= \frac{1}{C_n^1} \sum_{i=1}^{n} X_{i:n} \\ \hat{\lambda}_2 &= \frac{1}{2C_n^2} \sum_{i=1}^{n} \left(C_{i-1}^1 - C_{n-i}^1 \right) X_{i:n} \\ \hat{\lambda}_3 &= \frac{1}{3C_n^3} \sum_{i=1}^{n} \left(C_{i-1}^2 - 2C_{i-1}^1 C_{n-i}^1 + C_{n-i}^2 \right) X_{i:n} \end{aligned}$$

Then the L-moment ratio $\tau_k = \frac{\lambda_k}{\lambda_2}$ is naturally estimated by

$$\hat{\tau}_k = \frac{\hat{\lambda}_k}{\hat{\lambda}_2}$$

3. Maximum-Likelihood Estimator

The maximum-likelihood estimator of the GEV distribution has been developed by Jenkinson [61], Otten and Van Montfort [88], Prescott and Walden [95, 96],

and Hosking [56]. The log-likelihood is given by

$$\ln L(a,b,\kappa|\underline{x}) = -n\ln b - \sum_{i=1}^{n}\left(1-\kappa\frac{x_i-a}{b}\right)^{\frac{1}{\kappa}} + \left(\frac{1}{\kappa}-1\right)\sum_{i=1}^{n}\ln\left(1-\kappa\frac{x_i-a}{b}\right)$$

The MLE of a, b and κ can be obtained by solving the system of equations, which correspond to setting to zero the first derivatives of $\ln L(a,b,\kappa|\underline{x})$ with respect to each parameter.

Smith [103] has shown that the maximum likelihood is regular in the case $\kappa < 0.5$, and that when $0.5 < \kappa < 1$, maximum likelihood estimators exist but are non-regular. If $\kappa > 1$, corresponding to very short tailed distributions, the likelihood is unbounded and maximum likelihood fails.

A feasible solution is subject to constraints: $-0.5 < \kappa < 0.5$, which ensures that the $\{X_i\}$ have finite second moments and satisfy the maximum likelihood regularity conditions given in Smith [103].

4. Morrison and Smith's Mixed Method

Morrison and Smith [84] proposed the following mixed method: estimate κ by maximising $\ln(a,b,\kappa|\underline{x})$ with a and b given as functions of κ by the L-moments

$$\begin{aligned}\lambda_1 &= E(X_1) = a + \frac{b}{\kappa}\left[1-\Gamma(1+\kappa)\right]\\ \lambda_2 &= \frac{1}{2}E\left(|X_1-X_2|\right) = \frac{b}{\kappa}(1-2^{-\kappa})\Gamma(1+\kappa)\end{aligned}$$

with λ_1, λ_2 estimated by

$$\begin{aligned}\hat{\lambda}_1 &= \frac{1}{n}\sum_{i=1}^{n}X_i\\ \hat{\lambda}_2 &= \frac{1}{n(n-1)}\sum_{i<j}|X_i-X_j|\end{aligned}$$

Then

$$\begin{aligned} a &= \lambda_1 - \frac{\lambda_2[1-\Gamma(1+\lambda_3)]}{(1-2^{-\lambda_3})\Gamma(1+\lambda_3)} \\ b &= \frac{\lambda_2\lambda_3}{(1-2^{-\lambda_3})\Gamma(1+\lambda_3)} \end{aligned}$$

Let $\lambda_3 = \kappa$, then maximize the log–likelihood function $\ln\left(\hat{\lambda}_1, \hat{\lambda}_2, \lambda_3\right)$ to obtain $\hat{\kappa}_3$, that is,

$$\hat{\lambda}_3 = \arg\max_{\lambda_3}\left[\ln\left(\hat{\lambda}_1, \hat{\lambda}_2, \lambda_3\right)\right]$$

The only difference between two L–moment estimator for location a and scale b by Hosking et al. [56] and by Morrison and Smith [84] is that they choose a different estimator for λ_2.

We can obtain location and scale estimators using Theorem 2.13 and equation (3.6). Surprisingly, they are exactly the same as estimators by Morrison and Smith [84]:

$$\begin{aligned} \hat{a} &= \bar{X} - \frac{1-\Gamma(1+\hat{\kappa})}{(1-2^{-\kappa})\Gamma(1+\hat{\kappa})}\sum_{i<j}|X_j - X_i| \\ \hat{b} &= \frac{\hat{\kappa}}{(1-2^{-\kappa})\Gamma(1+\hat{\kappa})}\sum_{i<j}|X_j - X_i| \end{aligned}$$

5. New Mixed Method

Our new mixed method is based on our previous result. We first fix the shape parameter κ, then estimators of parameters a and b are given as function of κ:

$$\begin{aligned} \hat{a} &= \bar{X}_n - \left(\sqrt{\frac{n-1}{n}}\frac{\mu(\kappa)}{\sigma(\kappa)}\right) S_{X,n} \\ \hat{b} &= \left(\sqrt{\frac{n-1}{n}}\frac{1}{\sigma(\kappa)}\right) S_{X,n} \end{aligned}$$

Next we estimate the shape parameter κ by the maximum spacings (MPS) method.

Wong and Li [121] pointed out that the problem of an unbounded log-likelihood function is avoided, since it is always bounded from above by $-(n+1)\ln(n+1)$, and showed that efficient estimators can still be obtained by the maximum product of spacings for $\kappa \geq 1$ while the MLE fails.

The mean and variance of the GEV with $a = 0$ and $b = 1$ are given by:

$$\begin{aligned}
\mu(\kappa) &= \frac{1}{\kappa}\left[1-\Gamma(1+\kappa)\right], \quad \kappa > -1 \\
\sigma(\kappa) &= \frac{1}{|\kappa|}\sqrt{\Gamma(1+2\kappa)-\Gamma^2(1+\kappa)}, \quad \kappa > -\frac{1}{2}
\end{aligned}$$

Let

$$T_i = \left[1-\kappa\left(\frac{x_{(i)}-\hat{a}}{\hat{b}}\right)\right]^{\frac{1}{\kappa}}$$

For $\kappa > -\frac{1}{2}$, we have

$$\begin{aligned}
\kappa\frac{X_{(i)}-\hat{a}}{\hat{b}} &= \kappa\frac{X_{(i)}-\left[\bar{X}_n-\sqrt{\frac{n-1}{n}}\frac{\mu(\kappa)}{\sigma(\kappa)}S_{X,n}\right]}{\sqrt{\frac{n-1}{n}}\frac{1}{\sigma(\kappa)}S_{X,n}} \\
&= sign(\kappa)\sqrt{\frac{n}{n-1}}\frac{X_{(i)}-\bar{X}_n}{S_{X,n}}\sqrt{\Gamma(1+2\kappa)-\Gamma^2(1+\kappa)}+\left[1-\Gamma(1+\kappa)\right]
\end{aligned}$$

Then

$$1-\kappa\left(\frac{X_{(i)}-\hat{a}}{\hat{b}}\right) = \Gamma(1+\kappa)-sign(\kappa)\sqrt{\frac{n}{n-1}}\frac{X_{(i)}-\bar{x}_n}{S_{X,n}}\sqrt{\Gamma(1+2\kappa)-\Gamma^2(1+\kappa)}$$

This gives

$$T_i = \left[\Gamma(1+\kappa)-sign(\kappa)\sqrt{\frac{n}{n-1}}\frac{X_{(i)}-\bar{X}_n}{S_{X,n}}\sqrt{\Gamma(1+2\kappa)-\Gamma^2(1+\kappa)}\right]^{\frac{1}{\kappa}}$$

Now we have

$$M(\kappa) = -\sum_{i=1}^{n-1}\ln\left[e^{-T_{i+1}}-e^{-T_i}\right]-\ln\left(1-e^{-T_n}\right)+T_1$$

From that, we obtain the shape estimator by maximizing spacing product:

$$\hat{\kappa} = \arg\min_{\kappa} \left\{ -\sum_{i=1}^{n-1} \ln\left[e^{-T_{i+1}} - e^{-T_i}\right] - \ln\left(1 - e^{-T_n}\right) + T_1 \right\} \tag{3.7}$$

To improve the estimator of κ, we impose the constraint $\kappa(x-a) < b$ on the optimization problem in above equation. The procedure is implement by the function *fmincon*, which is a constrained nonlinear optimization tool provided by Matlab.

3.2.3 Simulation Results and Conclusions

This section will compare the proposed approach for estimating generalized extreme value distribution to L-moment. This will be done through a series of numerical experiments. Monte Carlo simulation are performed for parameters $a = 12,\ b = 5$, and $\kappa = 0.25,\ -0.25$ with sample sizes from 20 to 50. For each sample size, 10000 replicates are generated, and the mean square error (MSE) were calculated for the Mixed Method and L-moment method.

Figures 3.1-3.6 show how the MSE of the estimators vary as sample sizes are changed. These results are based on the average performance across the two methods and therefore provide a general idea about each technique's performance.

From Figures 3.1–3.3, we can see that the MSEs of the mixed estimators are smaller than that of L-moment estimators when $\kappa = 0.25$. Figures 3.4–3.6 show that MSE of estimators a and κ obtained from both methods are very close to each other respectively, and MSE of the mixed estimator of b is a little larger than that of the L-moment. There is a dramatic decrease in the MSE until sample size n reaches 25, except for the mixed estimator of κ when it is positive, whose MSE is small and stable. After that, the MSE decreases relatively slowly.

Figure 3.1: Location estimator of GEV with $a = 12, b = 5, \kappa = 0.25, 10000$ trials

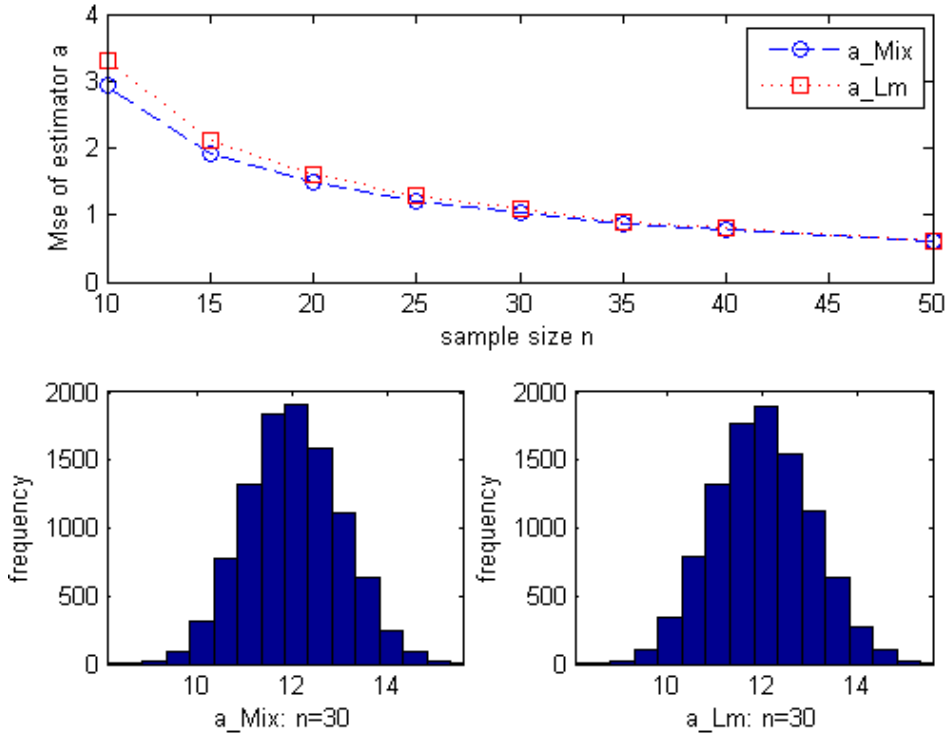


Figure 3.2: Scale estimator of GEV with $a = 12, b = 5, \kappa = 0.25, 10000$ trials

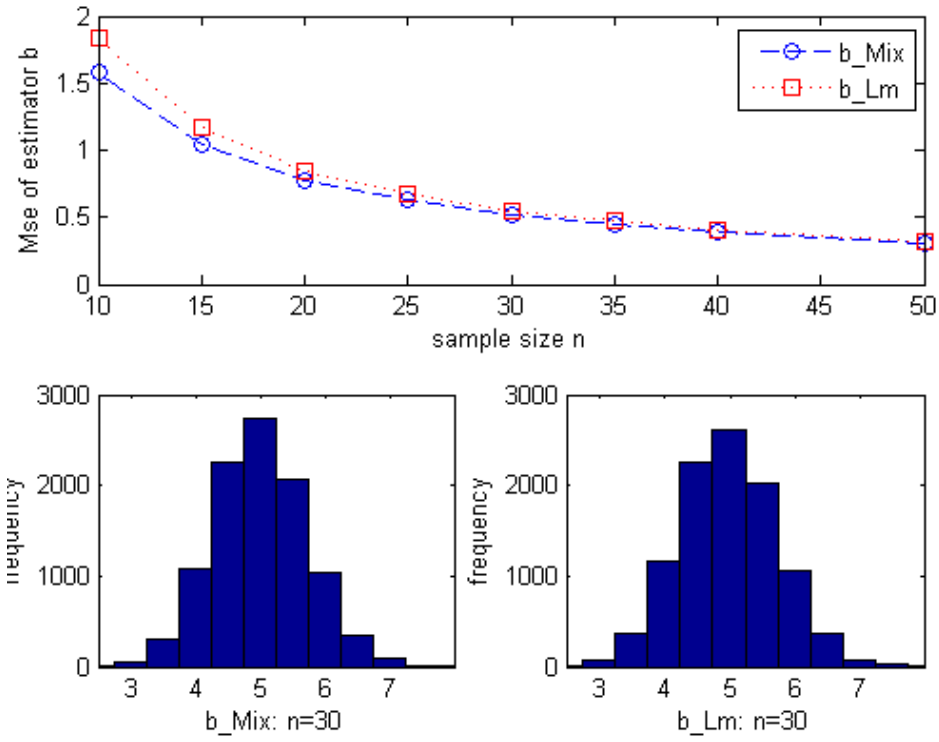


Figure 3.3: Shape estimator of GEV with $a = 12, b = 5, \kappa = 0.25, 10000$ trials

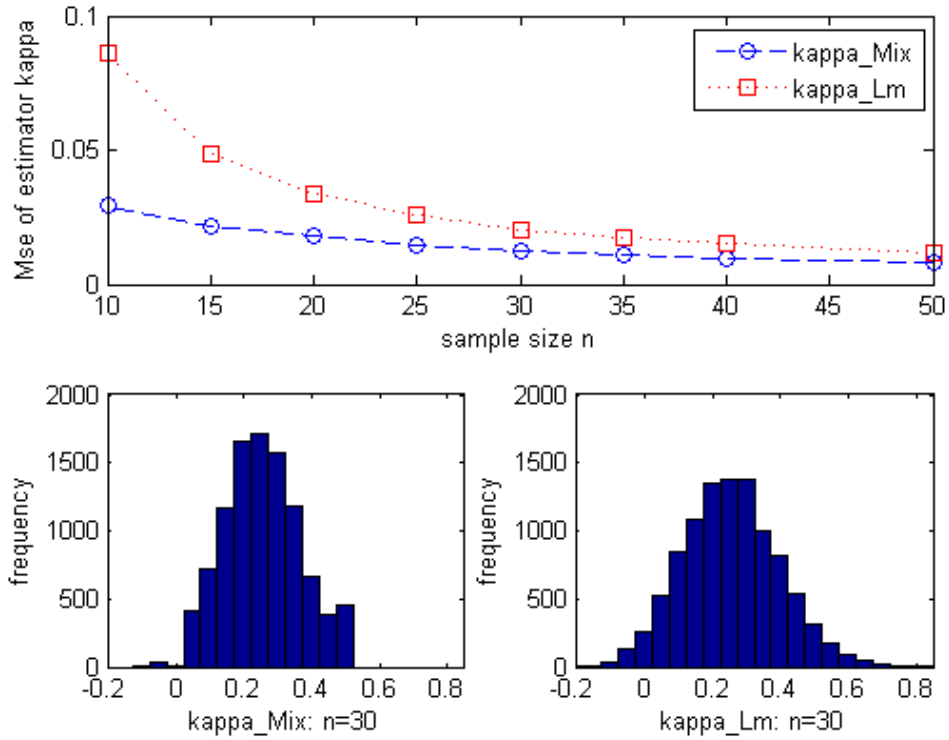


Figure 3.4: Location estimator of GEV with $a = 12, b = 5, \kappa = -0.25, 10000$ trials

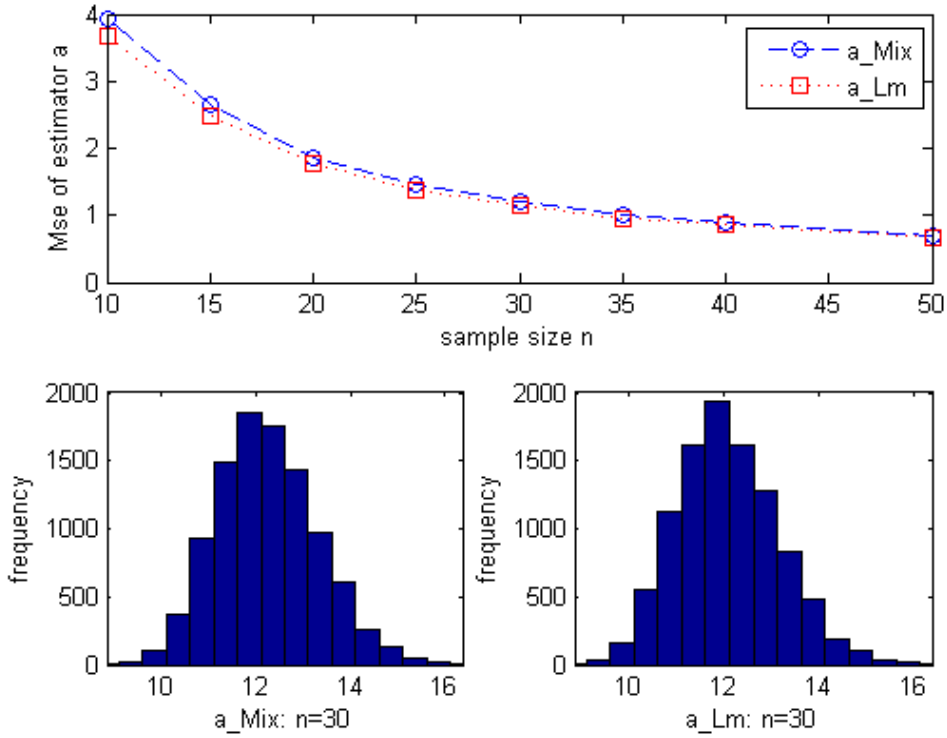


Figure 3.5: Scale estimator of GEV with $a = 12, b = 5, \kappa = -0.25, 10000$ trials

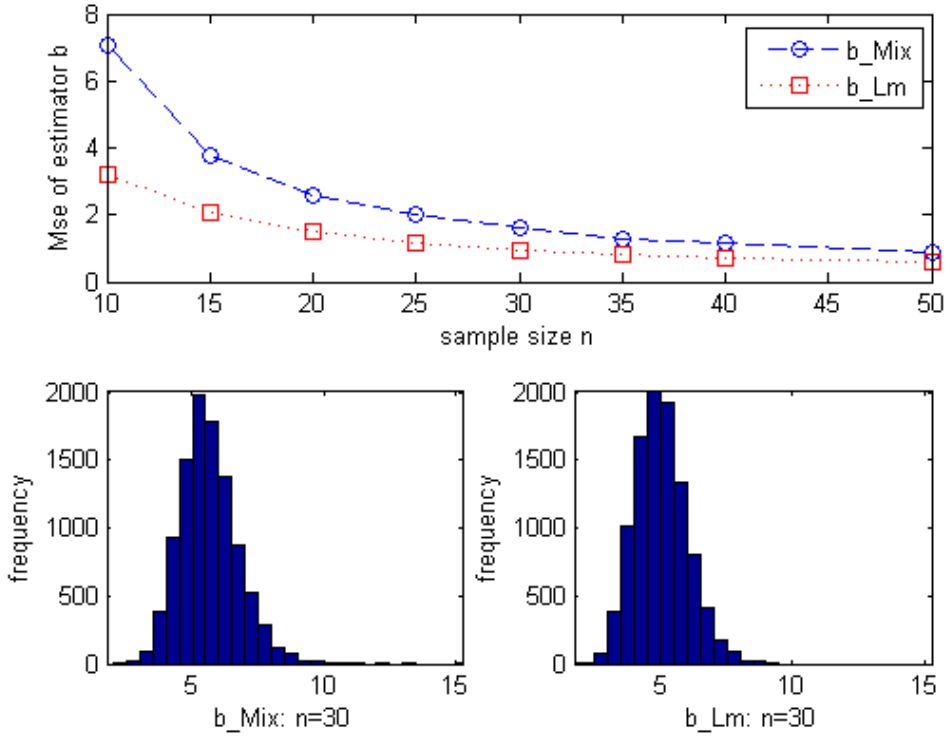
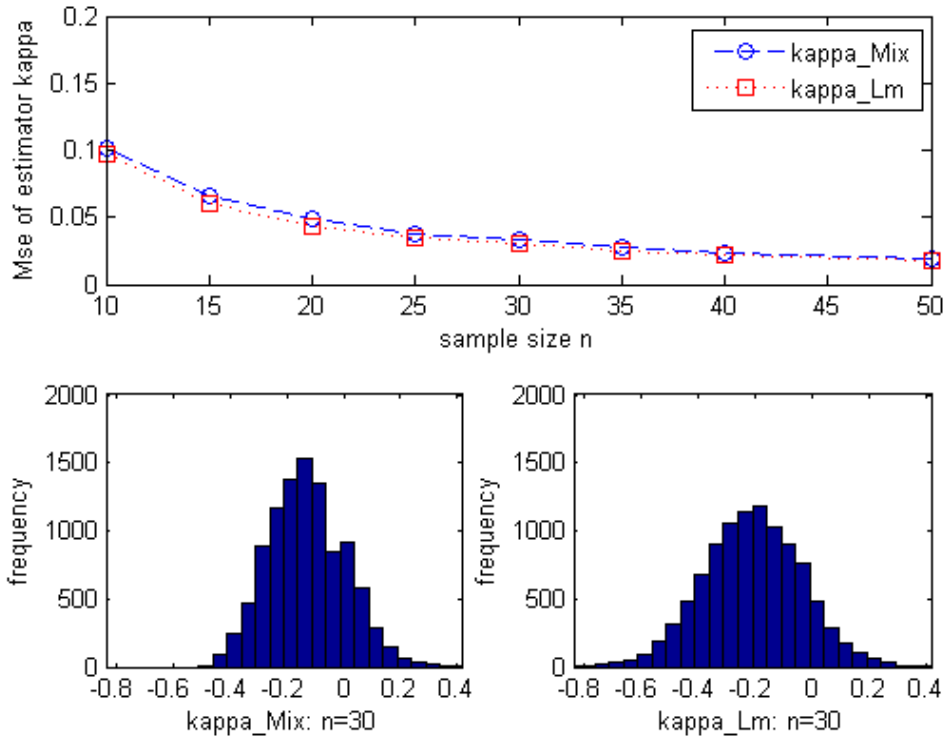


Figure 3.6: Shape estimator of GEV with $a = 12, b = 5, \kappa = -0.25, 10000$ trials



3.3 Example 2: Power Gumbel Distribution

As mentioned at the start of this chapter, examination of the power Gumbel distribution was motivated by problems in fitting bundle powers in CANDU reactors. The power Gumbel distribution is a generalization of the Gumbel distribution. By introducing a shape parameter r, the cdf of the random variable X is given by

$$F(x|a,b,r) = e^{-e^{-\left(\frac{x-a}{b}\right)^r}}, \quad -\infty < x < \infty,\ r \textit{ positive odd integer}$$

and the pdf by

$$f(x|a,b,r) = \frac{r}{b}\left(\frac{x-a}{b}\right)^{r-1} e^{-\left(\frac{x-a}{b}\right)^r} e^{-e^{-\left(\frac{x-a}{b}\right)^r}}$$

Letting$Z = \frac{X-a}{b}$, then

$$g(z) = rz^{r-1}e^{-z^r}e^{-e^{-z^r}}$$

and

$$f(x|a,b,r) = \frac{1}{b}g\left(\frac{x-a}{b}\right)$$

which is a location–scale family for fixed odd r.

3.3.1 Maximum Likelihood Method

Let $\{X_1, X_2, \cdots, X_n\}$ be a random sample from a power Gumbel distribution. The likelihood function is

$$L(a,b,r|\underline{x}) = \prod_{i=1}^{n} f(x_i|a,b,r) = \left(\frac{r}{b}\right)^n \prod_{i=1}^{n}\left(\frac{x_i-a}{b}\right)^{r-1} e^{-\left(\frac{x_i-a}{b}\right)^r} e^{-e^{-\left(\frac{x_i-a}{b}\right)^r}}$$

and the log–likelihood function is

$$\ln L(a,b,r|\underline{x}) = n(\ln r - \ln b) + \sum_{i=1}^{n}\ln\left(\frac{x_i-a}{b}\right)^{r-1} - \sum_{i=1}^{n}\left(\frac{x_i-a}{b}\right)^r - \sum_{i=1}^{n} e^{-\left(\frac{x_i-a}{b}\right)^r}$$

After some calculation, we get

$$\begin{aligned}
\frac{\partial L(a,b,r|\underline{x})}{\partial a} &= \sum_{i=1}^{n} \frac{1}{(x_i-a)^{r-1}}(r-1)(x_i-a)^{r-2}(-1) + \frac{r}{b}\sum_{i=1}^{n}\left(\frac{x_i-a}{b}\right)^{r-1} \\
&- \frac{r}{b}\sum_{i=1}^{n}\left(\frac{x_i-a}{b}\right)^{r-1} e^{-\left(\frac{x_i-a}{b}\right)^r} \\
&= \sum_{i=1}^{n}\frac{1-r}{x_i-a} + \frac{r}{b}\sum_{i=1}^{n}\left(\frac{x_i-a}{b}\right)^{r-1} - \frac{r}{b}\sum_{i=1}^{n}\left(\frac{x_i-a}{b}\right)^{r-1} e^{-\left(\frac{x_i-a}{b}\right)^r}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial L(a,b,r|\underline{x})}{\partial b} &= \frac{-n}{b} - \frac{n(r-1)}{b} + \frac{r}{b^{r+1}}\sum_{i=1}^{n}(x_i-a)^r - \frac{r}{b^{r+1}}\sum_{i=1}^{n}(x_i-a)^r e^{-\left(\frac{x_i-a}{b}\right)^r} \\
&= -\frac{nr}{b} + \frac{r}{b}\sum_{i=1}^{n}\left(\frac{x_i-a}{b}\right)^r - \frac{r}{b}\sum_{i=1}^{n}\left(\frac{x_i-a}{b}\right)^r e^{-\left(\frac{x_i-a}{b}\right)^r}
\end{aligned}$$

and, treating for the time being, r as a continuous parameter

$$\begin{aligned}
\frac{\partial L(a,b,r|\underline{x})}{\partial r} &= \frac{n}{r} + \sum_{i=1}^{n}\ln\left|\frac{x_i-a}{b}\right| - \sum_{i=1}^{n}\left(\frac{x_i-a}{b}\right)^r \ln\left|\frac{x_i-a}{b}\right| \\
&+ \sum_{i=1}^{n}\left(\frac{x_i-a}{b}\right)^r e^{-\left(\frac{x_i-a}{b}\right)^r}\ln\left|\frac{x_i-a}{b}\right|
\end{aligned}$$

The maximum likelihood estimators of a, b and r can be identified by solving the following system of equations, which correspond to setting to zero the first derivatives of $\ln L(a,b,r|\underline{x})$ with respect to each parameter:

$$\begin{cases}
\sum_{i=1}^{n}\frac{1-r}{x_i-a} + \frac{r}{b}\sum_{i=1}^{n}\left(\frac{x_i-a}{b}\right)^{r-1} - \frac{r}{b}\sum_{i=1}^{n}\left(\frac{x_i-a}{b}\right)^{r-1} e^{-\left(\frac{x_i-a}{b}\right)^r} = 0 \\
n - \sum_{i=1}^{n}\left(\frac{x_i-a}{b}\right)^r + \sum_{i=1}^{n}\left(\frac{x_i-a}{b}\right)^r e^{-\left(\frac{x_i-a}{b}\right)^r} = 0 \\
\frac{n}{r} + \sum_{i=1}^{n}\ln\left|\frac{x_i-a}{b}\right| - \sum_{i=1}^{n}\left(\frac{x_i-a}{b}\right)^r \ln\left|\frac{x_i-a}{b}\right| + \sum_{i=1}^{n}\left(\frac{x_i-a}{b}\right)^r e^{-\left(\frac{x_i-a}{b}\right)^r}\ln\left|\frac{x_i-a}{b}\right| = 0
\end{cases}$$

When solve above nonlinear systems, however, we may encounter local maximum and initial value problem, which may cause maximum-likelihood estimator unstable.

Below, we review the maximum likelihood estimation of an integer–valued parameter. Let $\{X_1, X_2, \cdots, X_n\}$ be a random sample with likelihood function $L(r, \underline{\theta})$, where r is an unknown integer and $\underline{\theta}$ is a parameter vector. Let $\hat{\underline{\theta}}(r)$ denote the value

of $\underline{\theta}$ which maximizes $L(r, \underline{\theta})$ for given r and let

$$L^*(r) = L(r, \hat{\underline{\theta}}(r)).$$

There are three ways to obtain a maximum likelihood estimator of the integer parameter r:

1. Solve $\frac{\partial L^*(r)}{\partial r} = 0$ and then insert the two nearest integers to $\hat{r}$ into the likelihood function with the integer yielding the largest value being the MLE of r.

2. MLE of r is the integer that satisfies

$$D(N+1) < 1 \leq D(r)$$

   where

$$D(r) = \frac{L^*(N)}{L^*(r-1)}$$

3. Dahiya (1981) If $L^*(r)$ is unimodal, then the MLE of r is

$$\hat{r} = \lceil V \rceil$$

   such that

$$L^*(V) = L^*(V-1)$$

   where $\lceil t \rceil$ denotes the greatest integer $\leq t$. If V happens to be an integer, then both V and $V-1$ are the integer MLEs of r.

3.3.2 Mixed Method

The basic idea is as follows:

1. Taking the location a and scale b to be functions of the shape r, respectively;

2. Estimate r by the maximum spacing product method;

3. Estimate a and b.

We present the details. Setting $t = e^{-z^r}$, for fixed odd r, we have

$$\begin{aligned}
\mu_Z(r) &= E(Z) = \int_{-\infty}^{\infty} zg(z)dz = \int_{-\infty}^{\infty} rz^r e^{-z^r} e^{-e^{-z^r}} dz \\
&= \int_{\infty}^{0} r(-\ln t)te^{-t} \left[\frac{1}{-rt(-\ln t)^{\frac{r-1}{r}}} \right] dt \\
&= \int_{0}^{\infty} (-\ln t)^{\frac{1}{r}} e^{-t} dt
\end{aligned}$$

and

$$\sigma_Z^2(r) = E(Z^2) - [E(Z)]^2 = \int_{-\infty}^{\infty} rz^{r+1} e^{-z^r} e^{-e^{-z^r}} dz - [\mu_Z(r)]^2$$

Similarly

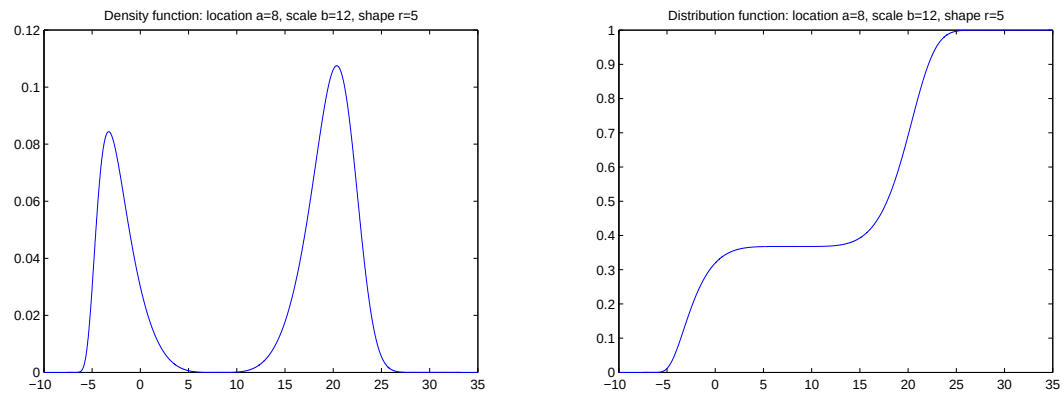
$$\begin{aligned}
\int_{-\infty}^{\infty} rz^{r+1} e^{-z^r} e^{-e^{-z^r}} dz &= \int_{-\infty}^{\infty} z^2 e^{-z^r} e^{-e^{-z^r}} dz^r = -\int_{-\infty}^{\infty} z^2 e^{-e^{-z^r}} de^{-z^r} \\
&= \int_{0}^{\infty} (-\ln t)^{\frac{2}{r}} e^{-t} dt
\end{aligned}$$

Using the previous result, we have

$$\begin{aligned}
\hat{a}(r) &= \bar{x}^* - \frac{2}{\sqrt{\frac{n_1}{n_1-1}} + \sqrt{\frac{n_2}{n_2-1}}} \frac{\mu_Z(r)}{\sigma_Z(r)} S_{X,n}^* \\
\hat{b}(r) &= \left(\sqrt{\frac{n-1}{n}} \frac{1}{\sigma_Z(r)} \right) S_{x,n}
\end{aligned}$$

Both expressions are functions of r. We then estimate r by using the maximum spacing product or maximum likelihood method. To find the MLE of r, we can search r from $1, 3, 5, \cdots, r_{max}$, where r_{max} is specified previously. Then $\hat{r}$ is a value that maximizes the likelihood function.

Below, we present a simple method to determine n_1. We first observe the graphs of the density and distribution functions.



(a) Density function (b) Distribution function

Figure 3.7: Distribution function and density of power Gumbel

We find that the probability for some intervals around the location is very small. This suggest determining n_1 as follow:

1. Arrange the data $\underline{x}$ in increasing order and check whether there exist any outliers. If so, remove them. The new dataset is denoted by $\underline{x}^{(1)}$. We only consider one single outlier here and let Lr denote

$$Lr = \begin{cases} 0 & \text{no outlier} \\ 1 & \text{otherwise} \end{cases}$$

2. To avoid $n_1 = 1$ or $n_2 = 1$, trim 5% of the highest and lowest data values from $\underline{x}^{(1)}$, and denote the remaining data by $\underline{x}^{(2)}$:

$$\underline{x}^{(2)} = \left(x^{(1)}_{(L)}, x^{(1)}_{(L+1)}, \cdots, x^{(1)}_{(U)} \right)$$

where $x^{(1)}_{(1)}, \cdots, x^{(1)}_{(L-1)}$ for $L \geq 2$ and $x^{(1)}_{(U+1)}, \cdots, x^{(1)}_{(n)}$ are removed by trimming.

3. Compute the differences in $\underline{x}^{(2)}$:

$$D_i = x^{(1)}_{(i+1)} - x^{(1)}_{(i)}, \quad i = L, L+1, \cdots, U-1$$

and write as

$$D = (D_1, \cdots, D_k, \cdots, D_{U-L})$$

So n_1 can be taken to be

$$n_1 = \min\{k | D_k = \max(D)\} + (L-1) + Lr \tag{3.8}$$

Then

$$\frac{x_{(i)} - \hat{a}}{\hat{b}} = \frac{x_{(i)} - \left[\bar{x}^* - \frac{2}{\sqrt{\frac{n_1}{n_1-1}} + \sqrt{\frac{n_2}{n_2-1}}} \frac{\mu_Z(r)}{\sigma_Z(r)} S^*_{X,n}\right]}{\sqrt{\frac{n-1}{n}} \frac{1}{\sigma(r)} S_{X,n}}$$

Let $T_i = e^{-\frac{x_{(i)} - \hat{a}}{\hat{b}}}$, so we have

$$M(r) = -\sum_{i=1}^{n-1} \ln\left[e^{-T_{i+1}} - e^{-T_i}\right] - \ln\left(1 - e^{-T_n}\right) + T_1$$

From that, we obtain the MSP estimator for shape r by maximizing spacing product:

$$r_{MSP} = \arg\min_{\kappa} \left\{ -\sum_{i=1}^{n-1} \ln\left[e^{-T_{i+1}} - e^{-T_i}\right] - \ln\left(1 - e^{-T_n}\right) + T_1 \right\}$$

and we take the estimator of r to be the odd number which is the closest to the MSP estimator r_{MSP}. After that, we can calculate $\mu_Z(\hat{r})$ and $\sigma_Z(\hat{r})$ by using Matlab, and then obtain $\hat{a}$ and $\hat{b}$.

3.3.3 Simulation Results and Conclusions

To illustrate the application of the mixed method we developed in this chapter, a simulation was conducted in order to compare the performance of maximum–likelihood estimators with the mixed method, based on the Bias, Mean Square Error (MSE), and sample statistics of the distribution. When experimenting with the mixed method and maximum likelihood method, different values for the parameter r, ranging between 1 and 11, were used for parameters $a = 12, b = 5$ and $a = 5, b = 12$, respectively.

The experimental results for the two groups of parameters a and b based on 100000 replications are presented in Tables 3.2 - 3.10 and Figures 3.8 - 3.9. In order to compare performance of the approaches we use Bias, MSE, mean, standard deviation, minimum and maximum of estimator.

Tables 3.2 - 3.10 summarize six such performance metrics. Some of the maximum likelihood estimators are very large, and attain 10^3 in the worse case. Thus, Bias and MSE in the rightmost column of Tables 3.3 and 3.5 are only computed for those estimators which are less than 30, and estimators with more than 30 are removed.

Figures 3.8 - 3.9 describe the distribution of estimators. The simulation results show that most mixed estimators of a and b are close to the true value of the parameters and the range is small too. On the contrary, the range of maximum likelihood estimators a and b are very wide. The figures show that mixed method also gives better accuracy rate for shape parameter r than the maximum likelihood estimator.

Table 3.1: $a = 12, b = 5, n = 30, 100000$ trials

	Estimator of location a							
	$\hat{a}$_mix				$\hat{a}$_mle			
r	Mean	Std	Min	Max	Mean	Std	Min	Max
3	12.0206	0.2719	9.5812	17.3333	13.6001	13.4683	0.3324	1.4162e+03
5	12.0164	0.1734	11.2913	15.9216	12.8965	30.7468	0.4903	227.5385
7	12.0161	0.1320	11.4829	16.6162	96.0672	266.6248	0.0035	6.7822e+03
9	12.0141	0.1082	11.5477	16.5579	92.1359	261.6533	0.0174	6.5949e+03
11	12.0118	0.0910	11.5365	16.7514	52.9689	199.6551	0.0119	5.5564e+03

Table 3.2: $a = 12, b = 5, n = 30, 100000$ trials

	Estimator of location a					
	$\hat{a}$_mix		$\hat{a}$_mle		$\hat{a}$_mle2 ($\hat{a}_mle < 30$)	
r	Bias	MSE	Bias	MSE	Bias	MSE
3	0.0206	0.0743	1.6001	183.9527	1.2457	3.9981
5	0.0164	0.0303	0.8965	946.1606	-0.6162	8.0179
7	0.0161	0.0177	84.0672	7.8155e+04	-3.1833	16.2461
9	0.0141	0.0119	80.1359	7.4884e+04	-4.1436	21.2169
11	0.0118	0.0084	40.9689	4.1540e+04	-4.4283	22.1025

Table 3.3: $a = 12, b = 5, n = 30, 100000$ trials

	Estimator of scale b							
	$\hat{b}$_mix				$\hat{b}$_mle			
r	Mean	Std	Min	Max	Mean	Std	Min	Max
3	4.9819	0.3412	1.5996	6.1714	5.8267	13.9480	1.1879	1.4272e+03
5	4.9936	0.1865	1.1046	5.7581	6.6835	31.5822	1.6809	242.9645
7	5.0057	0.1394	1.5887	5.5525	94.5657	271.1926	1.4745	6.7961e+03
9	5.0133	0.1208	1.1525	5.5008	91.2408	265.9397	1.3821	6.6087e+03
11	5.0183	0.1127	1.2303	5.5044	51.4653	202.8171	1.3884	5.5703e+03

Table 3.4: $a = 12, b = 5, n = 30, 100000$ trials

	Estimator of scale b					
	$\hat{b}$_mix		$\hat{b}$_mle		$\hat{b}$_mle2 ($\hat{b}_mle < 30$)	
r	Bias	MSE	Bias	MSE	Bias	MSE
3	-0.0181	0.1167	0.8267	195.2293	0.4367	0.7022
5	0.0064	0.0348	1.6835	1.0003e+03	0.0756	0.8493
7	0.0057	0.0195	89.5657	8.1567e+04	0.0307	2.8280
9	0.0133	0.0148	86.2408	7.8161e+04	0.2044	5.0187
11	0.0183	0.0130	46.4653	4.3293e+04	0.1712	5.8071

Table 3.5: $a = 12, b = 5, n = 30, 100000$ trials

	Estimator of shape r							
	$\hat{r}$_mix				$\hat{r}$_mle			
r	Mean	Std	Min	Max	Mean	Std	Min	Max
3	3.0367	0.3990	1	7	1.6074	2.8815	1	321
5	4.8611	1.1220	1	11	1.3919	6.1991	1	77
7	6.7382	1.6419	1	25	18.8844	53.7570	1	985
9	8.5604	2.1568	1	21	18.2189	52.5056	1	1135
11	10.2219	2.6591	1	25	10.4406	40.1993	1	975

Table 3.6: $a = 5, b = 12, n = 30, 100000$ trials

	Estimator of location a							
	$\hat{a}$_mix				$\hat{a}$_mle			
r	Mean	Std	Min	Max	Mean	Std	Min	Max
3	5.0494	0.6525	-0.8051	17.8000	9.9062	1.9741	0.0094	18.8497
5	5.0382	0.4213	2.6115	17.4711	11.0966	1.3263	0.0107	19.0519
7	5.0359	0.3192	3.3910	16.8912	11.4300	1.0390	0.1221	16.2673

Table 3.7: $a = 5, b = 12, n = 30, 100000$ trials

	Estimator of location a			
	$\hat{a}$_mix		$\hat{a}$_mle	
r	Bias	MSE	Bias	MSE
3	0.0494	0.4282	4.9062	27.9674
5	0.0382	0.1789	6.0966	38.9277
7	0.0359	0.1032	6.4300	42.4245

Table 3.8: $a = 5, b = 12, n = 30, 100000$ trials

	Estimator of scale b							
	$\hat{b}$_mix				$\hat{b}$_mle			
r	Mean	Std	Min	Max	Mean	Std	Min	Max
3	11.9565	0.8188	3.8391	14.8114	13.7772	1.6076	7.9525	23.7135
5	11.9983	0.4444	3.2039	13.7041	14.8264	1.5804	9.4967	23.1904
7	12.0209	0.3399	2.9902	13.4201	15.3398	1.5842	9.6682	23.9999

Table 3.9: $a = 5, b = 12, n = 30, 100000$ trials

	Estimator of scale b			
	$\hat{b}$_mix		$\hat{b}$_mle	
r	Bias	MSE	Bias	MSE
3	-0.0435	0.6723	2.8264	5.7429
5	-0.0017	0.1975	1.6835	10.4859
7	0.0209	0.1159	3.3398	13.6636

Table 3.10: $a = 5, b = 12, n = 30, 100000$ trials

	Estimator of shape r							
	$\hat{r}$_mix				$\hat{r}$_mle			
r	Mean	Std	Min	Max	Mean	Std	Min	Max
3	3.0367	0.3990	1	7	1.3448	0.7555	1	5
5	4.8594	1.1221	1	11	1.0761	0.3851	1	5
7	6.7331	1.6358	1	17	1.0138	0.1672	1	5

The following data are from simulation when $a = 12,\ b = 5\ and\ r = 7$:

```
18.0656, 16.5515, 17.4030, 17.5168, 16.8029, 17.8202,
18.0470, 17.1581, 16.7376, 16.4746, 18.3347, 16.5160,
16.3857, 18.1416, 18.5857, 16.8941, 16.9804, 17.0654,
16.2820,  8.7796, 17.7185, 17.9546, 16.4826, 16.8270,
16.1095, 17.3254, 17.0584, 16.2761, 16.3912, 17.6338
```

Our program gives mixed estimators as:

$$a_mix = 15.9216,\ b_mix = 1.1022,\ r_mix = 1$$

The MSE for parameters a and b are very large; this is caused by the split algorithm for n_1 and n_2. The split algorithm will trim the two smallest and two largest observation. Since there is only observation, 8.7796, smaller than location a, the split algorithm trims this observation and gives $n_1 = 10$. If we determine n_1 and n_2 by hand, that

is, $n_1 = 1$, then the mixed estimators are:

$$a_mix = 12.9681, \quad b_mix = 2.7901, \; r_mix = 1$$

We can see that mixed estimators of a and b have been improved.

Analyzing Tables 3.2 - 3.10 and Figures 3.8 - 3.9, it is apparent that the mixed estimators achieved smaller bias and MSE than the maximum likelihood estimators. The mixed method successfully utilizes the benefit from the new adaptive sample mean.

In short, the mixed method works satisfactorily. Not only does the mixed method yield closer estimates for data generated from a known parameter set, whereas the MLE does not, it also keeps a simple formulation and execution.

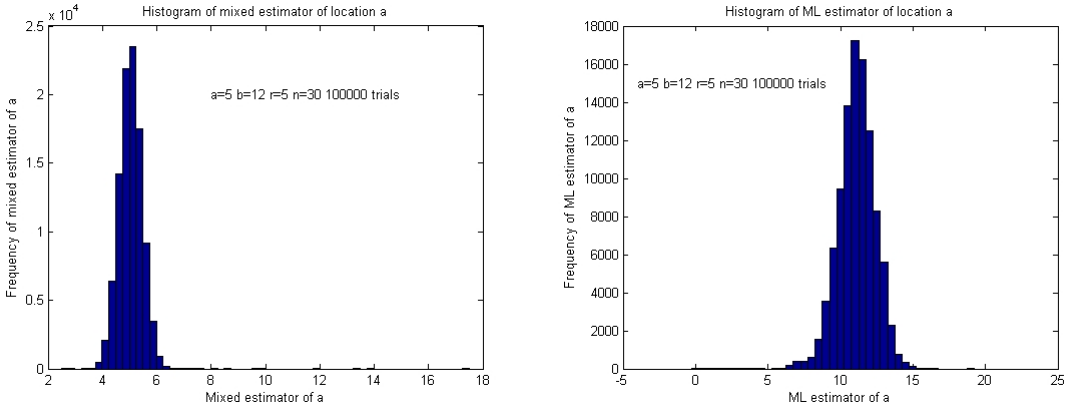


Figure 3.8: Location estimator of power Gumbel with $a = 5, b = 12, r = 5$

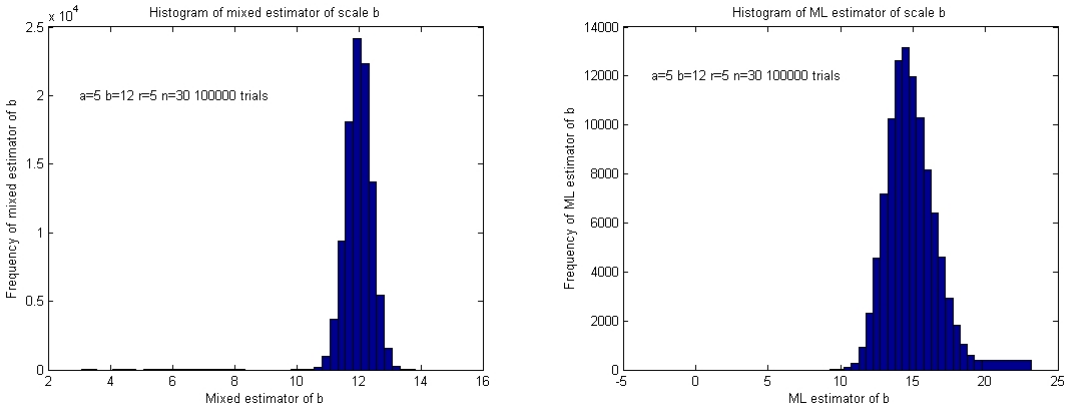


Figure 3.9: Scale estimator of power Gumbel with $a = 5, b = 12, r = 5$

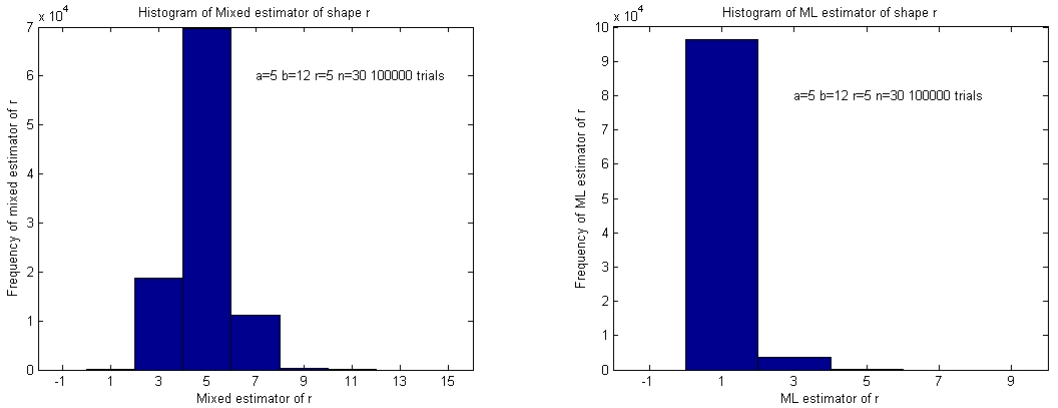


Figure 3.10: Shape estimator of power Gumbel with $a = 5, b = 12, r = 5$

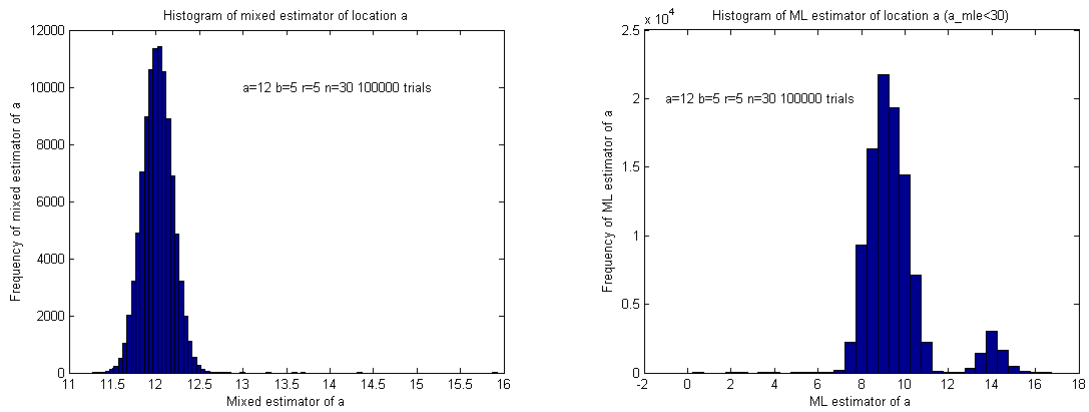


Figure 3.11: Location estimator of power Gumbel with $a = 1, b = 5, r = 5$

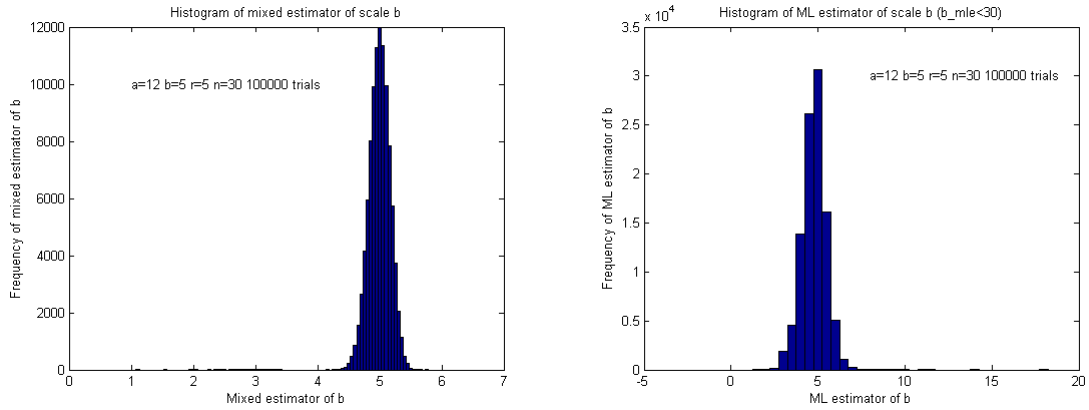


Figure 3.12: Scale estimator of power Gumbel with $a = 12, b = 5, r = 5$

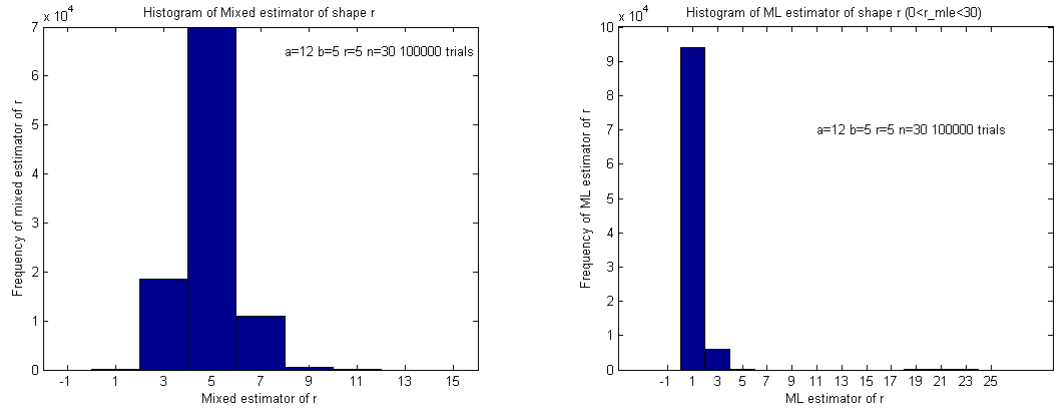


Figure 3.13: Shape estimator of power Gumbel with $a = 12, b = 5, r = 5$

Chapter 4

Bayesian Estimation for the Gumbel Minimum Distribution

4.1 Bayesian Method

Assume a random variable X has density function $f(x|\underline{\theta})$, where $\underline{\theta}$ is the vector parameter. Consider a random sample $\{X_1, \cdots, X_n\}$ of size n. The joint distribution is

$$f(\underline{x}|\underline{\theta}) = \prod_{i=1}^{n} f(x_i|\underline{\theta})$$

Bayesian inference differs from the classical statistical theory since unknown parameters are considered as random variables. For this reason, prior distributiona must be defined initially. This prior distribution expresses the information available to the researcher before any data are obtained.

Definition 4.1. The distribution of $\underline{\theta}$, $\pi(\underline{\theta})$, is called a prior distribution of $\underline{\theta}$. The conditional distribution of $\underline{\theta}$ given $\underline{X}$, $\pi(\underline{\theta}|\underline{x})$, is called the posterior distribution.

Interest lies in calculation of the posterior distribution $\pi(\underline{\theta}|\underline{x})$ of the parameters $\underline{\theta}$ given the observed data $\underline{x}$. According to Bayes theorem, the posterior distribution can be written as

$$\pi(\underline{\theta}|\underline{x}) = \frac{f(\underline{x}|\underline{\theta})\pi(\underline{\theta})}{f(\underline{x})} \propto f(\underline{x}|\underline{\theta})\pi(\underline{\theta})$$

The posterior distribution embodies both prior and observed data information, which is expressed by the prior distribution $\pi(\underline{\theta})$ and the likelihood $f(\underline{x}|\underline{\theta})$.

Definition 4.2. Let $\{X_1, \cdots, X_n\}$ denote a random sample of size n from a distribu-

tion that has pdf $f(x|\underline{\theta}), \underline{\theta} \in \Theta$. Let $t(\underline{X})$ denote an estimator of $g(\underline{\theta})$. The quality of the estimator $t(\underline{X})$ is measured by a real-valued loss function. For examples, quadratic loss function is defined to be

$$\mathcal{L}[t(\underline{x}), g(\underline{\theta})] = \|t(\underline{x}) - g(\underline{\theta})\|^2$$

Definition 4.3. Given a loss function $\mathcal{L}(\cdot)$ and an estimator $t(\underline{X})$, the risk is defined as the average loss over the density $f(\underline{x}|\underline{\theta})$, that is,

$$\begin{aligned} R[t(\underline{X}), g(\underline{\theta})] &= E_{\underline{X}}\{\mathcal{L}[t(\underline{X}), g(\underline{\theta})]\} \\ &= \int \mathcal{L}[t(\underline{x}), g(\underline{\theta})] f(\underline{x}|\underline{\theta}) d\underline{x} \end{aligned}$$

where $E_{\underline{X}}$ denotes that the expectation is over the distribution of $\underline{X}$ (with $\underline{\theta}$ fixed). Bayes risk is the risk averaged over the prior distribution on $\underline{\theta}$:

$$\begin{aligned} \mathcal{R}[\pi, t(\underline{X})] &= E_{\underline{\theta}}\{R[t(\underline{X}), g(\underline{\theta})]\} \\ &= \int\int \mathcal{L}[t(\underline{x}), g(\underline{\theta})] f(\underline{x}|\underline{\theta})\pi(\underline{\theta}) d\underline{\theta} d\underline{x} \end{aligned}$$

The Bayesian estimator of $g(\underline{\theta})$ is the value of $t(\underline{X})$ that minimizes the Bayes risk:

$$\hat{g}(\underline{\theta})_{Bay} = \arg\min_{t(\underline{X})} \mathcal{R}[\pi, t(\underline{X})]$$

The following result can be found in most statistics texts:

Theorem 4.1. *When the loss function is quadratic:*

$$\mathcal{L}[t(\underline{X}), g(\underline{\theta})] = \|t(\underline{X}) - g(\underline{\theta})\|^2$$

the Bayes estimator is given by

$$\begin{aligned} \hat{g}(\underline{\theta})_{Bay} &= E[g(\underline{\theta})|\underline{X}] = \int g(\underline{\theta})\pi(\underline{\theta}|\underline{x}) d\underline{\theta} \\ &= \frac{\int g(\underline{\theta}) f(\underline{x}|\underline{\theta})\pi(\underline{\theta}) d\underline{\theta}}{\int f(\underline{x}|\underline{\theta})\pi(\underline{\theta}) d\underline{\theta}} \end{aligned} \tag{4.1}$$

which is the posterior mean.

Let $L(\underline{\theta}) = \log f(\underline{x}|\underline{\theta})$ denote the log-likelihood function. Lindley [76] shows that the posterior mean of an arbitrary function $g(\underline{\theta})$ can be written as

$$E\left[g\left(\underline{\theta}\right)|\underline{X}\right] = \frac{\int g\left(\underline{\theta}\right)\pi(\underline{\theta})\exp\left[L(\underline{\theta})\right]d\underline{\theta}}{\int \pi(\underline{\theta})\exp\left[L(\underline{\theta})\right]d\underline{\theta}} \tag{4.2}$$

which can be asymptotically approximated by

$$\begin{aligned} E\left[g\left(\underline{\theta}\right)|\underline{X}\right] &= g\left(\underline{\theta}\right) + \frac{1}{2}\sum_i\sum_j\left(\frac{\partial^2 g}{\partial\theta_i\partial\theta_j} + 2\frac{\partial g}{\partial\theta_j}\frac{\partial\log\left[\pi(\underline{\theta})\right]}{\partial\theta_j}\right)\sigma_{ij} \\ &+ \frac{1}{2}\sum_i\sum_j\sum_k\sum_l\frac{\partial^3 L}{\partial\theta_i\partial\theta_j\partial\theta_k}\sigma_{ij}\sigma_{kl}\frac{\partial g}{\partial\theta_l} + O\left(\frac{1}{n^2}\right) \end{aligned}$$

where $i,\ ,j,\ k,\ l = 1,\ 2,\ \cdots,\ p$, $\underline{\theta} = (\theta_1,\ \cdots,\ \theta_p)$, and σ_{ij} denote the elements of the variance-covariance matrix of parameter estimators. All constants are evaluated at $\widehat{\underline{\theta}}$, the maximum likelihood estimate.

Tierney and Kadane [110] propose another approximation for the posterior mean:

$$E\left[g\left(\underline{\theta}\right)|\underline{X}\right] = \left(\frac{\det\Sigma^*}{\det\Sigma}\right)^{1/2} e^{n\left[L^*(\hat{\underline{\theta}}^*)-\tilde{L}(\hat{\underline{\theta}})\right]} + O\left(\frac{1}{n^2}\right) \tag{4.3}$$

where $\hat{\underline{\theta}}^*$ and $\hat{\underline{\theta}}$ maximize L^* and $\tilde{L}$, respectively, and $\det\Sigma^*$ and $\det\Sigma$ are negative the inverse Hessians of L^* and $\tilde{L}$ at $\hat{\underline{\theta}}^*$ and $\hat{\underline{\theta}}$, respectively,

$$\begin{aligned} \tilde{L} &= L\left(\underline{\theta}\right) + \frac{1}{n}\ln\left[\pi(\underline{\theta})\right] \\ L^* &= \frac{1}{n}\left[\tilde{L} + \ln g\left(\underline{\theta}\right)\right] \end{aligned}$$

But both above approximations are too complicated for Gumbel distribution, and not good for constructing prediction interval.

4.2 Full Bayesian Estimation for Gumbel Minimum Distribution

Consider a random variable X with pdf

$$f(x \mid a, b) = \frac{1}{b} e^{\frac{x-a}{b}} e^{-e^{\frac{x-a}{b}}}$$

where a and b are the location and scale parameters, respectively. In the Bayesian approach we regard a and b as random variables with a joint p.d.f. $\pi(a, b)$. We shall investigate the point estimator for both Jeffrey's non-informative prior and conjugate prior.

4.2.1 An Approximation Theorem for an Exponential Sum of Gumbel Random Variables

The Bayesian estimator for Gumbel minimum distribution obtained by the Lindley approximation or from Tierney and Kadane is implicit. In this section, we give a new approximation, which will be used to derive explicit full Bayesian estimators and Gibbs samplers in later chapter. First we introduce the following theorem, which is used to prove our result.

Theorem 4.2. *Korchevsky V.M., and Petrov V.V. [67].*

Let $\{X_n\}$ be a sequence of identically distributed random variables with finite mathematical expectation. Set $S_n = X_1 + X_2 + \cdots + X_n$. Suppose that

$$\begin{aligned} Var\left[\sum_{k=1}^{n}\left(X_k I_{\{0 \le X_k \le k\}}\right)\right] &\le C\sum_{k=1}^{n}\left[Var\left(X_k I_{\{0 \le X_k \le k\}}\right)\right] \quad (I) \\ Var\left[\sum_{k=1}^{n}\left(X_k I_{\{-k \le X_k \le 0\}}\right)\right] &\le C\sum_{k=1}^{n}\left[Var\left(X_k I_{\{-k \le X_k \le 0\}}\right)\right] \quad (II) \end{aligned}$$

where C is a constant. Then

$$\frac{S_n}{n} \longrightarrow E(X_1) \quad a.s.$$

Based on above result, we propose the following approximation.

Theorem 4.3. *Suppose that $\{X_1, X_2, \cdots, X_n\}$ is a random sample from the Gumbel minimum distribution with location parameter a and scale parameter b. Then for $p \geq 0$*

$$\frac{e^{(\sum_{i=1}^n x_i)/b}}{\left(\sum_{i=1}^n e^{x_i/b}\right)^{n+p}} \longrightarrow \frac{e^{-\left(\frac{p}{n}\sum_{i=1}^n x_i\right)/b}}{\left\{n\Gamma\left(\frac{1}{n}\right)\left[\Gamma\left(1-\frac{1}{n}\right)\right]^{n-1}\right\}^{n+p}}, \quad as\ n \longrightarrow \infty \tag{4.4}$$

Remark 4.1. For small n, we make following corrections:

$$\frac{e^{(\sum_{i=1}^n x_i)/b}}{\left(\sum_{i=1}^n e^{x_i/b}\right)^{n}} \approx \frac{1}{\left\{n\Gamma\left(\frac{1}{n}\right)\left[\Gamma\left(1-\frac{1}{n}\right)\right]^{n-1}\right\}^{nk_{\Delta_1}}}$$

and

$$\frac{e^{(\sum_{i=1}^n x_i)/b}}{\left(\sum_{i=1}^n e^{x_i/b}\right)^{n}} \approx \frac{e^{-\frac{q(\underline{x})}{b}}}{\left\{n\Gamma\left(\frac{1}{n}\right)\left[\Gamma\left(1-\frac{1}{n}\right)\right]^{n-1}\right\}^{nk_{\Delta_2}}} \tag{4.5}$$

where $k_{\Delta_1} > 0,\ k_{\Delta_2} > 0$ can be determined by Monte Carlo simulation, and $q(\underline{x}) > 0$. In this thesis, we will take

$$q(\underline{x}) = \frac{k_n S_{X,n}}{E S_{Z,n}} \tag{4.6}$$

where $X = a + bZ$, $E\left(\frac{S_{X,n}}{b}\right) = E(S_Z)$, and $k_n = k(n)$ is a function of sample size n. Again $E(S_Z)$ can be obtained by Monte Carlo. The following table shows that our result gives a good approximation.

Table 4.1: a=12 b= 5 $q(\underline{x}) = \frac{nS_{X,n}}{ES_{Z,n}}$ with 1000000 trials

n	$\frac{e^{(\sum_{i=1}^n x_i)/b}}{\left(\sum_{i=1}^n e^{x_i/b}\right)^n}$	$\frac{1}{\left\{n\Gamma\left(\frac{1}{n}\right)\left[\Gamma\left(1-\frac{1}{n}\right)\right]^{n-1}\right\}^{nk_{\Delta_1}}}$	$\frac{e^{-\frac{q(\underline{x})}{b}}}{\left\{n\Gamma\left(\frac{1}{n}\right)\left[\Gamma\left(1-\frac{1}{n}\right)\right]^{n-1}\right\}^{nk_{\Delta_2}}}$
10	3.3595×10^{-13}	5.1724×10^{-13}	5.3951×10^{-13}
20	2.5861×10^{-30}	1.5239×10^{-31}	1.5663×10^{-29}
30	9.3342×10^{-51}	2.4141×10^{-52}	2.6204×10^{-50}
50	2.1296×10^{-98}	5.4245×10^{-98}	3.4161×10^{-99}

Proof. We rewrite

$$\begin{aligned}
\frac{e^{(\sum_{i=1}^n x_i)t}}{\left(\sum_{i=1}^n e^{x_i t}\right)^{n+p}} &= \frac{\left[e^{\left(\frac{1}{n}\sum_{i=1}^n x_i\right)t}\right]^{n+p}\left[e^{\left(\frac{1}{n}\sum_{i=1}^n x_i\right)t}\right]^{-p}}{\left(\sum_{i=1}^n e^{x_i t}\right)^{n+p}} \\
&= \frac{\left[e^{\bar{x}t}\right]^{n+p}\left[e^{\left(\frac{1}{n}\sum_{i=1}^n x_i\right)t}\right]^{-p}}{\left(\sum_{i=1}^n e^{x_i t}\right)^{n+p}} \\
&= \frac{e^{-\left(\frac{p}{n}\sum_{i=1}^n x_i\right)t}}{\left(\sum_{i=1}^n e^{(x_i-\bar{x})t}\right)^{n+p}}
\end{aligned}$$

Let

$$R_i = X_i - \bar{X} = \frac{n-1}{n}X_i - \frac{1}{n}\sum_{j\neq i} X_j = \sum_{j=1}^n k_j X_j$$

with $k_i = \frac{n-1}{n}$ and $k_j = -\frac{1}{n}, j \neq i$, Then the moment generating function of R_i is given by

$$\begin{aligned}
\phi_{R_i}(t) &= E\left[e^{R_i t}\right] = E\left[e^{(X_i-\bar{X})t}\right] = E\left[e^{(X_i-\frac{1}{n}\sum_{i=1}^n X_i)t}\right] \\
&= \prod_{i=1}^n E\left[e^{k_j X_j t}\right] = \prod_{i=1}^n \left[\Gamma(1-k_j bt)e^{k_j at}\right] \\
&= e^{(\sum_{j=1}^n k_j)at}\prod_{i=1}^n \Gamma(1-k_j bt) \\
&= \Gamma\left(1-\frac{n-1}{n}bt\right)\left[\Gamma\left(1-\frac{1}{n}bt\right)\right]^{n-1}
\end{aligned}$$

where $\sum_{j=1}^n k_j = 0$. Since $e^{(X_i-\bar{X})t}$ is non-negative, condition II in Theorem 4.2 is

true for any non-negative number. On the other hand, there always exits

$$C \geq \frac{\sum_{k=1}^{n} Var\left(e^{R_k t}\right) + 2\sum_{i<j} Cov\left(e^{R_i t}, e^{R_j t}\right)}{\sum_{k=1}^{n} Var\left(e^{R_k t}\right)}$$

Then condition I in Theorem 4.2 is true. Therefore, we get

$$\begin{aligned}\sum_{i=1}^{n} e^{(x_i-\bar{x})t} &\approx nE\left[e^{R_i t}\right] \\ &= n\Gamma\left(1-\frac{n-1}{n}bt\right)\left[\Gamma\left(1-\frac{1}{n}bt\right)\right]^{n-1}\end{aligned}$$

Now taking and $b = \frac{1}{t}$, then

$$\frac{e^{(\sum_{i=1}^{n} x_i)/b}}{\left(\sum_{i=1}^{n} e^{x_i/b}\right)^{n+p}} \approx \frac{e^{-\left(\frac{p}{n}\sum_{i=1}^{n} x_i\right)/b}}{\left\{n\Gamma\left(\frac{1}{n}\right)\left[\Gamma\left(1-\frac{1}{n}\right)\right]^{n-1}\right\}^{n+p}}, \quad as\ n \longrightarrow \infty$$

In our following application, we are only interested in the case of $p = 0$. For small n, we have instead:

$$\frac{e^{(\sum_{i=1}^{n} x_i)/b}}{\left(\sum_{i=1}^{n} e^{x_i/b}\right)^{n}} \approx \frac{1}{\left\{n\Gamma\left(\frac{1}{n}\right)\left[\Gamma\left(1-\frac{1}{n}\right)\right]^{n-1}\right\}^{nk_{\Delta_1}}},$$

where $k_{\Delta_1} > 0$ is to be determined. We observe that

$$\frac{e^{(\sum_{i=1}^{n} x_i)/b}}{\left(\sum_{i=1}^{n} e^{x_i/b}\right)^{n}} = \frac{1}{\left(\sum_{i=1}^{n} e^{\frac{x_i-\bar{x}}{b}}\right)^{n}}$$

and $\frac{x_i-\bar{x}}{b}$ is independent of parameters a and b. To determine k_{Δ_1}, we generate B groups of random numbers from a standard Gumbel minimum distribution, and each group includes n data points, say $z_1^{(j)}, \cdots, z_n^{(j)}$, where superscript the j denoted j^{th} group. For the jth group, we have

$$\frac{e^{\sum_{i=1}^{n} z_i^{(j)}}}{\left(\sum_{i=1}^{n} e^{z_i^{(j)}}\right)^{n}} \approx \frac{1}{\left\{n\Gamma\left(\frac{1}{n}\right)\left[\Gamma\left(1-\frac{1}{n}\right)\right]^{n-1}\right\}^{nk_j}},$$

This gives

$$k_j = -\frac{\ln\left[\frac{e^{\frac{1}{n}\sum_{i=1}^{n} z_i^{(j)}}}{\left(\sum_{i=1}^{n} e^{z_i^{(j)}}\right)}\right]}{\ln\left\{n\Gamma\left(\frac{1}{n}\right)\left[\Gamma\left(1-\frac{1}{n}\right)\right]^{n-1}\right\}}$$

Then we take

$$k_{\Delta_1} = \frac{1}{B}\sum_{j=1}^{n} k_j \tag{4.7}$$

In the following Bayesian estimation and prediction, we also use this form:

$$\frac{e^{(\sum_{i=1}^{n} x_i)/b}}{\left(\sum_{i=1}^{n} e^{x_i/b}\right)^n} \approx \frac{e^{-\frac{q(\underline{x})}{b}}}{\left\{n\Gamma\left(\frac{1}{n}\right)\left[\Gamma\left(1-\frac{1}{n}\right)\right]^{n-1}\right\}^{nk_{\Delta_2}}}$$

where $q(\underline{x}) > 0$. As above, for the jth group sample, $z_1^{(j)}, \cdots, z_n^{(j)}$, we have

$$\frac{e^{\sum_{i=1}^{n} z_i^{(j)}}}{\left(\sum_{i=1}^{n} e^{z_i^{(j)}}\right)^n} \approx \frac{e^{-\frac{k_n S_{Z^{(j)},n}}{ES_{Z,n}}}}{\left\{n\Gamma\left(\frac{1}{n}\right)\left[\Gamma\left(1-\frac{1}{n}\right)\right]^{n-1}\right\}^{nk'_j}},$$

This gives

$$k'_j = -\frac{k_n S_{Z^{(j)},n}}{nE[S_{Z,n}]\ln\left\{n\Gamma\left(\frac{1}{n}\right)\left[\Gamma\left(1-\frac{1}{n}\right)\right]^{n-1}\right\}} - \frac{\ln\left[\frac{e^{\frac{1}{n}\sum_{i=1}^{n} z_i^{(j)}}}{\sum_{i=1}^{n} e^{z_i^{(j)}}}\right]}{\ln\left\{n\Gamma\left(\frac{1}{n}\right)\left[\Gamma\left(1-\frac{1}{n}\right)\right]^{n-1}\right\}}$$

Then we obtain

$$k_{\Delta_2} = \frac{1}{B}\sum_{j=1}^{n} k'_j \tag{4.8}$$

Therefore

$$\frac{e^{(\sum_{i=1}^{n} x_i)/b}}{\left(\sum_{i=1}^{n} e^{x_i/b}\right)^n} \approx \frac{e^{-\frac{k_n S_{X,n}}{bES_{Z,n}}}}{\left\{n\Gamma\left(\frac{1}{n}\right)\left[\Gamma\left(1-\frac{1}{n}\right)\right]^{n-1}\right\}^{nk_{\Delta_2}}}$$

4.2.2 Priors for the Gumbel Minimum Distribution

In Bayesian inference, the performance of the estimator or predictor depends on the prior distribution and also on the loss function used. An important problem in Bayesian analysis is how to define the prior distribution. If prior information about the parameter is available, it should be incorporated in the prior density. If we have no prior information, we want a prior with minimal influence on the inference. We call such a prior a noninformative prior.

Jeffrey's Non-informative Prior

Suppose that $\{X_1, X_2, \cdots, X_n\}$ is a random sample from a Gumbel minimum distribution with unknown location a and scale b. Then the likelihood function is given by

$$\begin{aligned} L(a, b \mid x_1, x_2, \cdots, x_n) &= \prod_{i=1}^{n} f(x_i \mid a, b) = \left(\frac{1}{b}\right)^n e^{\sum_{i=1}^{n} \frac{x_i - a}{b}} e^{-\sum_{i=1}^{n} e^{\frac{x_i - a}{b}}} \\ &= \left(\frac{1}{b}\right)^n e^{\sum_{i=1}^{n} \frac{x_i}{b}} e^{-\frac{na}{b}} e^{-e^{-\frac{a}{b}} \sum_{i=1}^{n} e^{\frac{x_i}{b}}} \end{aligned}$$

Jeffrey's non-informative prior chooses the prior $\pi(a, b)$ to be proportional to $|\det I(a, b)|^{\frac{1}{2}}$, where $I(a, b)$ is the expected Fisher information matrix. For a location-scale model $f(x) = \frac{1}{b} g\left(\frac{x-a}{b}\right)$, Shao (p.113, 2005) shows that

$$I(a, b) = \frac{n}{b^2} \begin{pmatrix} \int \frac{[g'(t)]^2}{g(t)} dt & \int \frac{g'(t)[tg'(t)+g(t)]}{g(t)} dt \\ \int \frac{g'(t)[tg'(t)+g(t)]}{g(t)} dt & \int \frac{[tg'(t)+g(t)]^2}{g(t)} dt \end{pmatrix}$$

implying that the Jeffrey's non-informative prior is given by

$$\pi(a, b) = \frac{1}{b^2}$$

Asymptotically Conjugate Prior

In Bayesian theory, if the posterior distributions $\pi(\underline{\theta}|\underline{x})$ are in the same family as the prior probability distribution $\pi(\underline{\theta})$ with different parameters, the prior is then called a conjugate prior for the likelihood. Finks [37] describe how to construct conjugate prior:

"A conjugate prior is constructed by first factoring the likelihood function into two parts. One factor must be independent of the parameter(s) of interest but may be dependent on the data. The second factor is a function dependent on the parameter(s) of interest and dependent on the data only through the sufficient statistics. The conjugate prior family is defined to be proportional to this second factor."

Raiffa & Schlaifer [97], and DeGroot [25] point out that when there exists a set of fixed-dimension sufficient statistics there must exist a conjugate prior family. By equality (4.5), we have

$$r \approx \Delta^{k\Delta_2} e^{\frac{q(\underline{x})+\sum_{i=1}^{n} x_i}{nb}}$$

where $\Delta = n\Gamma\left(\frac{1}{n}\right)\left[\Gamma\left(1-\frac{1}{n}\right)\right]^{n-1}$. We can then rewrite the log–likelihood function for the Gumbel minimum distribution as

$$\begin{aligned} L(a,b \mid x_1, x_2, \cdots, x_n) &= \prod_{i=1}^{n} f(x_i \mid a,b) = \left(\frac{1}{b}\right)^n e^{\sum_{i=1}^{n} \frac{x_i}{b}} e^{-n\frac{a}{b}} e^{-re^{-\frac{a}{b}}} \\ &\approx \left(\frac{1}{b}\right)^n e^{-\frac{\sum_{i=1}^{n} x_i}{b}} e^{-n\frac{a}{b}} e^{-(\Delta^{k\Delta_2} e^{\frac{q(\underline{x})+\sum_{i=1}^{n} x_i}{nb}})e^{-\frac{a}{b}}} \end{aligned}$$

According to the Neyman Factorization Theorem, $(q(\underline{x}), \sum_{i=1}^{n} x_i)$ is a joint sufficient statistic for parameters a and b.

We now consider the prior

$$\pi(a,b) = \frac{r^n[q(\underline{x})]^{2n-2}}{\Gamma(n)\Gamma(2n-2)} \frac{1}{b^{n+2}} e^{-\frac{q(\underline{x})}{b}} \exp\left[-n\frac{a}{b} - r\exp\left(-\frac{a}{b}\right)\right] \tag{4.9}$$

It can be approximated as

$$\pi(a,b) \approx \frac{\Delta^{nk_{\Delta_2}}[q(\underline{x})]^{2n-2}}{\Gamma(n)\Gamma(2n-2)} \frac{1}{b^{n+2}} e^{-\frac{\sum_{i=1}^n x_i}{b}} \exp\left[-n\frac{a}{b} - (\Delta^{k\Delta_2} e^{\frac{q(\underline{x})+\sum_{i=1}^n x_i}{nb}}) \exp\left(-\frac{a}{b}\right)\right]$$

which is a conjugate prior for parameters a and b according to Fink (1997). Next, we will verify the posterior prior $\pi(a,b|\underline{x})$ belong to the same family as $\pi(a,b)$. Let

$$\begin{aligned}
K &= \int_0^\infty \int_{-\infty}^\infty \pi(a,b) L(a,b \mid x_1, x_2, \cdots, x_n) da db \\
&= \int_0^\infty \int_{-\infty}^\infty \frac{r^n [q(\underline{x})]^{2n-2}}{\Gamma(n)\Gamma(2n-2)} \frac{1}{b^{n+2}} e^{-\frac{q(\underline{x})}{b}} \exp\left[-n\frac{a}{b} - r\exp\left(-\frac{a}{b}\right)\right] \frac{1}{b^n} e^{\sum_{i=1}^n \frac{x_i}{b}} e^{-\frac{na}{b}} e^{-re^{-\frac{a}{b}}} da db \\
&= \frac{[q(\underline{x})]^{2n-2}}{\Gamma(n)\Gamma(2n-2)} \int_0^\infty \frac{r^n}{b^{2n+2}} e^{-\frac{q(\underline{x})}{b}} e^{\frac{1}{b}\sum_{i=1}^n x_i} \left[\int_{-\infty}^\infty e^{-2n\frac{a}{b}} e^{-2re^{-\frac{a}{b}}} da\right] db \\
&= \frac{[q(\underline{x})]^{2n-2}}{\Gamma(n)\Gamma(2n-2)} \int_0^\infty \frac{r^n}{b^{2n+1}} e^{-\frac{q(\underline{x})}{b}} e^{\frac{1}{b}\sum_{i=1}^n x_i} \frac{\Gamma(2n)}{(2r)^{2n}} db \\
&= \frac{\Gamma(2n)[q(\underline{x})]^{2n-2}}{2^{2n}\Gamma(n)\Gamma(2n-2)} \int_0^\infty \frac{1}{b^{2n+1}} e^{-\frac{q(\underline{x})}{b}} \frac{e^{\frac{1}{b}\sum_{i=1}^n x_i}}{r^n} db \\
&\approx \frac{\Gamma(2n)[q(\underline{x})]^{2n-2}}{2^{2n}\Gamma(n)\Gamma(2n-2)} \times \frac{\Gamma(2n)}{\Delta^{k_{\Delta_2}}[2q(\underline{x})]^{2n}} \\
&= \frac{\Gamma^2(2n)}{\Delta^{k_{\Delta_2}} 4^{2n}\Gamma(n)\Gamma(2n-2)[q(\underline{x})]^2}
\end{aligned}$$

using approximation (4.5) in the second last equality. Therefore, we obtain the posterior distribution:

$$\begin{aligned}
\pi(a,b|x_1,\cdots,x_n) &= \frac{\pi(a,b)L(a,b \mid x_1,x_2,\cdots,x_n)}{\int_0^\infty \int_{-\infty}^\infty \pi(a,b)L(a,b \mid x_1,x_2,\cdots,x_n) da db} \\
&= \frac{\frac{r^n[q(\underline{x})]^{2n-2}}{\Gamma(n)\Gamma(2n-2)} \frac{1}{b^{n+2}} e^{-\frac{q(\underline{x})}{b}} \exp\left[-n\frac{a}{b} - r\exp\left(-\frac{a}{b}\right)\right] \frac{1}{b^n} e^{\sum_{i=1}^n \frac{x_i}{b}} e^{-\frac{na}{b}} e^{-re^{-\frac{a}{b}}}}{\frac{\Gamma^2(2n)}{\Delta^{k_{\Delta_2}} 4^{2n}\Gamma(n)\Gamma(2n-2)[q(\underline{x})]^2}} \\
&= \frac{(2r)^{2n}[2q(\underline{x})]^{2n}}{\Gamma(2n)\Gamma(2n)} \frac{1}{b^{2n+2}} e^{-\frac{2q(\underline{x})}{b}} \exp\left[-2n\frac{a}{b} - 2r\exp\left(-\frac{a}{b}\right)\right] \qquad (4.10)
\end{aligned}$$

which has a similar structure with $\pi(a,b)$, and thus shows that $\pi(a,b)$ is a conjugate prior.

4.2.3 Bayesian Estimators Based on Jeffrey's Non-informative Prior

By Bayes theorem, the posterior distribution of a and b is

$$\pi(a,b \mid x_1, x_2, \cdots, x_n) = \frac{\pi(a,b)L(a,b \mid x_1, x_2, \cdots, x_n)}{\int_0^\infty \int_{-\infty}^\infty \pi(a,b)L(a,b \mid x_1, x_2, \cdots, x_n)dadb}$$

Put

$$\begin{aligned} K &= \int_0^\infty \int_{-\infty}^\infty \pi(a,b)L(a,b \mid x_1, x_2, \cdots, x_n)dadb \\ &= \int_0^\infty \int_{-\infty}^\infty \left(\frac{1}{b}\right)^{n+2} e^{\sum_{i=1}^n \frac{x_i}{b}} e^{-\frac{na}{b}} e^{-e^{-\frac{a}{b}} \sum_{i=1}^n e^{\frac{x_i}{b}}} dadb \\ &= \int_0^\infty \left(\frac{1}{b}\right)^{n+2} e^{\sum_{i=1}^n \frac{x_i}{b}} \left[\int_{-\infty}^\infty e^{-\frac{n}{b}a} e^{-re^{-\frac{a}{b}}} da\right] db \end{aligned}$$

since

$$\int_{-\infty}^\infty e^{-\frac{n}{b}a} e^{-re^{-\frac{a}{b}}} da = \frac{\Gamma(n)}{r^n} b$$

Using Theorem 4.3 and Proposition A.8, we get

$$\begin{aligned} K &= \Gamma(n) \int_0^\infty \left(\frac{1}{b}\right)^{n+1} \frac{e^{\frac{1}{b}(\sum_{i=1}^n x_i)}}{(\sum_{i=1}^n e^{\frac{x_i}{b}})^n} db \\ &= \frac{\Gamma(n)}{\left\{n\Gamma\left(\frac{1}{n}\right)\left[\Gamma\left(1-\frac{1}{n}\right)\right]^{n-1}\right\}^{k_{\Delta_2} n}} \int_0^\infty \left(\frac{1}{b}\right)^{n+1} e^{-\frac{q(\underline{x})}{b}} db \\ &= \frac{\Gamma^2(n)}{\Delta^{nk_{\Delta_2}} [q(\underline{x})]^n} \end{aligned}$$

where

$$\Delta = n\Gamma\left(\frac{1}{n}\right)\left[\Gamma\left(1-\frac{1}{n}\right)\right]^{n-1}$$

We first estimate a and b under squared error (SE) loss. Then

$$\begin{aligned} &\int_0^\infty \int_{-\infty}^\infty a\pi(a,b)L(a,b \mid x_1, x_2, \cdots, x_n)dadb \\ = &\int_0^\infty \left(\frac{1}{b}\right)^{n+2} e^{\sum_{i=1}^n \frac{x_i}{b}} \left[\int_{-\infty}^\infty a e^{-\frac{n}{b}a} e^{-e^{-\frac{a}{b}} \sum_{i=1}^n e^{\frac{x_i}{b}}} da\right] db \end{aligned}$$

Let $t = e^{-\frac{a}{b}}$, and notice that

$$\frac{1}{\left(\sum_{i=1}^{n} e^{\frac{x_i}{b}}\right)^n} = \frac{e^{-\frac{1}{b}\sum_{i=1}^{n} x_i}}{\left\{n\Gamma\left(\frac{1}{n}\right)\left[\Gamma\left(1-\frac{1}{n}\right)\right]^{n-1}\right\}^{nk_{\Delta_1}}}$$

Thus

$$\ln\left(\sum_{i=1}^{n} e^{\frac{x_i}{b}}\right) = \frac{1}{bn}\sum_{i=1}^{n} x_i + k_{\Delta_1}\ln\Delta$$

Again let $t = e^{-\frac{a}{b}}$, then

$$\begin{aligned}
& e^{\frac{1}{b}\sum_{i=1}^{n} x_i}\int_{-\infty}^{\infty} a e^{-\frac{n}{b}a} e^{-e^{-\frac{a}{b}}\sum_{i=1}^{n} e^{\frac{x_i}{b}}} da \\
= & -b^2 e^{\frac{1}{b}\sum_{i=1}^{n} x_i}\int_{0}^{\infty} t^{n-1} e^{-\left(\sum_{i=1}^{n} e^{\frac{x_i}{b}}\right)t}\ln(t)dt \\
= & -b^2\frac{e^{\frac{1}{b}\sum_{i=1}^{n} x_i}}{(\sum_{i=1}^{n} e^{\frac{x_i}{b}})^n}\left[\Gamma'(n) - \Gamma(n)\ln\left(\sum_{i=1}^{n} e^{\frac{x_i}{b}}\right)\right] \\
= & b^2\frac{e^{\frac{1}{b}\sum_{i=1}^{n} x_i}}{(\sum_{i=1}^{n} e^{\frac{x_i}{b}})^n}\left[\Gamma(n)\left(\frac{1}{bn}\sum_{i=1}^{n} x_i + k_{\Delta_1}\ln\Delta\right) - \Gamma'(n)\right] \\
= & \frac{e^{-\frac{q(\underline{x})}{b}}}{\Delta^{nk_{\Delta_2}}}\left[\Gamma(n)\left(b\bar{x} + k_{\Delta_1} b^2\ln\Delta\right) - b^2\Gamma'(n)\right]
\end{aligned}$$

so

$$\begin{aligned}
& \int_0^{\infty}\int_{-\infty}^{\infty} a\pi(a,b)L(a,b \mid x_1, x_2, \cdots, x_n)dadb \\
= & \int_0^{\infty}\left(\frac{1}{b}\right)^{n+2}\frac{e^{-\frac{p(\underline{x})}{b}}}{\Delta^{nk_{\Delta_2}}}\left[\Gamma(n)\left(b\bar{x} + k_{\Delta_1} b^2\ln\Delta\right) - b^2\Gamma'(n)\right]db \\
= & \frac{1}{\Delta^{nk_{\Delta_2}}}\int_0^{\infty}\left(\frac{1}{b}\right)^{n+2} e^{-\frac{q(\underline{x})}{b}}\left\{b\Gamma(n)\bar{x} + b^2\left[k_{\Delta_1}\Gamma(n)\ln\Delta - \Gamma'(n)\right]\right\}db \\
= & \frac{1}{\Delta^{nk_{\Delta}}}\left[\bar{x}\Gamma(n)\int_0^{\infty}\left(\frac{1}{b}\right)^{n+1} e^{-\frac{q(\underline{x})}{b}}db + \left[k_{\Delta}\Gamma(n)\ln\Delta - \Gamma'(n)\right]\int_0^{\infty}\left(\frac{1}{b}\right)^{n} e^{-\frac{p(\underline{x})}{b}}db\right] \\
= & \frac{1}{\Delta^{nk_{\Delta_2}}}\left[\bar{x}\Gamma(n)\int_0^{\infty} t^{n-1}e^{-q(\underline{x})t}dt + \left[k_{\Delta_1}\Gamma(n)\ln\Delta - \Gamma'(n)\right]\int_0^{\infty} t^{n-2}e^{-q(\underline{x})t}dt\right] \\
= & \frac{\Gamma^2(n)}{\Delta^{nk_{\Delta_2}}[q(\underline{x})]^n}\bar{x} + \frac{k_{\Delta_1}\Gamma(n)\ln\Delta - \Gamma'(n)}{\Delta^{nk_{\Delta_2}}[q(\underline{x})]^{n-1}}\Gamma(n-1)
\end{aligned}$$

Therefore

$$\begin{aligned}\hat{a}_B &= \bar{x} + \frac{k_{\Delta_1}\Gamma(n)\ln\Delta - \Gamma'(n)}{(n-1)\Gamma(n)} q(\underline{x}) \\ &= \bar{x} + \frac{k_{\Delta_1}\ln\Delta - \psi(n)}{n-1} q(\underline{x})\end{aligned}$$

where $\psi(\cdot)$ denotes the digamma function. For $n \geq 0$

$$\begin{aligned}&\int_0^\infty \int_{-\infty}^\infty b\pi(a,b)L(a,b \mid x_1, x_2, \cdots, x_n)dadb \\ =\ & \int_0^\infty \left(\frac{1}{b}\right)^{n+1} e^{\sum_{i=1}^n \frac{x_i}{b}} \left[\int_{-\infty}^\infty e^{-\frac{n}{b}a} e^{-e^{-\frac{a}{b}}\sum_{i=1}^n e^{\frac{x_i}{b}}} da\right] db \\ =\ & \Gamma(n)\int_0^\infty \left(\frac{1}{b}\right)^n \frac{e^{\frac{1}{b}(\sum_{i=1}^n x_i)}}{(\sum_{i=1}^n e^{\frac{x_i}{b}})^n} db \\ =\ & \frac{\Gamma(n)}{\Delta^{nk_{\Delta_2}}} \int_0^\infty \left(\frac{1}{b}\right)^n e^{-\frac{q(\underline{x})}{b}} db \\ =\ & \frac{\Gamma(n)}{\Delta^{nk_{\Delta_2}}} \frac{\Gamma(n-1)}{[q(\underline{x})]^{n-1}}\end{aligned}$$

Then for $n \geq 0$,

$$\begin{aligned}\hat{b}_B &= \frac{1}{K}\int_0^\infty \int_{-\infty}^\infty b\pi(a,b)L(a,b \mid x_1, x_2, \cdots, x_n)dadb \\ &= \frac{1}{n-1} q(\underline{x})\end{aligned}$$

Recall that $q(\underline{x})$ has the form

$$q(\underline{x}) = \frac{k_n S_{X,n}}{ES_{Z,n}}$$

We choose $k_n = n-1$ so that $\hat{b}_B$ is unbiased, that is,

$$q(\underline{x}) = \frac{(n-1)S_{X,n}}{ES_{Z,n}}$$

Thus, the Bayesian estimators are given by

$$\hat{a}_B = \bar{x} + [k_{\Delta_1}\ln\Delta - \psi(n)]\frac{S_{X,n}}{ES_{Z,n}} \tag{4.11}$$

$$\hat{b}_B = \frac{S_{X,n}}{ES_{Z,n}} \tag{4.12}$$

4.2.4 Bayesian Estimators Based on an Asymptotically Conjugate Prior

To estimate the location parameter a, we need to calculate

$$\begin{aligned}
&\int_0^\infty \int_{-\infty}^\infty a\pi(a,b)L(a,b \mid x_1, x_2, \cdots, x_n)dadb \\
= &\int_0^\infty \int_{-\infty}^\infty a\frac{r^n[q(\underline{x})]^{2n-2}}{\Gamma(n)\Gamma(2n-2)}\frac{1}{b^{n+2}}e^{-\frac{q(\underline{x})}{b}}\exp\left[-n\frac{a}{b} - r\exp\left(-\frac{a}{b}\right)\right]\left(\frac{1}{b}\right)^n e^{\sum_{i=1}^n \frac{x_i}{b}}e^{-\frac{na}{b}}e^{-re^{-\frac{a}{b}}}dadb \\
= &\frac{[q(\underline{x})]^{2n-2}}{\Gamma(n)\Gamma(2n-2)}\int_0^\infty \int_{-\infty}^\infty \frac{ar^n}{b^{2n+2}}e^{-\frac{q(\underline{x})}{b}}e^{-2n\frac{a}{b}}e^{\sum_{i=1}^n \frac{x_i}{b}}e^{-2re^{-\frac{a}{b}}}dadb \\
= &\frac{[q(\underline{x})]^{2n-2}}{\Gamma(n)\Gamma(2n-2)}\int_0^\infty \frac{r^n}{b^{2n+2}}e^{-\frac{q(\underline{x})}{b}}e^{\sum_{i=1}^n \frac{x_i}{b}}\left[\int_{-\infty}^\infty ae^{-2n\frac{a}{b}}e^{-2re^{-\frac{a}{b}}}da\right]db
\end{aligned}$$

As in the last section,

$$\ln\left(\sum_{i=1}^n e^{\frac{x_i}{b}}\right) = \frac{1}{bn}\sum_{i=1}^n x_i + k_{\Delta_1}\ln\Delta$$

Let $t = e^{-\frac{a}{b}}$ and use Proposition A.8 to get

$$\begin{aligned}
&\int_{-\infty}^\infty ae^{-2n\frac{a}{b}}e^{-2re^{-\frac{a}{b}}}da = \int_\infty^0 (-b)(\ln t)t^{2n}e^{-2rt}\left(-\frac{b}{t}\right)dt \\
= &-b^2\int_0^\infty t^{2n-1}e^{-2rt}\ln t dt = -b^2\frac{1}{(2r)^{2n}}\left[\Gamma'(2n) - \Gamma(2n)\ln(2r)\right] \\
= &\frac{b^2}{(2r)^{2n}}\left[\Gamma(2n)\left(\frac{1}{bn}\sum_{i=1}^n x_i + k_{\Delta_1}\ln\Delta + \ln 2\right) - \Gamma'(2n)\right]
\end{aligned}$$

so then

$$\begin{aligned}
&\frac{r^n}{b^{2n+2}}e^{-\frac{q(\underline{x})}{b}}e^{\sum_{i=1}^n \frac{x_i}{b}} \times \frac{b^2}{(2r)^{2n}}\left[\Gamma(2n)\left(\frac{1}{bn}\sum_{i=1}^n x_i + k_{\Delta_1}\ln\Delta + \ln 2\right) - \Gamma'(2n)\right] \\
= &\frac{1}{2^{2n}}\frac{1}{b^{2n}}e^{-\frac{q(\underline{x})}{b}}\frac{e^{\sum_{i=1}^n \frac{x_i}{b}}}{r^n}\left[\Gamma(2n)\left(\frac{1}{b}\bar{x} + k_{\Delta_1}\ln\Delta + \ln 2\right) - \Gamma'(2n)\right] \\
= &\frac{\Gamma(2n)}{2^{2n}\Delta^{k_{\Delta_2}}}e^{-2\frac{q(\underline{x})}{b}}\left[\frac{\bar{x}}{b^{2n+1}} + \frac{k_{\Delta_1}\ln\Delta + \ln 2 - \psi(2n)}{b^{2n}}\right]
\end{aligned}$$

where $\psi(\cdot)$ denotes digamma function. And we have

$$
\begin{aligned}
& \int_0^\infty \int_{-\infty}^\infty a\pi(a,b)L(a,b \mid x_1,x_2,\cdots,x_n)dadb \\
= & \frac{[q(\underline{x})]^{2n-2}}{\Gamma(n)\Gamma(2n-2)} \int_0^\infty \frac{\Gamma(2n)}{2^{2n}\Delta^{k_{\Delta_2}}} e^{-2\frac{q(\underline{x})}{b}} \left[\frac{\bar{x}}{b^{2n+1}} + \frac{k_{\Delta_1}\ln\Delta + \ln 2 - \psi(2n)}{b^{2n}}\right] db \\
= & \frac{\Gamma(2n)[q(\underline{x})]^{2n-2}\bar{x}}{2^{2n}\Delta^{k_{\Delta_2}}\Gamma(n)\Gamma(2n-2)} \int_0^\infty \frac{1}{b^{2n+1}} e^{-2\frac{q(\underline{x})}{b}} db \\
+ & \frac{\Gamma(2n)[q(\underline{x})]^{2n-2}\left[k_{\Delta_1}\ln\Delta + \ln 2 - \psi(2n)\right]}{2^{2n}\Delta^{k_{\Delta_2}}\Gamma(n)\Gamma(2n-2)} \int_0^\infty \frac{1}{b^{2n}} e^{-2\frac{q(\underline{x})}{b}} db \\
= & \frac{\Gamma(2n)[q(\underline{x})]^{2n-2}\bar{x}}{2^{2n}\Delta^{k_{\Delta_2}}\Gamma(n)\Gamma(2n-2)} \int_0^\infty t^{2n-1} e^{-2q(\underline{x})t} dt \\
+ & \frac{\Gamma(2n)[q(\underline{x})]^{2n-2}\left[k_{\Delta_1}\ln\Delta + \ln 2 - \psi(2n)\right]}{2^{2n}\Delta^{k_{\Delta_2}}\Gamma(n)\Gamma(2n-2)} \int_0^\infty t^{2n-2} e^{-2q(\underline{x})t} dt \\
= & \frac{\Gamma(2n)[q(\underline{x})]^{2n-2}\bar{x}}{2^{2n}\Delta^{k_{\Delta_2}}\Gamma(n)\Gamma(2n-2)} \times \frac{\Gamma(2n)}{[2q(\underline{x})]^{2n}} + \frac{\Gamma(2n)[q(\underline{x})]^{2n-2}\left[k_{\Delta_1}\ln\Delta + \ln 2 - \psi(2n)\right]}{2^{2n}\Delta^{k_{\Delta_2}}\Gamma(n)\Gamma(2n-2)} \times \frac{\Gamma(2n-1)}{[2q(\underline{x})]^{2n-1}} \\
= & \frac{\Gamma^2(2n)}{4^{2n}\Delta^{k_{\Delta_2}}\Gamma(n)\Gamma(2n-2)[q(\underline{x})]^2}\bar{x} + \frac{2\Gamma(2n)\Gamma(2n-1)\left[k_{\Delta_1}\ln\Delta + \ln 2 - \psi(2n)\right]}{4^{2n}\Delta^{k_{\Delta_2}}\Gamma(n)\Gamma(2n-2)q(\underline{x})}
\end{aligned}
$$

Thus we have

$$
\begin{aligned}
\hat{a}_B &= \frac{\int_0^\infty \int_{-\infty}^\infty a\pi(a,b)L(a,b \mid x_1,x_2,\cdots,x_n)dadb}{\int_0^\infty \int_{-\infty}^\infty \pi(a,b)L(a,b \mid x_1,x_2,\cdots,x_n)dadb} \\
&= \frac{\frac{\Gamma^2(2n)}{4^{2n}\Delta^{k_{\Delta_2}}\Gamma(n)\Gamma(2n-2)[q(\underline{x})]^2}\bar{x}}{\frac{\Gamma^2(2n)}{4^{2n}\Delta^{k_{\Delta_2}}\Gamma(n)\Gamma(2n-2)[q(\underline{x})]^2}} + \frac{\frac{2\Gamma(2n)\Gamma(2n-1)\left[k_{\Delta_1}\ln\Delta + \ln 2 - \psi(2n)\right]}{4^{2n}\Delta^{k_{\Delta_2}}\Gamma(n)\Gamma(2n-2)q(\underline{x})}}{\frac{\Gamma^2(2n)}{4^{2n}\Delta^{k_{\Delta_2}}\Gamma(n)\Gamma(2n-2)[q(\underline{x})]^2}} \\
&= \bar{x} + \frac{2[k_{\Delta_1}\ln\Delta + \ln 2 - \psi(2n)]}{2n-1} q(\underline{x}) \qquad (4.13)
\end{aligned}
$$

where $\Delta = n\Gamma\left(\frac{1}{n}\right)\left[\Gamma\left(1-\frac{1}{n}\right)\right]^{n-1}$. Now we find the estimator of b.

$$
\begin{aligned}
& \int_0^\infty \int_{-\infty}^\infty b\pi(a,b)L(a,b \mid x_1,x_2,\cdots,x_n)dadb \\
= & \int_0^\infty \int_{-\infty}^\infty b\frac{r^n[q(\underline{x})]^{2n-2}}{\Gamma(n)\Gamma(2n-2)} \frac{1}{b^{n+2}} e^{-\frac{q(\underline{x})}{b}} \exp\left[-n\frac{a}{b} - r\exp\left(-\frac{a}{b}\right)\right] \left(\frac{1}{b}\right)^n e^{\sum_{i=1}^n \frac{x_i}{b}} e^{-\frac{na}{b}} e^{-re^{-\frac{a}{b}}} dadb \\
= & \frac{[q(\underline{x})]^{2n-2}}{\Gamma(n)\Gamma(2n-2)} \int_0^\infty \int_{-\infty}^\infty \frac{r^n}{b^{2n+1}} e^{-\frac{q(\underline{x})}{b}} e^{-2n\frac{a}{b}} e^{\sum_{i=1}^n \frac{x_i}{b}} e^{-2re^{-\frac{a}{b}}} dadb \\
= & \frac{[q(\underline{x})]^{2n-2}}{\Gamma(n)\Gamma(2n-2)} \int_0^\infty \frac{r^n}{b^{2n+1}} e^{-\frac{q(\underline{x})}{b}} e^{\sum_{i=1}^n \frac{x_i}{b}} \left[\int_{-\infty}^\infty e^{-2n\frac{a}{b}} e^{-2re^{-\frac{a}{b}}} da\right] db
\end{aligned}
$$

Recall that

$$\int_{-\infty}^{\infty} e^{-2n\frac{a}{b}} e^{-2re^{-\frac{a}{b}}} da = \frac{b\Gamma(2n)}{(2r)^{2n}}$$

Then we have

$$\begin{aligned}
& \int_0^{\infty}\int_{-\infty}^{\infty} b\pi(a,b)L(a,b \mid x_1, x_2, \cdots, x_n)dadb \\
= & \frac{\Gamma(2n)[q(\underline{x})]^{2n-2}}{2^{2n}\Gamma(n)\Gamma(2n-2)} \int_0^{\infty} \frac{1}{b^{2n}} e^{-\frac{q(\underline{x})}{b}} \frac{e^{\sum_{i=1}^{n}\frac{x_i}{b}}}{r^n} db \\
= & \frac{\Gamma(2n)[q(\underline{x})]^{2n-2}}{2^{2n}\Gamma(n)\Gamma(2n-2)\Delta^{k_{\Delta_2}}} \int_0^{\infty} \frac{1}{b^{2n}} e^{-2\frac{q(\underline{x})}{b}} db \\
= & \frac{\Gamma(2n)[q(\underline{x})]^{2n-2}}{2^{2n}\Gamma(n)\Gamma(2n-2)\Delta^{k_{\Delta_2}}} \int_0^{\infty} t^{2n-2} e^{-2q(\underline{x})t} dt \\
= & \frac{\Gamma(2n)[q(\underline{x})]^{2n-2}}{2^{2n}\Gamma(n)\Gamma(2n-2)\Delta^{k_{\Delta_2}}} \times \frac{\Gamma(2n-1)}{[2q(\underline{x})]^{2n-1}} \\
= & \frac{2\Gamma(2n-1)\Gamma(2n)}{4^{2n}\Gamma(n)\Gamma(2n-2)\Delta^{k_{\Delta_2}} q(\underline{x})}
\end{aligned}$$

and also

$$\begin{aligned}
\hat{b}_B &= \frac{\int_0^{\infty}\int_{-\infty}^{\infty} b\pi(a,b)L(a,b \mid x_1, x_2, \cdots, x_n)dadb}{\int_0^{\infty}\int_{-\infty}^{\infty} \pi(a,b)L(a,b \mid x_1, x_2, \cdots, x_n)dadb} \\
&= \frac{\frac{2\Gamma(2n-1)\Gamma(2n)}{4^{2n}\Gamma(n)\Gamma(2n-2)\Delta^{k_{\Delta_2}} q(\underline{x})}}{\frac{\Gamma^2(2n)}{\Delta^{k_{\Delta_2}} 4^{2n}\Gamma(n)\Gamma(2n-2)[q(\underline{x})]^2}} \\
&= \frac{2}{2n-1} q(\underline{x})
\end{aligned}$$

As in the last section, we choose $k_n = \frac{2n-1}{2}$ so that $\hat{b}_B$ is unbiased, that is,

$$q(\underline{x}) = \frac{(2n-1)S_{X,n}}{2ES_{Z,n}}$$

Thus

$$\hat{b}_B = \frac{S_{X,n}}{ES_{Z,n}} \tag{4.14}$$

4.3 Bayesian-Maximum Likelihood Estimation for the Gumbel Minimum Distribution

In this section, we first use a Bayesian method to estimate the location parameter for a fixed scale parameter b. Then we use maximum likelihood to estimate b.

4.3.1 Estimation for the Gumbel Minimum Distribution Using a Noninformative Prior

Suppose that $\{X_1, X_2, \cdots, X_n\}$ is a random sample from a Gumbel minimum distribution with location a and scale b. Then the likelihood function is given by

$$\begin{aligned} L(a, b \mid x_1, x_2, \cdots, x_n) &= \prod_{i=1}^{n} f(x_i \mid a, b) \\ &= \left(\frac{1}{b}\right)^n e^{\sum_{i=1}^{n} \frac{x_i - a}{b}} e^{-\sum_{i=1}^{n} e^{\frac{x_i - a}{b}}} \end{aligned}$$

Using the prior $\pi(a) \equiv 1$ we get the posterior

$$\pi(a \mid x_1, x_2, \cdots, x_n) \quad = \quad \frac{1}{b\Gamma(n)} \left(\sum_{i=1}^{n} e^{\frac{x_i}{b}}\right)^n e^{-\frac{na}{b}} e^{-e^{-\frac{a}{b}}\left(\sum_{i=1}^{n} e^{\frac{x_i}{b}}\right)}$$

Assuming a squared error loss function, the Bayesian estimate of a parameter is its posterior mean. Therefore, the Bayesian estimate of the parameter a is given by

$$\begin{aligned} \hat{a}_B &= \int_{-\infty}^{\infty} a\pi(a \mid x_1, x_2, \cdots, x_n) da \\ &= \frac{1}{b\Gamma(n)} \left(\sum_{i=1}^{n} e^{\frac{x_i}{b}}\right)^n \int_{-\infty}^{\infty} a e^{-\frac{na}{b}} e^{-e^{-\frac{a}{b}}\left(\sum_{i=1}^{n} e^{\frac{x_i}{b}}\right)} da \end{aligned}$$

Recall from Proposition A.2

$$\int_0^{\infty} x^n e^{-kx} (\ln x) dx = \frac{n!}{k^{n+1}} \left[1 + \frac{1}{2} + \frac{1}{3} + \cdots \frac{1}{n} - \gamma - \ln k\right], \quad Re(k) > 0$$

Taking $t = e^{-\frac{a}{b}}$, then $a = -b \ln t$ for $0 < t < \infty$, so

$$\begin{aligned}
\hat{a}_B &= \frac{1}{b\Gamma(n)} \left(\sum_{i=1}^{n} e^{\frac{x_i}{b}} \right)^n \int_{-\infty}^{\infty} a e^{-\frac{na}{b}} e^{-e^{-\frac{a}{b}} \left(\sum_{i=1}^{n} e^{\frac{x_i}{b}} \right)} da \\
&= \frac{1}{b\Gamma(n)} \left(\sum_{i=1}^{n} e^{\frac{x_i}{b}} \right)^n \int_{\infty}^{0} (-b \ln t) t^n e^{-\left(\sum_{i=1}^{n} e^{\frac{x_i}{b}} \right) t} \left(-\frac{b}{t} \right) dt \\
&= -\frac{b}{\Gamma(n)} \left(\sum_{i=1}^{n} e^{\frac{x_i}{b}} \right)^n \int_{0}^{\infty} t^{n-1} e^{-\left(\sum_{i=1}^{n} e^{\frac{x_i}{b}} \right) t} (\ln t) dt \\
&= -\frac{b}{\Gamma(n)} \left(\sum_{i=1}^{n} e^{\frac{x_i}{b}} \right)^n \frac{(n-1)!}{\left(\sum_{i=1}^{n} e^{\frac{x_i}{b}} \right)^n} \left[1 + \frac{1}{2} + \frac{1}{3} + \cdots \frac{1}{n} - \gamma - \ln \left(\sum_{i=1}^{n} e^{\frac{x_i}{b}} \right) \right] \\
&= b \left[\gamma + \ln \left(\sum_{i=1}^{n} e^{\frac{x_i}{b}} \right) - 1 - \frac{1}{2} - \frac{1}{3} - \cdots - \frac{1}{n} \right]
\end{aligned}$$

It is well known that

$$\gamma = \sum_{k=1}^{n} \frac{1}{n} - \ln n - \sum_{m=2}^{\infty} \frac{\zeta(m, n+1)}{m}$$

where $\zeta(s, k)$ is the Hurwitz zeta function. Then

$$\hat{a}_B = b \ln \left(\frac{1}{n} \sum_{i=1}^{n} e^{\frac{x_i}{b}} \right) - b \sum_{k=2}^{\infty} \frac{\zeta(k, n+1)}{k}$$

According to Choi and Srivastava [18]

$$\sum_{k=2}^{\infty} \frac{\zeta(k, a) t^k}{k} = \log \Gamma(a - t) - \log \Gamma(a) + t\psi(a), \quad (|t| < |a|)$$

where $\psi(a) = \frac{\Gamma'(a)}{\Gamma(a)}$ and Gradshteyn [44] gives

$$\sum_{k=1}^{n} \frac{1}{k} = \gamma + \ln n + \frac{1}{2n} - \sum_{k=2}^{\infty} \frac{A_k}{n(n+1) \cdots (n+k-1)}$$

where γ denotes Euler constant and

$$\begin{aligned}
A_k &= \frac{1}{k} \int_0^1 x(1-x)(2-x) \cdots (k-1-x) dx \\
A_2 &= \frac{1}{12}, \quad A_3 = \frac{1}{12}, \quad A_4 = \frac{19}{120}, \quad A_5 = \frac{9}{20}
\end{aligned}$$

Notice that

$$\psi(1) \;=\; -\gamma, \quad \psi(n+1) = \psi(1) + \sum_{k=1}^{n} \frac{1}{k}, \quad \gamma = \lim_{n\to\infty}\left(\sum_{k=1}^{n}\frac{1}{k} - \log n\right)$$

Therefore

$$\begin{aligned}
\sum_{k=2}^{\infty} \frac{\zeta(k, n+1)}{k} &= \log\Gamma(n) - \log\Gamma(n+1) + \psi(n+1) \\
&= \psi(n+1) - \log n = \sum_{k=1}^{n}\frac{1}{k} - \log n - \gamma \\
&= \frac{1}{2n} - \sum_{k=2}^{\infty} \frac{A_k}{n(n+1)\cdots(n+k-1)}
\end{aligned}$$

and then

$$\hat{a}_B = b\ln\left(\frac{1}{n}\sum_{i=1}^{n} e^{\frac{x_i}{b}}\right) - b\left(\frac{1}{2n} - \sum_{k=2}^{\infty}\frac{A_k}{n(n+1)\cdots(n+k-1)}\right) \tag{4.15}$$

We write

$$\begin{aligned}
M_n &= \left(\frac{1}{2n} - \sum_{k=2}^{\infty}\frac{A_k}{n(n+1)\cdots(n+k-1)}\right) \\
\frac{a}{b} &= \ln\left(\frac{1}{n}\sum_{i=1}^{n} e^{\frac{x_i}{b}}\right) - M_n \quad e^{-\frac{a}{b}} = \frac{n}{\sum_{i=1}^{n} e^{\frac{x_i}{b}}} e^{M_n}
\end{aligned}$$

and then the log-likelihood function can be written as

$$\begin{aligned}
l = \ln L(a, b \mid x_1, x_2, \cdots, x_n) &= \ln\left[\left(\frac{1}{b}\right)^n e^{\sum_{i=1}^{n}\frac{x_i - a}{b}} e^{-\sum_{i=1}^{n} e^{\frac{x_i - a}{b}}}\right] \\
&= -n\ln b + \sum_{i=1}^{n}\frac{x_i - a}{b} - \sum_{i=1}^{n} e^{\frac{x_i - a}{b}} \\
&= -n\ln b + \sum_{i=1}^{n}\frac{x_i}{b} - \frac{na}{b} - e^{-\frac{a}{b}}\sum_{i=1}^{n} e^{\frac{x_i}{b}} \\
&= -n\ln b + \sum_{i=1}^{n}\frac{x_i}{b} - n\left[\ln\left(\frac{1}{n}\sum_{i=1}^{n} e^{\frac{x_i}{b}}\right) - M_n\right] - ne^{M_n} \\
&= n\ln n - n\ln b + \sum_{i=1}^{n}\frac{x_i}{b} - n\ln\left(\sum_{i=1}^{n} e^{\frac{x_i}{b}}\right) - n\left(M_n - e^{M_n}\right)
\end{aligned}$$

Therefore

$$
\begin{aligned}
\frac{\partial l}{\partial b} &= -\frac{n}{b} - \frac{\sum_{i=1}^{n} x_i}{b^2} + \frac{n\sum_{i=1}^{n} x_i e^{\frac{x_i}{b}}}{b^2 \sum_{i=1}^{n} e^{\frac{x_i}{b}}} \\
\frac{\partial^2 l}{\partial b^2} &= \frac{n}{b^2} + \frac{2\sum_{i=1}^{n} x_i}{b^3} \\
&\quad + \frac{\left[n\sum_{i=1}^{n}\left(-\frac{x_i^2}{b^2}e^{\frac{x_i}{b}}\right)\right]\left(b^2\sum_{i=1}^{n} e^{\frac{x_i}{b}}\right) - \left(n\sum_{i=1}^{n} x_i e^{\frac{x_i}{b}}\right)\left[\left(2b\sum_{i=1}^{n} e^{\frac{x_i}{b}}\right) - \sum_{i=1}^{n} x_i e^{\frac{x_i}{b}}\right]}{\left(b^2\sum_{i=1}^{n} e^{\frac{x_i}{b}}\right)^2} \\
&= \frac{n}{b^2} + \frac{2\sum_{i=1}^{n} x_i}{b^3} \\
&\quad + \frac{\left[n\sum_{i=1}^{n}\left(-x_i^2 e^{\frac{x_i}{b}}\right)\right]\left(\sum_{i=1}^{n} e^{\frac{x_i}{b}}\right) - \left(n\sum_{i=1}^{n} x_i e^{\frac{x_i}{b}}\right)\left[\left(2b\sum_{i=1}^{n} e^{\frac{x_i}{b}}\right) - \sum_{i=1}^{n} x_i e^{\frac{x_i}{b}}\right]}{\left(b^2\sum_{i=1}^{n} e^{\frac{x_i}{b}}\right)^2}
\end{aligned}
$$

Setting $\frac{\partial l}{\partial b} = 0$, we have

$$
b = \frac{\sum_{i=1}^{n} x_i e^{\frac{x_i}{b}}}{\sum_{i=1}^{n} e^{\frac{x_i}{b}}} - \frac{1}{n}\sum_{i=1}^{n} x_i \tag{4.16}
$$

Then we can find b numerically.

4.3.2 Estimation for the Gumbel Maximum Distribution Using a Noninformative Prior

Consider a random variable X with pdf

$$
f(x \mid a, b) = \frac{1}{b} e^{-\frac{x-a}{b}} e^{-e^{-\frac{x-a}{b}}}
$$

where a and b are the location and scale parameters, respectively.

Suppose that $\{X_1, X_2, \cdots, X_n\}$ is a random sample from a Gumbel maximum distribution with location a and scale b. Then the likelihood function of $X_1, X_2, \cdots, X_n$ is given by

$$
\begin{aligned}
L(a, b \mid x_1, x_2, \cdots, x_n) &= \prod_{i=1}^{n} f(x_i \mid a, b) \\
&= \left(\frac{1}{b}\right)^n e^{-\sum_{i=1}^{n}\frac{x_i-a}{b}} e^{-\sum_{i=1}^{n} e^{-\frac{x_i-a}{b}}}
\end{aligned}
$$

Now use as prior for a

$$\pi(a) \quad = \quad 1$$

and so

$$\begin{aligned}
\int_{-\infty}^{\infty} \pi(a) L(a|x_1, \cdots, x_n) da &= \int_{-\infty}^{\infty} \left(\frac{1}{b}\right)^n e^{-\sum_{i=1}^n \frac{x_i - a}{b}} e^{-\sum_{i=1}^n e^{-\frac{x_i - a}{b}}} da \\
&= \left(\frac{1}{b}\right)^n e^{-\sum_{i=1}^n \frac{x_i}{b}} \int_{-\infty}^{\infty} e^{n\frac{a}{b}} e^{-e^{\frac{a}{b}} \sum_{i=1}^n e^{-\frac{x_i}{b}}} da \\
&= \left(\frac{1}{b}\right)^{n-1} e^{-\sum_{i=1}^n \frac{x_i}{b}} \int_0^{\infty} t^{n-1} e^{-t\left(\sum_{i=1}^n e^{-\frac{x_i}{b}}\right)} da \\
&= \Gamma(n) \left(\frac{1}{b}\right)^{n-1} e^{-\sum_{i=1}^n \frac{x_i}{b}} \left(\sum_{i=1}^n e^{-\frac{x_i}{b}}\right)^{-n}
\end{aligned}$$

then the posterior for a is

$$\begin{aligned}
\pi(a \mid x_1, x_2, \cdots, x_n) &= \frac{\pi(a) L(a|x_1, \cdots, x_n) da}{\int_{-\infty}^{\infty} \pi(a) L(a|x_1, \cdots, x_n) da} \\
&= \frac{1}{b\Gamma(n)} \left(\sum_{i=1}^n e^{-\frac{x_i}{b}}\right)^n e^{\frac{na}{b}} e^{-e^{\frac{a}{b}} \left(\sum_{i=1}^n e^{-\frac{x_i}{b}}\right)}
\end{aligned}$$

Assuming a squared error loss function, the Bayesian estimate of a parameter is its posterior mean. Therefore, the Bayesian estimate of the parameter a is given by

$$\begin{aligned}
\hat{a}_B &= \int_{-\infty}^{\infty} a\pi(a \mid x_1, x_2, \cdots, x_n) da \\
&= \frac{1}{b\Gamma(n)} \left(\sum_{i=1}^n e^{-\frac{x_i}{b}}\right)^n \int_{-\infty}^{\infty} a e^{\frac{na}{b}} e^{-e^{\frac{a}{b}} \left(\sum_{i=1}^n e^{-\frac{x_i}{b}}\right)} da
\end{aligned}$$

Recall from Proposition A.2

$$\int_0^{\infty} x^n e^{-kx} (\ln x) dx = \frac{n!}{k^{n+1}} \left[1 + \frac{1}{2} + \frac{1}{3} + \cdots \frac{1}{n} - \gamma - \ln k\right], \quad Re(k) > 0.$$

Taking $t = e^{\frac{a}{b}}$, then $a = b \ln t$ for $0 < t < \infty$, so

$$\begin{aligned}
\hat{a}_B &= \frac{1}{b\Gamma(n)} \left(\sum_{i=1}^{n} e^{-\frac{x_i}{b}} \right)^n \int_{-\infty}^{\infty} a e^{\frac{na}{b}} e^{-e^{\frac{a}{b}} \left(\sum_{i=1}^{n} e^{-\frac{x_i}{b}} \right)} da \\
&= \frac{1}{b\Gamma(n)} \left(\sum_{i=1}^{n} e^{-\frac{x_i}{b}} \right)^n \int_{0}^{\infty} (b \ln t) t^n e^{-\left(\sum_{i=1}^{n} e^{-\frac{x_i}{b}} \right) t} \left(\frac{b}{t} \right) dt \\
&= \frac{b}{\Gamma(n)} \left(\sum_{i=1}^{n} e^{-\frac{x_i}{b}} \right)^n \int_{0}^{\infty} t^{n-1} e^{-\left(\sum_{i=1}^{n} e^{-\frac{x_i}{b}} \right) t} (\ln t) dt \\
&= \frac{b}{\Gamma(n)} \left(\sum_{i=1}^{n} e^{-\frac{x_i}{b}} \right)^n \frac{(n-1)!}{\left(\sum_{i=1}^{n} e^{-\frac{x_i}{b}} \right)^n} \left[1 + \frac{1}{2} + \frac{1}{3} + \cdots \frac{1}{n} - \gamma - \ln \left(\sum_{i=1}^{n} e^{-\frac{x_i}{b}} \right) \right] \\
&= b \left[1 + \frac{1}{2} + \frac{1}{3} + \cdots \frac{1}{n} - \gamma - \ln \left(\sum_{i=1}^{n} e^{-\frac{x_i}{b}} \right) \right]
\end{aligned}$$

It is well known that

$$\gamma = \sum_{k=1}^{n} \frac{1}{n} - \ln n - \sum_{m=2}^{\infty} \frac{\zeta(m, n+1)}{m}$$

where $\zeta(s, k)$ is the Hurwitz zeta function. Then

$$\hat{a}_B = b \sum_{k=2}^{\infty} \frac{\zeta(k, n+1)}{k} - b \ln \left(\frac{1}{n} \sum_{i=1}^{n} e^{-\frac{x_i}{b}} \right)$$

As last section, we have

$$\hat{a}_B = b \left(\frac{1}{2n} - \sum_{k=2}^{\infty} \frac{A_k}{n(n+1)\cdots(n+k-1)} \right) - b \ln \left(\frac{1}{n} \sum_{i=1}^{n} e^{\frac{x_i}{b}} \right) \tag{4.17}$$

We write

$$\begin{aligned}
M_n &= \left(\frac{1}{2n} - \sum_{k=2}^{\infty} \frac{A_k}{n(n+1)\cdots(n+k-1)} \right) \\
\frac{\hat{a}}{b} &= M_n - \ln \left(\frac{1}{n} \sum_{i=1}^{n} e^{-\frac{x_i}{b}} \right) \\
e^{\frac{\hat{a}}{b}} &= \frac{n}{\sum_{i=1}^{n} e^{-\frac{x_i}{b}}} e^{M_n}
\end{aligned}$$

Since log-likelihood function can be written as

$$
\begin{aligned}
l = \ln L(a, b \mid x_1, x_2, \cdots, x_n) &= \ln \left[\left(\frac{1}{b}\right)^n e^{-\sum_{i=1}^n \frac{x_i - a}{b}} e^{-\sum_{i=1}^n e^{-\frac{x_i - a}{b}}} \right] \\
&= -n \ln b - \sum_{i=1}^n \frac{x_i - a}{b} - \sum_{i=1}^n e^{-\frac{x_i - a}{b}} \\
&= -n \ln b - \sum_{i=1}^n \frac{x_i}{b} + \frac{na}{b} - e^{\frac{a}{b}} \sum_{i=1}^n e^{-\frac{x_i}{b}} \\
&= -n \ln b - \sum_{i=1}^n \frac{x_i}{b} + n \left[\ln \left(\frac{1}{n} \sum_{i=1}^n e^{-\frac{x_i}{b}} \right) - M_n \right] - n e^{M_n} \\
&= n \ln n - n \ln b - \sum_{i=1}^n \frac{x_i}{b} + n \ln \left(\sum_{i=1}^n e^{-\frac{x_i}{b}} \right) - n \left(M_n + e^{M_n} \right)
\end{aligned}
$$

Therefore

$$
\frac{\partial l}{\partial b} = -\frac{n}{b} + \frac{\sum_{i=1}^n x_i}{b^2} - \frac{n \sum_{i=1}^n x_i e^{-\frac{x_i}{b}}}{b^2 \sum_{i=1}^n e^{-\frac{x_i}{b}}}
$$

$$
\begin{aligned}
\frac{\partial^2 l}{\partial b^2} &= \frac{n}{b^2} - \frac{2 \sum_{i=1}^n x_i}{b^3} \\
&+ \frac{\left[n \sum_{i=1}^n \left(\frac{x_i^2}{b^2} e^{-\frac{x_i}{b}} \right) \right] \left(b^2 \sum_{i=1}^n e^{-\frac{x_i}{b}} \right) - \left(n \sum_{i=1}^n x_i e^{-\frac{x_i}{b}} \right) \left[\left(2b \sum_{i=1}^n e^{-\frac{x_i}{b}} \right) + \sum_{i=1}^n x_i e^{-\frac{x_i}{b}} \right]}{\left(b^2 \sum_{i=1}^n e^{-\frac{x_i}{b}} \right)^2} \\
&= \frac{n}{b^2} + \frac{2 \sum_{i=1}^n x_i}{b^3} \\
&+ \frac{\left[n \sum_{i=1}^n \left(-x_i^2 e^{-\frac{x_i}{b}} \right) \right] \left(\sum_{i=1}^n e^{-\frac{x_i}{b}} \right) - \left(n \sum_{i=1}^n x_i e^{-\frac{x_i}{b}} \right) \left[\left(2b \sum_{i=1}^n e^{-\frac{x_i}{b}} \right) + \sum_{i=1}^n x_i e^{-\frac{x_i}{b}} \right]}{\left(b^2 \sum_{i=1}^n e^{-\frac{x_i}{b}} \right)^2}
\end{aligned}
$$

Setting $\frac{\partial l}{\partial b} = 0$, we have

$$
b = \frac{1}{n} \sum_{i=1}^n x_i - \frac{\sum_{i=1}^n x_i e^{-\frac{x_i}{b}}}{\sum_{i=1}^n e^{-\frac{x_i}{b}}} \tag{4.18}
$$

Then we can use numerical method to find b.

4.4 Another Example: Estimation for the Normal Distribution

Consider a random variable X with pdf

$$f(x \mid a, b) = \frac{1}{\sqrt{2\pi}b} e^{-\frac{(x-a)^2}{2b^2}} \tag{4.19}$$

where a and b are the location and scale parameters, respectively.

Suppose that $\{X_1, X_2, \cdots, X_n\}$ is a random sample from a normal distribution with location a and scale b. Then the likelihood function is given by

$$L(a, b \mid x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} f(x_i \mid a, b) = \left(\frac{1}{\sqrt{2\pi}b}\right)^n e^{-\sum_{i=1}^{n} \frac{(x_i-a)^2}{2b^2}}$$

Now use Jeffrey's prior for b such that $\pi(b) = \frac{1}{b}$. Let $t = \frac{1}{b}$, and so

$$\begin{aligned}
\int_0^\infty \pi(b) L(a|x_1, \cdots, x_n) db &= \int_0^\infty \frac{1}{b} \left(\frac{1}{\sqrt{2\pi}b}\right)^n e^{-\sum_{i=1}^{n} \frac{(x_i-a)^2}{2b^2}} db \\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^n \int_0^\infty \frac{1}{b^{n+1}} e^{-\sum_{i=1}^{n} \frac{(x_i-a)^2}{2b^2}} db \\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^n \int_\infty^0 t^{n+1} e^{-\left(\sum_{i=1}^{n} \frac{(x_i-a)^2}{2}\right)t} \left(-\frac{1}{t^2}\right) dt \\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^n \int_0^\infty t^{n-1} e^{-\left(\sum_{i=1}^{n} \frac{(x_i-a)^2}{2}\right)t^2} dt \\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{\Gamma(\frac{n}{2})}{2\left[\sum_{i=1}^{n} \frac{(x_i-a)^2}{2}\right]^{\frac{n}{2}}}.
\end{aligned}$$

Then the posterior for b is

$$\begin{aligned}
\pi(b \mid x_1, x_2, \cdots, x_n) &= \frac{\pi(b) L(b|x_1, \cdots, x_n) db}{\int_0^\infty \pi(b) L(b|x_1, \cdots, x_n) db} \\
&= \frac{2\left[\sum_{i=1}^{n} \frac{(x_i-a)^2}{2}\right]^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} \frac{1}{b^{n+1}} e^{-\sum_{i=1}^{n} \frac{(x_i-a)^2}{2b^2}}
\end{aligned}$$

Assuming a squared error loss function, the Bayesian estimate of a parameter is its

posterior mean. Therefore, the Bayesian estimate of the parameter b is given by

$$\begin{aligned}
\hat{b}_B &= \int_0^\infty b\pi(b \mid x_1, x_2, \cdots, x_n)db = \int_0^\infty b\frac{2\left[\sum_{i=1}^n \frac{(x_i-a)^2}{2}\right]^{\frac{n}{2}}}{\Gamma(\frac{n}{2})}\frac{1}{b^{n+1}}e^{-\sum_{i=1}^n \frac{(x_i-a)^2}{2b^2}}db \\
&= \frac{2\left[\sum_{i=1}^n \frac{(x_i-a)^2}{2}\right]^{\frac{n}{2}}}{\Gamma(\frac{n}{2})}\int_0^\infty \frac{1}{b^n}e^{-\sum_{i=1}^n \frac{(x_i-a)^2}{2b^2}}db \\
&= \frac{2\left[\sum_{i=1}^n \frac{(x_i-a)^2}{2}\right]^{\frac{n}{2}}}{\Gamma(\frac{n}{2})}\int_0^\infty t^{n-2}e^{-\sum_{i=1}^n \frac{(x_i-a)^2}{2}t^2}dt \\
&= \frac{2\left[\sum_{i=1}^n \frac{(x_i-a)^2}{2}\right]^{\frac{n}{2}}}{\Gamma(\frac{n}{2})}\frac{\Gamma(\frac{n-1}{2})}{2\left[\sum_{i=1}^n \frac{(x_i-a)^2}{2}\right]^{\frac{n-1}{2}}} \\
&= \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})}\left[\sum_{i=1}^n \frac{(x_i-a)^2}{2}\right]^{\frac{1}{2}} \\
&= \sqrt{\frac{n-1}{2}}\frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})}S_n
\end{aligned}$$

Now use Jeffrey's prior for a such that $\pi(a) = 1$, and so

$$\begin{aligned}
\int_{-\infty}^\infty \pi(a)L(a|x_1, \cdots, x_n)da &= \int_{-\infty}^\infty \left(\frac{1}{\sqrt{2\pi}b}\right)^n e^{-\sum_{i=1}^n \frac{(x_i-a)^2}{2b^2}}da \\
&= \left(\frac{1}{\sqrt{2\pi}b}\right)^n \int_{-\infty}^\infty e^{-\sum_{i=1}^n \frac{(x_i-a)^2}{2b^2}}da \\
&= \left(\frac{1}{\sqrt{2\pi}b}\right)^n \int_{-\infty}^\infty e^{-\frac{na^2-(2\sum_{i=1}^n x_i)a+\sum_{i=1}^n x_i^2}{2b^2}}da \\
&= \left(\frac{1}{\sqrt{2\pi}b}\right)^n e^{-\frac{\sum_{i=1}^n x_i^2}{2b^2}}\int_{-\infty}^\infty e^{-\frac{na^2-(2\sum_{i=1}^n x_i)a}{2b^2}}da \\
&= \left(\frac{1}{\sqrt{2\pi}b}\right)^n e^{-\frac{\sum_{i=1}^n x_i^2}{2b^2}}e^{\frac{(\sum_{i=1}^n x_i)^2}{2nb^2}}\sqrt{\frac{2\pi}{n}}b
\end{aligned}$$

then posterior for a is

$$\begin{aligned}
\pi(a \mid x_1, x_2, \cdots, x_n) &= \frac{\pi(a)L(a|x_1, \cdots, x_n)da}{\int_{-\infty}^{\infty} \pi(b)L(b|x_1, \cdots, x_n)da} \\
&= \sqrt{\frac{n}{2\pi}}\frac{1}{b}e^{-\frac{1}{2nb^2}(\sum_{i=1}^{n} x_i)^2}e^{-\frac{na^2-(2\sum_{i=1}^{n} x_i)a}{2b^2}} \\
&= \frac{1}{\sqrt{2\pi}(b/\sqrt{n})}e^{-\frac{n}{2b^2}\bar{x}_n^2}e^{-\frac{n}{2b^2}(a^2-2\bar{x}_n a+\bar{x}_n^2)} \\
&= \frac{1}{\sqrt{2\pi}(b/\sqrt{n})}e^{-\frac{(a-\bar{x})^2}{2(b/\sqrt{n})^2}}
\end{aligned}$$

Assuming a squared error loss function, the Bayesian estimate of a parameter is its posterior mean. Therefore, the Bayesian estimate of the parameter a is given by

$$\begin{aligned}
\hat{a}_B &= \int_{-\infty}^{\infty} a\pi(a \mid x_1, x_2, \cdots, x_n)da = \sqrt{\frac{n}{2\pi}}\frac{1}{b}\int_{-\infty}^{\infty} ae^{-\frac{n}{2b^2}(a-\bar{x}_n)^2}da \\
&= \sqrt{\frac{n}{2\pi}}\frac{1}{b}\sqrt{\frac{2}{n}}b\bar{x}_n\Gamma\left(\frac{1}{2}, \frac{n\bar{x}_n^2}{2b^2}\right) = \sqrt{\frac{n}{2\pi}}\frac{1}{b}\sqrt{\frac{2}{n}}b\bar{x}_n\sqrt{\pi}\left[1-\Phi\left(\frac{\sqrt{n}\bar{x}_n}{\sqrt{2}b}\right)\right] \\
&= \bar{x}_n\left[1-\Phi\left(\frac{\sqrt{n}\bar{x}_n}{\sqrt{2}b}\right)\right]
\end{aligned}$$

where $\Phi(x)$ denotes the error function

$$\Phi(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-\frac{t^2}{2}}dt$$

Chapter 5

Frequentist Prediction Interval for a Location–Scale Family

In this chapter, we first construct prediction intervals using two existing methods and then present a simple Monte Carlo method. All three methods are based on our previous estimators for a location–scale model.

5.1 Introduction

A prediction interval (PI) is an interval constructed using the values of a past sample to contain the values of a future observations. It can take several forms.

Suppose that $\underline{Y} = (Y_1, Y_2, \cdots, Y_m)$ are future observations from a population with density function $f_X(x)$ and $T(\underline{Y})$ is a function of $\underline{Y}$. Let $L(\underline{X}) = L(X_1, X_2, \cdots, X_n)$ and $U(\underline{X}) = U(X_1, X_2, \cdots, X_n)$ be two statistics which are functions of the observed sample $\underline{X} = (X_1, X_2, \cdots, X_n)$. Then a typical PI can be defined as follows:

Definition 5.1. An interval $(-\infty, U(\underline{X}))$ is called a $100\gamma\%$ one–sided upper prediction interval for $T(\underline{Y})$ if

$$P\left[T(\underline{Y}) < U(\underline{X})\right] = \gamma$$

where γ is called the prediction probability. Similarly, an interval $(L(\underline{X}), \infty)$ is called a $100\gamma\%$ one–sided lower prediction interval for $T(\underline{Y})$ if

$$P\left[T(\underline{Y}) > L(\underline{X})\right] = \gamma$$

and an interval $(L(\underline{X}), U(\underline{X}))$ is called a $100\gamma\%$ two–sided prediction interval for

$T(\underline{Y})$ if

$$P\left[L(\underline{X}) < T(\underline{Y}) < U(\underline{X})\right] = \gamma$$

The goal here is to predict the jth future order statistic $T(\underline{Y}) = Y_{(j)}$ based on a random sample from a location-scale family with density

$$f_X(x) = f(x|a,b) = \frac{1}{b} g\left(\frac{x-a}{b}\right)$$

We often take the following form in practice:

$$L(\underline{X}) = \hat{a} + k_1\hat{b}, \quad U(\underline{X}) = \hat{a} + k_2\hat{b}$$

The construction of an exact prediction interval for a future observation is typically carried out by inverting an appropriate pivotal quantity. Except for some simple situations, however, it is generally impossible to obtain the exact distribution of the pivotal quantity. In the following two sections, we consider the pivotal statistic

$$W = \frac{Y_{(j)} - a}{b}$$

Several methods have been proposed in the literature to determine distribution of W. A very common method to obtain approximate quantiles, k_1 and k_2, is Monte Carlo simulation, since W is pivotal.

5.2 Construction of a Conditional Prediction Interval

Lawless [72] provides a general method for deriving prediction intervals based on the idea of conditioning on a ancillary statistic $A = (a_1, a_2, \cdots, a_n)$ with $a_i = \frac{X_i - \hat{a}}{\hat{b}}$. He uses a conditional method to obtain an explicit formula for a conditional cumulative

distribution of W such that

$$\begin{aligned} P\left(Y_{(j)} < \hat{a} + k_2\hat{b}\middle| A\right) &= P\left(\frac{Y_{(j)} - a}{b} < \frac{\hat{a} - a}{b} + k_2\frac{\hat{b}}{b}\middle| A\right) \\ &= P\left(W < U + k_2 V \middle| A\right) \\ &= E\left[P\left(W < U + k_2 V \middle| U, V\right)\middle| A\right] \end{aligned} \tag{5.1}$$

where $U = \frac{\hat{a}-a}{b}$, $V = \frac{\hat{b}}{b}$, $\hat{a}$ and $\hat{b}$ are equivariant estimators of a and b respectively. Lawless [72] also gives the general form of the joint conditional density

$$f_{U,V|A}(u, v|A) = Kv^{n-2}\prod_{i=1}^{n} g\left(u + a_i v\right)$$

Given a configuration ancillary statistic:

$$A = (a_1, \cdots, a_n) = \left(\frac{x_1 - \bar{x}}{s_x}, \cdots, \frac{x_n - \bar{x}}{s_x}\right)$$

Fraser [40] shows that the conditional probability for $(\bar{x}, s)$, given A, has the form

$$f_{\bar{X},S_x|A}(\bar{x}, s_x|A; a, b) = \frac{K}{b^n}s_x^{n-2}\prod_{i=1}^{n} g\left(\frac{x_i - a}{b}\right)$$

where K is normalizing constant. Notice that $a_i = \frac{x_i - \bar{x}}{s_x}$ and then rewrite

$$f_{\bar{X},S_x|A}(\bar{x}, s_x|A; a, b) = \frac{K}{b^n}s_x^{n-2}\prod_{i=1}^{n} g\left(\frac{\bar{x} + a_i s_x - a}{b}\right)$$

Recall that $Z = \frac{X-a}{b}$, so then

$$f_{\bar{Z},S_Z|A}(\bar{z}, s_z|A; a, b) = Ks_z^{n-2}\prod_{i=1}^{n} g\left(\bar{z} + a_i s_z\right)$$

As before, we write

$$\hat{a} = \bar{X}_n + t_n S_X \quad \textit{and} \quad \hat{b} = k_n S_X$$

Then

$$\begin{aligned} U &= \frac{\hat{a}-a}{b} = \frac{\bar{X}_n + t_n S_X - a}{b} = \bar{Z} + t_n S_Z \\ V &= \frac{\hat{b}}{b} = k_n S_Z \end{aligned}$$

Suppose that the joint distribution of U and V given $A = (a_1, \cdots, a_n)$ has the form:

$$\begin{aligned} f_{U,V|A}(u,v|A) &= f_{\bar{Z},S_Z|A}\left(u - \frac{t_n}{k_n}v, \frac{1}{k_n}v|A\right)\frac{\partial\left(\bar{Z},S_Z\right)}{\partial\left(U,V\right)} \\ &= \frac{1}{k_n}f_{\bar{Z},S_Z|A}\left(u - \frac{t_n}{k_n}v, \frac{1}{k_n}v|A\right) \\ &= \frac{K}{k_n^{n-1}}v^{n-2}\prod_{i=1}^{n} g\left(u + \frac{a_i - t_n}{k_n}v\right) \end{aligned} \tag{5.2}$$

Let $Y_{(j)}$ denote j^{th} order statistics, recall that the *pdf* of $Y_{(j)}$ is

$$f_{Y_{(j)}}(y) = \frac{m!}{(j-1)!(m-j)!}f_X(y)\left[F_X(y)\right]^{j-1}\left[1-F_X(y)\right]^{m-j}$$

Then we have

$$\begin{aligned} f_W(w) &= \frac{bm!}{(j-1)!(m-j)!}f_X(bw+a)\left[F_X(bw+a)\right]^{j-1}\left[1-F_X(bw+a)\right]^{m-j} \\ &= \frac{bm!}{(j-1)!(m-j)!} \times \frac{1}{b}g(w) \times \left[\int_{-\infty}^{bw+a} f_X(t)dt\right]^{j-1}\left[1-\int_{-\infty}^{bw+a} f_X(t)dt\right]^{m-j} \\ &= \frac{m!}{(j-1)!(m-j)!}g(w)\left[\int_{-\infty}^{bw+a}\frac{1}{b}g\left(\frac{t-a}{b}\right)dt\right]^{j-1}\left[1-\int_{-\infty}^{bw+a}\frac{1}{b}g\left(\frac{t-a}{b}\right)dt\right]^{m-j} \\ &= \frac{m!}{(j-1)!(m-j)!}g(w)\left[\int_{-\infty}^{w} g(s)ds\right]^{j-1}\left[1-\int_{-\infty}^{w} g(s)ds\right]^{m-j} \\ &= \frac{m!}{(j-1)!(m-j)!}g(w)\left[G(w)\right]^{j-1}\left[1-G(w)\right]^{m-j} \end{aligned}$$

which is independent of a and b. Let $F_W(w)$ denote the cdf of W and p the w_p–quantile such that $G(w) = w_p$, so

$$\begin{aligned}
F_W(w) &= \int_{-\infty}^{w} f_W(t)dt = \int_{-\infty}^{w} \frac{m!}{(j-1)!(m-j)!} g(t)\,[G(t)]^{j-1}\,[1-G(t)]^{m-j}\,dt \\
&= \int_{-\infty}^{w} \frac{m!}{(j-1)!(m-j)!}\,[G(t)]^{j-1}\,[1-G(t)]^{m-j}\,dG(t) \\
&= \frac{m!}{(j-1)!(m-j)!} \int_{-\infty}^{w} [G(t)]^{j-1} \sum_{i=0}^{m-j} \frac{(m-j)!}{i!(m-j-i)!}(-1)^i\,[G(t)]^i\,dG(t) \\
&= \frac{m!}{(j-1)!(m-j)!} \sum_{i=0}^{m-j} \left[\frac{(m-j)!}{i!(m-j-i)!}(-1)^i \int_{-\infty}^{w} [G(t)]^{j+i-1} dG(t)\right] \\
&= \frac{m!}{(j-1)!} \sum_{i=0}^{m-j} \left[\frac{(-1)^i}{i!(m-j-i)!(i+j)}[G(w)]^{i+j}\right] \qquad (5.3)
\end{aligned}$$

Proposition 5.1. *Let $\hat{a}$ and $\hat{b}$ be the two location–scale invariant estimators of a and b obtained from observations $x_1, \cdots, x_n$. Define $U = \frac{\hat{a}-a}{b}$ and $V = \frac{\hat{b}}{b}$. A $100\gamma\%$ one–sided upper prediction interval for $Y_{(j)}$, future order observation of $Y_1, \cdots, Y_m$, is given by $(-\infty, \hat{a} + k_2\hat{b})$ with k_2 satisfying*

$$\int_0^{\infty} \int_{-\infty}^{\infty} F_W\left(u + k_2 v\right) f_{U,V|A}(u, v|A)dudv = \gamma$$

and we can evaluate k_2 numerically.

We now extend the result to two-sided prediction intervals. Let $(L(\underline{X}), U(\underline{X})) = (\hat{a} + k_1\hat{b}, \hat{a} + k_2\hat{b})$ be the $100\gamma\%$ prediction interval for $T(\underline{Y})$. For a given functional form of $\hat{a}$ and $\hat{b}$ all we need is the pair (k_1, k_2) such that (5.3) is satisfied. But it is obvious that there will be more than one such pair which can satisfy the equation.

To find an optimum pair we need to impose one more condition. One of criteria is $f_W(k_1) = f_W(k_2)$. This is in the similar spirit of finding the shortest interval. Another criterion is motivated from the equal tail area criteria of confidence interval. For this type of prediction interval we solve for k_l and k_2 from two equations of one-sided prediction interval, i.e. $\int_0^{\infty} \int_{-\infty}^{\infty} F_W\left(u + k_1 v\right) f_{U,V|A}(u, v|A)dudv = \frac{1-\gamma}{2}$

and $\int_0^\infty \int_{-\infty}^\infty F_W\left(u+k_2 v\right) f_{U,V|A}(u,v|A)dudv = \frac{1+\gamma}{2}$. We summarize the procedure as follow:

Proposition 5.2. *Let $\hat{a}$ and $\hat{b}$ be the two location–scale invariant estimators of a and b obtained from observations $x_1, \cdots, x_n$. Define $U = \frac{\hat{a}-a}{b}$ and $V = \frac{\hat{b}}{b}$. A $100\gamma\%$ two-sided optimum prediction interval for $Y_{(j)}$, future order observation of $Y_1, \cdots, Y_m$, is given by $(\hat{a}+k_1\hat{b}, \hat{a}+k_2\hat{b})$ with k_1 and k_2 satisfying*

$$\begin{aligned}
\int_0^\infty \int_{-\infty}^\infty F_W\left(u+k_1 v\right) f_{U,V|A}(u,v|A)dudv &= \frac{1-\gamma}{2} \\
\int_0^\infty \int_{-\infty}^\infty F_W\left(u+k_2 v\right) f_{U,V|A}(u,v|A)dudv &= \frac{1+\gamma}{2}
\end{aligned}$$

5.3 Construction of an Unconditional Prediction Interval

Instead of an exact analytic approach, Kushary [70] uses a Taylor series expansion and Monte Carlo simulation to obtain an approximate formula for a unconditional cumulative density of W such that

$$\begin{aligned}
P\left(Y_{(j)} < \hat{a}+k_2\hat{b}\right) &= P\left(\frac{Y_{(j)}-a}{b} < \frac{\hat{a}-a}{b} + k_2\frac{\hat{b}}{b}\right) \\
&= P\left(W < U + k_2 V\right) \\
&= E\left[P\left(W < U + k_2 V \middle| U, V\right)\right] \qquad (5.4)
\end{aligned}$$

where $\hat{a}$ and $\hat{b}$ are invariant estimators.

Recall that the Taylor series expansion of $f(x,y)$ for a point in the neighborhood of (x_0, y_0) is

$$\begin{aligned}
f(x,y) &= f(x_0,y_0) + \left[(x-x_0)\frac{\partial}{\partial x} + (y-y_0)\frac{\partial}{\partial y}\right] f \\
&+ \frac{1}{2!}\left[(x-x_0)\frac{\partial}{\partial x} + (y-y_0)\frac{\partial}{\partial y}\right]^2 f + \frac{1}{3!}\left[(x-x_0)\frac{\partial}{\partial x} + (y-y_0)\frac{\partial}{\partial y}\right]^3 f + \cdots
\end{aligned}$$

Let $W = \frac{Y_{(i)}-a}{b}$. Since

$$\begin{aligned}
\gamma &= P\left(Y_{(j)} < \hat{a} + k_2\hat{b}\right) = P\left(\frac{Y_{(j)}-a}{b} < \frac{\hat{a}-a}{b} + k_2\frac{\hat{b}}{b}\right) \\
&= P\left(W < U + k_2 V\right) = E\left[P\left(W < U + k_2 V \middle| U, V\right)\right] \\
&= E_{U,V}\left[F_W\left(U + k_2 V\right)\right]
\end{aligned}$$

Noticing that W is pivotal, we expand $F_W\left(U + k_2 V\right)$ at $(0,1)$ using Taylor's formula. This gives

$$\begin{aligned}
F_W\left(U + k_2 V\right) &\approx F_W\left(k_2\right) + \left[U\frac{\partial}{\partial U} + (V-1)\frac{\partial}{\partial V}\right] F_W\left(k_2\right) \\
&+ \frac{1}{2!}\left[U\frac{\partial}{\partial U} + (V-1)\frac{\partial}{\partial V}\right]^2 F_W\left(k_2\right) \\
&= F_W\left(k_2\right) + f_W\left(k_2\right)\left[U + k_2(V-1)\right] \\
&+ \frac{1}{2} f'_W(k_2)\left[U^2 + 2k_2 U(V-1) + k_2^2(V-1)^2\right]
\end{aligned}$$

so we have

$$\begin{aligned}
\gamma &\approx F_W\left(k_2\right) + f_W\left(k_2\right)\left[E(U) + k_2 E(V-1)\right] \\
&+ \frac{1}{2} f'_W(k_2) E\left[U^2 + 2k_2 U(V-1) + k_2^2(V-1)^2\right]
\end{aligned}$$

Obviously, we can find $E(U),\ E(V-1),\ E(U^2),\ E\left[U(V-1)\right]$, and $E\left[(V-1)^2\right]$ by Monte Carlo simulation, since U, V are independent of parameters, and then solve above equation to obtain k_2. From previous results:

$$\hat{a} = \bar{X}_n - \frac{\mu_Z}{E(S_{Z,n})} S_{X,n} \quad and \quad \hat{b} = \frac{1}{E(S_{Z,n})} S_{X,n}$$

Let $Z_i = \frac{X_i - a}{b}$, so we have

$$\begin{aligned}
U &= \frac{\hat{a}-a}{b} = \bar{Z}_n - \frac{\mu_Z}{E(S_Z)} S_Z \\
V &= \frac{\hat{b}}{b} = \frac{1}{E(S_Z)} S_Z
\end{aligned}$$

Let $W = \frac{Y_{(i)}-a}{b}$ and notice that

$$E(U) = 0 \;\; and \;\; E(V) = 1$$

so we have

$$\gamma \;\approx\; F_W(k_2) + \frac{1}{2} f'_W(k_2) E\left[U^2 + 2k_2 U(V-1) + k_2^2 (V-1)^2\right] \tag{5.5}$$

where

$$\begin{aligned}
E[U^2] &= E\left[\bar{Z}_n - \frac{\mu_Z}{E(S_{Z,n})} S_{Z,n}\right]^2 \\
&= E[\bar{Z}_n^2] - \frac{2\mu_Z}{E(S_{Z,n})} E[\bar{Z}_n S_{Z,n}] + \left[\frac{\mu_Z}{E(S_{Z,n})}\right]^2 E(S_{Z,n}^2) \\
E[U(V-1)] &= E[UV] = E\left[\left(\bar{Z}_n - \frac{\mu_Z}{E(S_{Z,n})} S_{Z,n}\right) \frac{1}{E(S_{Z,n})} S_{Z,n}\right] \\
&= \frac{1}{E(S_{Z,n})} E[\bar{Z}_n S_{Z,n}] - \frac{\mu_Z}{[E(S_{Z,n})]^2} E(\bar{Z}_n S_{Z,n}^2) \\
E[(V-1)^2] &= E[V^2] - 1 = E\left[\left(\frac{1}{E(S_{Z,n})} S_{Z,n}\right)^2\right] - 1 \\
&= \frac{1}{[E(S_{Z,n})]^2} E[S_{Z,n}^2] - 1
\end{aligned}$$

5.4 Construction of Prediction Intervals Using a Monte Carlo Method

Suppose that $\{X_1, X_2, \cdots, X_n\}$ is a random sample from a location-scale family, and and let $\underline{Y} = (Y_1, Y_2, \cdots, Y_m)$ denote future observations with the same distribution.

In this section, we construct prediction intervals for the location-scale family when both a and b are unknown with pivotal

$$U = \frac{Y_{(j)} - X_{(i)}}{\hat{b}}, \quad i = 1, 2, \cdots, n, \quad j = 1, 2, \cdots, m \tag{5.6}$$

The exact distribution of the pivotal U is too hard to obtain, so we use Monte Carlo

to estimate the percentiles of U.

Let $X_i = a + bZ_i^X$ and $Y_j = a + bZ_j^Y$. Using our previous estimator $\hat{b} = \frac{1}{E(S_{Z^X,n})} S_{X,n}$, then the procedure for generating the percentiles of U is as follows:

1. Generate random variables $Z_1^X, \cdots, Z_n^X$ and $Z_1^Y, \cdots, Z_m^Y$ to be identically and independently standard distribution $(a = 0, b = 1)$.

2. Calculate the statistic:

$$\begin{aligned} \hat{U} &= \frac{Y_{(j)} - X_{(i)}}{\hat{b}} = \frac{b\left[Z_{(j)}^Y - Z_{(i)}^X\right]}{S_{X,n}} E(S_{Z^X,n}) \\ &= \frac{Z_{(j)}^Y - Z_{(i)}^X}{S_{Z^X,n}} E(S_{Z^X,n}) \end{aligned}$$

3. Repeat Step 1 and Step 2 for B times.

4. Estimate the desired percentile of U by using $\hat{U}_1, \cdots, \hat{U}_B$.

Returning to our case and setting $\gamma = 1 - \alpha$

$$\begin{aligned} 1 - \alpha &= P\left(\widehat{Q}_{\frac{\alpha}{2}} < \frac{Y_{(j)} - X_{(i)}}{\hat{b}} \leq \widehat{Q}_{1-\frac{\alpha}{2}}\right) \\ &= P\left(X_{(i)} + \widehat{Q}_{\frac{\alpha}{2}}\hat{b} < Y_{(j)} \leq X_{(i)} + \widehat{Q}_{1-\frac{\alpha}{2}}\hat{b}\right) \end{aligned}$$

that is, a $100(1-\alpha)\%$ two-sided prediction interval for $Y_{(j)}$ is

$$[L(\underline{X}), U(\underline{X})] = \left[X_{(i)} + \widehat{Q}_{\frac{\alpha}{2}}\hat{b}, X_{(i)} + \widehat{Q}_{1-\frac{\alpha}{2}}\hat{b}\right] \tag{5.7}$$

Then the length of the above prediction interval is

$$L_{PI} = \left(\widehat{Q}_{1-\frac{\alpha}{2}} - \widehat{Q}_{\frac{\alpha}{2}}\right)\hat{b} \tag{5.8}$$

If we simulate D times, then the coverage probability can be estimated using

$$Pr = \frac{1}{D}\sum_{k=1}^{D} c_k$$

where

$$c_k = \begin{cases} 1, & y_{(j)}^{(k)} \in [L(\underline{X}), U(\underline{X})] \\ 0, & else \end{cases}$$

In the Section 7.5, we will give an example to illustrate how to choose $X_{(i)}$ in order to obtain an optimal prediction interval.

Chapter 6

Bayesian Prediction Intervals for the Gumbel Minimum Distribution

Bayesian prediction, in particular, for the minimum, maximum and mean of a future sample, are common types of prediction. In this chapter, predictions are first proposed that are based on a Bayesian predictive density function when the scale parameter is known. When both location and scale parameters are unknown, it is impossible to obtain an explicit form for a Bayesian predictive density function. Later in this chapter, a Gibbs sampler, a special case of Markov Chain Monte Carlo (MCMC), is suggested for calculating prediction intervals for future observation from a Gumbel minimum distribution.

Suppose that $\{X_1, X_2, \cdots, X_n\}$ is a random sample from the Gumbel minimum distribution $f_X(x|\theta)$ with $\theta = (a, b) \in \Theta$. $\underline{Y} = (Y_1, Y_2, \cdots, Y_m)$ are future observations from same distribution. Then the predictive density $f(\underline{Y}|\underline{X})$ describes the uncertainty in the values that $\underline{Y}$ will take, given the information provided by $\underline{X}$ and any other available knowledge.

Let $\pi(\theta)$ be a prior distribution describing the available information on the value of θ. It then follows that

$$f(\underline{y}|\underline{x}) = \int_{\Theta} f_Y(y|\theta)\pi(\theta|\underline{x})d\theta \tag{6.1}$$

which is an average of the probability density of $\underline{Y}$ conditional on the (unknown) value of θ, weighted with $\pi(\theta|\underline{x})$, the posterior distribution of θ given $\underline{X}$.

6.1 Construction of a Bayesian Prediction Interval Using a Noninformative Prior on a when b is Known

When little or no information about location a is available, a noninformative uniform prior of the following form is often used with

$$\pi(a) = 1$$

In this case, the posterior for a is given by

$$\pi(a \mid x_1, x_2, \cdots, x_n) = \frac{1}{b\Gamma(n)} \left(\sum_{i=1}^{n} e^{\frac{x_i}{b}} \right)^n e^{-\frac{na}{b}} e^{-e^{-\frac{a}{b}} \left(\sum_{i=1}^{n} e^{\frac{x_i}{b}} \right)}$$

Hoppe and Lin [58] develop a Bayesian prediction interval for the minimum of any specified number of future observation from a Gumbel distribution based on previous observations and a noninformative uniform prior. They give the $100(1-\alpha)\%$ lower prediction limit for Y_{min}

$$y_L = b \ln \left[\frac{(1-\alpha)^{-\frac{1}{n}} - 1}{m} \sum_{i=1}^{m} e^{\frac{x_i}{b}} \right]$$

However, there is an error in their proof when b is unknown. In below, we construct a Bayesian prediction interval for any future order statistic.

6.1.1 Prediction Interval for Y_{max}

The lower prediction limit for Y_{max} is obtained by evaluating $P(Y_{max} > k \mid \underline{x})$ for some constant k. We have been unable to get an explicit form for k, and so a numerical

solution is required. Write

$$
\begin{aligned}
&f_{Y_{max}}(y \mid a)\pi(a \mid x_1, \cdots, x_n) \\
&= \left\{ m\frac{1}{b}e^{\frac{y-a}{b}}e^{-e^{\frac{y-a}{b}}}\sum_{k=0}^{m-1}(-1)^k\tbinom{m-1}{k}\left[e^{-e^{\frac{y-a}{b}}}\right]^k \right\} \times \left\{ \frac{1}{b\Gamma(n)}\left(\sum_{i=1}^{n}e^{\frac{x_i}{b}}\right)^n e^{-\frac{na}{b}}e^{-e^{-\frac{a}{b}}\left(\sum_{i=1}^{n}e^{\frac{x_i}{b}}\right)} \right\} \\
&= \frac{m}{b^2\Gamma(n)}\left(\sum_{i=1}^{n}e^{\frac{y_i}{b}}\right)^n \sum_{k=0}^{m-1}\left[(-1)^k\tbinom{m-1}{k}e^{\frac{y-a}{b}}e^{-\frac{na}{b}}e^{-e^{\frac{y-a}{b}}}\left(e^{-e^{\frac{y-a}{b}}}\right)^k e^{-e^{-\frac{a}{b}}\left(\sum_{i=1}^{n}e^{\frac{x_i}{b}}\right)}\right]
\end{aligned}
$$

Now let $t = e^{\frac{-a}{b}}$, $A = \sum_{i=1}^{n} e^{\frac{x_i}{b}}$, and $B = e^{\frac{y}{b}}$, so

$$
f_{Y_{max}}(y \mid a)\pi(a \mid x_1, \cdots, x_n) = \frac{mA^nB}{b^2\Gamma(n)}\sum_{k=0}^{m-1}\left[(-1)^k\tbinom{m-1}{k}t^{n+1}e^{-[A+(k+1)B]t}\right]
$$

Now using Proposition A.3,

$$
\int_0^\infty x^n e^{-ax}dx = \Gamma(n+1)a^{-n-1}, \quad a > 0
$$

And then the predictive density function for Y_{max} is

$$
\begin{aligned}
f_{Y_{max}}(y|x_1, \cdots, x_n) &= \int_{-\infty}^{\infty} f_{Y_{max}}(y \mid a)\pi(a \mid x_1, \cdots, x_n)da \\
&= \int_{\infty}^{0} \frac{mA^nB}{b^2\Gamma(n)}\sum_{k=0}^{m-1}\left[(-1)^k\tbinom{m-1}{k}t^{n+1}e^{-[A+(k+1)B]t}\right]\left(-\frac{b}{t}\right)dt \\
&= \frac{mA^nB}{b\Gamma(n)}\sum_{k=0}^{m-1}\left[(-1)^k\tbinom{m-1}{k}\int_0^\infty t^n e^{-[A+(k+1)B]t}dt\right] \\
&= \frac{mnA^nB}{b}\sum_{k=0}^{m-1}\left[(-1)^k\tbinom{m-1}{k}\left(\frac{1}{A+(k+1)B}\right)^{n+1}\right]
\end{aligned}
$$

Suppose the $100(1-\alpha)\%$ one-sided upper prediction interval for Y_{max} is $I_B = (y_L, \infty)$. We then have

$$
\begin{aligned}
\alpha &= \int_{-\infty}^{y_L} f_{Y_{max}}(y|x_1, \cdots, x_n)dy \\
&= \frac{mnA^n}{b}\sum_{k=0}^{m-1}(-1)^k\tbinom{m-1}{k}\int_{-\infty}^{y_L}\frac{e^{\frac{y}{b}}}{\left[A+(k+1)e^{\frac{y}{b}}\right]^{n+1}}dy \\
&= mnA^n\sum_{k=0}^{m-1}(-1)^k\tbinom{m-1}{k}\int_{-\infty}^{y_L}\frac{1}{\left[A+(k+1)e^{\frac{y}{b}}\right]^{n+1}}de^{\frac{y}{b}} \\
&= mnA^n\sum_{k=0}^{m-1}(-1)^k\binom{m-1}{k}\frac{1}{k+1}\int_{-\infty}^{y_L}\frac{1}{\left[A+(k+1)e^{\frac{y}{b}}\right]^{n+1}}d\left[A+(k+1)e^{\frac{y}{b}}\right] \\
&= \sum_{k=0}^{m-1}(-1)^k\binom{m-1}{k}\frac{mA^n}{k+1}\left\{\frac{1}{A^n}-\frac{1}{\left[A+(k+1)e^{\frac{y_L}{b}}\right]^n}\right\} \\
&= \sum_{k=1}^{m}(-1)^{k-1}\binom{m}{k}\left[1-\left(\frac{\sum_{i=1}^n e^{\frac{x_i}{b}}}{\sum_{i=1}^n e^{\frac{x_i}{b}}+ke^{\frac{y_L}{b}}}\right)^n\right]
\end{aligned}
$$

Recall that (Alan, p.34): $\sum_{k=0}^m (-1)^k\binom{m}{k} = 0$, therefore

$$
\begin{aligned}
\sum_{k=1}^{m}(-1)^k\binom{m}{k}\left(\frac{\sum_{i=1}^n e^{\frac{x_i}{b}}}{\sum_{i=1}^n e^{\frac{x_i}{b}}+ke^{\frac{y_L}{b}}}\right)^n &= \alpha+\sum_{k=1}^{m}(-1)^k\binom{m}{k} = \alpha+\sum_{k=0}^{m}(-1)^k\binom{m}{k}-1 \\
&= \alpha-1
\end{aligned}
$$

This gives

$$
\sum_{k=0}^{m}(-1)^k\binom{m}{k}\left(\frac{\sum_{i=1}^n e^{\frac{x_i}{b}}}{\sum_{i=1}^n e^{\frac{y_i}{b}}+ke^{\frac{y_L}{b}}}\right)^n = \alpha
$$

Next, we will prove that there is a unique y_L for the above equation. Let $A = \sum_{i=1}^n e^{\frac{x_i}{b}}$ and $B = e^{\frac{y_L}{b}}$, then, we have

$$
\sum_{k=0}^{m}(-1)^k\binom{m}{k}\left(\frac{A}{A+kB}\right)^n = \alpha
$$

Setting

$$H(B) = \sum_{k=0}^{m} (-1)^k \binom{m}{k} \frac{1}{(A+kB)^n}$$

Obviously, $H(B) \geq 0$, then

$$\begin{aligned}
H'(B) &= \sum_{k=1}^{m} (-1)^k \binom{m}{k} \frac{-kn}{(A+kB)^{n+1}} = n \sum_{k=1}^{m} (-1)^{k-1} \frac{m!}{(k-1)!(m-k)!} \frac{1}{(A+kB)^{n+1}} \\
&= mn \sum_{i=0}^{m-1} (-1)^i \frac{(m-1)!}{i!(m-1-i)!} \frac{1}{[A+(i+1)B]^{n+1}} \\
&= mn \sum_{k=0}^{m-1} (-1)^k \binom{m-1}{k} \frac{1}{[(A+B)+kB]^{n+1}} > 0
\end{aligned}$$

This shows that $H(B)$ is monotone increasing for B and so that y_L is the unique solution for $H(B)A^n = \alpha$.

6.1.2 Prediction Intervals for $Y_{(j)}$

In constructing a lower prediction limit, we need to determine the predictive density for $Y_{(j)}$. As in the last section, we first evaluate

$$\begin{aligned}
&f_{Y_{(j)}}(y \mid a)\pi(a \mid \underline{x}) = \left\{ \frac{m!}{(j-1)!(m-j)!b} e^{\frac{y-a}{b}} \left[1 - e^{-e^{\frac{y-a}{b}}}\right]^{j-1} e^{-(m-j+1)e^{\frac{x-a}{b}}} \right\} \\
&\times \left\{ \frac{1}{b\Gamma(n)} \left(\sum_{i=1}^{n} e^{\frac{x_i}{b}} \right)^n e^{-\frac{na}{b}} e^{-e^{-\frac{a}{b}}\left(\sum_{i=1}^{n} e^{\frac{x_i}{b}}\right)} \right\} \\
&= \frac{m!}{b^2\Gamma(n)(j-1)!(m-j)!} \left(\sum_{i=1}^{n} e^{\frac{x_i}{b}} \right)^n e^{\frac{y-a}{b}} e^{-\frac{na}{b}} \left(1 - e^{-e^{\frac{y-a}{b}}}\right)^{j-1} e^{-(m-j+1)e^{\frac{y-a}{b}}} \\
&\times e^{-e^{-\frac{a}{b}}\left(\sum_{i=1}^{n} e^{\frac{x_i}{b}}\right)}
\end{aligned}$$

Let $t = e^{\frac{-a}{b}}, \quad A = \sum_{i=1}^{n} e^{\frac{x_i}{b}}, \quad B = e^{\frac{y}{b}}$, then

$$\begin{aligned}
f_{Y_{(j)}}(y \mid a)\pi(a \mid \underline{x}) &= \frac{m!}{b^2\Gamma(n)(j-1)!(m-j)!} A^n B t^{n+1} \left(1 - e^{-Bt}\right)^{j-1} e^{-(m-j+1)Bt} e^{-At} \\
&= \frac{m!}{b^2\Gamma(n)(j-1)!(m-j)!} A^n B t^{n+1} e^{-[A+(m-j+1)B]t} \sum_{i=0}^{j-1} (-1)^i \tbinom{j-1}{i} e^{-iBt} \\
&= \frac{m! A^n B}{b^2\Gamma(n)(j-1)!(m-j)!} \sum_{i=0}^{j-1} (-1)^i \tbinom{j-1}{i} t^{n+1} e^{-[A+(m-j+1+i)B]t}
\end{aligned}$$

Now using Proposition A.3, the predictive density function for $Y_{(j)}$ is then given by

$$\begin{aligned}
f_{Y_{(j)}}(y|\underline{x}) &= \int_{-\infty}^{\infty} f_{Y_{(j)}}(y \mid a)\pi(a \mid \underline{x}) da \\
&= \int_{\infty}^{0} \frac{m! A^n B}{b^2\Gamma(n)(j-1)!(m-j)!} \sum_{i=0}^{j-1} (-1)^i \tbinom{j-1}{i} t^{n+1} e^{-[A+(m-j+1+i)B]t} \left(-\frac{b}{t}\right) dt \\
&= \frac{m! A^n B}{b\Gamma(n)(j-1)!(m-j)!} \sum_{i=0}^{j-1} (-1)^i \tbinom{j-1}{i} \int_{0}^{\infty} t^n e^{-[A+(m-j+1+i)B]t} dt \\
&= \frac{m! A^n B}{b\Gamma(n)(j-1)!(m-j)!} \sum_{i=0}^{j-1} (-1)^i \tbinom{j-1}{i} \frac{\Gamma(n+1)}{[A+(m-j+1+i)B]^{n+1}} \\
&= \frac{m! n A^n B}{b(j-1)!(m-j)!} \sum_{i=0}^{j-1} (-1)^i \tbinom{j-1}{i} \frac{1}{[A+(m-j+1+i)B]^{n+1}}
\end{aligned}$$

Suppose the one-sided $100(1-\alpha)\%$ upper prediction interval $Y_{(j)}$ is $I_B = (y_L, \infty)$, we then have

$$
\begin{aligned}
\alpha &= \int_{-\infty}^{y_L} f_{Y_{(j)}}(y|x_1, \cdots, x_n)dy \\
&= \int_{-\infty}^{y_L} \frac{m!nA^nB}{b(j-1)!(m-j)!} \sum_{i=0}^{j-1} (-1)^i \tbinom{j-1}{i} \frac{1}{[A+(m-j+1+i)B]^{n+1}} dy \\
&= \frac{m!nA^n}{b(j-1)!(m-j)!} \sum_{i=0}^{j-1} (-1)^i \tbinom{j-1}{i} \int_{-\infty}^{y_L} \frac{B}{[A+(m-j+1+i)B]^{n+1}} dy \\
&= \frac{m!nA^n}{(j-1)!(m-j)!} \sum_{i=0}^{j-1} (-1)^i \tbinom{j-1}{i} \int_{-\infty}^{y_L} \frac{e^{\frac{y}{b}}}{[A+(m-j+1+i)e^{\frac{y}{b}}]^{n+1}} d\left(\frac{y}{b}\right) \\
&= \frac{m!nA^n}{(j-1)!(m-j)!} \sum_{i=0}^{j-1} (-1)^i \tbinom{j-1}{i} \int_{-\infty}^{y_L} \frac{1}{[A+(m-j+1+i)e^{\frac{y}{b}}]^{n+1}} de^{\frac{y}{b}} \\
&= \frac{m!nA^n}{(j-1)!(m-j)!} \sum_{i=0}^{j-1} (-1)^i \tbinom{j-1}{i} \frac{1}{m-j+1+i} \\
&\times \int_{-\infty}^{y_L} \frac{1}{[A+(m-j+1+i)e^{\frac{y}{b}}]^{n+1}} d[A+(m-j+1+i)e^{\frac{y}{b}}] \\
&= \frac{m!}{(j-1)!(m-j)!} \sum_{i=0}^{j-1} (-1)^i \tbinom{j-1}{i} \frac{1}{m-j+1+i} \left[1 - \frac{A^n}{[A+(m-j+1+i)e^{\frac{y_L}{b}}]^n}\right] \\
&= \frac{m!}{(j-1)!(m-j)!} \sum_{i=0}^{j-1} (-1)^i \tbinom{j-1}{i} \frac{1}{m-j+1+i} \left[1 - \left(\frac{A}{A+(m-j+1+i)e^{\frac{y_L}{b}}}\right)^n\right] \\
&= \frac{m!}{(j-1)!(m-j)!} \sum_{i=0}^{j-1} (-1)^i \tbinom{j-1}{i} \frac{1}{m-j+1+i} \left[1 - \left(\frac{e^{\frac{x_i}{b}}}{e^{\frac{x_i}{b}}+(m-j+1+i)e^{\frac{y_L}{b}}}\right)^n\right]
\end{aligned}
$$

We determine y_L by solving the following equation

$$
\alpha = \frac{m!}{(j-1)!(m-j)!} \sum_{i=0}^{j-1} (-1)^i \tbinom{j-1}{i} \frac{1}{m-j+1+i} \left[1 - \left(\frac{e^{\frac{x_i}{b}}}{e^{\frac{x_i}{b}}+(m-j+1+i)e^{\frac{y_L}{b}}}\right)^n\right] \tag{6.2}
$$

which can be solved numerically.

6.2 Construction of a Bayesian Prediction Interval Using a Conjugate Prior for a with Known b

Engelund and Rackwitz [33] propose the following conjugate prior for location a with b fixed for Gumbel maximum distribution

$$\pi'(a) = \frac{r'^{n+p}}{b\Gamma(n+p)} \exp\left[(n+p)\frac{a}{b} - r' \exp\left(\frac{a}{b}\right)\right]$$

where p is constant and $r' = \sum_{i=1}^{n} e^{-\frac{x_i}{b}}$ is sufficient statistics for parameter a. We will use the prior

$$\pi(a) = \frac{r^{n+p}}{b\Gamma(n+p)} \exp\left[-(n+p)\frac{a}{b} - r \exp\left(-\frac{a}{b}\right)\right] \tag{6.3}$$

with $r = \sum_{i=1}^{n} e^{\frac{x_i}{b}}$. Using Proposition A.1, we have

$$\begin{aligned}
& \int_{-\infty}^{\infty} \pi(a) L(a|x_1, \cdots, x_n) da \\
= & \int_{-\infty}^{\infty} \frac{r^{n+p}}{b\Gamma(n+p)} \exp\left[-(n+p)\frac{a}{b} - r \exp\left(-\frac{a}{b}\right)\right] \left(\frac{1}{b}\right)^n e^{\sum_{i=1}^{n} \frac{x_i}{b}} e^{-\frac{na}{b}} e^{-e^{-\frac{a}{b}} \sum_{i=1}^{n} e^{\frac{x_i}{b}}} da \\
= & \frac{r^{n+p} e^{\sum_{i=1}^{n} \frac{x_i}{b}}}{b^{n+1}\Gamma(n+p)} \int_{-\infty}^{\infty} e^{(2n+p)(-\frac{a}{b})} e^{-2re^{-\frac{a}{b}}} da \\
= & -\frac{r^{n+p} e^{\sum_{i=1}^{n} \frac{x_i}{b}}}{b^n\Gamma(n+p)} e^{-(2n+p)\ln(2r)} \int_{-\infty}^{\infty} e^{(2n+p)[-\frac{a}{b}+\ln(2r)]} e^{-e^{-\frac{a}{b}+\ln(2r)}} d\left[-\frac{a}{b} + \ln(2r)\right] \\
= & \frac{r^{n+p} e^{\sum_{i=1}^{n} \frac{x_i}{b}}}{b^n\Gamma(n+p)} \times \frac{1}{(2r)^{2n+p}} \times \Gamma(2n+p) = \frac{\Gamma(2n+p) e^{\sum_{i=1}^{n} \frac{x_i}{b}}}{2^{2n+p} b^n \Gamma(n+p) \left(\sum_{i=1}^{n} e^{\frac{x_i}{b}}\right)^n}
\end{aligned}$$

where $2n + p > 0$. Then the posterior distribution is

$$\begin{aligned}
\pi(a|x_1, \cdots, x_n) &= \frac{\pi(a) L(a|x_1, \cdots, x_n)}{\int_{-\infty}^{\infty} \pi(a) L(a|x_1, \cdots, x_n) da} \\
&= \frac{\frac{r^{n+p}}{b\Gamma(n+p)} \exp\left[-(n+p)\frac{a}{b} - r \exp\left(-\frac{a}{b}\right)\right] \left(\frac{1}{b}\right)^n e^{\sum_{i=1}^{n} \frac{x_i}{b}} e^{-\frac{na}{b}} e^{-e^{-\frac{a}{b}} \sum_{i=1}^{n} e^{\frac{x_i}{b}}}}{\frac{\Gamma(2n+p) e^{\sum_{i=1}^{n} \frac{x_i}{b}}}{2^{2n+p} b^n \Gamma(n+p) \left(\sum_{i=1}^{n} e^{\frac{x_i}{b}}\right)^n}} \\
&= \frac{(2r)^{2n+p}}{b\Gamma(2n+p)} \exp\left[-(2n+p)\frac{a}{b} - 2r \exp\left(-\frac{a}{b}\right)\right] \qquad (6.4)
\end{aligned}$$

Notice that $f_{Y_{min}}(y|a) = \frac{m}{b} e^{\frac{y-a}{b}} e^{-me^{\frac{y-a}{b}}}$, therefore

$$
\begin{aligned}
& f_{Y_{min}}(y|a)\pi(a|x_1, \cdots, x_n) \\
= \; & \frac{m}{b} e^{\frac{y-a}{b}} e^{-me^{\frac{y-a}{b}}} \times \frac{(2r)^{2n+p}}{b\Gamma(2n+p)} \exp\left[-(2n+p)\frac{a}{b} - 2r\exp\left(-\frac{a}{b}\right)\right] \\
= \; & \frac{m(2r)^{2n+p}}{b^2\Gamma(2n+p)} e^{\frac{y}{b}} e^{-(2n+p+1)\frac{a}{b}} e^{-(2r+me^{\frac{y}{b}})e^{-\frac{a}{b}}}
\end{aligned}
$$

Then for $2n+p>0$, the predictive density for Y_{min} is

$$
\begin{aligned}
& f_{Y_{min}}(y|x_1, \cdots, x_n) \\
= \; & \int_{-\infty}^{\infty} f_{Y_{min}}(y|a)\pi(a|x_1, \cdots, x_n) da \\
= \; & \frac{m(2r)^{2n+p}}{b^2\Gamma(2n+p)} e^{\frac{y}{b}} \int_{-\infty}^{\infty} e^{-(2n+p+1)\frac{a}{b}} e^{-(2r+me^{\frac{y}{b}})e^{-\frac{a}{b}}} da \\
= \; & \frac{m(2r)^{2n+p}}{b\Gamma(2n+p)} e^{\frac{y}{b}} \times (-1) e^{-(2n+p+1)\ln(2r+me^{\frac{y}{b}}))} \\
\times \; & \int_{-\infty}^{\infty} e^{(2n+p+1)(-\frac{a}{b}+\ln(2r+me^{\frac{y}{b}}))} e^{-e^{-\frac{a}{b}+\ln(2r+me^{\frac{x}{b}})}} d\left[-\frac{a}{b} + \ln(2r+me^{\frac{y}{b}})\right] \\
= \; & \frac{m(2r)^{2n+p}}{b\Gamma(2n+p)(2r+me^{\frac{y}{b}})^{2n+p+1}} e^{\frac{y}{b}} \\
\times \; & (-1) \int_{-\infty}^{\infty} e^{(2n+p+1)(-\frac{a}{b}+\ln(2r+me^{\frac{y}{b}}))} e^{-e^{-\frac{a}{b}+\ln(2r+me^{\frac{y}{b}})}} d\left[-\frac{a}{b} + \ln(2r+me^{\frac{y}{b}})\right] \\
= \; & \frac{m(2r)^{2n+p}}{b\Gamma(2n+p)(2r+me^{\frac{y}{b}})^{2n+p+1}} e^{\frac{y}{b}} \times (-1) \int_{\infty}^{-\infty} e^{(2n+p+1)t} e^{-e^t} dt \\
= \; & \frac{m(2r)^{2n+p}}{b\Gamma(2n+p)(2r+me^{\frac{y}{b}})^{2n+p+1}} e^{\frac{y}{b}} \times \Gamma(2n+p+1) \\
= \; & \frac{m(2n+p)(2r)^{2n+p}}{b(2r+me^{\frac{y}{b}})^{2n+p+1}} e^{\frac{y}{b}}
\end{aligned}
$$

Suppose the $100(1-\alpha)\%$ one-sided upper prediction interval is $I_B = (y_L, \infty)$. We have

$$\begin{aligned}
\alpha &= \int_{-\infty}^{y_L} f_{Y_{min}}(y|x_1, \cdots, x_n)dx = \int_{-\infty}^{y_L} \frac{m(2n+p)(2r)^{2n+p}}{b(2r+me^{\frac{y}{b}})^{2n+p+1}} e^{\frac{y}{b}} dy \\
&= m(2n+p)(2r)^{2n+p} \int_{-\infty}^{y_L} \frac{e^{\frac{y}{b}}}{(2r+me^{\frac{y}{b}})^{2n+p+1}} d\left(\frac{y}{b}\right) \\
&= (2n+p)(2r)^{2n+p} \int_{-\infty}^{y_L} \frac{1}{(2r+me^{\frac{y}{b}})^{2n+p+1}} d\left(me^{\frac{y}{b}}\right) \\
&= (2r)^{2n+p} \times \left[(2r)^{-(2n+p)} - (2r+me^{\frac{y_L}{b}})^{-(2n+p)}\right] \\
&= 1 - \frac{1}{[1+\frac{m}{2r}e^{\frac{y_L}{b}}]^{(2n+p)}}
\end{aligned}$$

This gives

$$y_L = b \ln \left[\frac{2[(1-\alpha)^{-\frac{1}{2n+p}} - 1]}{m} \sum_{i=1}^{m} e^{\frac{x_i}{b}} \right] \tag{6.5}$$

6.2.1 Prediction Interval for Y_{max}

Notice that

$$\begin{aligned}
f_{Y_{max}}(y \mid a) &= m f_Y(y) \left[F_Y(y)\right]^{m-1} \\
&= m\frac{1}{b} e^{\frac{y-a}{b}} e^{-e^{\frac{y-a}{b}}} \left[1 - e^{-e^{\frac{y-a}{b}}}\right]^{m-1} \\
&= m\frac{1}{b} e^{\frac{y-a}{b}} e^{-e^{\frac{y-a}{b}}} \sum_{k=0}^{m-1} (-1)^k \binom{m-1}{k} \left[e^{-e^{\frac{y-a}{b}}}\right]^k
\end{aligned}$$

Therefore

$$\begin{aligned}
&f_{Y_{max}}(y|a)\pi(a|x_1, \cdots, x_n) \\
&= m\frac{1}{b} e^{\frac{y-a}{b}} e^{-e^{\frac{y-a}{b}}} \sum_{k=0}^{m-1} (-1)^k \binom{m-1}{k} \left[e^{-e^{\frac{y-a}{b}}}\right]^k \times \frac{(2r)^{2n+p}}{b\Gamma(2n+p)} \exp\left[-(2n+p)\frac{a}{b} - 2r\exp\left(-\frac{a}{b}\right)\right] \\
&= \frac{m(2r)^{2n+p}}{b^2\Gamma(2n+p)} e^{\frac{y}{b}} \sum_{k=0}^{m-1} (-1)^k \binom{m-1}{k} e^{-(2n+p+1)\frac{a}{b}} e^{-[2r+(k+1)e^{\frac{y}{b}}]e^{-\frac{a}{b}}}
\end{aligned}$$

Then for $2n+p>0$, the predictive density of Y_{max} is

$$
\begin{aligned}
f_{Y_{max}}(y|x_1,\cdots,x_n) &= \int_{-\infty}^{\infty} f_{Y_{max}}(y \mid a)\pi(a \mid x_1,\cdots,x_n)da \\
&= \int_{-\infty}^{\infty} \frac{m(2r)^{2n+p}}{b^2\Gamma(2n+p)} e^{\frac{y}{b}} \sum_{k=0}^{m-1} (-1)^k \tbinom{m-1}{k} e^{-(2n+p+1)\frac{a}{b}} e^{-[2r+(k+1)e^{\frac{y}{b}}]e^{-\frac{a}{b}}} da \\
&= \frac{m(2r)^{2n+p}}{b^2\Gamma(2n+p)} e^{\frac{y}{b}} \sum_{k=0}^{m-1} \left[(-1)^k \tbinom{m-1}{k} \int_{-\infty}^{\infty} e^{-(2n+p+1)\frac{a}{b}} e^{-[2r+(k+1)e^{\frac{y}{b}}]e^{-\frac{a}{b}}} da \right]
\end{aligned}
$$

Since

$$
\begin{aligned}
&\int_{-\infty}^{\infty} e^{-(2n+p+1)\frac{a}{b}} e^{-[2r+(k+1)e^{\frac{y}{b}}]e^{-\frac{a}{b}}} da \\
&= -b e^{-(2n+p+1)\left\{\ln[2r+(k+1)e^{\frac{y}{b}}]\right\}} \\
&\times \int_{-\infty}^{\infty} e^{(2n+p+1)\left\{-\frac{a}{b}+\ln[2r+(k+1)e^{\frac{y}{b}}]\right\}} e^{-e^{-\frac{a}{b}+\ln[2r+(k+1)e^{\frac{y}{b}}]}} d\left\{-\frac{a}{b}+\ln[2r+(k+1)e^{\frac{y}{b}}]\right\} \\
&= -b e^{-(2n+p+1)\left\{\ln[2r+(k+1)e^{\frac{x}{b}}]\right\}} \int_{\infty}^{-\infty} e^{(2n+p+1)t} e^{-e^t} dt \\
&= b\Gamma(2n+p+1) e^{-(2n+p+1)\left\{\ln[2r+(k+1)e^{\frac{y}{b}}]\right\}} \\
&= b\Gamma(2n+p+1) \frac{1}{[2r+(k+1)e^{\frac{y}{b}}]^{2n+p+1}}
\end{aligned}
$$

and then

$$
f_{Y_{max}}(y|x_1,\cdots,x_n) = \frac{m(2n+p)(2r)^{2n+p}}{b} e^{\frac{y}{b}} \sum_{k=0}^{m-1} \left[(-1)^k \tbinom{m-1}{k} \frac{1}{[2r+(k+1)e^{\frac{y}{b}}]^{2n+p+1}} \right] \tag{6.6}
$$

Suppose the one-sided $100(1-\alpha)\%$ upper prediction interval is $I_B = (y_L, \infty)$. We then have

$$
\begin{aligned}
\alpha &= \int_{-\infty}^{y_L} f_{Y_{max}}(y|x_1, \cdots, x_n)dy \\
&= \int_{-\infty}^{y_L} \frac{m(2n+p)(2r)^{2n+p}}{b} e^{\frac{y}{b}} \sum_{k=0}^{m-1} \left[(-1)^k \tbinom{m-1}{k} \frac{1}{[2r+(k+1)e^{\frac{y}{b}}]^{2n+p+1}} \right] dy \\
&= m(2n+p)(2r)^{2n+p} \sum_{k=0}^{m-1} \left[(-1)^k \tbinom{m-1}{k} \int_{-\infty}^{y_L} \frac{1}{\left[2r+(k+1)e^{\frac{y}{b}}\right]^{2n+p+1}} de^{\frac{y}{b}} \right] \\
&= m(2n+p)(2r)^{2n+p} \sum_{k=0}^{m-1} \left\{ (-1)^k \binom{m-1}{k} \frac{1}{k+1} \int_{-\infty}^{y_L} \frac{1}{\left[2r+(k+1)e^{\frac{y}{b}}\right]^{2n+p+1}} d\left[2r+(k+1)e^{\frac{y}{b}}\right] \right\} \\
&= \sum_{k=0}^{m-1} (-1)^k \binom{m-1}{k} \frac{m(2r)^{2n+p}}{k+1} \left\{ \frac{1}{r^{n+p}} - \frac{1}{\left[2r+(k+1)e^{\frac{y_L}{b}}\right]^{2n+p}} \right\} \\
&= \sum_{k=1}^{m} (-1)^{k-1} \binom{m}{k} \left[1 - \left(\frac{2\sum_{i=1}^{n} e^{\frac{x_i}{b}}}{2\sum_{i=1}^{n} e^{\frac{x_i}{b}} + ke^{\frac{y_L}{b}}} \right)^{2n+p} \right]
\end{aligned}
$$

Recall that

$$
\sum_{k=0}^{m} (-1)^k \binom{m}{k} = 0
$$

Therefore

$$
\begin{aligned}
\sum_{k=1}^{m} (-1)^k \binom{m}{k} \left(\frac{2\sum_{i=1}^{n} e^{\frac{x_i}{b}}}{2\sum_{i=1}^{n} e^{\frac{x_i}{b}} + ke^{\frac{y_L}{b}}} \right)^{2n+p} &= \alpha + \sum_{k=1}^{m} (-1)^k \binom{m}{k} \\
&= \alpha + \sum_{k=0}^{m} (-1)^k \binom{m}{k} - 1 \\
&= \alpha - 1
\end{aligned}
$$

This gives

$$
\sum_{k=0}^{m} (-1)^k \binom{m}{k} \left(\frac{2\sum_{i=1}^{n} e^{\frac{x_i}{b}}}{2\sum_{i=1}^{n} e^{\frac{x_i}{b}} + ke^{\frac{y_L}{b}}} \right)^{n+p} = \alpha \tag{6.7}
$$

6.2.2 Prediction Interval for $Y_{(j)}$

We first calculate the posterior distribution

$$\begin{aligned}
f_{Y_{(j)}}(y \mid a) &= \frac{m!}{(j-1)!(m-j)!} f_Y(y) \left[F_Y(y)\right]^{j-1} \left[1 - F_Y(y)\right]^{m-j} \\
&= \frac{m!}{(j-1)!(m-j)!b} e^{\frac{y-a}{b}} e^{-e^{\frac{y-a}{b}}} \left[1 - e^{-e^{\frac{y-a}{b}}}\right]^{j-1} \left(e^{-e^{\frac{y-a}{b}}}\right)^{m-j} \\
&= \frac{m!}{(j-1)!(m-j)!b} e^{\frac{y-a}{b}} \left[1 - e^{-e^{\frac{y-a}{b}}}\right]^{j-1} e^{-(m-j+1)e^{\frac{y-a}{b}}}
\end{aligned}$$

Therefore

$$\begin{aligned}
&f_{Y_{(j)}}(y \mid a)\pi(a \mid x_1, \cdots, x_n) \\
&= \left\{ \frac{m!}{(j-1)!(m-j)!b} e^{\frac{y-a}{b}} \left[1 - e^{-e^{\frac{y-a}{b}}}\right]^{j-1} e^{-(m-j+1)e^{\frac{y-a}{b}}} \right\} \\
&\times \frac{(2r)^{2n+p}}{b\Gamma(2n+p)} \exp\left[-(2n+p)\frac{a}{b} - 2r\exp\left(-\frac{a}{b}\right)\right] \\
&= \frac{m!(2r)^{2n+p}}{(j-1)!(m-j)!b^2\Gamma(2n+p)} e^{\frac{y}{b}} e^{-(2n+p+1)\frac{a}{b}} \left(1 - e^{-e^{\frac{y-a}{b}}}\right)^{j-1} e^{-[2r+(m-j+1)e^{\frac{y}{b}}]e^{-\frac{a}{b}}} \\
&= \frac{m!(2r)^{2n+p}}{(j-1)!(m-j)!b^2\Gamma(2n+p)} e^{\frac{y}{b}} e^{-(2n+p+1)\frac{a}{b}} e^{-[2r+(m-j+1)e^{\frac{y}{b}}]e^{-\frac{a}{b}}} \sum_{i=0}^{j-1} (-1)^i \binom{j-1}{i} e^{-ie^{\frac{y}{b}}e^{-\frac{a}{b}}} \\
&= \frac{m!(2r)^{2n+p}}{(j-1)!(m-j)!b^2\Gamma(2n+p)} e^{\frac{y}{b}} e^{-(2n+p+1)\frac{a}{b}} \sum_{i=0}^{j-1} (-1)^i \binom{j-1}{i} e^{-[2r+(m-j+i+1)e^{\frac{y}{b}}]e^{-\frac{a}{b}}}
\end{aligned}$$

Then the predictive density function for $Y_{(j)}$ is

$$
\begin{aligned}
f_{Y_{(j)}}(y|x_1,\cdots,x_n) &= \int_{-\infty}^{\infty} f_{Y_{(j)}}(y \mid a)\pi(a \mid x_1,\cdots,x_n)da \\
&= \int_{-\infty}^{\infty} \frac{m!(2r)^{2n+p}}{(j-1)!(m-j)!b^2\Gamma(2n+p)} e^{\frac{y}{b}} e^{-(2n+p+1)\frac{a}{b}} \sum_{i=0}^{j-1} (-1)^i \tbinom{j-1}{i} e^{-[2r+(m-j+i+1)e^{\frac{y}{b}}]e^{-\frac{a}{b}}} da \\
&= \frac{m!(2r)^{2n+p}}{(j-1)!(m-j)!b\Gamma(2n+p)} e^{\frac{y}{b}} \sum_{i=0}^{j-1} \left[(-1)^i \tbinom{j-1}{i} \int_{-\infty}^{\infty} e^{-(2n+p+1)\frac{a}{b}} e^{-[2r+(m-j+i+1)e^{\frac{y}{b}}]e^{-\frac{a}{b}}} d\left(\frac{a}{b}\right) \right] \\
&= \frac{m!(2r)^{2n+p}}{(j-1)!(m-j)!b\Gamma(2n+p)} e^{\frac{y}{b}} \sum_{i=0}^{j-1} \left[(-1)^i \tbinom{j-1}{i} \Gamma(2n+p+1) \frac{1}{\left[2r+(m-j+i+1)e^{\frac{y}{b}}\right]^{2n+p+1}} \right] \\
&= \frac{m!(2n+p)(2r)^{2n+p}}{(j-1)!(m-j)!b} e^{\frac{y}{b}} \sum_{i=0}^{j-1} \left[(-1)^i \tbinom{j-1}{i} \frac{1}{\left[2r+(m-j+i+1)e^{\frac{y}{b}}\right]^{2n+p+1}} \right] \qquad (6.8)
\end{aligned}
$$

Suppose the one-sided $100(1-\alpha)\%$ upper prediction interval is $I_B = (y_L, \infty)$. We than have

$$
\begin{aligned}
\alpha &= \int_{-\infty}^{y_L} f_{Y_{(j)}}(y|x_1,\cdots,x_n)dy \\
&= \int_{-\infty}^{y_L} \frac{m!(2n+p)(2r)^{2n+p}}{(j-1)!(m-j)!b} e^{\frac{y}{b}} \sum_{i=0}^{j-1} \left[(-1)^i \tbinom{j-1}{i} \frac{1}{\left[2r+(m-j+i+1)e^{\frac{y}{b}}\right]^{2n+p+1}} \right] dy \\
&= \frac{m!(2n+p)(2r)^{2n+p}}{(j-1)!(m-j)!} \sum_{i=0}^{j-1} (-1)^i \tbinom{j-1}{i} \int_{-\infty}^{y_L} \frac{e^{\frac{y}{b}}}{[2r+(m-j+1+i)e^{\frac{y}{b}}]^{2n+p+1}} d\left(\frac{y}{b}\right) \\
&= \frac{m!(2n+p)(2r)^{2n+p}}{(j-1)!(m-j)!} \sum_{i=0}^{j-1} (-1)^i \tbinom{j-1}{i} \int_{-\infty}^{y_L} \frac{1}{[2r+(m-j+1+i)e^{\frac{y}{b}}]^{2n+p+1}} d\left(e^{\frac{y}{b}}\right) \\
&= \frac{m!(2n+p)(2r)^{2n+p}}{(j-1)!(m-j)!} \\
&\times \sum_{i=0}^{j-1} (-1)^i \tbinom{j-1}{i} \frac{1}{m-j+1+i} \int_{-\infty}^{y_L} \frac{1}{[2r+(m-j+1+i)e^{\frac{y}{b}}]^{n+p+1}} d[2r+(m-j+1+i)e^{\frac{y}{b}}] \\
&= \frac{m!}{(j-1)!(m-j)!} \sum_{i=0}^{j-1} (-1)^i \tbinom{j-1}{i} \frac{1}{m-j+1+i} \left[1 - \frac{(2r)^{2n+p}}{[2r+(m-j+1+i)e^{\frac{y_L}{b}}]^{2n+p}} \right] \\
&= \frac{m!}{(j-1)!(m-j)!} \sum_{i=0}^{j-1} (-1)^i \tbinom{j-1}{i} \frac{1}{m-j+1+i} \left[1 - \left(\frac{2\sum_{k=1}^{n} e^{\frac{x_k}{b}}}{2\sum_{k=1}^{n} e^{\frac{x_k}{b}} + (m-j+1+i)e^{\frac{y_L}{b}}} \right)^{2n+p} \right] \\
&\qquad (6.9)
\end{aligned}
$$

6.3 Constructing Bayesian Prediction Intervals Using Gibbs Sampling

In this section, we construct Bayesian prediction intervals for the Gumbel minimum distribution by Gibbs sampling, a special case of a single–component Metropolis–Hastings algorithm, to estimate the percentiles for distribution of future observations.

The Metropolis-Hastings (M-H) algorithm aims at obtaining simulations from the target density $p(x)$ such that $p(x) = \frac{1}{K} f(x)$, where the normalizing constant K may not be known, and is very difficult to compute. The M-H algorithm generates a sequence of draws from this distribution as follows:

1. Start with any initial value $x^{(0)}$ satisfying $f(x^{(0)}) > 0$.

2. Using the current x value, sample a candidate point x^* from an arbitrary proposed conditional distribution $q(y|x)$.

3. Calculate the acceptance probability for the candidate point x^* as:

$$r = \min\left\{1, \frac{f(x^*)q[x^*|x^{(k-1)}]}{f[x^{(k-1)}]q[x^{(k-1)}|x^*]}\right\}$$

4. Set

$$X^{(k+1)} = \begin{cases} x^* & \text{with probability } r \\ x^{(k-1)} & \text{with probability } 1-r \end{cases}$$

An easy way to implement step (4) is to generate $U \sim U(0,1)$ and to set $X^{(k)} = x^*$ if $U < r$, and to set $X^{(k)} = x^{(k-1)}$ otherwise.

The convergence of the resulting chain is guaranteed for every proposed $q(y|x)$ whose support includes the support of $p(x)$, that is

$$supp\,\{p(x)\} \subseteq \cup_x supp\,\{q(\cdot|x)\}$$

Following a sufficient burn-in period (of, say, k steps), the chain approaches its stationary distribution, and samples from the vector $x^{(t+1)}, \cdots, x^{(t+s)}$ approximate samples from $p(x)$. A key issue in the successful implementation of Metropolis-Hastings or any other MCMC sampler is the number of runs (steps) until the chain approaches stationarity (the length of the burn-in period). Typically the first 1000 to 5000 elements are thrown out. There are various convergence tests used to assess whether stationarity has indeed been reached.

A poor choice of initial values can greatly increase the required burn-in time. One suggestion is to start the chain as close to the centre of the distribution as possible, for example taking a value close to the distribution's mode (such as using an approximate MLE as the starting value).

To demonstrate that the Metropolis-Hasting (M-H) sampling generates a Markov Chain whose equilibrium density is the candidate density $p(x)$, it is sufficient to show that the M-H transition kernel satisfy the detailed balance equation with $p(x)$.

Recall that a Markov chain may reach a stationary distribution π^* with stationary distribution

$$\pi^* = \pi^* \mathbf{P}$$

where $\mathbf{P}$ denotes the probability transition matrix. In other words, π^* is the left eigenvector associated with the eigenvalue $\lambda = 1$ of $\mathbf{P}$.

A sufficient condition for a unique stationary distribution is that the detailed balance equation holds (for all i and j)

$$P(j \longrightarrow i)\pi_j^* = P(i \longrightarrow j)\pi_i^*$$

where we use the notation $P(i \longrightarrow j)$ to imply a move from state i to j.

The Gibbs sampler is a special case of a single-component Metropolis-Hastings

algorithm, which updates the sample vector one component at a time instead of updating the whole vector. Gibbs sampling uses

$$q\left(x^{*}|x\right)=p\left(x_{i}^{*}|x_{-i}\right)$$

and so the acceptance rate is always 1, that is

$$\frac{f(x^{*})q[x^{*}|x^{(k-1)}]}{f[x^{(k-1)}]q[x^{(k-1)}|x^{*}]}=1$$

where x^{*} is identical to x except for its value along the i-dimension x_{i}^{*} is sampled from the conditional $p(x_{i}|x_{-i})$ where $x_{-i}=(x_{1},\cdots,x_{i-1},x_{i+1}\cdots,x_{n})$.

To generate $x^{(k+1)}$, we sample each component in turn:

$$\begin{aligned}
x_{1}^{(k+1)} &\sim p\left(x_{1}|x_{2}^{(k)},\cdots,x_{n}^{(k)}\right)\\
x_{2}^{(k+1)} &\sim p\left(x_{2}|x_{1}^{(k+1)},x_{3}^{(k)},\cdots,x_{n}^{(k)}\right)\\
&\vdots\\
x_{j}^{(k+1)} &\sim p\left(x_{j}|x_{1}^{(k+1)},\cdots,x_{j-1}^{(k+1)},x_{j+1}^{(k)},\cdots,x_{n}^{(k)}\right)\\
&\vdots\\
x_{n}^{(k+1)} &\sim p\left(x_{n}|x_{1}^{(k+1)},\cdots,x_{n-1}^{(k+1)}\right)
\end{aligned}$$

where $p(x_{i}|\cdot)$ are conditionals of the target distribution $p(x)$.

Suppose that $\{X_{1},X_{2},\cdots,X_{n}\}$ is a random sample from the Gumbel minimum distribution with $\underline{\theta}=(a,b)$. $\underline{Y}=(Y_{1},Y_{2},\cdots,Y_{m})$ are future observations with same distribution. Then the Bayesian predictive density function is defined as

$$h_{B}(\underline{y}|\underline{x})=\int L\left(\underline{y}|\underline{\theta}\right)\pi\left(\underline{\theta}|\underline{x}\right)d\underline{\theta}$$

where $L\left(\underline{y}|\underline{\theta}\right)$ denotes the likelihood and $\pi\left(\underline{\theta}|\underline{x}\right)$ denotes the posterior distribution of $\underline{\theta}$ given $\underline{x}$.

In general, it is impossible to obtain the analytic form of the posterior distri-

bution of a general statistic $T(\underline{Y})$. Next, we use

$$h_B(y_j|\underline{x}, \underline{y}_{-j})$$

to generate Gibbs samplers and then estimate the percentiles of $T(\underline{Y})$. Using this result, we can easily construct Bayesian prediction intervals for $T(\underline{Y})$.

Since the joint density of $(\underline{X}, \underline{Y})$ is

$$h_B\left(\underline{x}, \underline{y}\right) = \int L\left(\underline{x}|\underline{\theta}\right) L\left(\underline{y}|\underline{\theta}\right) \pi\left(\underline{\theta}\right) d\underline{\theta}$$

and

$$h_B\left(\underline{y}|\underline{x}\right) = \frac{h_B\left(\underline{x}, \underline{y}\right)}{\int h_B\left(\underline{x}, \underline{y}\right) d\underline{y}}$$

then the full conditional predictive density of Y_j is

$$h_B(y_j|\underline{x}, \underline{y}_{-j}) = \frac{h_B(\underline{y}|\underline{x})}{h_B(\underline{y}_{-j}|\underline{x})} = \frac{h_B(\underline{x}, \underline{y})}{h_B(\underline{y}_{-j}|\underline{x}) \int h_B(\underline{x}, \underline{y}) d\underline{y}} \propto h_B(\underline{x}, \underline{y}) \tag{6.10}$$

Here, proportionality follows that the denominator does not depend on Y_j. Kim (2007) obtains a similar result for the lognormal distribution.

6.3.1 Case 1: a Unknown and b Known

In the following, we consider the Gumbel minimum distribution. Suppose that the prior and posterior for a are given by

$$\begin{aligned} \pi(a) &= 1 \\ \pi(a \mid x_1, x_2, \cdots, x_n) &= \frac{1}{b\Gamma(n)} \left(\sum_{i=1}^{n} e^{\frac{x_i}{b}} \right)^n e^{-\frac{na}{b}} e^{-e^{-\frac{a}{b}} \left(\sum_{i=1}^{n} e^{\frac{x_i}{b}} \right)} \end{aligned}$$

Therefore

$$
\begin{aligned}
&f_{\underline{Y}}(y_1, \cdots, y_m \mid a)\pi(a \mid x_1, \cdots, x_n) \\
&= \left\{ \left(\frac{1}{b}\right)^m e^{\sum_{i=1}^m \frac{y_i}{b}} e^{-\frac{ma}{b}} e^{-e^{-\frac{a}{b}}\left(\sum_{i=1}^m e^{\frac{y_i}{b}}\right)} \right\} \times \left\{ \frac{1}{b\Gamma(n)} \left(\sum_{i=1}^n e^{\frac{x_i}{b}}\right)^n e^{-\frac{na}{b}} e^{-e^{-\frac{a}{b}}\left(\sum_{i=1}^n e^{\frac{x_i}{b}}\right)} \right\} \\
&= \frac{1}{b^{m+1}\Gamma(n)} \left(\sum_{i=1}^n e^{\frac{x_i}{b}}\right)^n e^{\sum_{i=1}^m \frac{y_i}{b}} e^{-\frac{m+n}{b}a} e^{-e^{-\frac{a}{b}}\left(\sum_{i=1}^n e^{\frac{x_i}{b}} + \sum_{i=1}^m e^{\frac{y_i}{b}}\right)}
\end{aligned}
$$

Now let

$$
\begin{aligned}
A &= \sum_{i=1}^n e^{\frac{x_i}{b}} + \sum_{i=1}^m e^{\frac{y_i}{b}} \\
t &= e^{-\frac{a}{b}}
\end{aligned}
$$

Now using Proposition A.4, the Bayesian predictive density function for $\underline{Y}$ is

$$
\begin{aligned}
h_B(\underline{y}|\underline{x}) &= \int_{-\infty}^{\infty} f_{\underline{Y}}(\underline{y} \mid a)\pi(a \mid x_1, \cdots, x_n)da \\
&= \frac{1}{b^{m-1}\Gamma(n)} \left(\sum_{i=1}^n e^{\frac{x_i}{b}}\right)^n e^{\sum_{i=1}^m \frac{y_i}{b}} \int_0^{\infty} t^{m+n-1} e^{-At} dt \\
&= \frac{\Gamma(m+n)}{b^{m-1}\Gamma(n)} \frac{e^{\sum_{i=1}^m \frac{y_i}{b}}}{\left(\sum_{i=1}^n e^{\frac{x_i}{b}} + \sum_{i=1}^m e^{\frac{y_i}{b}}\right)^{m+n}} \left(\sum_{i=1}^n e^{\frac{x_i}{b}}\right)^n \qquad (6.11)
\end{aligned}
$$

However, it is not easy to calculate percentiles of $\underline{Y}$ when m is large. Below, we use Gibbs sampling algorithm to generate $\underline{Y}$, and then estimate the percentiles of $T(\underline{Y})$. The joint density of $\underline{X}$ and $\underline{Y}$ is

$$
\begin{aligned}
h_B(\underline{x}, \underline{y}) &= \int_{-\infty}^{\infty} f_{\underline{X}}(\underline{x} \mid a) f_{\underline{Y}}(\underline{y} \mid a)\pi(a)da \\
&= \int_{-\infty}^{\infty} \left(\frac{1}{b}\right)^n e^{\sum_{i=1}^n \frac{x_i}{b}} e^{-\frac{na}{b}} e^{-e^{-\frac{a}{b}}\left(\sum_{i=1}^n e^{\frac{x_i}{b}}\right)} \left(\frac{1}{b}\right)^m e^{\sum_{i=1}^m \frac{y_i}{b}} e^{-\frac{ma}{b}} e^{-e^{-\frac{a}{b}}\left(\sum_{i=1}^m e^{\frac{y_i}{b}}\right)} da \\
&= \frac{1}{b^{m+n}} e^{\sum_{i=1}^n \frac{x_i}{b} + \sum_{i=1}^m \frac{y_i}{b}} \int_{-\infty}^{\infty} e^{-\frac{m+n}{b}a} e^{-e^{-\frac{a}{b}}\left(\sum_{i=1}^n e^{\frac{x_i}{b}} + \sum_{i=1}^m e^{\frac{y_i}{b}}\right)} da \\
&= \frac{\Gamma(m+n)}{b^{m+n-1}} \frac{e^{\sum_{i=1}^n \frac{x_i}{b} + \sum_{i=1}^m \frac{y_i}{b}}}{\left(\sum_{i=1}^n e^{\frac{x_i}{b}} + \sum_{i=1}^m e^{\frac{y_i}{b}}\right)^{m+n}} \qquad (6.12)
\end{aligned}
$$

If we use the approximation from Theorem 4.3,

$$\frac{e^{\sum_{i=1}^{m+n}\frac{v_i}{b}}}{\left(\sum_{i=1}^{m+n} e^{\frac{v_i}{b}}\right)^{m+n}} \approx \frac{e^{-K_n\frac{S^2_{V,m+n}}{b^2}}}{\Delta^{k_\Delta}}$$

and notice that

$$\begin{aligned}
S^2_{n+m}(v) &= \frac{1}{n+m-1}\sum_{i=1}^{n+m} v_i^2 - \frac{n+m}{n+m-1}\left(\frac{1}{n+m}\sum_{i=1}^{n+m} v_i\right)^2 \\
&\propto (n+m)\sum_{i=1}^{n+m} v_i^2 - \left(\sum_{i=1}^{n+m} v_i\right)^2 \\
&= (n+m)\left(y_j^2 + \sum_{i=1}^{n} x_i^2 + \sum_{i=1,i\neq j}^{m} y_i^2\right) - \left(\sum_{i=1}^{n} x_i + \sum_{i=1,i\neq j}^{m} y_i + y_j\right)^2 \\
&= (n+m-1)y_j^2 + (n+m)\left(\sum_{i=1}^{n} x_i^2 + \sum_{i=1,i\neq j}^{m} y_i^2\right) - \left(\sum_{i=1}^{n} x_i + \sum_{i=1,i\neq j}^{m} y_i\right)^2 \\
&- 2\left(\sum_{i=1}^{n} x_i + \sum_{i=1,i\neq j}^{m} y_i\right) y_j
\end{aligned}$$

therefore

$$h_B(\underline{x},\underline{y}) \propto e^{-K_n\frac{S^2_{V,m+n}}{b^2}}$$

We only pick out the terms in $h_B(\underline{x},\underline{y})$ which involve y_j. Then

$$h_B(y_j|\underline{x},\underline{y}_{-j}) \propto \exp\left\{-\frac{K_n}{b^2}\left[(n+m-1)y_j^2 - 2\left(\sum_{i=1}^{n} x_i + \sum_{i=1,i\neq j}^{m} y_i\right) y_j\right]\right\} \; j = 1,\cdots,m \tag{6.13}$$

Setting

$$\begin{aligned}
A &= m+n-1 \\
B_j &= \sum_{i=1}^{n} x_i + \sum_{i=1}^{j-1} y_i^{(k+1)} + \sum_{i=j+1}^{m} y_i^{(k)}
\end{aligned}$$

Then

$$
\begin{aligned}
-\frac{K_n}{b^2}\left(Ay_j^2 - 2B_j y_j\right) &= -\frac{K_n A}{b^2}\left[\left(y_j - \frac{B_j}{A}\right)^2 - \left(\frac{B_j}{A}\right)^2\right] \\
&= -\frac{\left(y_j - \frac{B_j}{A}\right)^2}{2\left(\frac{b}{\sqrt{2K_n A}}\right)^2} + \frac{K_n B_j^2}{b^2 A}
\end{aligned}
$$

Therefore we have

$$
h_B(y_j|\underline{x}, \underline{y}_{-j}) \propto \exp\left\{-\frac{\left(y_j - \frac{B_j}{A}\right)^2}{2\left(\frac{b}{\sqrt{2K_n A}}\right)^2}\right\} \tag{6.14}
$$

which is a normal distribution. We will generate D such Gibbs samplers, and so obtain:

$$
Y_{(j)}^1, Y_{(j)}^2, \cdots, Y_{(j)}^D
$$

We then can estimate the percentile of $Y_{(j)}$ and then calculate the prediction interval:

$$
1 - \alpha = P\left(\widehat{Q}_{\frac{\alpha}{2}} < Y_{(j)} \leq \widehat{Q}_{1-\frac{\alpha}{2}}\right)
$$

Hence, a $100(1-\alpha)\%$ two-sided prediction interval for $Y_{(j)}$ is

$$
[L(\underline{X}), U(\underline{X})] = \left[\widehat{Q}_{\frac{\alpha}{2}}, \widehat{Q}_{1-\frac{\alpha}{2}}\right]
$$

and the coverage probability is

$$
Pr = \frac{1}{D}\sum_{k=1}^{D} c_k
$$

where

$$
c_k = \begin{cases} 1, & y_{(j)}^{(k)} \in [L(\underline{X}), U(\underline{X})] \\ 0, & else \end{cases}
$$

6.3.2 Case 2: Both a and b Unknown

Let $v = (x_1, \cdots, x_n, y_1, \cdots, y_m)$ and $r = \sum_{i=1}^{m+n} e^{\frac{v_i}{b}}$. In this section, we first consider a prior with $q(\underline{v}) > 0$:

$$\pi(a,b) = \frac{r^{m+n}[q(\underline{v})]^{2(m+n)-2}}{\Gamma(m+n)\Gamma[2(m+n)-2]} \frac{1}{b^{(m+n)+2}} e^{-\frac{q(\underline{v})}{b}} \exp\left[-(m+n)\frac{a}{b} - r\exp\left(-\frac{a}{b}\right)\right] \tag{6.15}$$

Let $S_{m+n}(\underline{v})$ denote the sample standard deviation. We take $q(\underline{v}) > 0$ such that

$$p(\underline{v}) \propto S_{m+n}(\underline{v})$$

Then the joint density of $\underline{X}$ and $\underline{Y}$ is given by

$$\begin{aligned}
& h_B(\underline{x}, \underline{y}) = \int_0^\infty \int_{-\infty}^\infty f_{\underline{X}}(\underline{x} \mid a,b) f_{\underline{Y}}(\underline{y} \mid a,b)\pi(a,b)dadb \\
= & \frac{r^{m+n}[q(\underline{v})]^{2(m+n)-2}}{\Gamma(m+n)\Gamma[2(m+n)-2]} \int_0^\infty \int_{-\infty}^\infty r^{m+n} \left(\frac{1}{b}\right)^{2(n+m)+2} e^{\sum_{i=1}^{n+m} \frac{v_i}{b}} e^{-\frac{q(\underline{v})}{b}} e^{-\frac{2(n+m)a}{b}} e^{-2re^{-\frac{a}{b}}} dadb \\
= & \frac{[q(\underline{v})]^{2(m+n)-2}}{\Gamma(m+n)\Gamma[2(m+n)-2]} \\
\times & \int_0^\infty r^{m+n} \left(\frac{1}{b}\right)^{2(m+nm)+2} e^{\sum_{i=1}^{(m+n)} \frac{v_i}{b}} e^{-\frac{q(\underline{v})}{b}} \left[\int_{-\infty}^\infty e^{-\frac{2(n+m)}{b}a} e^{-2re^{-\frac{a}{b}}} da\right] db
\end{aligned}$$

since

$$\int_{-\infty}^\infty e^{-\frac{2(n+m)}{b}a} e^{-2re^{-\frac{a}{b}}} da = \frac{\Gamma[2(n+m)]}{(2r)^{2(n+m)}} b$$

Then

$$\begin{aligned}
h_B(\underline{x}, \underline{y}) &= \frac{\Gamma[2(m+n)][q(\underline{v})]^{2(m+n)-2}}{2^{2(m+n)}\Gamma(m+n)\Gamma[2(m+n)-2]} \int_0^\infty \left(\frac{1}{b}\right)^{2(n+m)+1} \frac{e^{\sum_{i=1}^{(m+n)} \frac{v_i}{b}}}{r^{m+n}} e^{-\frac{q(\underline{v})}{b}} db \\
&= \frac{\Gamma[2(m+n)][q(\underline{v})]^{2(m+n)-2}}{2^{2(m+n)}\Gamma(m+n)\Gamma[2(m+n)-2]\Delta^{k_{\Delta_2}}} \int_0^\infty \left(\frac{1}{b}\right)^{2(n+m)+1} e^{-2\frac{q(\underline{v})}{b}} db \\
&= \frac{\Gamma[2(m+n)][q(\underline{v})]^{2(m+n)-2}}{2^{2(m+n)}\Gamma(m+n)\Gamma[2(m+n)-2]\Delta^{k_{\Delta_2}}} \int_0^\infty t^{2(n+m)-1} e^{-2q(\underline{v})t} dt \\
&= \frac{\Gamma^2[2(m+n)]}{4^{2(m+n)}\Gamma(m+n)\Gamma[2(m+n)-2]\Delta^{k_{\Delta_2}}[q(\underline{v})]^2}
\end{aligned} \tag{6.16}$$

We only pick out the terms in $h_B(\underline{x}, \underline{y})$ which involve y_j and from equation (6.13), then

$$h_B(y_j|\underline{x}, \underline{y}_{-j}) \quad \propto \quad \frac{1}{[q(\underline{v})]^2} \propto \frac{1}{S^2_{n+m}(v)} \quad j = 1, \cdots, m$$

We can use this to obtain the Gibbs sampler given $\underline{y}^{(k)} = \left(y_1^{(k)}, \cdots, y_m^{(k)}\right)$:

$$\begin{aligned}
Y_1^{(k+1)} &\propto \frac{1}{A\left[y_1^{(k+1)}\right]^2 + (n+m)\left(\sum_{i=1}^{n} x_i^2 + \sum_{i=2}^{m}\left[y_i^{(k)}\right]^2\right) - \left(\sum_{i=1}^{n} x_i + \sum_{i=2}^{m} y_i^{(k)}\right)^2 - 2\left(\sum_{i=1}^{n} x_i + \sum_{i=2}^{m} y_i^{(k)}\right) y_1^{(k+1)}} \\
&\vdots \\
Y_j^{(k+1)} &\propto \frac{1}{A\left[y_j^{(k+1)}\right]^2 - 2B_j y_j^{(k+1)} + C_j} \\
&\vdots \\
Y_m^{(k+1)} &\propto \frac{1}{A\left[y_m^{(k+1)}\right]^2 + (n+m)\left(\sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{m-1}\left[y_i^{(k+1)}\right]^2\right) - \left(\sum_{i=1}^{n} x_i + \sum_{i=1}^{m-1} y_i^{(k+1)}\right)^2 - 2\left(\sum_{i=1}^{n} x_i + \sum_{i=1}^{m-1} y_i^{(k+1)}\right) y_m^{(k+1)}}
\end{aligned}$$

where

$$\begin{aligned}
A &= n + m - 1 \\
B_j &= \sum_{i=1}^{n} x_i + \sum_{i=1}^{j-1} y_i^{(k+1)} + \sum_{i=j+1}^{m} y_i^{(k)} \\
C_j &= (n+m)\left(\sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{j-1}\left[y_i^{(k+1)}\right]^2 + \sum_{i=j+1}^{m}\left[y_i^{(k)}\right]^2\right) - \left(\sum_{i=1}^{n} x_i + \sum_{i=1}^{j-1} y_i^{(k+1)} + \sum_{i=j+1}^{m} y_i^{(k)}\right)^2
\end{aligned}$$

Set

$$\underline{V} = \left(x_1, \cdots, x_n, y_1^{(k+1)}, \cdots, y_{j-1}^{(k+1)}, y_{j+1}^{(k)}, \cdots, y_{j+1}^{(k)}\right)$$

and let V_i be ith element of $\underline{V}$. Recall the Cauchy-Schwarz inequality

$$\left(\sum_{i=1}^{n} x_i y_i\right)^2 \leq \left(\sum_{i=1}^{n} x_i\right)^2 \left(\sum_{i=1}^{n} y_i\right)^2, x_i, y_i \in \mathbb{R}$$

Equality holds if and only if X and Y are linearly dependent, that is, one is a scalar multiple of the other (which includes the case when one or both are zero). Taking

$y_i = 1$, we have

$$\left(\sum_{i=1}^{n} x_i\right)^2 \leq n\left(\sum_{i=1}^{n} x_i^2\right), \quad x_i,\ y_i \in \mathbb{R}$$

In our case, X_i and Y_j are independent. It is obvious that

$$\begin{aligned}
AC_j - B_j^2 &= (m+n)(m+n-1)\sum_{i=1}^{m+n-1} V_i^2 - (m+n-1)\left(\sum_{i=1}^{m+n-1} V_i\right)^2 \\
&= (n+m-1)\left[(n+m)\sum_{i=1}^{m+n-1} V_i^2 - \left(\sum_{i=1}^{m+n-1} V_i\right)^2\right] \\
&= (n+m-1)\left[\sum_{i=1}^{m+n-1} V_i^2 + (n+m-1)\sum_{i=1}^{m+n-1} V_i^2 - \left(\sum_{i=1}^{m+n-1} V_i\right)^2\right] \\
&> 0
\end{aligned}$$

Using Proposition A.10, we have the normalized constant for Y_j

$$K_j = \int_{-\infty}^{\infty} \frac{1}{Ay^2 - 2B_j y + C_j} dy = \frac{\pi}{\sqrt{AC_j - B_j^2}}$$

We rewrite,

$$Ay^2 - 2B_j y + C_j = A\left[\left(y - \frac{B_j}{A}\right)^2 + \frac{AC_j - B_j^2}{A^2}\right]$$

For $A > 0$ and $AC_j - B_j^2 > 0$, set

$$\begin{aligned}
\theta &= \frac{B_j}{A} \\
\sigma &= \frac{\sqrt{AC_j - B_j^2}}{A}
\end{aligned}$$

Then the predictive density for Y_j is given by

$$\begin{aligned} f_{(Y_j|\underline{X},\underline{Y}_{-j})}(y_j) &= \frac{1}{K_j}\frac{1}{A\left[\left(y_j-\frac{B_j}{A}\right)^2+\frac{AC_j-B_j^2}{A^2}\right]} \\ &= \frac{1}{\pi\frac{\sqrt{AC_j-B_j^2}}{A}}\frac{1}{1+\left(\frac{y_j-\frac{B_j}{A}}{\frac{\sqrt{AC_j-B_j^2}}{A}}\right)^2} \\ &= \frac{1}{\pi\sigma}\frac{1}{1+\left(\frac{y_j-\theta}{\sigma}\right)^2} \end{aligned}$$

which shows that $Y_j|\underline{X},\underline{Y}_{-j}$ has a Cauchy distribution with location $\frac{B_j}{A}$ and scale $\frac{\sqrt{AC_j-B_j^2}}{A}$.

The changes of variables $x = y - \frac{B_j}{A}$ and $u = \frac{\sqrt{AC_j-B_j^2}}{A}$ give

$$\begin{aligned} \int_{-\infty}^{y_j}\frac{1}{Ay^2-2B_jy+C_j}dy &= \frac{1}{A}\int_{-\infty}^{y_j}\frac{1}{\left(y-\frac{B_j}{A}\right)^2+\frac{AC_j-B_j^2}{A^2}}dy \\ &= \frac{1}{A}\int_{-\infty}^{y_j-\frac{B_j}{A}}\frac{1}{x^2+u^2}dx \\ &= \frac{1}{\sqrt{AC_j-B_j^2}}\left[\arctan\left(\frac{Ay_j-B_j}{\sqrt{AC_j-B_j^2}}\right)+\frac{\pi}{2}\right] \end{aligned}$$

Now we have

$$\begin{aligned} F_{(Y_j|\underline{X},\underline{Y}_{-j})}(y_j) &= \frac{1}{K_j}\int_{-\infty}^{y_j}\frac{1}{Ay^2-2B_jy+C_j}dy \\ &= \frac{1}{\pi}\left[\arctan\left(\frac{Ay_j-B_j}{\sqrt{AC_j-B_j^2}}\right)+\frac{\pi}{2}\right] \end{aligned} \tag{6.17}$$

To generate $Y_j^{(k+1)}$, we use the inverse *cdf* method to generate Cauchy random variables in three steps:

1. Generate a group of random numbers as initial value $y_1^{(0)},\cdots,\ y_m^{(0)}$, for example drawn from the standard uniform distribution on the open interval (0,1).

2. Sample t from a uniform distribution $U(k_1, k_2)$ with $0 \leq k_1 < k_2 \leq 1$. The method to determine value of k_1 and k_2 will be given later.

3. Compute $Y_j^{(k+1)}$ by following equation:

$$y_j^{(k+1)} = \frac{\sqrt{AC_j - B_j^2}}{A} \tan[(t - 0.5)\pi] + \frac{B_j}{A} \tag{6.18}$$

In general, we sample t from $U(0, 1)$, however, it is not reasonable if one only generate one random number. When t is too close to zero or one, the absolute value of the random number will be very large in the case of Cauchy distribution. This is not acceptable according to the "single law of chance" (Borel [11]).

We only sample t one time for each Gibbs sampling procedure, since θ and σ will be updated in each iteration. This suggests that we draw t from a narrower uniform distribution $U(k_1, k_2)$.

1. Find an estimator of location a and scale b based on data $\underline{x}$.

2. Generate D samples by Monte Carlo method such as

$$x_1^{(k)}, \cdots, x_n^{(k)}, y_1^{(k)}, \cdots, y_m^{(k)}, \quad k = 1,\ 2,\ \cdots,\ D$$

Then we compute

$$t_k = \frac{1}{\pi}\left(\arctan \frac{Ay_j^{(k)} - B_j}{\sqrt{AC_j - B_j^2}} + \frac{\pi}{2}\right)$$

We choose $k_1^*,\ k_2^*$ such that frequencies of $t \leq k_1^*$ and $t \geq k_2^*$ are zero or very small.

3. Sample t from $U(k_1^*, k_2^*)$ and compute y_j, then check:

$$\frac{|y_j - \bar{x}|}{S_X} = K^*$$

If $K^* \geq 5$, in general, y_j will be treated as an outlier. In this case, we increase k_1^* to k_1 or/and decrease k_2^* to k_2 until there is no outlier and still satisfy

$$P\left(k_1 \leq t \leq k_2\right) \geq 0.95$$

An example will be given in the next chapter.

Chapter 7

Comparative Study of Parameter Estimation and Prediction Intervals for the Gumbel Distribution

The main focus of this chapter is to study and compare different parameter estimation and prediction interval algorithms for the Gumbel minimum distribution. The results of different methods applied in this case study are then analyzed and compared.

7.1 Introduction

The Gumbel minimum distribution with location parameter a and scale parameter b is given by

$$\begin{aligned} F(x) &= 1 - e^{-e^{\frac{x-a}{b}}}, \quad -\infty < x < \infty, \\ f(x) &= \frac{1}{b} e^{\frac{x-a}{b}} e^{-e^{\frac{x-a}{b}}} \end{aligned}$$

The mean and variance are

$$\mu_X = a - \gamma b \quad \sigma_X^2 = \frac{1}{6}\pi^2 b^2$$

where γ is Euler constant. The Gumbel maximum distribution with location parameter a and scale parameter b is given by

$$\begin{aligned} F(x) &= e^{-e^{-\frac{x-a}{b}}}, \quad -\infty < x < \infty, \\ f(x) &= \frac{1}{b} e^{-\frac{x-a}{b}} e^{-e^{-\frac{x-a}{b}}} \end{aligned}$$

The mean and variance are

$$\mu_X = a + \gamma b, \quad \sigma_X^2 = \frac{1}{6}\pi^2 b^2$$

Let $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ denote the order statistics of a random sample, $\{X_1, X_2, \cdots, X_n\}$, from a continuous population with *cdf* $F_X(x)$ and *pdf* $f_X(x)$. Then the *pdf* of $X_{(j)}$ is

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j}$$

Next, we give some introduction to order statistics. Owing to the equality

$$\min(X_1, \cdots, X_n) = \max(-X_1, \cdots, -X_n)$$

it suffices to consider only the Gumbel minimum distribution. Suppose that $\{X_1, X_2, \cdots, X_n\}$ is a random sample from a Gumbel minimum distribution. Then the pdf for order statistics are given by

$$\begin{aligned} f_{X_{max}}(x \mid a) &= n f_X(x) \left[F_X(x)\right]^{n-1} = n\frac{1}{b} e^{\frac{x-a}{b}} e^{-e^{\frac{x-a}{b}}} \left[1 - e^{-e^{\frac{x-a}{b}}}\right]^{n-1} \\ &= n\frac{1}{b} e^{\frac{x-a}{b}} e^{-e^{\frac{x-a}{b}}} \sum_{k=0}^{n-1} (-1)^k \binom{n-1}{k} \left[e^{-e^{\frac{x-a}{b}}}\right]^k \end{aligned}$$

$$\begin{aligned} f_{X_{(2)}}(x \mid a) &= n(n-1) f_X(x) F_X(x) \left[1 - F_X(x)\right]^{n-2} \\ &= \frac{n(n-1)}{b} e^{\frac{x-a}{b}} e^{-e^{\frac{x-a}{b}}} \left[1 - e^{-e^{\frac{x-a}{b}}}\right] \left(e^{-e^{\frac{x-a}{b}}}\right)^{n-2} \\ &= \frac{n(n-1)}{b} e^{\frac{x-a}{b}} e^{-e^{\frac{x-a}{b}}} \left[1 - e^{-e^{\frac{x-a}{b}}}\right] e^{-(n-2)e^{\frac{x-a}{b}}} \end{aligned}$$

and

$$\begin{aligned} f_{X_{(j)}}(x \mid a) &= \frac{n!}{(j-1)!(n-j)!} f_X(x) \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j} \\ &= \frac{n!}{(j-1)!(n-j)!b} e^{\frac{x-a}{b}} e^{-e^{\frac{x-a}{b}}} \left[1 - e^{-e^{\frac{x-a}{b}}}\right]^{j-1} \left(e^{-e^{\frac{x-a}{b}}}\right)^{n-j} \\ &= \frac{n!}{(j-1)!(n-j)!b} e^{\frac{x-a}{b}} \left[1 - e^{-e^{\frac{x-a}{b}}}\right]^{j-1} e^{-(n-j+1)e^{\frac{x-a}{b}}} \end{aligned}$$

7.2 Methods of Parameter Estimation for the Gumbel Distribution

In the literature various estimators can be found that discuss the moments (MOM), probability weighted moments (PWM), maximum likelihood (MLE), principle of maximum entropy (POME) and Bayesian methods to determine the parameters in the Gumbel distribution.

Method of Moments

For the Gumbel minimum distribution

$$\begin{aligned} \hat{a} &= \bar{X}_n + \frac{\sqrt{6}\gamma}{\pi} S_n \\ \hat{b} &= \frac{\sqrt{6}}{\pi} S_n \end{aligned}$$

where γ is Euler constant. For the Gumbel maximum distribution

$$\hat{a} = \bar{X}_n - \frac{\sqrt{6}\gamma}{\pi} S_n \tag{7.1}$$

$$\hat{b} = \frac{\sqrt{6}}{\pi} S_n \tag{7.2}$$

Method of Maximum Likelihood

For the Gumbel minimum distribution, the maximum likelihood estimator for b is a solution to the following equation

$$b = \frac{\sum_{i=1}^{n} x_i \exp\left(\frac{x_i}{b}\right)}{\sum_{i=1}^{n} \exp\left(\frac{x_i}{b}\right)} - \bar{x} \tag{7.3}$$

and then we can obtain MLE for a from

$$\hat{a} = b \ln\left[\frac{\sum_{i=1}^{n} \exp\left(\frac{x_i}{b}\right)}{n}\right] \tag{7.4}$$

Corsini, Gini, Greco and Verrazzani [20] provide Cramer-Rao (CR) bounds for the location and scale parameters of the Gumbel minimum distribution

$$\begin{aligned} Var(\hat{a}) &\geq \frac{6}{n\pi^2}\left(1+\frac{\pi^2}{6}+\gamma^2-2\gamma\right)b^2 \approx \frac{1.1}{n}b^2 \\ Var(\hat{b}) &\geq \frac{6}{n\pi^2}b^2 \approx \frac{0.61}{n}b^2 \end{aligned}$$

For the Gumbel maximum distribution, the maximum likehood estimator for b is the solution of the following equation

$$b = \bar{x} - \frac{\sum_{i=1}^{n} x_i \exp\left(-\frac{x_i}{b}\right)}{\sum_{i=1}^{n} \exp\left(-\frac{x_i}{b}\right)} \tag{7.5}$$

and then we can obtain the MLE for a from

$$\hat{a} = b \ln\left[\frac{n}{\sum_{i=1}^{n} \exp\left(-\frac{x_i}{b}\right)}\right] \tag{7.6}$$

Fiorentino and Gabriele [38] constructed bias–corrected maximum–likelihood estimators:

$$\hat{a}^* = \hat{b}^* \ln\left[\frac{n}{\sum_{i=1}^{n} \exp\left(-\frac{x_i}{\hat{b}^*}\right)}\right] - \frac{0.7\hat{b}^*}{n} \tag{7.7}$$

$$\hat{b}^* = \frac{n}{n-0.8}\hat{b} \tag{7.8}$$

Hosking [56] proposed another correction for maximum–likelihood estimators:

$$\hat{a}' = \hat{a} - \frac{0.3698}{n}\hat{b} \tag{7.9}$$

$$\hat{b}' = \frac{n+0.7716}{n}\hat{b} \tag{7.10}$$

where $\hat{a}$ and $\hat{b}$ are maximum likelihood estimators for a and b.

Method of Probability Weighted Moment (PWMs)

The probability weighted moments (Landwehr et al., [71]) of a distribution are defined as

$$M_{i,j,k} = E\left[X^i F^j (1-F)^k\right] = \int_0^1 [x(F)]^i F^j (1-F)^k dF \tag{7.11}$$

where $F = P(X \leq x)$ is the cumulative distribution function, $x(F)$ is the inverse cumulative distribution function, and i, j and k are non-negative integers. For estimation of a distribution with p parameters, p PWMs are needed and the typical choice is to consider:

$$\alpha_k = M_{1,0,k} = E\left[X(1-F)^k\right] = \int_0^1 [x(F)]\,(1-F)^k dF, \quad k = 0, 1, 2, \cdots \tag{7.12}$$

$$\beta_j = M_{1,j,0} = E\left[X^i F^j\right] = \int_0^1 x(F) F^j dF \quad j = 0, 1, 2, \cdots \tag{7.13}$$

The CDF of the Gumbel minimum distribution and its inverse are given by:

$$\begin{aligned} F(x) &= 1 - e^{-e^{\frac{x-a}{b}}}, \quad -\infty < x < \infty, \\ x(F) &= a + b \ln\left[-\ln(1-F)\right] \end{aligned}$$

Considering the moment α_k, set $t = -\ln(1-F)$ and using Proposition A.12, we have

$$\begin{aligned} \alpha_k &= \int_0^1 x(F)[1-F]^k dF \\ &= \int_0^1 \{a + b\ln\left[-\ln(1-F)\right]\}\,[1-F]^k dF \\ &= a\int_0^1 [1-F]^k dF + b\int_0^1 \ln\left[-\ln(1-F)\right][1-F]^k dF \\ &= a\int_0^1 [1-F]^k dF + b\int_0^\infty e^{-(k+1)t}\ln(t) dt \\ &= a\int_0^1 [1-F]^k dF - \frac{b}{k+1}\left[\gamma + \ln(k+1)\right] \end{aligned}$$

where γ is Euler constant. If choosing $k = 0$ and $k = 1$, then

$$\begin{aligned}\alpha_0 &= a - b\gamma \\ \alpha_1 &= \frac{a}{2} - \frac{b}{2}(\gamma + \ln 2)\end{aligned}$$

the estimators of parameters a and b of the Gumbel minimum distribution from the two equations above, that is

$$\hat{a} = \frac{1}{\ln 2}\left[(\gamma + \ln 2)\hat{\alpha}_0 - 2\gamma\hat{\alpha}_1\right] \tag{7.14}$$

$$\hat{b} = \frac{1}{\ln 2}(\hat{\alpha}_0 - 2\hat{\alpha}_1) \tag{7.15}$$

The CDF of Gumbel maximum distribution and its inverse are given by:

$$\begin{aligned}F(x) &= e^{-e^{-\frac{x-a}{b}}}, \quad -\infty < x < \infty, \\ x(F) &= a - b\log\left[-\log(F)\right]\end{aligned}$$

Considering the moment β_j, we have

$$\beta_j = \frac{1}{1+j}\left\{a + b\left[\gamma + \log(1+j)\right]\right\}, \quad j > -1$$

For different values of j, say r and s, we have

$$\begin{aligned}\beta_r &= \frac{1}{1+r}\left\{a + b\left[\gamma + \log(1+r)\right]\right\} \\ \beta_s &= \frac{1}{1+s}\left\{a + b\left[\gamma + \log(1+s)\right]\right\}\end{aligned}$$

It is then easy to solve

$$\begin{aligned}a &= (r+1)\beta_r - b\left[\gamma + \log(1+r)\right] \\ b &= \frac{(s+1)\beta_s - (r+1)\beta_r}{\log(1+s) + \log(1+r)}\end{aligned}$$

If choosing $r = 0$ and $s = 1$, we can obtain the classical PWM estimators for the Gumbel maximum distribution:

$$\hat{a} = \frac{1}{\ln 2}\left[(\gamma + \ln 2)\hat{\beta}_0 - 2\gamma\hat{\beta}_1\right] \tag{7.16}$$

$$\hat{b} = \frac{2\hat{\beta}_1 - \hat{\beta}_0}{\log 2} \tag{7.17}$$

Landwehr et al.(1979) gave unbiased estimators of α_k and β_j:

$$\begin{aligned}\hat{\beta}_j &= \frac{1}{n}\sum_{i=1}^{n}\frac{(i-1)(i-2)\cdots(i-j)}{(n-1)(n-2)\cdots(n-j)}x_{(i)} \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{i-1}{n-1}\right)^j x_{(i)}\end{aligned}$$

In general, α_j and β_j are related through

$$\begin{aligned}\alpha_j &= \sum_{k=0}^{j}(-1)^k\binom{j}{k}\beta_k \\ \beta_j &= \sum_{k=0}^{j}(-1)^k\binom{j}{k}\alpha_k\end{aligned}$$

Therefore

$$\begin{aligned}\alpha_0 &= \beta_0 \\ \alpha_1 &= \beta_0 - \beta_1\end{aligned}$$

Another estimator of β_j is:

$$\hat{\beta}_j = \frac{1}{n}\sum_{i=1}^{n} x_{(i)}p_i^j$$

where p_i^j is a plotting position, a distribution– free estimate of $F(x_i)$. In general, we choose

$$p_i^j = \frac{i - D}{n + 1 - 2D}$$

where D is a constant, A popular choice for the Gumbel distribution is $D = 0.44$ (Guo [49]). Another reasonable choice for p_i^j is given by

$$p_i^j = \frac{\binom{i-1}{j}}{\binom{n-1}{j}}$$

The L-moments (LMOM), introduced by Hosking [57], are defined as

$$\lambda_r = \frac{1}{r}\sum_{k=0}^{r-1}(-1)^k C_{r-1}^k E(X_{r-k:r}), \quad r = 1, 2, \cdots$$

The first few L–moments are:

$$\begin{aligned}
\lambda_1 &= \int_0^1 x(F)dF \\
\lambda_2 &= \int_0^1 x(F)(2F-1)dF
\end{aligned}$$

For the Gumbel minimum distribution,

$$\lambda_1 = a - \gamma b$$

and let $t = -\ln(1-F)$. Then by Proposition A.12

$$\begin{aligned}
\lambda_2 &= \int_0^1 \{a + b\ln[-\ln(1-F)]\}(2F-1)dF \\
&= b\int_0^1 \ln[-\ln(1-F)](2F-1)dF \\
&= b\int_0^\infty \left(e^{-t} - e^{-3t}\right)\ln t dt \\
&= \frac{1}{3}(\ln 3 - 2\gamma)b
\end{aligned}$$

On the other hand, using results of probability weighted moments

$$\begin{aligned}
\hat{\lambda}_1 &= \hat{\beta}_0 \\
\hat{\lambda}_2 &= 2\hat{\beta}_1 - \hat{\beta}_0
\end{aligned}$$

Thus

$$\hat{a} = \bar{x} + \gamma\hat{b} \tag{7.18}$$

$$\hat{b} = \frac{3}{\ln 3 - 2\gamma}\left(2\widehat{\beta}_1 - \widehat{\beta}_0\right) \tag{7.19}$$

For the Gumbel maximum distribution,

$$\lambda_1 = a + \gamma b$$

and let $t = -\ln F$. Then by Proposition A.12

$$\begin{aligned}
\lambda_2 &= \int_0^1 \left[a - b\ln\left(-\ln F\right)\right](2F-1)dF \\
&= -b\int_0^1 \ln\left(-\ln F\right)(2F-1)dF \\
&= b\int_0^\infty \left(e^{-t} - 2e^{-2t}\right)\ln t dt \\
&= b\ln 2
\end{aligned}$$

Thus

$$\hat{a} = \bar{x} - \gamma\hat{b} \tag{7.20}$$

$$\hat{b} = \frac{1}{\ln 2}\left(2\widehat{\beta}_1 - \widehat{\beta}_0\right) \tag{7.21}$$

Best Linear Unbiased Estimators

Lloyd [77] obtained the best linear unbiased estimators (BLUEs) of the location a and scale b using least squares:

$$\hat{a} = \frac{1}{\Delta}\left(\alpha^T\Sigma^{-1}\alpha\underline{1}^T\Sigma^{-1} - \alpha^T\Sigma^{-1}\underline{1}\alpha^T\Sigma^{-1}\right)\underline{X}' \tag{7.22}$$

$$\hat{b} = \frac{1}{\Delta}\left(\underline{1}^T\Sigma^{-1}\underline{1}\alpha^T\Sigma^{-1} - \underline{1}^T\Sigma^{-1}\alpha\underline{1}^T\Sigma^{-1}\right)\underline{X}' \tag{7.23}$$

where $\underline{X}'$ denotes the vector of available order statistics from a location and scale distribution, α and Σ denote the mean vector and the variance–covariance matrix of

the order statistics from the standardized distribution respectively, and

$$\Delta = \left(\alpha^T \Sigma^{-1} \alpha\right) \left(\underline{1}^T \Sigma^{-1} \underline{1}\right) - \left(\alpha^T \Sigma^{-1} \underline{1}\right)^2$$

The variances and covariance of these estimators are given by

$$\begin{aligned} Var(\hat{a}) &= \frac{\alpha^T \Sigma^{-1} \alpha}{\Delta} b^2 \\ Var(\hat{b}) &= \frac{\underline{1}^T \Sigma^{-1} \underline{1}}{\Delta} b^2 \end{aligned}$$

and

$$Cov(\hat{a}, \hat{b}) = -\frac{\alpha^T \Sigma^{-1} \underline{1}}{\Delta} b^2$$

so, a and b can be expressed as linear functions of the order statistics

$$\begin{aligned} \hat{a} &= \sum_{i=1}^{n} \alpha_i x_{(i)} \\ \hat{b} &= \sum_{i=1}^{n} \beta_i x_{(i)} \end{aligned}$$

with coefficients maybe tabulated. See David [23]. They tabulated the coefficients of order statistics in the BLUEs of a and b.

Downton [26] obtained unbiased estimators in the form;

$$\hat{a} = \frac{(n-1)\ln 2 - (n+1)\gamma}{n(n-1)\ln 2} v + \frac{2\gamma}{n(n-1)\ln 2} w \tag{7.24}$$

$$\hat{b} = = -\frac{n+1}{n(n-1)\ln 2} v + \frac{2}{n(n-1)\ln 2} w \tag{7.25}$$

where $v = \sum_{i=1}^{n} x_{(i)}$ and $w = \sum_{i=1}^{n} i x_{(i)}$. The explicit variances are given by

$$\begin{aligned} Var(\hat{a}) &= \frac{b^2}{n(n-1)} [1.112825n - 0.906557] \\ Var(\hat{b}) &= \frac{b^2}{n(n-1)} [0.804621n - 0.185527] \end{aligned}$$

Hassanein [54] derived the best linear unbiased estimators of a and b for Gumbel minimum distribution by using two or three order statistics. Then the best

linear unbiased estimator of a and b are

$$\begin{aligned}\hat{a} &= \sum_{i=1}^{k} c_i x_{(n_i)} \\ \hat{b} &= \sum_{i=1}^{k} c_i' x_{(n_i)}\end{aligned}$$

Hassanein tabulated the coefficients c_i and c_i' for the sample size up to 20 and $k = 2, 3$.

Mann and Fertig [82] proposed small sample unbiased corrections to the asymptotically unbiased optimum estimator of Hassanein [54]:

$$\begin{aligned}\hat{a} &= \sum_{i=1}^{k} l_{i,n} x_{(n_i)} - \bar{l}_{k,n}\hat{b} \\ \hat{b} &= \sum_{i=1}^{k} \frac{L_{i,n}}{\bar{L}_{k,n}} x_{(n_i)}\end{aligned}$$

and the best invariant estimator

$$\tilde{a} = \hat{a} - B_{k,n}\tilde{b} \tag{7.26}$$

$$\tilde{b} = \frac{1}{1 + C_{k,n}}\hat{b} \tag{7.27}$$

and they tabulated the coefficients $l_{i,n}, L_{i,n}, \bar{l}_{k,n}, \bar{L}_{k,n}, B_{k,n}$ and $B_{k,n}$ for $n = 20, \cdots, 40$ and $k = 2, \cdots, 10$.

Let $X_i \overset{\text{iid}}{\sim} F, h(X_1, X_2, \cdots, X_r)$ a symmetric function, and $E|h(X_1, X_2, \cdots, X_r)| < \infty$. Let $\theta(F) = E[h(X_1, X_2, \cdots, X_r)]$, a U-statistic U_n with kernel h of order r is defined as

$$U_n = \frac{1}{C_n^r} \sum_C h(X_1, X_2, \cdots, X_r) \tag{7.28}$$

where $\sum_C$ ranges over all subsets of size r chosen from $1, 2, \cdots, n$. $E[U_n] = \theta(F)$, that is, U-statistics are unbiased. Let the BLUE of a as given at the end of last section be written as

$$\hat{a}_{blue} = d_1 x_{(1)} + d_2 x_{(2)} + \cdots + d_k x_{(k)} \tag{7.29}$$

and that of b be written as

$$\hat{b}_{blue} = h_1 x_{(1)} + h_2 x_{(2)} + \cdots + h_k x_{(k)} \tag{7.30}$$

Thomas and Sreekumar [109] obtained U–statistic estimators for a and b:

$$\hat{a}_n^k = \frac{1}{C_n^k} \sum_{r=1}^{n} \left[\sum_{j=1}^{k} C_{n-r}^{k-j} C_{r-1}^{j-1} d_j \right] x_{(r)} \tag{7.31}$$

$$\hat{b}_n^k = \frac{1}{C_n^k} \sum_{r=1}^{n} \left[\sum_{j=1}^{k} C_{n-r}^{k-j} C_{r-1}^{j-1} h_j \right] x_{(r)} \tag{7.32}$$

Method of Maximum Entropy

Jowitt [64] obtained estimators for the Gumbel maximum distribution from the principle of maximum entropy by solving following equations,

$$\frac{1}{n} \sum_{i=1}^{n} \frac{x_i - a}{b} = \gamma \tag{7.33}$$

$$\frac{1}{n} \sum_{i=1}^{n} e^{\frac{x_i - a}{b}} = 1 \tag{7.34}$$

Phien [92] derived expressions for the variances-covariances of the MME estimators as follow:

$$\begin{aligned} Var(\hat{a}) &= \frac{1.115}{n} b^2 \\ Var(\hat{b}) &= \frac{0.645}{n} b^2 \\ Cov(\hat{a}, \hat{b}) &= \frac{0.273}{n} b^2 \end{aligned}$$

Linear Regression Method

A simple method to estimate parameters is linear regression by using a $\log\log$ transformation. For the Gumbel maximum distribution, we have

$$\log\left[-\log F(x)\right] = -\frac{1}{b} x + \frac{a}{b} \tag{7.35}$$

This is a straight line with slope $-\frac{1}{b}$ and intercept $\frac{a}{b}$.

New Method 1

This result comes from Theorem 2.5:

$$\begin{aligned}\hat{a}_{est} &= \bar{X}_n - \frac{\mu_Z}{E(S_{Z,n})}S_n \\ \hat{b}_{est} &= \frac{1}{E(S_{Z,n})}S_n\end{aligned}$$

where Z denotes the standard Gumbel distribution.

New Bayesian Method

Bayesian–Maximum Likelihood Method with Prior $\pi(a) = 1$

For the Gumbel maximum distribution, we can use a numerical method to find b.

$$b = \frac{1}{n}\sum_{i=1}^{n} x_i - \frac{\sum_{i=1}^{n} x_i e^{-\frac{x_i}{b}}}{\sum_{i=1}^{n} e^{-\frac{x_i}{b}}}$$

Then find a Bayesian estimator for a

$$\hat{a}_B = b\left(\frac{1}{2n} - \sum_{k=2}^{\infty}\frac{A_k}{n(n+1)\cdots(n+k-1)}\right) - b\ln\left(\frac{1}{n}\sum_{i=1}^{n} e^{-\frac{x_i}{b}}\right)$$

For the Gumbel minimum distribution, we use

$$b = \frac{\sum_{i=1}^{n} x_i e^{\frac{x_i}{b}}}{\sum_{i=1}^{n} e^{\frac{x_i}{b}}} - \frac{1}{n}\sum_{i=1}^{n} x_i$$

Then find the Bayesian estimator for a,

$$\hat{a}_B = b\ln\left(\frac{1}{n}\sum_{i=1}^{n} e^{\frac{x_i}{b}}\right) - b\left(\frac{1}{2n} - \sum_{k=2}^{\infty}\frac{A_k}{n(n+1)\cdots(n+k-1)}\right)$$

Full Bayesian Estimator for Gumbel Minimum Distribution

Let $\psi(\cdot)$ denotes digamma function, $q(\underline{x}) = \frac{nS_{X,n}}{ES_{Z,n}}$, and $r = \sum_{i=1}^{n} e^{\frac{x_i}{b}}$

1. Jeffrey's non-informative prior: $\pi(a, b) = \frac{1}{b^2}$

$$\begin{aligned}\hat{a}_B &= \bar{x} + \frac{k_{\Delta_1}\ln\Delta - \psi(n)}{n-1}q(\underline{x}) \\ \hat{b}_B &= \frac{1}{n-1}q(\underline{x})\end{aligned}$$

2. Conjugate prior: $\pi(a,b) = \frac{r^n[q(\underline{x})]^{2n-2}}{\Gamma(n)\Gamma(2n-2)}\frac{1}{b^{n+2}}e^{-\frac{q(\underline{x})}{b}}\exp\left[-n\frac{a}{b} - r\exp\left(-\frac{a}{b}\right)\right]$

$$\begin{aligned}\hat{a}_B &= \bar{x} + \frac{2[k_{\Delta_1}\ln\Delta + \ln 2 - \psi(2n)]}{2n-1}q(\underline{x}) \\ \hat{b}_B &= \frac{2}{2n-1}q(\underline{x})\end{aligned}$$

where

$$\Delta = n\Gamma\left(1 - \frac{n-1}{n}\right)\left[\Gamma\left(1 - \frac{1}{n}\right)\right]^{n-1}$$

and k_{Δ_1} can be determined by Monte Carlo method with

$$k_{\Delta_1} = \frac{1}{B}\sum_{j=1}^{n} k_j$$

with

$$k_j = -\frac{\ln\left[\frac{e^{\frac{1}{n}\sum_{i=1}^{n} z_i^{(j)}}}{\left(\sum_{i=1}^{n} e^{z_i^{(j)}}\right)}\right]}{\ln\left\{n\Gamma\left(1 - \frac{n-1}{n}\right)\left[\Gamma\left(1 - \frac{1}{n}\right)\right]^{n-1}\right\}}$$

where $z_1^{(j)}, \cdots, z_n^{(j)}$ are generated from the standard Gumbel minimum distribution, and superscript j denotes j^{th} group.

7.3 Simulation Results and Conclusions

In this section, we provide a comprehensive comparison of the main estimation methods for the Gumbel minimum distribution. The comparison is based on: the bias and

the mean-squared error (MSE), which are given by

$$\begin{aligned} Bias &= \frac{1}{K}\sum_{i=1}^{K}\left(\hat{\theta}_i - \theta\right) \\ MSE &= \frac{1}{K}\sum_{i=1}^{K}\left(\hat{\theta}_i - \theta\right)^2 \end{aligned}$$

where K is the number of replications. Larger values of the bias and the mean-squared error correspond to less efficient estimators.

The bias and the mean-squared error were computed for sample sizes 20, 25, 30, $\cdots$, 80, 100, 150, 200, 300 with number of replications $K = 2,000,000$, and the parameter values $a = 12$ and $b = 5$. The plots of the computed biases and the mean squared errors for all five estimators are shown in Figures 7.1 - 7.4.

The PWM, Bayesian and new method two yield smaller biases of estimators a than the two other methods. However, the biases for a are approximately equal for all sufficiently large n. Figure 7.2 shows MSEs for all estimators of a far different sample sizes n. There are no obvious differences in MSE among all methods. The MLE estimator yields the worse performance for the biases of b, a, however, it gives the best performance for the MSE.

Of the five methods considered, the best performance is by the PWM estimators. The worst performance is by the MM estimators.

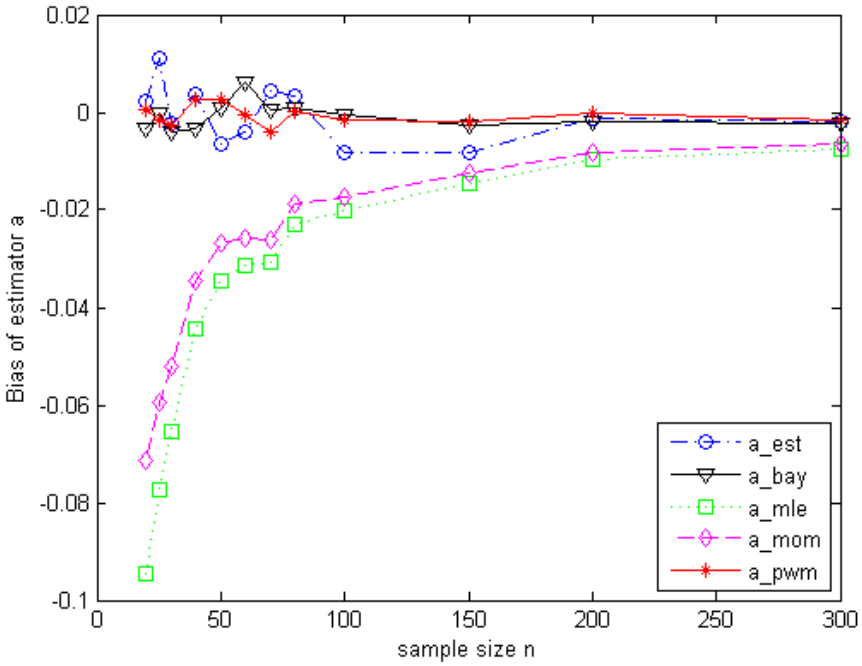
Table 7.1: $a = 12, b = 5, 2000000$ trials

	Estimator of location a									
	$\hat{a}_est$		$\hat{a}_bay$		$\hat{a}_mle$		$\hat{a}_mom$		$\hat{a}_pwm$	
n	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
20	0.0023	1.4471	-0.0035	1.4475	-0.0944	1.4192	-0.0714	1.4572	0.0005	1.4057
25	0.0110	1.1559	-0.0002	1.1564	-0.0772	1.1315	-0.0595	1.1631	-0.0016	1.1224
30	-0.0022	0.9640	-0.0042	0.9641	-0.0653	0.9392	-0.0522	0.9689	-0.0025	0.9337
40	0.0035	0.7228	-0.0032	0.7230	-0.0443	0.6989	-0.0345	0.7252	0.0025	0.6970
50	-0.0065	0.5815	0.0009	0.5813	-0.0345	0.5606	-0.0269	0.5827	0.0028	0.5597
60	-0.0041	0.4831	0.0062	0.4830	-0.0315	0.4644	-0.0257	0.4842	-0.0005	0.4641
70	0.0046	0.4139	0.0004	0.4140	-0.0306	0.3976	-0.0262	0.4151	-0.0040	0.3976
80	0.0035	0.3632	0.0009	0.3632	-0.0231	0.3487	-0.0187	0.3638	0.0001	0.3487
100	-0.0084	0.2926	-0.0006	0.2925	-0.0203	0.2801	-0.0172	0.2930	-0.0017	0.2803
150	-0.0083	0.1949	-0.0027	0.1948	-0.0145	0.1859	-0.0124	0.1951	-0.0021	0.1864
200	-0.0011	0.1453	-0.0020	0.1453	-0.0095	0.1386	-0.0083	0.1454	-0.0003	0.1389
300	-0.0021	0.0973	-0.0025	0.0973	-0.0077	0.0925	-0.0065	0.0973	-0.0015	0.0927

Table 7.2: $a = 12, b = 5, 2000000$ trials

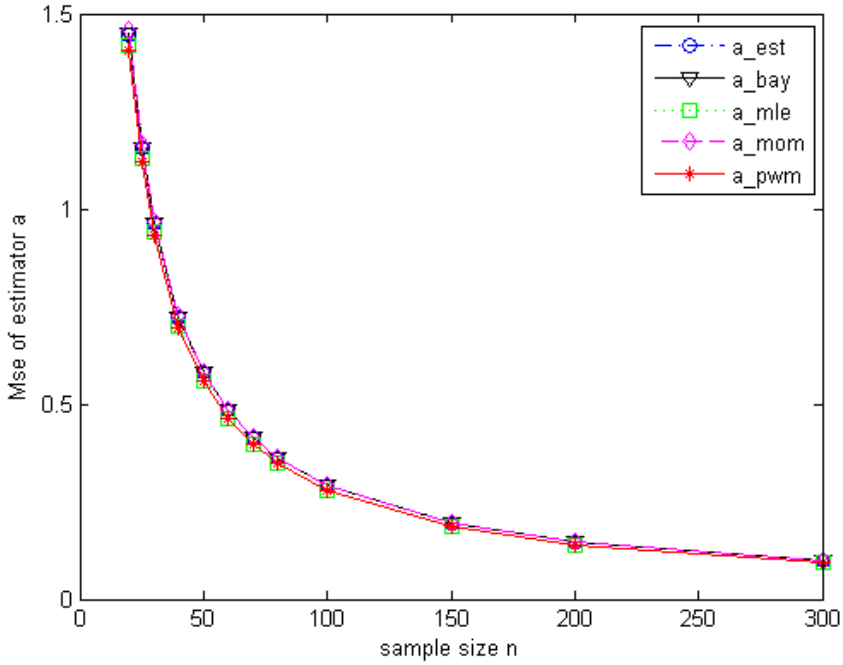
	Estimator of scale b									
	$\hat{b}_est$		$\hat{b}_bay$		$\hat{b}_mle$		$\hat{b}_mom$		$\hat{b}_pwm$	
n	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
20	0.0026	1.2852	-0.0033	1.2822	-0.1925	0.7987	-0.1251	1.2360	-0.0007	1.0466
25	0.0227	1.0515	-0.0062	1.0390	-0.1538	0.6330	-0.0995	1.0104	0.0009	0.8318
30	0.0017	0.8681	-0.0028	0.8666	-0.1268	0.5250	-0.0850	0.8455	0.0011	0.6895
40	0.0011	0.6572	-0.0010	0.6566	-0.0971	0.3908	-0.0648	0.6442	-0.0007	0.5139
50	-0.0164	0.5271	-0.0048	0.5293	-0.0773	0.3109	-0.0519	0.5220	-0.0004	0.4086
60	-0.0066	0.4431	-0.0017	0.4439	-0.0642	0.2565	-0.0441	0.4383	-0.0005	0.3386
70	0.0153	0.3836	-0.0053	0.3803	-0.0542	0.2209	-0.0379	0.3767	0.0005	0.2908
80	0.0058	0.3378	0.0111	0.3386	-0.0484	0.1917	-0.0326	0.3336	-0.0001	0.2540
100	-0.0110	0.2686	0.0035	0.2700	-0.0384	0.1533	-0.0264	0.2675	0.0004	0.2024
150	-0.0094	0.1801	0.0088	0.1814	-0.0246	0.1025	-0.0164	0.1797	0.0014	0.1351
200	-0.0018	0.1361	-0.0044	0.1360	-0.0195	0.0764	-0.0142	0.1356	-0.0004	0.1012
300	0.0004	0.0913	-0.0004	0.0913	-0.0117	0.0508	-0.0073	0.0911	0.0013	0.0673

Figure 7.1: Bias of location estimator of Gumbel



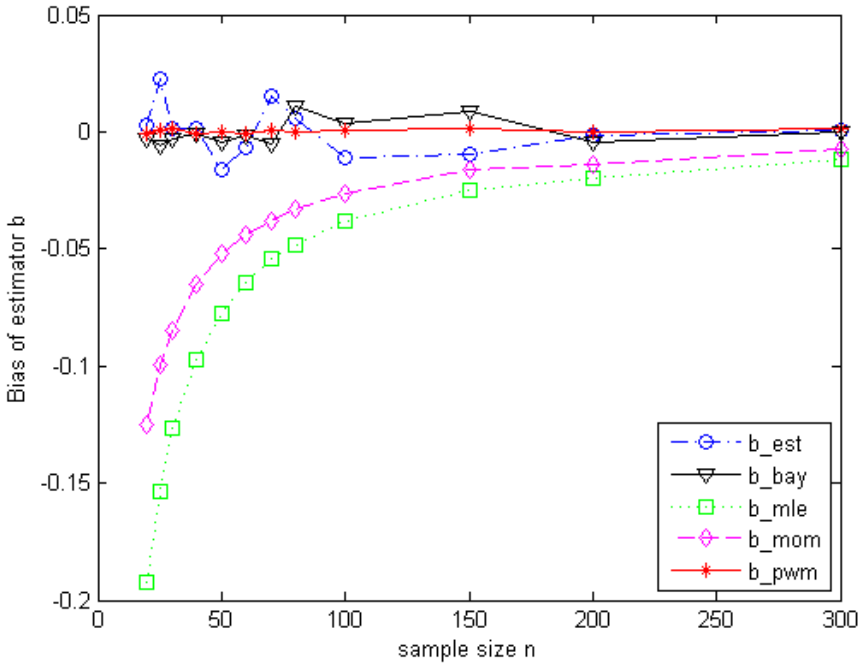
;

Figure 7.2: MSE of location estimator of Gumbel



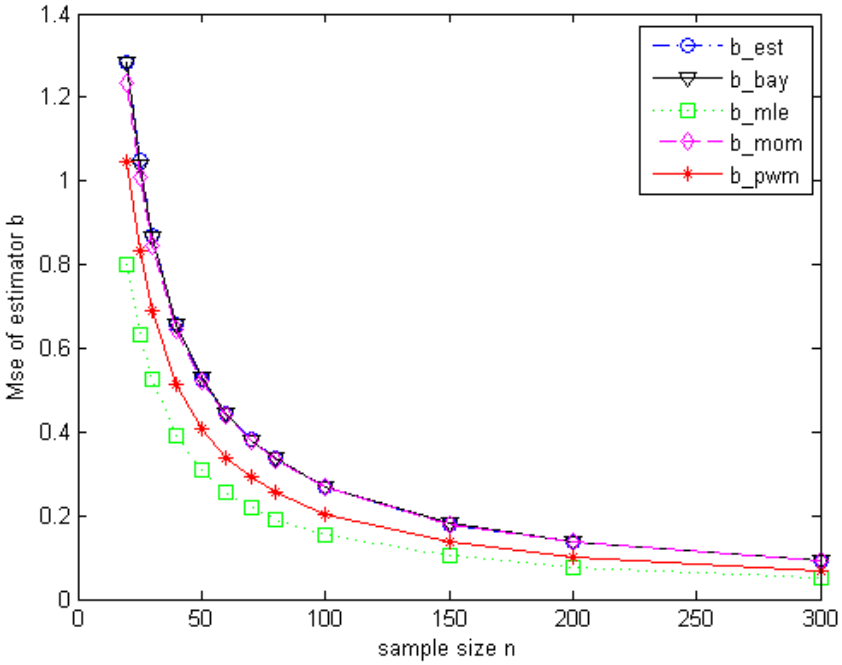
; ;

Figure 7.3: Bias of scale estimator of Gumbel



; ;

Figure 7.4: MSE of scale estimator of Gumbel



; ;

7.4 Methods of Prediction Interval Construction for the Gumbel Minimum Distribution

Our literature survey suggests several techniques for constructing prediction intervals for the Gumbel minimum distribution. These approaches are mainly divided into three methods: exact, approximation, and simulation. Two methods proposed in this thesis are based on simulation: Method 1 is based on Monte Carlo simulation and details are given in Section 5.4; Method 2 is based on Markov Chain Monte Carlo simulation and details are provided in Section 6.3.

Suppose that $\{X_1, X_2, \cdots, X_n\}$ is a random sample from the Gumbel minimum distribution, let $\{Y_1, Y_2, \cdots, Y_m\}$ be m future independent observations from the same distribution. Consider the invariant statistics

$$T = \frac{\hat{a} - Y_{(1)}}{\hat{b}}$$

A $100\gamma\%$ lower prediction limit on $Y_{(1)}$ is then given by

$$\hat{a} - t_\gamma \hat{b}$$

such that

$$P\left(Y_{(1)} > \hat{a} - t_\gamma \hat{b}\right) = \gamma$$

7.4.1 The Approximation for the Distribution

For any invariant estimators $\hat{a}$ and $\hat{b}$, Engelhardt and Bain [30] derived that $L(t) = -\ln\left[\frac{M(t)\chi^2_\gamma(L)}{L}\right]$ is an approximate solution to the equation

$$\gamma = P\left[L(t) \leq \frac{\hat{b}}{b}t - \frac{\hat{a} - a}{b}\right] \tag{7.36}$$

with

$$v = v(t) = Var\left(\frac{\hat{a}-a}{b} - \frac{\hat{b}}{b}t\right) = Var\left(\frac{\hat{a}}{b}\right) + t^2 Var\left(\frac{\hat{b}}{b}\right) - 2tCov\left(\frac{\hat{a}}{b}, \frac{\hat{b}}{b}\right) \quad (7.37)$$

and

$$\begin{aligned} L &= L(t) = \frac{8v+12}{v^2+6v} \\ H(L) &= \frac{15L^2+5L+6}{15L^3+6L} \\ M(t) &= e^{H(L)-t} \end{aligned}$$

Engelhardt and Bain [30] provide estimators

$$\begin{aligned} \hat{a} &= \bar{X} + \gamma\hat{b} \\ \hat{b} &= \frac{1}{nk_n}\left[-\sum_{i=1}^{s} X_{(i)} + \left(\frac{s}{n-s}\right)\sum_{i=s+1}^{n} X_{(i)}\right] \end{aligned}$$

where $\gamma = 0.5772$, $s = \lfloor 0.84n \rfloor$, the largest integer not exceeding $0.84n$, and

$$k_n = \frac{1}{n}E\left[-\sum_{i=1}^{s}\frac{X_i - a}{b} + \left(\frac{s}{n-s}\right)\sum_{i=s+1}^{n}\frac{X_i - a}{b}\right]$$

In the same paper, they also provide values of k_n, $Var\left(\frac{\hat{b}}{b}\right)$, $Var\left(\frac{\hat{a}}{b}\right)$, and $Cov\left(\frac{\hat{a}}{b}, \frac{\hat{b}}{b}\right)$ for n from 2 to 60 in Table 6.

Engelhardt and Bain [32] showed that

$$P(T < t) \approx P[F(2, L) > mM(t)]$$

where $T = \frac{\hat{a}-Y_{(1)}}{\hat{b}}$. Thus, the value t_γ is approximated by the value t such that

$$F_{1-\gamma}(d, L) = mM(t)$$

Since the $(1-\gamma)$th quantile of $F(2, L)$ is

$$F_{1-\gamma}(2, L) = \frac{L(\gamma^{-\frac{2}{L}} - 1)}{2}$$

so t_γ is a solution to the equation

$$2mM(t) = L(t)\left[\gamma^{-\frac{2}{L(t)}} - 1\right]$$

Similar, for the jth smallest of m future observations, the value of t_γ such that

$$P\left(Y_{(j)} > \hat{a} - t_\gamma \hat{b}\right) = \gamma$$

is a solution to the equation

$$F_{1-\gamma}(d, L) = \frac{M(t)}{c} \tag{7.38}$$

where

$$\begin{aligned} c &\doteq \ln(m+0.5) - \ln(m-j+0.5) \\ d &\doteq \frac{2\left[\ln(m+0.5) - \ln(m-j+0.5)\right]^2}{\frac{1}{k-j+0.5} - \frac{1}{k+0.5}} \end{aligned}$$

Let $\hat{a}$ and $\hat{b}$ be the best linear invariant estimators. Fertig, Meyer and Mann [36] provide a table for percentiles of $T = \frac{\hat{a}-Y_{(1)}}{\hat{b}}$ with sample size $n = 5(5)25$ and future sample size $m = 1$. For $n > 25$ and $m > 1$, they suggest a different approximation for the distribution of the statistic T such that

$$P\left(T \le t\right) \approx P\left(\frac{v_3(\ln F_{v_1,v_2} + \lambda)}{\chi^2_{v_3}} \le \frac{t - \frac{B}{C}}{1+C}\right) \tag{7.39}$$

with

$$\begin{aligned} \psi'\left(\frac{v_1}{2}\right) &= A - \frac{B^2}{C}, \qquad v_2 = 2, \qquad v_3 = \frac{2}{C} \\ \lambda &= -\frac{B}{C} + \ln\left(\frac{v_1}{2}\right) - \psi\left(\frac{v_1}{2}\right) + \ln(m) \\ a^* &= \hat{a} - E\left(\frac{\hat{a}-a}{b}\right)b^*, \qquad b^* = \frac{1}{E\left(\frac{\hat{b}}{b}\right)}\hat{b}, \qquad \tilde{a} = a^* - \frac{B}{C}b^* \\ A &= Var\left(\frac{a^*-a}{b}\right), \quad B = Cov\left(\frac{a^*-a}{b}, \frac{b^*}{b}\right), \quad C = Var\left(\frac{b^*}{b}\right) \end{aligned}$$

where ψ and ψ' are the digamma and trigamma functions. Let Q_γ be 100γth percentile

of $\frac{v_3(\ln F_{v_1,v_2}+\lambda)}{\chi^2_{v_3}}$ such that

$$\gamma = \frac{v_3[\ln F_{v_1,v_2}(Q_\gamma)+\lambda]}{\chi^2_{v_3}(Q_\gamma)}$$

Then t_γ is a solution to the equation

$$Q_\gamma = \frac{t-\frac{B}{C}}{1+C}$$

Mann (1976) shows that the statistic

$$W^* = \frac{Y_{(k)} - \hat{a}^* + \frac{B^*}{C^*}\hat{b}^*}{\left[E\left(\frac{Y_{(k)}-a}{b}\right) + \frac{B^*}{C^*}\right]\hat{b}^*} \tag{7.40}$$

is distributed approximately by an F distribution with degrees of freedom

$$\begin{aligned} v_1 &= \frac{2\left[E\left(\frac{Y_{(k)}-a}{b}\right)+\frac{B^*}{C^*}\right]^2}{Var\left(\frac{Y_{(k)}-a}{b}\right)+A^*-\frac{B^{*2}}{C^*}} \\ v_2 &= \frac{2}{C^*} \end{aligned}$$

where $\hat{a}^*$ and $\hat{b}^*$ denote the best (or approximately best) linear unbiased estimator of a and b with

$$\begin{aligned} Var\left(\hat{a}^*\right) &= A^* b^2 \\ Var\left(\hat{b}^*\right) &= C^* b^2 \\ Cov\left(\hat{a}^*, \hat{b}^*\right) &= B^* b^2 \end{aligned}$$

7.4.2 Conditional and Unconditional Probability Method

Except for some simple cases, it is impossible to obtain exact prediction intervals for the kth future observation. Lawless [72] constructs prediction intervals for the smallest future observation from the Gumbel minimum distribution by conditioning $U = \frac{Y_{(1)}-\hat{a}}{\hat{b}}$ on the ancillary statistics $A = a_i$ with $a_i = \frac{X_i-\hat{a}}{\hat{b}}$ for $i = 1, 2, \cdots, n$. He

gives the conditional density of U given A as follow:

$$P(U \geq t \mid A) = K(A) \int_0^\infty \frac{z^{n-2} e^{z \sum_{i=1}^n a_i}}{(me^{tz} + \sum_{i=1}^n e^{a_i z})^n} dz \tag{7.41}$$

and approximates the unconditional density of U as

$$P(U \geq t) = K(A) \int_0^\infty \frac{z^{n-2} e^{-n\gamma z}}{\left[\frac{m}{n} e^{tz} + \Gamma(1+z)\right]^n} dz$$

where $K(A)$ is a normalizing constant and γ is Euler's constant.

Lawless' approach requires numerical integration for each new sample so one can not construct tables in which table values are only dependent on the sample size of the observed data and future sample. Kushary [70] provides another unconditional probability method to construct prediction intervals by using a Taylor expansion such that

$$\alpha \approx F_{W|U,V}(k_2) + \frac{1}{2} f'_{W|U,V}(k_2) E\left[U^2 + 2k_2 U(V-1) + k_2^2 (V-1)^2\right]$$

where $W = \frac{Y_{(j)} - a}{b}$, $U = \frac{\hat{a} - a}{b}$ and $V = \frac{\hat{b}}{b}$. Kushary uses the least squares estimator given by Smith & Bain (1975):

$$\begin{aligned} \hat{a} &= \bar{x} - \bar{c}\hat{b} \\ \hat{b} &= \frac{S_{cx}}{S_c^2} \end{aligned}$$

where for $r \leq n$

$$\begin{aligned} c_i &= F_W^{-1}\left(\frac{i}{n+1}\right) \\ S_{cx} &= \sum_{i=1}^r (x_{(i)} - \bar{x})(c_i - \bar{c}) \\ S_c^2 &= \sum_{i=1}^r (c_i - \bar{c})^2 \end{aligned}$$

7.4.3 Monte Carlo Estimation of the Distribution Percentiles and Method 1

Antle and Rademaker [5] obtained lower prediction intervals for the largest of a set of future observations from the Gumbel maximum distribution using maximum likelihood. They used Monte Carlo to obtain tables for percentiles of the statistic

$$\frac{Y_{(m)} - \hat{a}}{\hat{b}} - \ln m$$

for given random samples of size $n = 10(10)70, 100, \infty$.

Hoppe and Fang [58] consider a Bayesian approach to predicting the minimum of any specified number of future observations from a Gumbel distribution based on previous observations. The $(1-\alpha)\%$ lower prediction limit with known b is

$$L_B = c_{1-\alpha} b \ln \left[\frac{[(1-\alpha)^{-\frac{1}{n}} - 1]}{m} \sum_{i=1}^{n} e^{\frac{Y_i}{b}} \right]$$

where $c_{1-\alpha}$ are tabulated in Fang [35].

In Section 5.4, we construct prediction interval based on the pivotal

$$\hat{U} = \frac{Y_{(j)} - X_{(i)}}{\hat{b}} = \frac{Z^Y_{(j)} - Z^X_{(i)}}{S_{Z^X,n}} E(S_{Z^X,n})$$

where $\hat{b}$ is from our result in Theorem 2.5. This method refers to Method 1 in the next section.

7.4.4 Markov Chain Monte Carlo Estimation of the Distribution Percentiles and Method 2

In Section 6.3, we constructed Bayesian prediction intervals for the Gumbel minimum distribution by using Gibbs sampling, a special case of single–component Metropolis–Hastings, to estimate the percentiles of statistics of future observations, and this method refers to Method 2 in the next section.

7.5 Numerical Examples

Two important criteria to evaluate and compare prediction intervals are coverage probability and length of prediction interval. A prediction region for Y is a random set $D_n = D_n(\gamma, \underline{X})$, depending on the sample, that contains Y with a prescribed probability γ. For a level γ prediction interval $D_n = [L(\underline{X}), U(\underline{X})]$, the coverage probability of D_n for Y given $\underline{X}$ is

$$CP(D_n) = P\left[L(\underline{X}) \leq Y \leq U(\underline{X})\right]$$

In general, we hope that $CP(D_n)$ is exactly the same as γ, or converges to γ as the sample increases.

The length of a prediction interval is also an important consideration. We often hope that the length is short. However, the coverage probability is large when the length is long, and coverage probability is small when the length is short. In this work, we compare prediction intervals by simultaneously considering coverage probability and length, that is, comparing coverage probability with the same length, or comparing the length with the same coverage probability.

Below, we compare $(1-p)\%$ two-sided prediction intervals with $p = 0.1$ for a future order observation $Y_{(j)}$ constructed by Engelhardt and Bain [31] and by the method in Section 5.4, where we consider the pivotal:

$$U = \frac{Y_{(j)} - X_{(i)}}{\hat{b}}$$

According to our previous results, the length of a $(1-p)\%$ two-sided prediction intervals is given by

$$Length_PI = \left(Q_{1-\frac{p}{2}} - Q_{\frac{p}{2}}\right)\hat{b}$$

Since $\hat{b}$ is fixed for a given data set, we will choose i such that $Q_{1-\frac{p}{2}} - Q_{\frac{p}{2}}$ obtains the

smallest value, so that the prediction interval has the shortest length. This can be done by Monte Carlo method. The following graphs show the differences for different values of i.

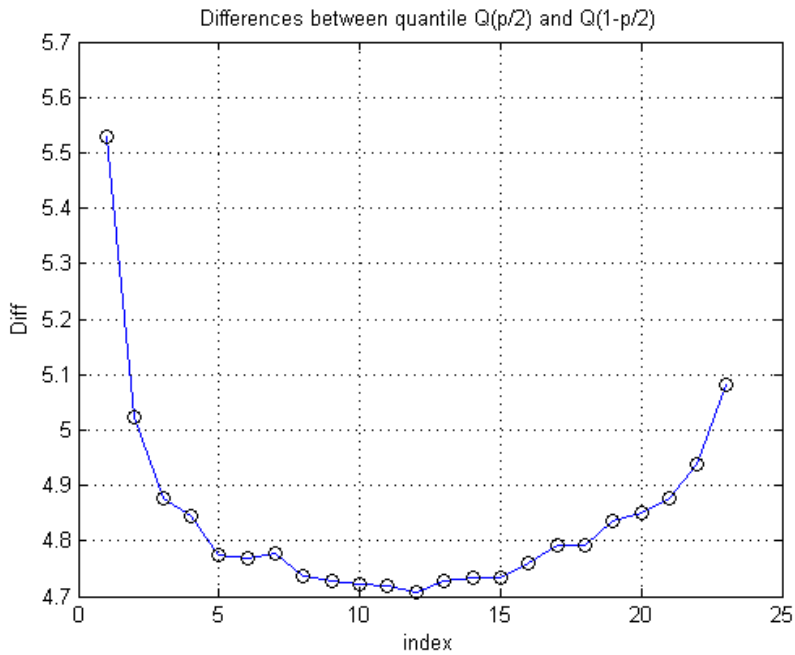
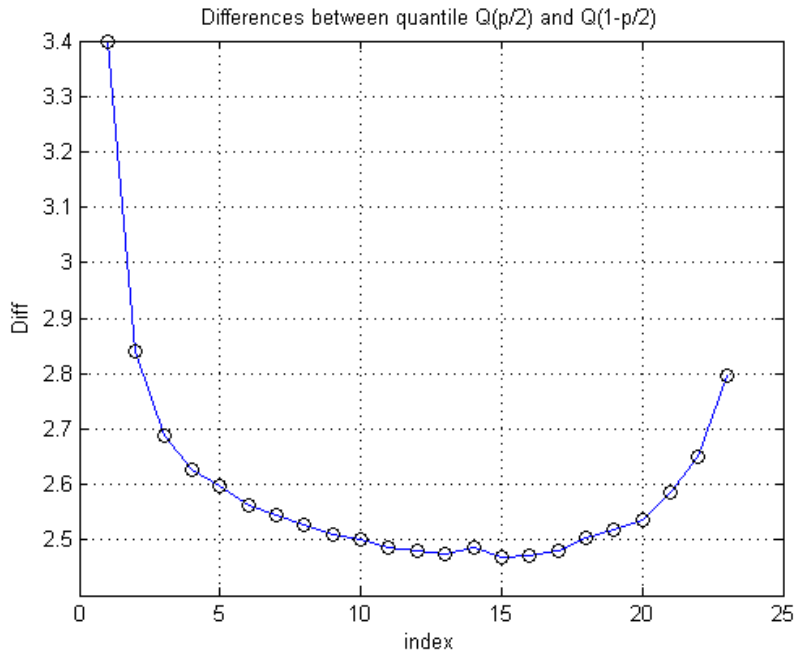


Figure 7.5: Difference between two quantiles for $n = 23, m = 8, j = 1$ with 2000000 simulations

From Figures 7.5-7.7, we can see that the optimum value of i will be 12, 15, 21 corresponding to Y_{min}, $Y_{(3)}$ and Y_{max}, respectively.

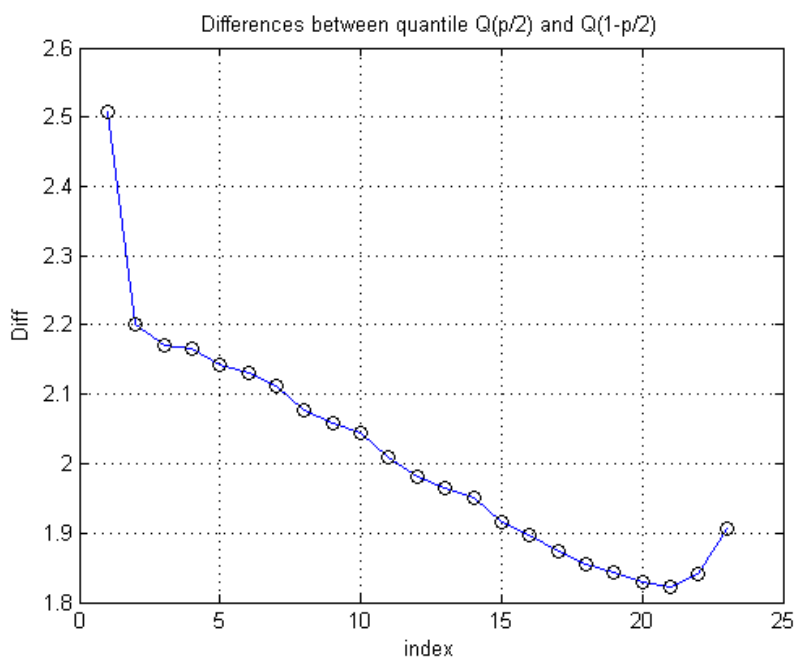
In the following, we study the effect of different values of n and m on coverage probability and average length of two-sided prediction intervals. Experiments were run 100,000 times and the results are shown in Tables 7.3 and 7.4. The tables present coverage probabilities and average length of prediction intervals obtained from two methods.

Figure 7.6: Difference between two quantiles for $n = 23, m = 8, j = 3$ with 2000000 simulations



;

Figure 7.7: Difference between two quantiles for $n = 23, m = 8, j = 8$ with 2000000 simulations



;

Tables 7.3 and 7.4 show that coverage probabilities of our approach, called Method 1, are closer to the pre-specified probability than those of Engelhardt and Bain. Although the length for Y_{min} from Method 1 is a little larger than that of theirs, the difference is very small. The experimental results demonstrate that the performance of Method 1 is better than that of Engelhardt and Bain [31].

We now present a numerical example to compare prediction limits obtained from different methods. The data denoted by t come from a result of a test on endurance of deep groove ball bearings and was originally discussed by Lieblein and Zelen (1956).

17.88	28.92	33.00	41.52	42.12	45.60	48.48	51.84	51.96
54.12	55.56	67.80	68.64	68.64	68.88	84.12	93.12	98.64
105.12	105.84	127.92	128.04	173.40				

The data have been used by many authors.

After a log transformation $x = \ln(t)$, the data $\underline{x}$ could be described by a Gumbel minimum distribution (Lieblein and Zelen,[74]).

2.8837	3.3645	3.4965	3.7262	3.7405	3.8199	3.8812	3.9482	3.9505
3.9912	4.0175	4.2166	4.2289	4.2289	4.2324	4.4322	4.5339	4.5915
4.6551	4.6619	4.8514	4.8523	5.1556				

Let $Y_{(1)}$ be minimum of m future observations respect to the log transformation and $R_{(1)}$ be minimum of m future observations respect to the original data. Consider a level α lower prediction limit (PL) such that

$$P\left[Y_{(1)} > k_\alpha(\underline{x})\right] = \alpha$$

Engelhardt and Bain [32] obtain a 90% lower prediction limit $k_\alpha(\underline{x}) = 0.775$ for $Y_{(1)}$ and $e^{0.775} = 2.17$ for $R_{(1)}$ when $m = 100$. Using maximum likelihood estimators conditional on a set of ancillary statistics, Lawless [72] gives a lower prediction limit

$k_\alpha(\underline{x}) = 0.732$ for $Y_{(1)}$ and 2.08 for $R_{(1)}$. Mann et al. [80] yields the prediction limit 0.695 for $Y_{(1)}$ and 2.00 for $R_{(1)}$.

In the same paper, Engelhardt and Bain [32] provide prediction limit for the 5th smallest future observation $Y_{(5)}$, they also calculate prediction limit for $Y_{(5)}$ by using method of Mann [81].

Using Method 1, proposed in Section 6.4, we find the frequentist prediction limit 1.0034 for $Y_{(1)}$, 2.7275 for $R_{(1)}$, and 2.5470 for $Y_{(5)}$, 13.12 for $R_{(5)}$.

	$m=1$		$m=2$		$m=3$		$m=4$		$m=5$	
	Length	*CP*	*Length*	*CP*	*Length*	*CP*	*Length*	*CP*	*Length*	*CP*
Method 1										
$n=10$	30.9918	0.9489	31.6498	0.9485	32.4410	0.9499	33.0223	0.9483	33.6505	0.9497
20	27.8951	0.9504	28.1801	0.9502	28.5287	0.9502	28.7842	0.9488	29.1332	0.9493
30	26.8579	0.9459	26.9894	0.9480	27.2805	0.9486	27.4727	0.9496	27.6372	0.9499
50	26.0126	0.9496	26.1293	0.9492	26.2505	0.9490	26.3144	0.9486	26.5109	0.9489
Eng.&Bain										
$n=10$	29.3028	0.8565	30.0310	0.8509	30.7010	0.8565	31.2625	0.8554	31.7306	0.8652
20	27.1092	0.8448	27.5652	0.8418	27.9232	0.8465	28.2021	0.8487	28.4630	0.8495
30	26.3812	0.8401	26.7013	0.8393	26.9502	0.8410	27.1583	0.8446	27.3544	0.8457
50	25.7957	0.8365	25.9999	0.8335	26.1664	0.8377	26.2950	0.8355	26.4342	0.8402

Table 7.3: 95% two-sided prediction intervals for Y_{min} with 100000 trials: $a = 12, b = 5$

	$m=2$		$m=3$		$m=4$		$m=5$	
	Length	*CP*	*Length*	*CP*	*Length*	*CP*	*Length*	*CP*
Method 1								
$n=10$	21.2290	0.9494	18.0613	0.9501	16.4283	0.9498	15.5449	0.9487
20	18.7156	0.9512	15.5058	0.9489	13.9585	0.9491	12.9427	0.9490
30	17.8242	0.9498	14.7352	0.9489	13.1209	0.9492	12.1519	0.9496
50	17.1470	0.9492	14.0969	0.9491	12.4548	0.9485	11.4783	0.9491
Eng.&Bain								
$n=10$	22.4895	0.9661	19.6958	0.9705	18.3532	0.9731	17.6466	0.9742
20	19.8327	0.9672	16.7000	0.9685	15.0477	0.9691	14.0323	0.9705
30	19.1488	0.9669	15.9793	0.9684	14.3050	0.9698	13.2362	0.9701
50	18.6411	0.9678	15.4660	0.9687	13.7601	0.9705	12.6777	0.9711

Table 7.4: 95% two-sided prediction intervals for Y_{max} with 100000 trials: $a = 12$ $b = 5$

Below, we construct 90% lower Bayesian prediction limits for $Y_{(1)}$ and $Y_{(5)}$, and for $R_{(1)}$ and $R_{(5)}$. This method, Method 2, is described in Section 7.3.2. We only discuss how to determine k_1 and k_2 in Step 1.

1. Find the Bayesian estimator for location a and scale b using (4.35) and (4.38):

$$\hat{a}_{Bay} = 4.3861 \quad \hat{b}_{Bay} = 0.41504$$

   The estimators are close to estimators given by Engelhardt and Bain [32]:

$$\hat{a} = 4.435 \quad \hat{b} = 0.493$$

   and to the MLE given by Lawless [72]

$$\hat{a}_{ML} = 4.41 \quad \hat{b}_{ML} = 0.48$$

2. Generate 100000 samples by the Monte Carlo method with parameter $\hat{a}_{Bay}$ and $\hat{b}_{Bay}$:

$$x_1^{(k)}, \cdots, x_n^{(k)}, y_1^{(k)}, \cdots, y_m^{(k)}, \quad k = 1,\ 2,\ \cdots,\ 100000$$

   where $n = 23$, $m = 100$, k denote the kth sample. Then we compute

$$t = \frac{1}{\pi}\left(\arctan\frac{Ay_j - B_j}{\sqrt{AC_j - B_j^2}} + \frac{\pi}{2}\right)$$

Here is the table of frequencies of t:

Table 7.5: Distribution of t respect to $Y_{(1)}$

t	≤ 0.26	0.27	0.28	0.29	0.3	0.31	0.32	0.33
frequency	0	3	1	2	3	7	7	16
t	0.34	0.35	0.36	0.37	0.38	0.39	0.4	0.41
frequency	24	44	47	101	166	252	397	577
t	0.42	0.43	0.44	0.45	0.46	0.47	0.48	0.49
frequency	906	1371	2193	3148	4937	6944	9485	12552
t	0.5	0.51	0.52	0.53	0.54	0.55	0.56	≥ 0.57
frequency	15107	16268	13494	8172	3073	650	49	0

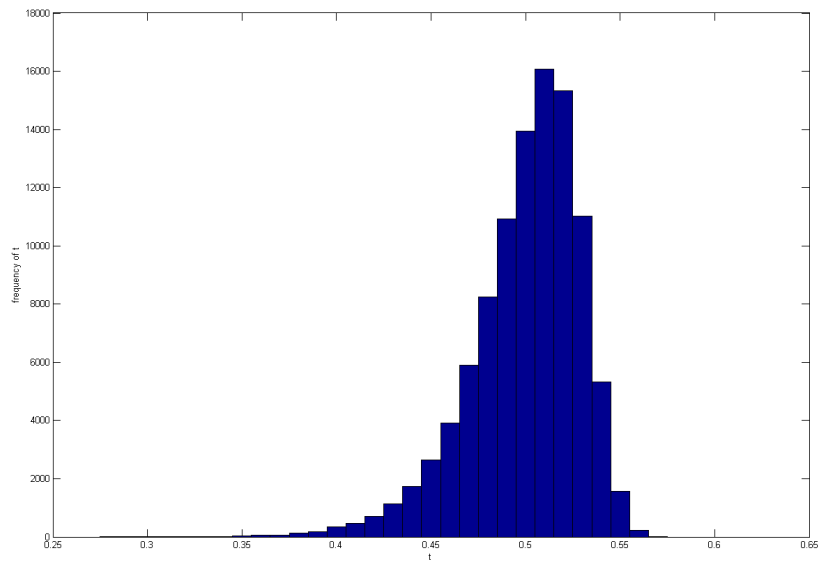


Figure 7.8: Predictive distribution of Gumbel

According to the simulation results, we choose $k_1 = 0.45$ and $k_2 = 0.54$. Now we sample t from $U(0.45, 0.54)$, where U denotes the uniform distribution, and then generate Y_j by following equation:

$$y_j = \frac{\sqrt{AC_j - B_j^2}\tan[(t-0.5)\pi] + B_j}{A}$$

By such method with a given *Lpb*, the length of the burn-in period, we obtain a Gibbs

sample with size D for $Y_{(1)}$,

$$y_{(1)}^{(1)},\ y_{(1)}^{(2)},\ \cdots,\ y_{(1)}^{(D)}$$

We then obtain the estimator $\hat{Q}_\alpha$ of Q_α such that:

$$P\left(Y_{(1)} \leq Q_\alpha\right) = \alpha$$

Taking $Lpb = 1000$ and $D = 300$, we have $\hat{Q}_\alpha = 2.6550$ and then obtain 090% lower prediction limit 14.225 for $R_{(1)}$. Similarly, 90% lower prediction limit for $R_{(5)}$ is 15.726.

The results are summarized in table 7.6 and 7.7. From the tables, we can conclude that PL obtained from our methods for "Ball Bearings" data have been improved.

Table 7.6: 90% lower Prediction Limit for future minimum observation with $n = 23, m = 100$

	Engelhardt & Bain	Lawless	Mann et al.	Method 1	Method 2
$Y_{(1)}$	0.775	0.732	0.695	1.0034	2.6550
$R_{(1)}$	2.17	2.08	2.00	2.7275	14.2250

Table 7.7: 90% lower Prediction Limit for future 5th observation with $n = 23, m = 100$

	Engelhardt & Bain	Mann et al.	Method 1	Method 2
$Y_{(5)}$	2.3599	2.3292	2.5740	2.7553
$R_{(5)}$	10.59	10.27	13.12	15.726

Hoppe and Fang [58] compare performances of one-sided upper prediction intervals obtained from three methods for the following datasets:

CANDU Data								
Group 1	3.772	3.790	3.934	4.108	4.386	4.408	4.412	4.424
	4.425	4.454	4.517	4.517	4.627	4.666	4.674	4.701
	4.711	4.742	4.794	4.875	4.999	5.052	5.064	
Group 2	3.230	3.799	3.800	3.839	3.915	4.033	4.373	4.529
	4.572	4.583	4.713	4.744	4.877	4.953	5.004	5.149
Group 3	4.157	4.268	4.268	4.276	4.291	4.314	4.362	4.380
	4.411	4.419	4.421	4.446	4.453	4.479	4.483	4.493
	4.497	4.497	4.503	4.531	4.591	4.683		
Group 4	4.318	4.319	4.338	4.375	4.433	4.598	4.674	4.695
	4.700	4.764	4.779	4.844	4.880	4.905		

which describe the minimum wall thickness of feeder pipes from Ontario Power Generation (OPG).

The first method is Hahn's approach, which was originally designed for normal data and gives a $100(1-\alpha)\%$ prediction interval (L_H, ∞) for the minimum of m future observations from a Gumbel population such that lower prediction limit is

$$L_H = \hat{a} + t_{1-\alpha}\hat{b}$$

where $\hat{a}$ and $\hat{b}$ are the MLEs of a, b, respectively.

The second method, presented in Fang [35], is a hybrid Bayesian approach, in which the location parameter is estimated using a noninformative (flat) prior and the scale parameter is estimated by maximum likelihood. Fang gives the lower prediction limit as

$$L_{B^*} = c_{1-\alpha}\hat{b}\ln\left[\frac{(1-\alpha)^{-\frac{1}{n}} - 1}{m}\sum_{i=1}^{n} e^{\frac{x_i}{c_{1-\alpha}\hat{b}}}\right]$$

The value $t_{1-\alpha}$ and $c_{1-\alpha}$ can be computed by simulations and tabulated in Fang [35].

Table 7.8 shows one-sided lower prediction limits for the Groups 1, 2, 3, 4, respectively. These correspond to predicting the minimum of $m = 13,\ 36,\ 178,\ 32$ unobserved feeder thicknesses. Based on the table, we can conclude that there are no significant differences between Hahn's approach and a hybrid Bayesian approach. However, the values in the table show that both Method 1 and Method 2 give a larger lower prediction limit than other three methods, and larger is better for a one-sided lower limit.

Table 7.8: $100(1-\alpha)\%$ One-sided Lower Prediction Limit forx Y_{min} for CANDU Data

	Group 1		*Group* 2		*Group* 3		*Group* 4	
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$
Method 1	2.265	2.882	0.172	1.281	3.312	3.542	2.904	3.364
Method 2	2.802	3.030	0.458	1.342	4.100	4.328	3.654	3.802
L_H	2.256	2.865	0.136	1.203	3.110	3.382	2.861	3.320
L_{B^*}	2.251	2.860	0.136	1.203	3.101	3.373	2.860	3.320

Chapter 8

Resampling and Prediction Interval Construction

The Bootstrap is an appealing nonparametric approach to the construction of prediction intervals when observations $\{X_1, \cdots, X_n\}$ are drawn from an unknown distribution. In this chapter, we propose a new generalized bootstrap resampling scheme.

8.1 Introduction

Let $\{X_1, \cdots, X_n, Y\}$ be independently and identically distributed random variables with a unknown common distribution F. Here the X's represent the given sample and Y is a single unobserved future value. Put

$$T = \frac{\bar{X} - Y}{S}, \tag{8.1}$$

where $\bar{X}$ and S^2 are the sample mean and variance. Then a $100(1-\alpha)\%$ one-sided upper prediction interval for Y is

$$\left(\bar{X} - t_{1-\alpha} S, \infty\right)$$

where $t_{1-\alpha}$ be the $(1-\alpha)$ quantile of T;

Since the distribution F is unknown, one may use the bootstrap method to estimate $t_{1-\alpha}$ (Bai and Olshen [7]). Let $\{X_1^*, \cdots, X_n^*\}$ and Y^* be random samples of sizes n and 1 drawn with replacement from $\{X_1, \cdots, X_n\}$. Define

$$T = \frac{\bar{X}^* - Y^*}{S^*},$$

where $\bar{X}^*$ and S^* are the bootstrap versions of $\bar{X}$ and S. Then the estimate $t_{1-\alpha}^*$ is

just the $(1-\alpha)$ quantile of T^*. The bootstrap version of the above interval is

$$\left(\bar{X} - t^*_{1-\alpha} S, \infty\right) \tag{8.2}$$

8.2 Bootstrap Resampling

The bootstrap method, introduced by Efron [29], is a very general resampling technique for estimating a parameter without making assumptions about the distribution of the data. It gained widespread use because it is better than some other asymptotic methods.

The basic idea behind the bootstrap is: given a sample $\{X_1,\ X_2, \cdots,\ X_n\}$ of size n, a bootstrap resample of $\underline{X}$ is a sample

$$\underline{X}^* = (X_1^*,\ X_2^*, \cdots,\ X_n^*) \tag{8.3}$$

where each value X_j^* in the above equation is a random sample from $\underline{X} = (X_1,\ X_2, \cdots,\ X_n)$ with replacement, that is, given distinct values $X_1,\ X_2, \cdots,\ X_n$,

$$P\left(X_j^* = X_i\right) = \frac{1}{n}$$

with independent choices of X_j^* for $1 \le j \le n$. In particular, repeated values $X_{j_1}^* = X_{j_2}^* = X_i$ are allowed.

We repeat this process B times to form B independent resamples: ${}^1\underline{X}^*, \cdots, {}^B\underline{X}^*$. The bootstrap estimator of a statistic $\theta = \theta\left(\underline{X}\right)$ for ${}^k\underline{X}^*$ is then given by

$$\widehat{\theta}_k^* = \widehat{\theta}\left({}^k\underline{X}^*\right), \quad 1 \le k \le B$$

Thus, we can obtain the mean and variance for $\hat{\theta}$ as follow:

$$\begin{aligned}
\widehat{\theta}_{mean} &= \frac{1}{B}\sum_{k=1}^{B} \widehat{\theta}_k^* \\
\widehat{Var}(\theta) &= \frac{1}{B-1}\sum_{k=1}^{B} \left(\widehat{\theta}_k^* - \widehat{\theta}_{mean}\right)^2
\end{aligned}$$

The generalized bootstrap (GB) was introduced by Dudewicz [27]. The essence of the GB is to fit a suitable distribution to the available data, and then take samples from the fitted distribution.

Suppose given a random sample $\{X_1, X_2, \cdots, X_n\}$ from $F(\cdot)$ we want to estimate $\theta(X_1, X_2, \cdots, X_n)$. Dudewicz (1992) proposed following algorithm:

1. Estimate F by $\hat{F}$
2. Independently generate N random samples of size n from $\hat{F}$ and obtain $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_N$.
3. Estimate θ using $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_N$

Dudewicz [27] also presents following result.

Theorem 8.1. *The Bootstrap Method is a special case of the Generalized Bootstrap in which one takes $\hat{F}$ to be the empirical distribution function F_n in Step 1.*

Sun and Muller-Schwarz [106] compared the performances of the Bootstrap Method and the Generalized Bootstrap and reported two important conclusions:

1. The Generalized Bootstrap is more consistent in parameter estimation than is the Bootstrap Method.
2. The number of bootstrap samples N should be at least 500.

Lin [75] has shown that GB methods is better than bootstrap method for small to moderate sample sizes, and also gives good asymptotic performance.

Karian and Dudewicz [65] suggested fitting the Generalized Lambda Distribution, which can match any mean, any variance, skewness and kurtosis in a very broad range.

8.3 A New Resampling Method

If the random variable X has probability density $f(x)$ over a domain S, and if X is constrained to $A \in S$, the probability density of the constrained random variable is

$$f_c(x) = \begin{cases} \frac{f(x)}{Pr(x \in A)}, & x \in A \\ 0, & \text{elsewhere} \end{cases} \tag{8.4}$$

The most common types of constraints are truncations.

In this section, we develop a new resampling scheme and details are given below:

Theorem 8.2. *Suppose that X is a random variable with cdf $F(x)$. Given a random sample of size n from X with values $\underline{x} = (x_1, \cdots, x_n)$ of size n we build up $n+1$ truncated blocks, that is, $A_0 = \left\{x : x \leq x_{(1)}\right\}$, $A_i = \left\{x : x_{(i)} < x \leq x_{(i+1)}\right\}, i = 1, \cdots, n-1$ and $A_n = \left\{x : x > x_{(n)}\right\}$.*

Let $\hat{F}(x)$ be estimator of $F(x)$ and $\hat{F}_{c_i}(x)$ be estimated distribution of block $A_i, 1 \leq i \leq n$. Then a generalized bootstrap resample is a sample

$$\underline{X}^* = (X_1^*, \cdots, X_m^*)$$

where $X_j^*, 1 \leq j \leq m$ is constructed as follow:

1. *If $x_{(i)} = x_{(i+1)} = \cdots = x_{(j)}$ for some $i < j$, we take $x_{(i+k)} = x_{(i)} + kd, k = 1, 2, \cdots, (j-i)$ for some small positive number d such that $x_{(j)} + d < x_{(j+1)}$.*

2. *Randomly select a block from $A_0, A_1, \cdots, A_n$ with replacement according to prob-*

ability:

$$\begin{aligned} P(A_0) &= \hat{F}(x_{(1)}) \\ P(A_i) &= \hat{F}(x_{(i+1)}) - \hat{F}(x_{(i)}), \quad 1 \le i \le n-1 \\ P(A_n) &= 1 - \hat{F}(x_{(n)}) \end{aligned}$$

3. *Generate a sample point according to $\hat{F}_{c_i}(x)$ if block $A_i, 1 \le i \le n$, is selected in the jth random selection. Or more simply, X_j^* is taken to be mean of the selected block A_i.*

If we sample a point according to $\hat{F}_{c_i}(x)$ in step two, the generalized bootstrap resample $\underline{X}$ are actually equivalent to the sample generated from estimated distribution $\hat{F}(x)$.

Now we discuss truncated means. First consider the double truncated block $x_{(i)} < X \le x_{(i+1)}$,

$$f_{X_{T_D}}(x) = f(x|x_{(i)} < X \le x_{(i+1)}) = \begin{cases} \frac{f(x)}{F(x_{(i+1)}) - F(x_{(i)})}, & x_{(i)} < X \le x_{(i+1)} \\ 0, & \text{elsewhere} \end{cases}$$

right truncated block $X \le x_{(1)}$,

$$f_{X_{T_R}}(x) = f(x|X \le x_{(1)}) = \begin{cases} \frac{f(x)}{F(x_{(1)})}, & X \le x_{(1)} \\ 0, & \text{elsewhere} \end{cases}$$

and left truncated block $X > x_{(n)}$,

$$f_{X_{T_L}}(x) = f(x|X > x_{(n)}) = \begin{cases} \frac{f(x)}{1 - F(x_{(n)})}, & X > x_{(n)} \\ 0, & \text{elsewhere} \end{cases}$$

We first discuss the block $x_{(i)} \leq X < x_{(i+1)}$. By integral mean value theorem we have

$$\begin{aligned} E\left(X_{T_D}\right) &= \frac{\int_{x_{(i)}}^{x_{(i+1)}} x dF(x)}{F(x_{(i+1)}) - F(x_{(i)})} = \frac{x_{(i+1)}F(x_{(i+1)}) - x_{(i)}F(x_{(i)}) - \int_{x_{(i)}}^{x_{(i+1)}} F(x)dx}{F(x_{(i+1)}) - F(x_{(i)})} \\ &= \frac{x_{(i+1)}F(x_{(i+1)}) - x_{(i)}F(x_{(i)}) - [x_{(i+1)} - x_{(i)}]F(\xi_1)}{F(x_{(i+1)}) - F(x_{(i)})} \end{aligned}$$

where $x_{(i)} < \xi_1 < x_{(i+1)}$.

Now we discuss the block $X > x_{(n)}$. Since there exists a point $M_{x_{(n)}}$ with $x_{(n)} < M_{x_{(n)}} < \infty$ such that $\int_{M_{x_{(n)}}}^{\infty} x dF(x) \approx 0$, then

$$\int_{x_{(n)}}^{\infty} x dF(x) = \int_{M_{x_{(n)}}}^{\infty} x dF(x) + \int_{x_{(n)}}^{M_{x_{(n)}}} x dF(x) \approx \int_{x_{(n)}}^{M_{x_{(n)}}} x dF(x)$$

$$\begin{aligned} E\left(X_{T_L}\right) &= \frac{\int_{x_{(n)}}^{\infty} x dF(x)}{1 - F(x_{(n)})} = \frac{\int_{x_{(n)}}^{M_{x_{(n)}}} x dF(x)}{1 - F(x_{(n)})} \\ &= \frac{M_{x_{(n)}}F(M_{x_{(n)}}) - x_{(n)}F(x_{(n)}) - \int_{x_{(n)}}^{M_{x_{(n)}}} F(x)dx}{1 - F(x_{(n)})} \\ &= \frac{M_{x_{(n)}}F(M_{x_{(n)}}) - x_{(n)}F(x_{(n)}) - [M_{x_{(n)}} - x_{(n)}]F(\xi_2)}{1 - F(x_{(n)})} \end{aligned}$$

where $x_{(n)} < \xi_2 < M_{x_{(n)}}$.

Last we consider block $X \leq x_{(1)}$. Since there exist a point $M_{x_{(1)}}$ with $M_{x_{(1)}} < x_{(1)}$ such that $\int_{-\infty}^{M_{x_{(1)}}} x dF(x) \approx 0$, then

$$\int_{-\infty}^{x_{(1)}} x dF(x) = \int_{-\infty}^{M_{x_{(1)}}} x dF(x) + \int_{M_{x_{(1)}}}^{x_{(1)}} x dF(x) \approx \int_{M_{x_{(1)}}}^{x_{(1)}} x dF(x)$$

$$\begin{aligned} E\left(X_{T_R}\right) &= \frac{\int_{-\infty}^{x_{(1)}} x dF(x)}{F(x_{(1)})} = \frac{\int_{M_{x_{(1)}}}^{x_{(1)}} x dF(x)}{F(x_{(1)})} \\ &= \frac{x_{(1)}F(x_{(1)}) - M_{x_{(1)}}F(M_{x_{(1)}}) - \int_{M_{x_{(1)}}}^{x_{(1)}} F(x)dx}{F(x_{(1)})} \\ &= \frac{x_{(1)}F(x_{(1)}) - M_{x_{(1)}}F(M_{x_{(1)}}) - [x_{(1)} - M_{x_{(1)}}]F(\xi_3)}{F(x_{(1)})} \end{aligned}$$

where $M_{x_{(1)}} < \xi_3 < x_{(1)}$.

Lemma 8.3. *Suppose that X is a random variable with cdf $F(x)$. Given sample values $\underline{x} = (x_1, \cdots, x_n)$ of size n from X, we build up $n+1$ truncated blocks, that is, $A_0 = \left\{x : x \leq x_{(1)}\right\}$, $A_i = \left\{x : x_{(i)} < x \leq x_{(i+1)}\right\}, i = 1, \cdots, n-1$ and $A_n = \left\{x : x > x_{(n)}\right\}$.*

1. *For a block $A_i, i = 1, \cdots, n-1$, there exists a point ξ_{T_D} such that $x_{(i)} < \xi_{T_D} < x_{(i+1)}$, then*

$$E\left(X_{T_D}\right) = \frac{x_{(i+1)}F(x_{(i+1)}) - x_{(i)}F(x_{(i)}) - [x_{(i+1)} - x_{(i)}]F(\xi_{T_D})}{F(x_{(i+1)}) - F(x_{(i)})}$$

2. *For a block A_0, choose a point $M_{x_{(1)}}$ such that $M_{x_{(1)}} < x_{(1)}$ and $\int_{-\infty}^{M_{x_{(1)}}} x dF(x) \approx 0$, there exists a point ξ_{T_R} such that $M_{x_{(1)}} < \xi_{T_R} < x_{(1)}$, then*

$$E\left(X_{T_R}\right) = \frac{x_{(1)}F(x_{(1)}) - M_{x_{(1)}}F(M_{x_{(1)}}) - [x_{(1)} - M_{x_{(1)}}]F(\xi_{T_R})}{F(x_{(1)})}$$

3. *For a block A_n, choose a point $M_{x_{(n)}}$ such that $M_{x_{(n)}} > x_{(n)}$ and $\int_{M_{x_{(n)}}}^{\infty} x dF(x) \approx 0$, there exists a point ξ_{T_L} such that $x_{(n)} < \xi_{T_L} < M_{x_{(n)}}$, then*

$$E\left(X_{T_L}\right) = \frac{M_{x_{(n)}}F(M_{x_{(n)}}) - x_{(n)}F(x_{(n)}) - [M_{x_{(n)}} - x_{(n)}]F(\xi_{T_L})}{1 - F(x_{(n)})}$$

8.4 Another Look at the Bootstrap Method

Let $\{x_1, \cdots, x_n\}$ be the values of a random sample of size n from a population with cdf $F(t)$. Then the empirical distribution function is defined as

$$\hat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} 1_{\left\{x_{(i)} \leq t\right\}}$$

where 1_A is the indicator of event A and $x_{(1)}, \cdots, x_{(n)}$ are the order statistics.

From the previous section, by taking $F(\xi_{T_D}) = F(x_{(i)})$, $F(\xi_{T_L}) = F(x_{(n)})$, and

$F(\xi_{T_R}) = F(M_{x_{(1)}})$, we have the approximate means:

$$\begin{aligned} E\left(X_{T_D}\right) &\approx x_{(i+1)}, \quad x_{(i)} < X \leq x_{(i+1)}, \; i = 1, 2, \cdots, n-1 \\ E\left(X_{T_L}\right) &\approx M_{x_{(n)}}, \quad X > x_{(n)} \\ E\left(X_{T_R}\right) &\approx x_{(1)}, \quad X \leq x_{(1)} \end{aligned}$$

Now let $m = n$, and let each block be chosen according to empirical distribution, that is,

$$\begin{aligned} P\left(X \in A_0\right) &= \cdots = P\left(X \in A_{n-1}\right) = \frac{1}{n} \\ P\left(X \in A_n\right) &= 0 \end{aligned}$$

we can see that that for distinct values $x_1, \cdots, x_n$, a bootstrap sample point X_j^* can be thought of as sampling from the chosen truncated block,

$$X_j^* = x_i$$

with $x_i \in A_i, \; P\left(A_i\right) = \frac{1}{n}, \; j = 1, 2, \cdots, n-1, \; i = 0, 1, \cdots, n-1.$

8.5 Smoothed Resampling

Kroese ([68], pp. 41–42) states that "A disadvantage of the Bootstrap procedure is that discrete distributions are generated." Criticism of the Bootstrap method has been growing in recent years (Young [122]). Several paper have discussed replacing the discrete distribution F_n by a smoothed version of that empirical distribution.

Chernick ([17], p. 109) points out "$\cdots$ the generalized bootstrap $\cdots$ [is] a promising alternative to the $\cdots$ bootstrap since it has the advantage of taking account of the fact that the data are continuous but it does not seem to suffer the drawbacks $\cdots$."

Let $\{X_1, X_2, \cdots, X_n\}$ denote a random sample of size n from F with density

f, with F completely unspecified. A kernel function K which satisfies

$$\int x^i K(x)dx = \begin{cases} 1 & i = 0 \\ 0 & 1 \leq i \leq r-1 \\ \kappa & i = r \end{cases}$$

with $\kappa \neq 0$, for positive integer $r \geq 2$, is called an rth order kernel function.

An rth order kernel estimator $\hat{f}$ of f based on K is define by

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \tag{8.5}$$

and

$$\hat{F}_h(x) = \int_{-\infty}^{x} \hat{f}_h(t)dt$$

with h a smoothing bandwidth to be specified. In any case, there is the question of the choice of h. If it is too small, then the resulting distribution will not be very smooth, if it is too large, then the smoothed portions overlap and we loose information given in the original data.

It is well known that K necessarily takes negative values if $r \geq 3$, and so $\hat{f}(x)$ will also be negative for some values of x, and thus it is not a proper density function.

Given data $\{x_1, \cdots, x_n\}$, a smooth bootstrap can be constructed as follows:

$$\tilde{X} = x_I + h\varepsilon \tag{8.6}$$

where I is uniformly distribution on $\{1, 2, \cdots, n\}$ and ε has density function K, then $\tilde{X}$ has distribution $\hat{F}_h$. In general, we take $\varepsilon \sim N(0, 1)$, then,

$$\begin{aligned} P\left(\tilde{X} \leq x\right) &= P\left(\frac{x - X_i}{h}\right) \\ &= \frac{1}{n} \sum_{i=1}^{n} \Phi\left(\frac{x - x_i}{h}\right) \end{aligned}$$

then

$$\hat{f}_{\tilde{X}}(x) = \frac{1}{nh}\sum_{i=1}^{n}\phi\left(\frac{x-x_i}{h}\right)$$

where Φ and ϕ denote distribution and density of standard normal distribution. Thus, we take K as the Gaussian kernel.

In case of the Gaussian kernel, the rule of thumb (Silverman [102]) yields the estimator

$$\hat{h} = 1.06\sigma n^{-\frac{1}{5}}$$

where σ^2 is the variance of $\underline{X}$. Jones, Marron and Sheather [63] studied the Monte Carlo performance of the normal reference bandwidth based on the standard deviation and they suggested

$$\hat{h} = 1.06Sn^{-\frac{1}{5}}$$

where S is the sample standard deviation.

When a computing truncated mean for each block we suppose that $\{x_1, x_2, \cdots, x_n\}$ are fixed. In order to introducing randomness to each truncated block, we assume $F(\xi_i)$ is uniform on $\left[F(x_{(i)}), F(x_{(i+1)})\right]$. Then we can take mean of each $F(\xi_i)$ to be

$$\begin{aligned}
E\left[F(\xi_3)\right] &= \frac{F(x_{(1)}) + F(M_{x_{(1)}})}{2}, \quad X \le x_{(1)} \\
E\left[F(\xi_1)\right] &= \frac{F(x_{(i+1)}) + F(x_{(i)})}{2}, \quad x_{(i)} < X \le x_{(i+1)} \\
E\left[F(\xi_2)\right] &= \frac{F(x_{(n)}) + F(M_{x_{(n)}})}{2}, \quad X > x_{(n)}
\end{aligned}$$

so we have

$$\begin{aligned} x_1^T &= E\left(X_{T_R}\right) = \frac{x_{(1)} + M_{x_{(1)}}}{2}, \quad X \leq x_{(1)} \\ x_i^T &= E\left(X_{T_D}\right) = \frac{x_{(i)} + x_{(i+1)}}{2}, \quad x_{(i)} < X \leq x_{(i+1)} \\ x_n^T &= E\left(X_{T_L}\right) = \frac{x_{(n)} + M_{x_{(n)}}}{2}, \quad X > x_{(n)} \end{aligned}$$

Using a similar method, we can construct our resamples as follows:

$$\tilde{X} = x_I^T + h\varepsilon$$

where I has density such that $P(I = i) = P(A_i)$.

8.6 A Simple Version of the Generalized Bootstrap

In this section, Efron's bootstrap and generalized bootstrap are utilized to construct prediction intervals. Efron's bootstrap approach has been introduced before. Here we first describe how to use generalized bootstrap for constructing prediction intervals.

For given sample $X_1, X_2, \cdots, X_n$, the t-quantile of X, denoted by Q_t, is defined by

$$Q_t = \min\left\{Q : F_X(Q) \geq t\right\}$$

When F_X is continuous and increasing, then Q_t is simply the unique root of the equation $F_X\left(Q_t\right) = t$. The well known direct–simulation quantile estimator is

$$\widehat{Q}_t = X_{(\lceil nt \rceil)} \tag{8.7}$$

and $\lceil t \rceil$ denotes the smallest integer that is greater than or equal to t. However, it might be possible to improve by using the following estimator (Avramidis and Wilson

[6])

$$\widehat{Q}_t = \begin{cases} X_{(1)}, & t \leq \frac{0.5}{n} \\ q_n X_{(\lceil nt+0.5\rceil-1)} + (1-q_n) X_{(\lceil nt+0.5\rceil)}, & \frac{0.5}{n} < t < \frac{n-0.5}{n} \\ X_{(n)}, & t \geq \frac{n-0.5}{n} \end{cases} \tag{8.8}$$

where

$$q_n = \lceil nt + 0.5\rceil - (nt + 0.5)$$

If we take $t = \frac{1.5}{n}, \frac{2.5}{n}, \cdots, \frac{n-0.5}{n}$, then $q_n = 0$. As before we take two points $M_{x_{(1)}}$ and $M_{x_{(n)}}$ such that $M_{x_{(1)}} < x_{(1)}$, $\int_{-\infty}^{M_{x_{(1)}}} x dF(x) \approx 0$, $M_{x_{(n)}} > x_{(n)}$ and $\int_{M_{x_{(n)}}}^{\infty} x dF(x) \approx 0$, So

$$\hat{F}_n(x) \quad = \quad \begin{cases} 0, & x \leq M_{x_{(1)}} \\ \frac{0.5}{n}, & M_{x_{(1)}} < x \leq x_{(1)} \\ \frac{1.5}{n}, & x_{(1)} < x \leq x_{(2)} \\ \frac{2.5}{n}, & x_{(2)} < x \leq x_{(3)} \\ \vdots & \\ \frac{i-0.5}{n}, & x_{(i-1)} < x \leq x_{(i)} \\ \vdots & \\ \frac{n-0.5}{n}, & x_{(n-1)} < x \leq x_{(n)} \\ \frac{0.5}{n}, & x_{(n)} < x \leq M_{x_{(n)}} \\ 0, & x > M_{x_{(n)}} \end{cases}$$

In the following, we give a simplified version of a generalized bootstrap resampling scheme:

1. Select n blocks with replacement according to probabilities

$$P\left(X \in A_0\right) = P\left(X \in A_n\right) = \frac{0.5}{n},\ P\left(X \in A_1\right) = \cdots = P\left(X \in A_{n-1}\right) = \frac{1}{n}$$

2. Take the mean of the selected block as a resample point

$$\begin{aligned} X_0^T &= E\left(X_{T_L}\right) = \frac{X_{(1)} + M_{X_{(1)}}}{2}, \quad X \in A_0 \\ X_i^T &= E\left(X_{T_D}\right) = \frac{X_{(i)} + X_{(i+1)}}{2}, \quad X \in A_i,\ 1 \le i \le n-1 \\ X_n^T &= E\left(X_{T_R}\right) = \frac{X_{(n)} + M_{X_{(n)}}}{2}, \quad X \in A_n \end{aligned}$$

3. Smooth a resample point (optional)

$$\tilde{X} = X_i^T + \hat{h}\varepsilon$$

where we take

$$\hat{h} = 1.06 n^{-\frac{1}{5}} \frac{1}{n+1} S$$

which is one-n^{th} of bandwidth obtained by Jones, Marron and Sheather [63], since their width is for the whole part distribution, but our is only for one-$(n+1)^{th}$ portion of the distribution.

In Step 3, we actually need to simulate a two-sided truncated normal variable. Robert [100] proposes an accept-reject algorithm for two-sided truncated normal distributions. Let $N([\mu, \mu^-, \mu^+], \sigma^2)$ denotes a truncated normal distribution with mean μ, variance σ^2, left truncation point μ^- and right truncation point μ^+. Without loss of generality take $\mu = 0$ and $\sigma^2 = 1$, The accept-reject algorithm based on uniform $U[\mu^-, \mu^+]$ is

1. Generate $z \sim U[\mu^-, \mu^+]$

2. Compute

$$\varrho(z) = \begin{cases} e^{-\frac{z^2}{2}} & \text{if } 0 \in U[\mu^-, \mu^+] \\ e^{\frac{(\mu^+)^2 - z^2}{2}} & \text{if } \mu^+ < 0 \\ e^{\frac{(\mu^-)^2 - z^2}{2}} & \text{if } 0 < \mu^- \end{cases}$$

3. Generate $u \sim U[0,1]$ and take $x = z$ if $u \leq \varrho(z)$; otherwise, return to Step 1.

Assume that

$$\begin{aligned} Y_0 &\sim N\left(0, M_{x_{(1)}} - x_0^T, x_{(1)} - x_0^T, 1\right) \\ Y_i &\sim N\left(0, x_{(i)} - x_i^T, x_{(i+1)} - x_i^T, 1\right), \quad 1 \leq i \leq n-1 \\ Y_n &\sim N\left(0, x_{(n)} - x_n^T, M_{x_{(n)}} - x_n^T, 1\right) \end{aligned}$$

then

$$\begin{aligned} \tilde{X} &= x_0^T + hY_0, \quad if\ X \in A_0 \\ \tilde{X} &= x_i^T + hY_i, \quad if\ X \in A_i,\ 1 \leq i \leq n-1 \\ \tilde{X} &= x_n^T + hY_n, \quad if\ X \in A_n \end{aligned}$$

In our case, we always have $0 \in U[\mu^-, \mu^+]$, so we only need to compute $\varrho(z) = e^{-\frac{z^2}{2}}$ in Step 2 of the accept-reject algorithm.

Now we turn to how to determine $M_{x_{(1)}}$ and $M_{x_{(n)}}$ by using outlier detection technique. Johnson [62] defines an outlier as an observation that appears to be inconsistent with other observations in the data set. An outlier has a low probability that it originates from the same statistical distribution as the rest observations in the data set.

A two-sided Grubb's [46] test is often used to evaluate observations, coming from a normal distribution of size n, which are suspiciously far from the main body

of the data. For a two-sided Grubb's test, the test statistic is defined as:

$$G = \frac{\max |X_i - \bar{X}|}{s}$$

with $\bar{X}$ and S denoting the sample mean and standard deviation, respectively, calculated with the suspected outlier included. The critical value of the Grubb's test is calculated as

$$C = \frac{n-1}{\sqrt{n}} \sqrt{\frac{t^2_{(\alpha/2n,n-2)}}{n-2+t^2_{(\alpha/2n,n-2)}}}$$

where $t_{(\alpha/2n,n-2)}$ denotes the upper critical value of the t-distribution with $(n-2)$ degrees of freedom and a significance level of $\alpha/(2n)$. If $G \geq C$, then the suspected observation is confirmed as an outlier at significance level α.

The Grubbs test can also be defined as a one-sided test. To test whether the minimum value is an outlier, the test statistic, based on the data $\{x_1, \cdots, x_n\}$ is

$$G_1 = \frac{\bar{x} - x_{(1)}}{s}$$

To test whether the maximum value is an outlier, the test statistic is

$$G_n = \frac{x_{(n)} - \bar{x}}{s}.$$

Then we take

$$M_{x_{(1)}} = \begin{cases} \bar{x} - Cs & \text{if } G_1 < C \\ \bar{x} - (G_1 + k)s & \text{else} \end{cases}$$

and

$$M_{x_{(n)}} = \begin{cases} \bar{x} + Cs & \text{if } G_n < C \\ \bar{x} + (G_n + k)s & \text{else} \end{cases}$$

where $k > 0$.

8.7 Numerical Examples

At beginning of this chapter, we introduced the bootstrap method to construct prediction intervals. We gave a simple version of a generalized bootstrap(Gbootstrap) and a smoothed generalized bootstrap(SGbootstrap). Since bootstrap techniques can be used to reduce bias, or to reduce variance of estimators, we conduct a comparative study of three bootstrap methods for prediction intervals.

Suppose that $\{X_1, X_2, \cdots, X_n\}$ is a random sample and Y is a future observation with same distribution. As in Bai and Olshen [7], we consider a pivotal

$$T^* = \frac{\bar{X}^* - Y^*}{S^*}$$

where $\{X_1^*, \cdots, X_n^*\}$ and Y^* be random samples of sizes n and 1, respectively, drawn with replacement from $\{X_1, \cdots, X_n\}$. When computing T^*, Bai and Olshen [7] draw Y^*, denoted by Y^*_{efron} from $\{X_1, \cdots, X_n\}$ with equally probabilities such that

$$P\left(Y^* = X_{(1)}\right) = \cdots = P\left(Y^* = X_{(n)}\right) = \frac{1}{n}$$

Instead, we generate Y^*, denoted by $Y^*_{Gbootstrap}$, in the Gbootstrap method according to probabilities:

$$\begin{aligned} P\left(Y^* = X_0^T\right) &= P\left(Y^* = X_n^T\right) = \frac{0.5}{n}, \\ P\left(Y^* = X_1^T\right) &= \cdots = P\left(Y^* = X_{n-1}^T\right) = \frac{1}{n}, \end{aligned}$$

where $X_0^T, X_1^T, \cdots, X_n^T$ are the midpoints of blocks $A_0, \cdots, A_n$. For the SGbootstrap method, we take

$$Y^*_{SGbootstrap} = Y^*_{Gbootstrap} + h\varepsilon \tag{8.9}$$

Then the procedure for estimating the percentiles of T^* is as follows:

1. Generate a generalized bootstrap resample $X_1^*, \cdots, X_n^*$ and Y^* from $X_1, \cdots, X_n$

2. Calculate the statistic:

$$\hat{T}^* = \frac{\bar{X}^* - Y^*}{S^*}$$

3. Repeat Step 1 and Step 2 for B_q times.

4. Estimate a q-percentile of T^* by using $\hat{T}_1^*, \cdots, \hat{T}_{B_q}^*$: $\widehat{Q}_q^*$.

Then a $100(1-\alpha)\%$ two-sided prediction interval for Y is

$$[L(\underline{X}), U(\underline{X})] = \left[\bar{X} - \widehat{Q}^*_{1-\frac{\alpha}{2}} S, \bar{X} - \widehat{Q}^*_{\frac{\alpha}{2}} S\right] \tag{8.10}$$

and the length of the prediction interval is given by

$$L_{PI} = \left(\widehat{Q}^*_{1-\frac{\alpha}{2}} - \widehat{Q}^*_{\frac{\alpha}{2}}\right) S$$

We now introduce Bootstrap aggregating (bagging), which is a useful technique to improve performance. In the following, we describe how to apply bagging to improve confidence intervals:

1. Generate B bootstrap resamples of $\underline{X}$

$$^{(k)}\underline{X}^* = \left(^{(k)}X_1^*,\ ^{(k)}X_2^*, \cdots,\ ^{(k)}X_n^*\right), \quad k = 1, 2, \cdots, B$$

2. Calculate a $100(1-\alpha)\%$ two-sided confidence interval of Y for each $^{(k)}\underline{X}^*$

$$\left[L^{(k)}(\underline{X}), U^{(k)}(\underline{X})\right] = \left[\bar{X} -^{(k)} \widehat{Q}^*_{1-\frac{\alpha}{2}} S, \bar{X} -^{(k)} \widehat{Q}^*_{\frac{\alpha}{2}} S\right]$$

3. Aggregate the bootstrap confidence intervals and obtain a $100(1-\alpha)\%$ two-sided confidence interval for Y based on $\underline{X}$, $[L(\underline{X}), U(\underline{X})]$, such that

$$\begin{aligned} L(\underline{X}) &= \frac{1}{B}\sum_{k=1}^{B} L^{(k)}(\underline{X}) \\ U(\underline{X}) &= \frac{1}{B}\sum_{k=1}^{B} U^{(k)}(\underline{X}) \end{aligned}$$

Example 8.1. A simulation study was undertaken to examine the behaviour of the prediction intervals given by the generalized bootstrap and smoothed generalized bootstrap. We take sample sizes $n = 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50$ from a Gumbel minimum distribution with location $a = 5$ and scale $b = 2$. For each n, 5000 independent random samples were generated and the coverage probabilities and average length of prediction intervals were calculated with corresponding to probabilities 0.90, 0.95 and 0.99.

From the tables below, we can see that generalized bootstrap performs better in both coverage probability and average length of prediction intervals than bootstrap when sample size is less than 20.

When the sample size n increases, as we expected, the performances of the bootstrap and generalized bootstrap are close to each other. Since n is large, we have more truncated blocks, that is, the blocks are becoming narrower, then the left end-point of each block is very close to middle point of that block.

	GB		*Smooth GB*		*Bai & Olshen*	
n	*Length*	*CP*	*Length*	*CP*	*Length*	*CP*
5	11.3475	0.8682	11.5032	0.8728	13.2256	0.8696
6	10.0866	0.8740	10.1842	0.8784	11.0004	0.8730
7	9.4903	0.8732	9.5562	0.8780	9.9792	0.8722
8	9.3913	0.8742	9.4406	0.8768	9.6134	0.8744
9	9.1930	0.8770	9.2290	0.8788	9.2661	0.8764
10	9.0599	0.8782	9.0894	0.8788	9.0991	0.8790
15	8.9760	0.8880	8.9870	0.8890	9.0866	0.8906
20	8.8344	0.8948	8.8373	0.8946	8.8466	0.8908
25	8.8136	0.9078	8.8164	0.9082	8.7563	0.9030
30	8.7498	0.8974	8.7506	0.8972	8.6761	0.8908
40	8.6604	0.9052	8.6614	0.9048	8.5784	0.9032

Table 8.1: 90% two-sided PIs for Y with 5000 trials: $a = 5, b = 2$

	GB		*Smooth GB*		*Bai & Olshen*	
n	*Length*	*CP*	*Length*	*CP*	*Length*	*CP*
5	16.4923	0.9196	16.7382	0.9232	20.5931	0.9204
6	13.7661	0.9248	13.9095	0.9282	15.4415	0.9226
7	12.4805	0.9200	12.5636	0.9224	13.7139	0.9198
8	12.0027	0.9146	12.0609	0.9164	13.0394	0.9200
9	11.5112	0.9188	11.5513	0.9202	12.2503	0.9204
10	11.1680	0.9168	11.2024	0.9188	11.7927	0.9192
15	10.7322	0.9258	10.7484	0.9264	11.0521	0.9262
20	10.5037	0.9354	10.5075	0.9348	10.6794	0.9360
25	10.4837	0.9442	10.4868	0.9450	10.6192	0.9476
30	10.4702	0.9410	10.4738	0.9408	10.5885	0.9406
40	10.4724	0.9438	10.4742	0.9434	10.4939	0.9440

Table 8.2: 95% two-sided PIs for Y with 5000 trials: $a = 5$ $b = 2$

	GB		*Smooth GB*		*Bai & Olshen*	
n	*Length*	*CP*	*Length*	*CP*	*Length*	*CP*
5	NaN	0.8464	39.5644	0.9722	NaN	0.8292
6	NaN	0.9662	26.5094	0.9738	NaN	0.9628
7	21.5106	0.9674	21.6681	0.9682	24.7572	0.9682
8	19.4535	0.9636	19.5482	0.9640	21.3650	0.9648
9	17.7814	0.9646	17.8468	0.9646	19.1830	0.9636
10	16.7017	0.9618	16.7579	0.9636	17.7256	0.9638
15	14.7914	0.9642	14.8090	0.9646	15.5453	0.9660
20	13.9381	0.9712	13.9494	0.9708	14.4390	0.9690
25	13.6076	0.9744	13.6134	0.9744	14.0220	0.9774
30	13.4626	0.9702	13.4672	0.9706	13.8512	0.9716
40	13.3695	0.9740	13.3718	0.9742	13.7088	0.9748

Table 8.3: 99% two-sided PIs for Y with 5000 trials: $a = 5, b = 2$

Example 8.2. We conducted leave-one-out experiments for CANDU data (Fang [35]) below, that is, use all-but-one observation in each group data as past observations and the left-out observation as future observation.

CANDU feeder wall thickness Group 1 data:

```
3.97, 4.00, 4.06, 4.12, 4.12, 4.15, 4.21, 4.39, 4.42, 4.42,
4.47, 4.47, 4.57, 4.62, 4.62, 4.74, 4.74, 4.80, 4.86, 4.86,
4.89, 4.92, 5.12
```

CANDU feeder wall thickness Group 2 data:

```
4.12, 4.38, 4.38, 4.39, 4.40, 4.41, 4.42, 4.42, 4.56, 4.62,
4.74, 4.80, 4.86, 5.10, 5.57, 5.73
```

CANDU feeder wall thickness Group 3 data:

```
4.21,  4.27,  4.27,  4.27,  4.30,  4.30,  4.39,  4.39,  4.40,  4.42,
4.45,  4.45,  4.45,  4.50,  4.50,  4.50,  4.52,  4.54,  4.56,  4.57,
4.59,  4.62
```

CANDU feeder wall thickness Group 4 data:

```
4.24,  4.26,  4.48,  4.59,  4.59,  4.62,  4.62,  4.74,  4.86,  4.89,
5.01,  5.01,  5.07,  5.19
```

Tables 8.4–8.15 show the various prediction intervals and their coverage probabilities for methods of generalized bootstrap (GB), smoothing generalized bootstrap and bootstrap with different prescribed probabilities: 0.90, 0.95 and 0.99.

Both generalized bootstrap and smoothed generalized bootstrap have shorter prediction interval lengths than the bootstrap.

	GB		*Smooth GB*		*Bai & Olshen*	
Y	Length	CP	Length	CP	Length	CP
4.00	1.0594	0.7730	1.0597	0.7850	1.0607	0.7890
4.06	1.0764	1.0000	1.0772	1.0000	1.0795	1.0000
4.12	1.0907	1.0000	1.0913	1.0000	1.0885	1.0000
4.12	1.0888	1.0000	1.0894	1.0000	1.0891	1.0000
4.15	1.0915	1.0000	1.0914	1.0000	1.0917	1.0000
4.21	1.0951	1.0000	1.0961	1.0000	1.0923	1.0000
4.39	1.0928	1.0000	1.0933	1.0000	1.0934	1.0000
4.42	1.0916	1.0000	1.0916	1.0000	1.0929	1.0000
4.42	1.0903	1.0000	1.0914	1.0000	1.0922	1.0000
4.47	1.0908	1.0000	1.0917	1.0000	1.0928	1.0000
4.47	1.0908	1.0000	1.0915	1.0000	1.0931	1.0000
4.57	1.0935	1.0000	1.0939	1.0000	1.0943	1.0000
4.62	1.0893	1.0000	1.0895	1.0000	1.0947	1.0000
4.62	1.0918	1.0000	1.0924	1.0000	1.0939	1.0000
4.74	1.0930	1.0000	1.0936	1.0000	1.0947	1.0000
4.74	1.0919	1.0000	1.0921	1.0000	1.0913	1.0000
4.80	1.0896	1.0000	1.0904	1.0000	1.0931	1.0000
4.86	1.0905	1.0000	1.0910	1.0000	1.0880	1.0000
4.86	1.0901	1.0000	1.0902	1.0000	1.0879	1.0000
4.89	1.0842	1.0000	1.0842	1.0000	1.0851	1.0000
4.92	1.0730	1.0000	1.0739	1.0000	1.0737	1.0000
5.12	0.9896	0	0.9901	0	0.9873	0

Table 8.4: 90% two-sided PIs for Y for CANDU group 1 data with 1000 trials

	GB		*Smooth GB*		*Bai & Olshen*	
Y	Length	CP	Length	CP	Length	CP
3.97	1.1812	0.8910	1.1817	0.8830	1.1915	0.8960
4.00	1.1985	0.9990	1.1995	0.9990	1.2143	0.9990
4.06	1.2181	1.0000	1.2184	1.0000	1.2310	1.0000
4.12	1.2291	1.0000	1.2303	1.0000	1.2399	1.0000
4.12	1.2304	1.0000	1.2308	1.0000	1.2421	1.0000
4.15	1.2313	1.0000	1.2321	1.0000	1.2422	1.0000
4.21	1.2333	1.0000	1.2341	1.0000	1.2431	1.0000
4.39	1.2295	1.0000	1.2303	1.0000	1.2424	1.0000
4.42	1.2270	1.0000	1.2279	1.0000	1.2375	1.0000
4.42	1.2249	1.0000	1.2258	1.0000	1.2400	1.0000
4.47	1.2248	1.0000	1.2256	1.0000	1.2374	1.0000
4.47	1.2269	1.0000	1.2276	1.0000	1.2382	1.0000
4.57	1.2280	1.0000	1.2289	1.0000	1.2368	1.0000
4.62	1.2245	1.0000	1.2249	1.0000	1.2377	1.0000
4.62	1.2252	1.0000	1.2261	1.0000	1.2362	1.0000
4.74	1.2283	1.0000	1.2293	1.0000	1.2379	1.0000
4.74	1.2269	1.0000	1.2276	1.0000	1.2329	1.0000
4.80	1.2255	1.0000	1.2264	1.0000	1.2365	1.0000
4.86	1.2270	1.0000	1.2271	1.0000	1.2337	1.0000
4.86	1.2282	1.0000	1.2287	1.0000	1.2363	1.0000
4.89	1.2240	1.0000	1.2247	1.0000	1.2351	1.0000
4.92	1.2150	1.0000	1.2159	1.0000	1.2283	1.0000
5.12	1.0905	0	1.0917	0	1.0864	0

Table 8.5: 95% two-sided PIs for Y for CANDU group 1 data with 1000 trials

	GB		*Smooth GB*		*Bai & Olshen*	
Y	Length	CP	Length	CP	Length	CP
3.97	1.4287	1.0000	1.4296	1.0000	1.4465	1.0000
4.00	1.4545	1.0000	1.4555	1.0000	1.4735	1.0000
4.06	1.4815	1.0000	1.4823	1.0000	1.4941	1.0000
4.12	1.4913	1.0000	1.4912	1.0000	1.5029	1.0000
4.12	1.4960	1.0000	1.4971	1.0000	1.5031	1.0000
4.15	1.4926	1.0000	1.4940	1.0000	1.5069	1.0000
4.21	1.4939	1.0000	1.4945	1.0000	1.5041	1.0000
4.39	1.4843	1.0000	1.4850	1.0000	1.4989	1.0000
4.42	1.4836	1.0000	1.4848	1.0000	1.4908	1.0000
4.42	1.4824	1.0000	1.4823	1.0000	1.4928	1.0000
4.47	1.4783	1.0000	1.4794	1.0000	1.4898	1.0000
4.47	1.4791	1.0000	1.4798	1.0000	1.4901	1.0000
4.57	1.4766	1.0000	1.4781	1.0000	1.4913	1.0000
4.62	1.4743	1.0000	1.4755	1.0000	1.4922	1.0000
4.62	1.4818	1.0000	1.4828	1.0000	1.4876	1.0000
4.74	1.4818	1.0000	1.4837	1.0000	1.4918	1.0000
4.74	1.4790	1.0000	1.4799	1.0000	1.4857	1.0000
4.80	1.4854	1.0000	1.4874	1.0000	1.4992	1.0000
4.86	1.4838	1.0000	1.4853	1.0000	1.4971	1.0000
4.86	1.4886	1.0000	1.4891	1.0000	1.4961	1.0000
4.89	1.4781	1.0000	1.4791	1.0000	1.4963	1.0000
4.92	1.4724	1.0000	1.4737	1.0000	1.4919	1.0000
5.12	1.2954	0.1210	1.2959	0.1340	1.2898	0.1210

Table 8.6: 99% two-sided PIs for Y for CANDU group 1 data with 1000 trials

	GB		*Smooth GB*		*Bai & Olshen*	
Y	Length	CP	Length	CP	Length	CP
4.12	1.5254	0	1.5288	0	1.5454	0
4.38	1.6475	1.0000	1.6484	1.0000	1.6799	1.0000
4.38	1.6419	1.0000	1.6432	1.0000	1.6790	1.0000
4.39	1.6484	1.0000	1.6502	1.0000	1.6744	1.0000
4.40	1.6484	1.0000	1.6507	1.0000	1.6807	1.0000
4.41	1.6563	1.0000	1.6582	1.0000	1.6759	1.0000
4.42	1.6561	1.0000	1.6581	1.0000	1.6785	1.0000
4.42	1.6515	1.0000	1.6532	1.0000	1.6901	1.0000
4.56	1.6471	1.0000	1.6492	1.0000	1.6869	1.0000
4.62	1.6522	1.0000	1.6543	1.0000	1.6890	1.0000
4.74	1.6626	1.0000	1.6630	1.0000	1.7009	1.0000
4.80	1.6627	1.0000	1.6644	1.0000	1.7027	1.0000
4.86	1.6704	1.0000	1.6716	1.0000	1.7070	1.0000
5.10	1.6702	1.0000	1.6715	1.0000	1.7030	1.0000
5.57	1.4766	0.7040	1.4787	0.7030	1.5169	0.7670
5.73	1.3485	0.0180	1.3496	0.0170	1.3841	0.0220

Table 8.7: 90% two-sided PIs for Y for CANDU group 2 data with 1000 trials

	GB		*Smooth GB*		*Bai & Olshen*	
Y	Length	CP	Length	CP	Length	CP
4.12	1.8331	0	1.8362	0	1.8534	0
4.38	1.9950	1.0000	1.9971	1.0000	2.0325	1.0000
4.38	1.9875	1.0000	1.9889	1.0000	2.0192	1.0000
4.39	1.9910	1.0000	1.9921	1.0000	2.0260	1.0000
4.40	1.9936	1.0000	1.9971	1.0000	2.0306	1.0000
4.41	2.0070	1.0000	2.0083	1.0000	2.0206	1.0000
4.42	2.0009	1.0000	2.0026	1.0000	2.0233	1.0000
4.42	1.9950	1.0000	1.9978	1.0000	2.0418	1.0000
4.56	1.9968	1.0000	1.9986	1.0000	2.0352	1.0000
4.62	2.0077	1.0000	2.0087	1.0000	2.0364	1.0000
4.74	2.0297	1.0000	2.0318	1.0000	2.0661	1.0000
4.80	2.0366	1.0000	2.0384	1.0000	2.0757	1.0000
4.86	2.0511	1.0000	2.0533	1.0000	2.0952	1.0000
5.10	2.0871	1.0000	2.0886	1.0000	2.1269	1.0000
5.57	1.8976	0.9960	1.8992	0.9960	1.9906	0.9980
5.73	1.6919	0.5970	1.6924	0.5920	1.7652	0.6510

Table 8.8: 95% two-sided PIs for Y for CANDU group 2 data with 1000 trials

	GB		*Smooth GB*		*Bai & Olshen*	
Y	Length	CP	Length	CP	Length	CP
4.12	2.7943	0.0670	2.7994	0.0850	2.8907	0.0380
4.38	2.9148	1.0000	2.9216	1.0000	3.0281	1.0000
4.38	2.9195	1.0000	2.9201	1.0000	3.0055	1.0000
4.39	2.9106	1.0000	2.9139	1.0000	3.0315	1.0000
4.40	2.9432	1.0000	2.9457	1.0000	3.0408	1.0000
4.41	2.9571	1.0000	2.9601	1.0000	3.0164	1.0000
4.42	2.9121	1.0000	2.9150	1.0000	3.0111	1.0000
4.42	2.9209	1.0000	2.9258	1.0000	3.0435	1.0000
4.56	2.9410	1.0000	2.9428	1.0000	3.0525	1.0000
4.62	2.9660	1.0000	2.9716	1.0000	3.0430	1.0000
4.74	3.0690	1.0000	3.0721	1.0000	3.1934	1.0000
4.80	3.1265	1.0000	3.1298	1.0000	3.2426	1.0000
4.86	3.1999	1.0000	3.1993	1.0000	3.3633	1.0000
5.10	3.3539	1.0000	3.3587	1.0000	3.6337	1.0000
5.57	2.8941	1.0000	2.8950	1.0000	3.1219	1.0000
5.73	2.4941	0.9960	2.4959	0.9970	2.6399	1.0000

Table 8.9: 99% two-sided PIs for Y for CANDU group 2 data with 1000 trials

	GB		*Smooth GB*		*Bai & Olshen*	
Y	Length	CP	Length	CP	Length	CP
4.21	0.3669	0	0.3670	0	0.3666	0
4.27	0.3917	1.0000	0.3920	1.0000	0.3919	1.0000
4.27	0.3925	1.0000	0.3926	1.0000	0.3926	1.0000
4.27	0.3921	1.0000	0.3922	1.0000	0.3917	1.0000
4.30	0.3960	1.0000	0.3963	1.0000	0.3959	1.0000
4.30	0.3961	1.0000	0.3965	1.0000	0.3964	1.0000
4.39	0.3955	1.0000	0.3958	1.0000	0.3959	1.0000
4.39	0.3955	1.0000	0.3957	1.0000	0.3958	1.0000
4.40	0.3954	1.0000	0.3956	1.0000	0.3965	1.0000
4.42	0.3954	1.0000	0.3957	1.0000	0.3955	1.0000
4.45	0.3945	1.0000	0.3947	1.0000	0.3952	1.0000
4.45	0.3950	1.0000	0.3951	1.0000	0.3953	1.0000
4.45	0.3953	1.0000	0.3956	1.0000	0.3957	1.0000
4.50	0.3954	1.0000	0.3957	1.0000	0.3950	1.0000
4.50	0.3946	1.0000	0.3950	1.0000	0.3955	1.0000
4.50	0.3956	1.0000	0.3958	1.0000	0.3952	1.0000
4.52	0.3957	1.0000	0.3959	1.0000	0.3950	1.0000
4.54	0.3954	1.0000	0.3957	1.0000	0.3949	1.0000
4.56	0.3930	1.0000	0.3932	1.0000	0.3941	1.0000
4.57	0.3912	1.0000	0.3912	1.0000	0.3920	1.0000
4.59	0.3835	0.9320	0.3835	0.9390	0.3848	0.9170
4.62	0.3701	0	0.3704	0	0.3700	0

Table 8.10: 90% two-sided PIs for Y for CANDU group 3 data with 1000 trials

	GB		*Smooth GB*		*Bai & Olshen*	
Y	Length	CP	Length	CP	Length	CP
4.21	0.4081	0.0930	0.4085	0.0970	0.4082	0.0910
4.27	0.4408	1.0000	0.4412	1.0000	0.4426	1.0000
4.27	0.4417	1.0000	0.4419	1.0000	0.4442	1.0000
4.27	0.4420	1.0000	0.4424	1.0000	0.4426	1.0000
4.30	0.4440	1.0000	0.4441	1.0000	0.4455	1.0000
4.30	0.4450	1.0000	0.4452	1.0000	0.4464	1.0000
4.39	0.4430	1.0000	0.4432	1.0000	0.4438	1.0000
4.39	0.4419	1.0000	0.4422	1.0000	0.4436	1.0000
4.40	0.4421	1.0000	0.4420	1.0000	0.4437	1.0000
4.42	0.4424	1.0000	0.4425	1.0000	0.4431	1.0000
4.45	0.4413	1.0000	0.4415	1.0000	0.4427	1.0000
4.45	0.4419	1.0000	0.4422	1.0000	0.4429	1.0000
4.45	0.4413	1.0000	0.4417	1.0000	0.4430	1.0000
4.50	0.4418	1.0000	0.4419	1.0000	0.4425	1.0000
4.50	0.4415	1.0000	0.4416	1.0000	0.4435	1.0000
4.50	0.4422	1.0000	0.4425	1.0000	0.4429	1.0000
4.52	0.4418	1.0000	0.4420	1.0000	0.4427	1.0000
4.54	0.4420	1.0000	0.4421	1.0000	0.4425	1.0000
4.56	0.4412	1.0000	0.4413	1.0000	0.4425	1.0000
4.57	0.4396	1.0000	0.4401	1.0000	0.4419	1.0000
4.59	0.4323	1.0000	0.4326	1.0000	0.4366	0.9990
4.62	0.4124	0.0660	0.4127	0.0730	0.4134	0.0530

Table 8.11: 95% two-sided PIs for Y for CANDU group 3 data with 1000 trials

	GB		*Smooth GB*		*Bai & Olshen*	
Y	Length	CP	Length	CP	Length	CP
4.21	0.4902	0.9430	0.4907	0.9490	0.4905	0.9520
4.27	0.5389	1.0000	0.5394	1.0000	0.5402	1.0000
4.27	0.5364	1.0000	0.5369	1.0000	0.5413	1.0000
4.27	0.5397	1.0000	0.5403	1.0000	0.5410	1.0000
4.30	0.5403	1.0000	0.5410	1.0000	0.5429	1.0000
4.30	0.5398	1.0000	0.5402	1.0000	0.5443	1.0000
4.39	0.5367	1.0000	0.5372	1.0000	0.5388	1.0000
4.39	0.5359	1.0000	0.5365	1.0000	0.5366	1.0000
4.40	0.5350	1.0000	0.5353	1.0000	0.5388	1.0000
4.42	0.5352	1.0000	0.5354	1.0000	0.5348	1.0000
4.45	0.5334	1.0000	0.5341	1.0000	0.5350	1.0000
4.45	0.5342	1.0000	0.5346	1.0000	0.5338	1.0000
4.45	0.5326	1.0000	0.5331	1.0000	0.5362	1.0000
4.50	0.5343	1.0000	0.5345	1.0000	0.5355	1.0000
4.50	0.5325	1.0000	0.5331	1.0000	0.5357	1.0000
4.50	0.5342	1.0000	0.5347	1.0000	0.5359	1.0000
4.52	0.5342	1.0000	0.5348	1.0000	0.5346	1.0000
4.54	0.5344	1.0000	0.5350	1.0000	0.5350	1.0000
4.56	0.5341	1.0000	0.5346	1.0000	0.5356	1.0000
4.57	0.5319	1.0000	0.5325	1.0000	0.5354	1.0000
4.59	0.5264	1.0000	0.5268	1.0000	0.5315	1.0000
4.62	0.4997	0.9370	0.5004	0.9460	0.5006	0.8920

Table 8.12: 99% two-sided PIs for Y for CANDU group 3 data with 1000 trials

	GB		*Smooth GB*		*Bai & Olshen*	
Y	Length	CP	Length	CP	Length	CP
4.24	0.9382	0.1490	0.9386	0.1430	0.9516	0.2960
4.26	0.9533	0.4920	0.9543	0.4910	0.9654	0.6050
4.48	1.0390	1.0000	1.0404	1.0000	1.0532	1.0000
4.59	1.0419	1.0000	1.0435	1.0000	1.0504	1.0000
4.59	1.0441	1.0000	1.0457	1.0000	1.0516	1.0000
4.62	1.0442	1.0000	1.0453	1.0000	1.0501	1.0000
4.62	1.0418	1.0000	1.0431	1.0000	1.0487	1.0000
4.74	1.0443	1.0000	1.0464	1.0000	1.0503	1.0000
4.86	1.0448	1.0000	1.0463	1.0000	1.0522	1.0000
4.89	1.0475	1.0000	1.0485	1.0000	1.0531	1.0000
5.01	1.0440	1.0000	1.0446	1.0000	1.0491	1.0000
5.01	1.0430	1.0000	1.0439	1.0000	1.0487	1.0000
5.07	1.0238	1.0000	1.0238	1.0000	1.0388	1.0000
5.19	0.9454	0.0030	0.9464	0.0030	0.9503	0.0020

Table 8.13: 90% two-sided PIs for Y for CANDU group 4 data with 1000 trials

	GB		*Smooth GB*		*Bai & Olshen*	
Y	Length	CP	Length	CP	Length	CP
4.24	1.0890	0.8870	1.0905	0.8850	1.1154	0.9470
4.26	1.1092	0.9890	1.1104	0.9840	1.1353	0.9950
4.48	1.1948	1.0000	1.1969	1.0000	1.2113	1.0000
4.59	1.1932	1.0000	1.1951	1.0000	1.2041	1.0000
4.59	1.1979	1.0000	1.2000	1.0000	1.2061	1.0000
4.62	1.1942	1.0000	1.1957	1.0000	1.2002	1.0000
4.62	1.1911	1.0000	1.1934	1.0000	1.2015	1.0000
4.74	1.1947	1.0000	1.1962	1.0000	1.1970	1.0000
4.86	1.1925	1.0000	1.1953	1.0000	1.2030	1.0000
4.89	1.1977	1.0000	1.1997	1.0000	1.2029	1.0000
5.01	1.1977	1.0000	1.1996	1.0000	1.2059	1.0000
5.01	1.1970	1.0000	1.1985	1.0000	1.2048	1.0000
5.07	1.1796	1.0000	1.1806	1.0000	1.1969	1.0000
5.19	1.0794	0.3380	1.0805	0.3310	1.0844	0.2910

Table 8.14: 95% two-sided PIs for Y for CANDU group 4 data with 1000 trials

	GB		*Smooth GB*		*Bai & Olshen*	
Y	Length	CP	Length	CP	Length	CP
4.24	1.3984	1.0000	1.4006	1.0000	1.4303	1.0000
4.26	1.4204	1.0000	1.4235	1.0000	1.4608	1.0000
4.48	1.5422	1.0000	1.5455	1.0000	1.5646	1.0000
4.59	1.5234	1.0000	1.5277	1.0000	1.5429	1.0000
4.59	1.5322	1.0000	1.5353	1.0000	1.5471	1.0000
4.62	1.5260	1.0000	1.5300	1.0000	1.5356	1.0000
4.62	1.5158	1.0000	1.5195	1.0000	1.5380	1.0000
4.74	1.5166	1.0000	1.5189	1.0000	1.5196	1.0000
4.86	1.5230	1.0000	1.5264	1.0000	1.5321	1.0000
4.89	1.5290	1.0000	1.5327	1.0000	1.5347	1.0000
5.01	1.5350	1.0000	1.5377	1.0000	1.5460	1.0000
5.01	1.5286	1.0000	1.5315	1.0000	1.5447	1.0000
5.07	1.5133	1.0000	1.5156	1.0000	1.5368	1.0000
5.19	1.3796	0.9890	1.3829	0.9860	1.3847	0.9780

Table 8.15: 99% two-sided PIs for Y for CANDU group 4 data with 1000 trials

8.8 Wilks' Formula for Symmetric Distributions in the GRS Method

Wilks' Formula [119] is a non-parametric procedure in which an upper percentile is estimated by the maximum of a sample, although other order statistics may also be used. Although this formula is an old result, it has found an application in determining the number of code runs required to achieved a specified confidence level in estimating a percentile, and is known in the nuclear industry as the GRS method (Glaeser [42]). Its importance in the nuclear industry lies in the reduction of the number of simulations that are used, each requiring a long time to complete.

Before explaining the innovation described in this thesis, we first describe the

method. Let $F(x)$ denote the distribution function of a random variable X and for specified $0 < \gamma < 1$ let x_γ be the 100γ percentile of the distribution of X, that is $F(x_\gamma) = \gamma$. For purposes of uniqueness assume further that F can be prescribed by a probability density function f.

Given a random sample $\{X_1, \ldots, X_n\}$ from F, let $\{X_{(1)}, \ldots, X_{(n)}\}$ denote the increasing order statistics. Starting with

$$\begin{aligned} P[X_{(n)} > x_\gamma] &= 1 - P[X_{(n)} \leq x_\gamma] \\ &= 1 - P[X_1 \leq x_\gamma, \ldots, X_n \leq x_\gamma] \\ &= 1 - F(x_\gamma)^n \\ &= 1 - \gamma^n \end{aligned}$$

set $1 - \gamma^n = 1 - \alpha$ to give $X_{(n)}$ as an upper $100(1-\alpha)\%$ confidence limit for x_γ and solve for the required sample size

$$n = \frac{\log \alpha}{\log \gamma} \tag{8.11}$$

In nuclear applications, the values $\alpha = 0.05, \gamma = 0.95$ are typically used and then $n = 59$.

There are circumstances when the density function f is symmetric, for instance in Popov [94] who considers the distribution of maximum centreline temperature. His plot shows a distribution that is symmetric. For symmetric densities, Hinkley [55], in a completely unrelated context, has shown that the symmetrized empirical distribution function provides a more efficient estimator of location. This suggests the possibility of using a symmetrized distribution function to more efficiently estimate x_γ, or equivalently, to decrease the sample size n needed while maintaining the same confidence $(1-\alpha)100\%$.

The purpose of this section is to introduce and explore this idea. The main

result is that the sample size can be reduced by roughly a factor of 2 in the presence of symmetry.

8.8.1 Symmetrization

Assume that the distribution F is symmetric about its median m. Then for any i the random variable X_i has the same distribution as $2m - X_i$ and thus $X_{(n)} \equiv \max_i\{X_i\}$ has the same distribution as $\max_i\{2m - X_i\}$ that is $2m - \min_i\{X_i\} \equiv 2m - X_{(1)}$. As a result, both $X_{(n)}$ and $2m - X_{(1)}$ contain information about x_γ. Therefore the random variable

$$T = \max\{X_{(n)}, 2m - X_{(1)}\} \tag{8.12}$$

should be a more efficient estimator of x_γ.

For large n Gumbel [47] has shown that the maximum and minimum of a sample are approximately independent, under mild conditions on the density f. Assuming independence

$$\begin{aligned} P[T > x_\gamma] &= 1 - P[T \leq x_\gamma] \\ &= 1 - P[X_{(n)} \leq x_\gamma, 2a - X_{(1)} \leq \gamma] \\ &= 1 - F(x_\gamma)^n F(x_\gamma)^n \\ &= 1 - \gamma^{2n} \end{aligned}$$

Solving for n in $1 - \gamma^{2n} = 1 - \alpha$ we get

$$n = \frac{\log \alpha}{2 \log \gamma} \tag{8.13}$$

showing that the required sample size has been reduced by a factor of 2 when the median is known.

A random variable X with density $f(x)$ is symmetric if X and $-X$ have the

same distribution, that is

$$f(x) = f(-x) \tag{8.14}$$

Analogously, X is symmetric with respect to some location parameter a, if $X - a$ and $a - X$ have the same distribution such that

$$f(a + x) = f(a - x) \tag{8.15}$$

When the median m is unknown, it can be estimated by the sample median X_{med}. In the symmetric case rather than use $X_{(n)}$ to determine the sample size, we may use $W = \max\{X_{(n)}, 2X_{med} - X_{(1)}\}$ and determine n such that $P[W > x_\gamma] = 1 - P[T \leq x_\gamma] = 1 - \gamma^{2n}$ If $\{X_{(n)}, 2X_{med} - X_{(1)}\}$ are independent then the sample size can be reduced by a factor of one-half. However, the independence only applies exactly if X_{med} is the true median. Based on Gumbel's asymptotic result, for finite n, we expect only approximate independence and therefore the improvement should be less than one-half.

For $\alpha = 0.05, \gamma = 0.95$ we have carried out extensive simulations and have found that for a variety of symmetric distributions including normal, t, Laplace, logistic, uniform, that $n = 34$ gives coverage at least 0.95, with the exception of the uniform. Table 8.16 shows the correlations and Figures 8.1 - 8.7 are scatterplots. What is interesting is the variety of shapes, despite correlations being close to zero.

The last column of Table 8.17 shows the confidence levels achieved using only a sample size of $n = 34$ for symmetric distributions.

	Correlation		*Correlation*	
Distribution	$X_{(1)}$	$X_{(n)}$	$2X_{med} - X_{(1)}$	$X_{(n)}$
Norm(0,1)	1.0000	0.0194	1.0000	0.2267
	0.0194	1.0000	0.2267	1.0000
standard	1.0000	0.0000	1.0000	-0.0000
Cauchy	0.0000	1.0000	-0.0000	1.0000
Laplace	1.0000	0.0167	1.0000	-0.0010
	0.0167	1.0000	-0.0010	1.0000
student	1.0000	0.0145	1.0000	0.0260
df=5	0.0145	1.0000	0.0260	1.0000
student	1.0000	0.0192	1.0000	0.0542
df=5	0.0192	1.0000	0.0542	1.0000
Lagistic(0,5)	1.0000	0.0184	1.0000	0.0503
	0.0184	1.0000	0.0503	1.0000
Uniform(-1,1)	1.0000	0.0291	1.0000	0.1721
	0.0291	1.0000	0.1721	1.0000

Table 8.16: Correlation with 1000000 trials

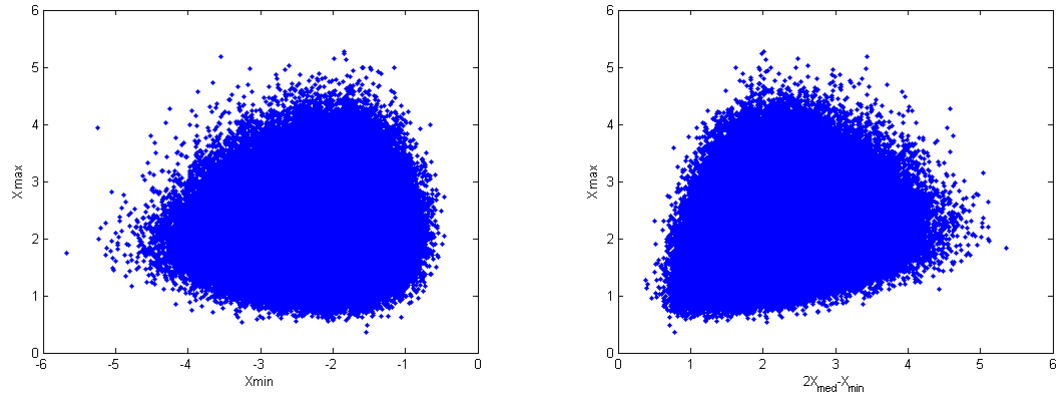


Figure 8.1: Scatter plot of standard normal distribution

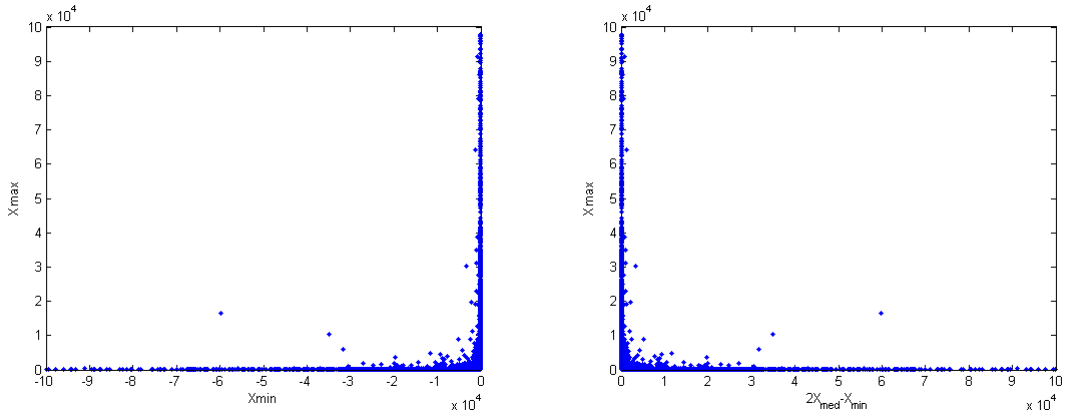


Figure 8.2: Scatter plot of standard Cauchy distribution

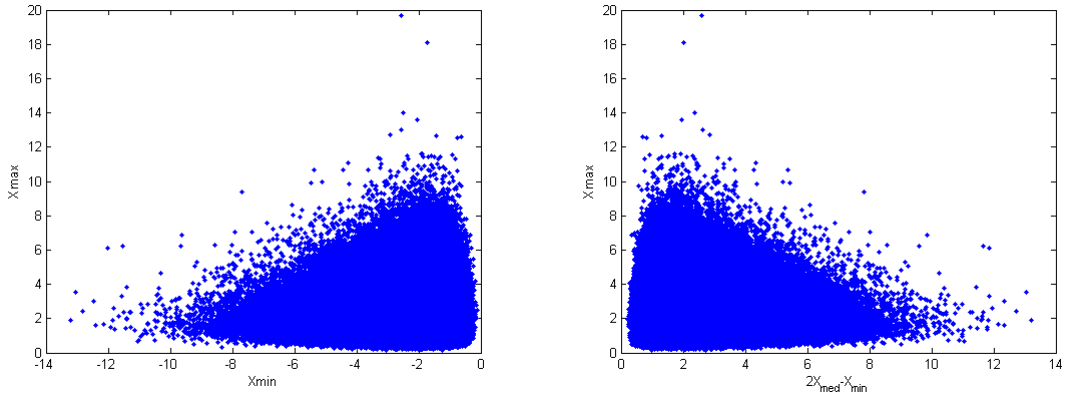


Figure 8.3: Scatter plot of Laplace distribution

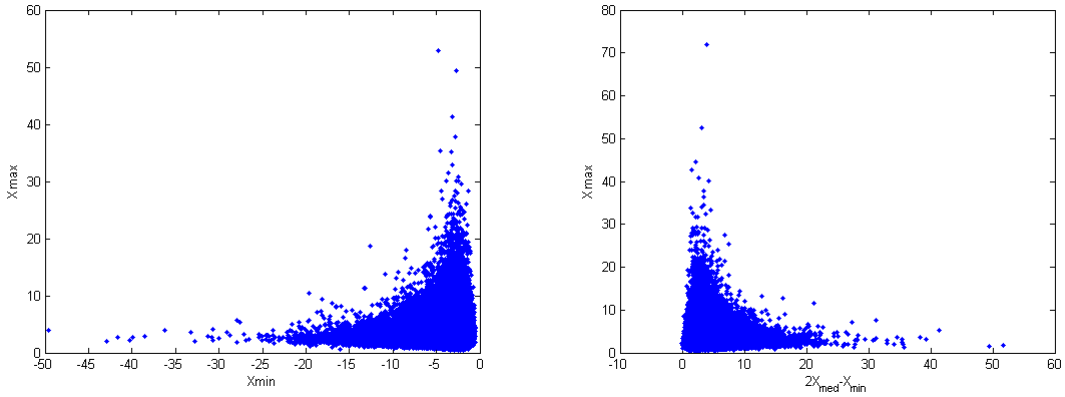


Figure 8.4: Scatter plot of Student distribution with df=5

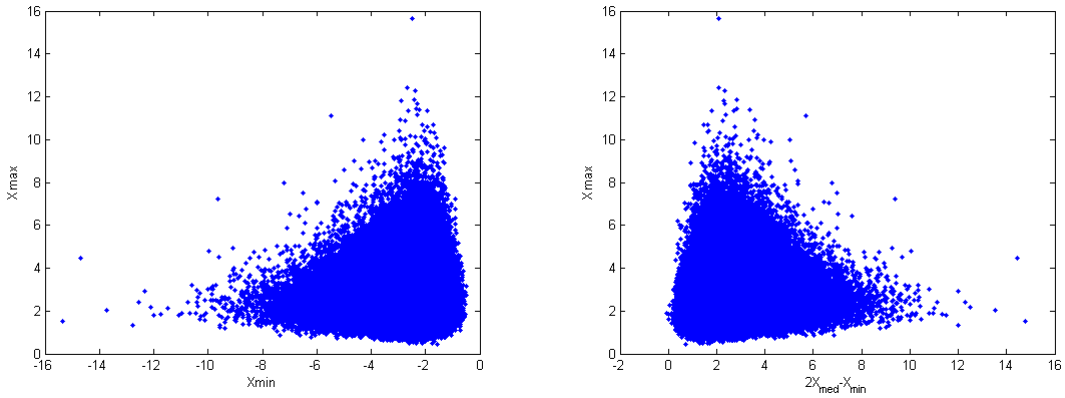


Figure 8.5: Scatter plot of Student distribution with df=10

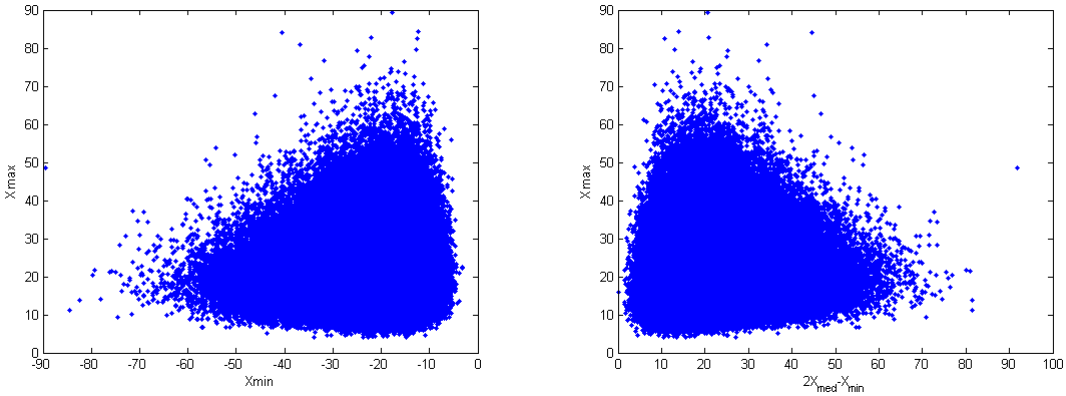


Figure 8.6: Scatter plot of logistic distribution

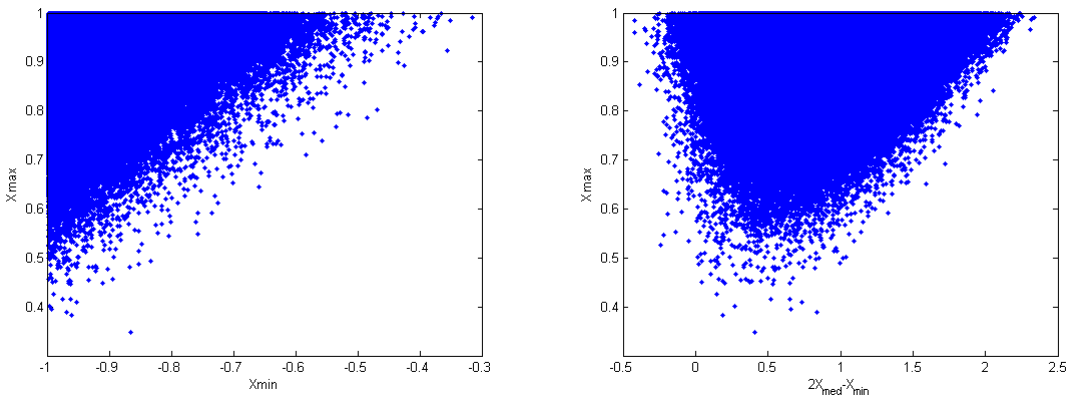


Figure 8.7: Scatter plot of uniform distribution

8.8.2 Comparison with the Generalized Bootstrap

Wilks' Formula provides a useful setting to examine the behaviour of the generalized bootstrap proposed in the previous section for estimating the probability $P(W > \chi_\gamma)$. The detail algorithm is given below:

1. Use $S_i = 2X_{med} - X_i$ to symmetrize $X_1, \cdots, X_n$ to $\underline{S} = \{X_1, \cdots, X_n, S_1, \cdots, S_n\}$ if n is even or to $\underline{S} = \{X_1, \cdots, X_n, S_1, \cdots, S_{n-1}\}$ with i not equal to middle number of n if n is odd.

2. Generate a generalized smoothing bootstrap sample with any length based on $\underline{S}$, then find estimator of χ_γ^B by using the empirical cdf. Repeat this step H times and take mean of χ_γ^B as our final estimator of χ_γ.

3. Generate another generalized smoothing bootstrap sample $x_1^B, x_2^B, \cdots, x_n^B$ based on $\underline{S}$ and define

$$w_i = \max\left[X_{max}^B, 2X_{med}^B - X_{min}^B\right]$$

   Repeat this step N times to get $w = \{w_1, w_2, \ldots, w_N\}$. Estimate $P(W > \chi_\gamma)$ by using the empirical cdf.

4. Repeat step 2–3 M times and take mean as our final estimator of $P(W > \chi_\gamma)$

	$\hat{P}(W > \chi_\gamma)$	
Distribution	Smoothed GB	Exact
normal(0,1)	0.9474	0.9485
Cauchy(0,1)	0.9626	0.9710
t-distribution (df=5)	0.9564	0.9607
t-distribution (df=10)	0.9528	0.9571
Laplace(0,1)	0.9609	0.9696
Logistic(0,5)	0.9560	0.9544
Uniform(-1,1)	0.9229	0.9005

Table 8.17: Improvement Using Wilks' Formula for Symmetric Distributions $n = 34$

Table 8.17 compares the confidence levels using the generalized bootstrap with the exact values based on simulation knowing the distribution. The results suggest that generalized bootstrap method can provide a precise estimator for $P(W > \chi_\gamma)$ when the distribution is unknown.

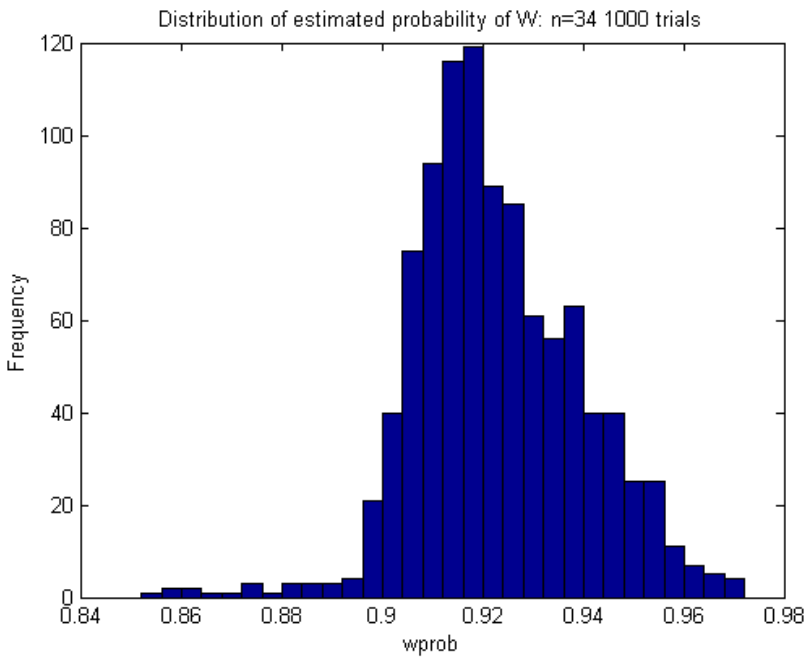


Figure 8.8: Estimated Probability of $P(W > \chi_\gamma)$ with $P(X_1 \leq \chi_\gamma) = 0.95$ and $X_1 \sim Uniform(-1, 1)$

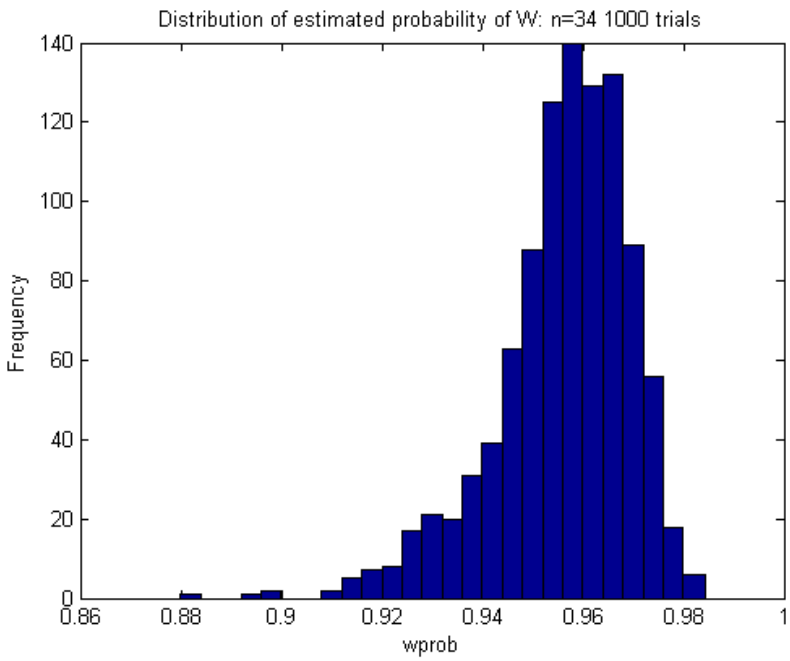


Figure 8.9: Estimated Probability of $P(W > \chi_\gamma)$ with $P(X_1 \leq \chi_\gamma) = 0.95$ and $X_1 \sim$ *logistic distribution with* $a = 0$ $b = 5$

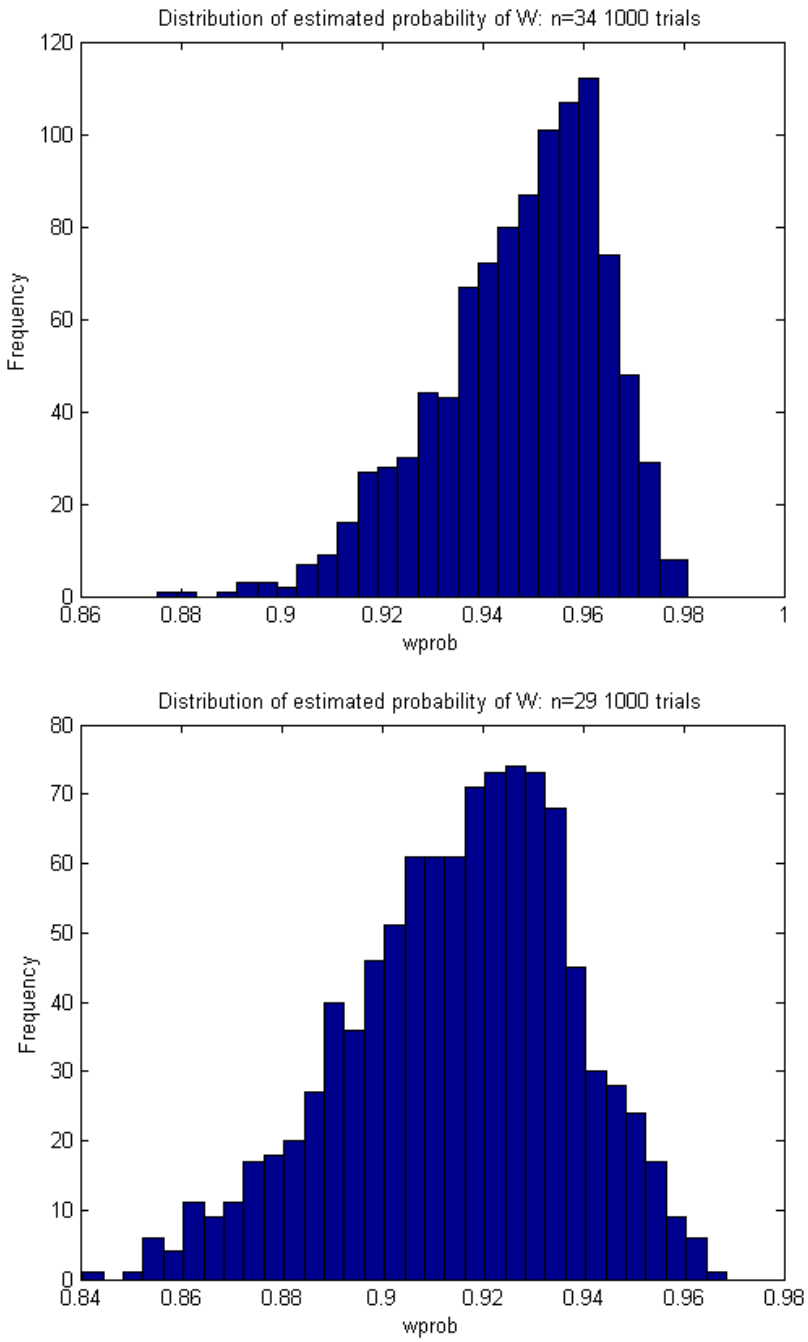


Figure 8.10: Estimated Probability of $P(W > \chi_\gamma)$ with $P(X_1 \leq \chi_\gamma) = 0.95$ and $X_1 \sim Norm(0,1)$

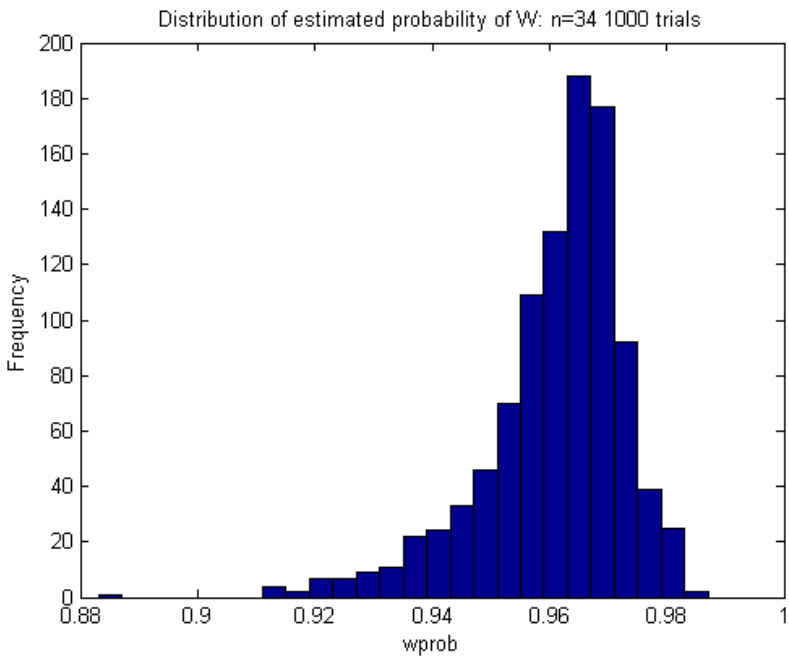


Figure 8.11: Estimated Probability of $P(W > \chi_\gamma)$ with $P(X_1 \leq \chi_\gamma) = 0.95$ and $X_1 \sim Laplace(0, 1)$

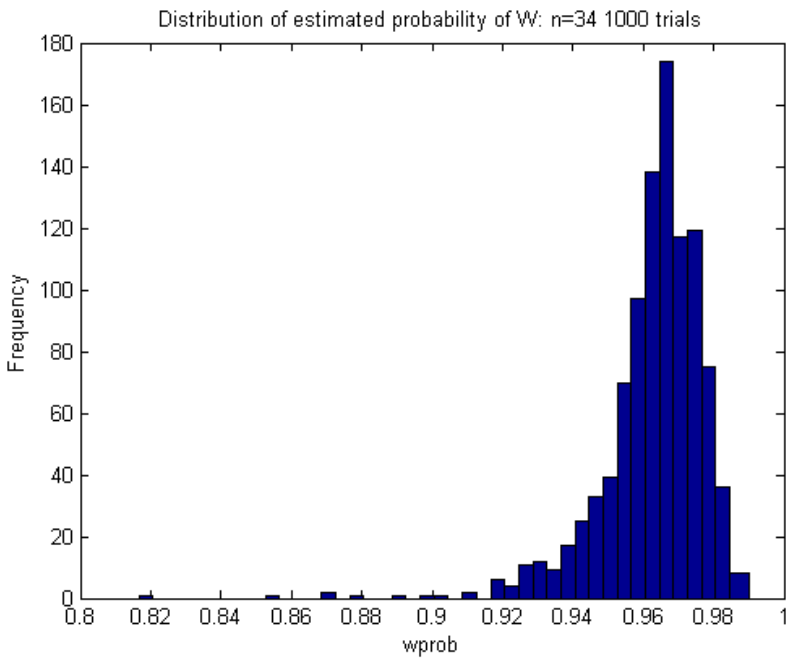


Figure 8.12: Estimated Probability of $P(W > \chi_\gamma)$ with $P(X_1 \leq \chi_\gamma) = 0.95$ and $X_1 \sim t - distribution\ with\ df = 1$

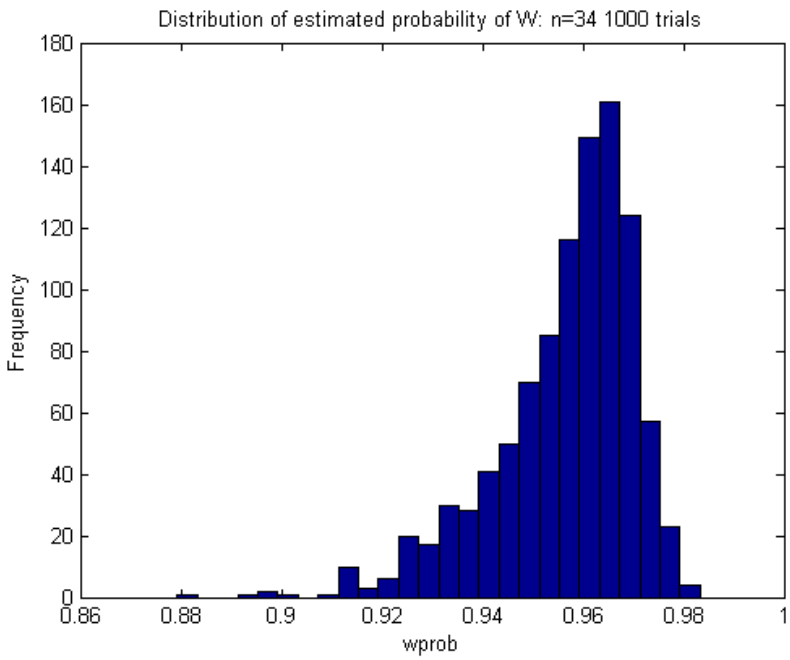


Figure 8.13: Estimated Probability of $P(W > \chi_\gamma)$ with $P(X_1 \leq \chi_\gamma) = 0.95$ and $X_1 \sim t - distribution\ with\ df = 5$

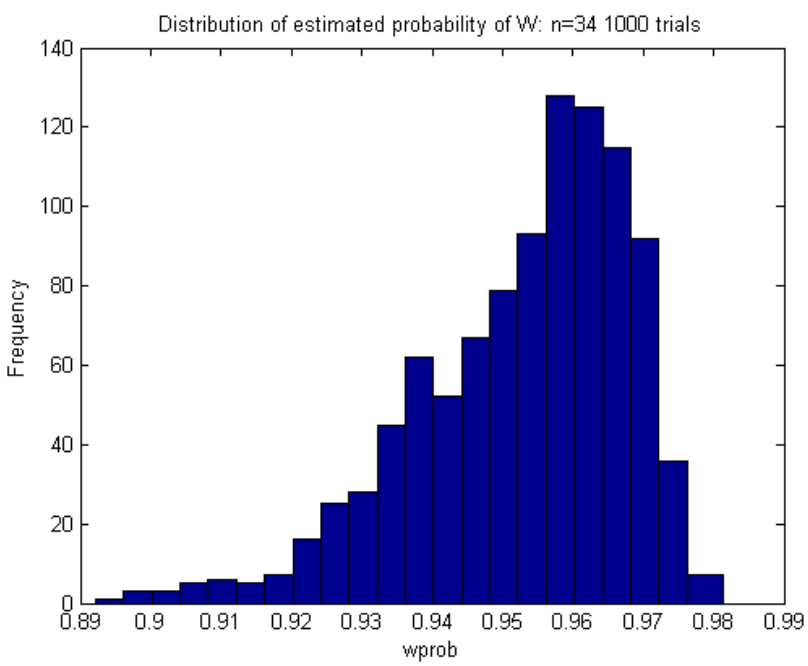


Figure 8.14: Estimated Probability of $P(W > \chi_\gamma)$ with $P(X_1 \leq \chi_\gamma) = 0.95$ and $X_1 \sim t - distribution\ with\ df = 10$

Appendix A

Useful Formulas

Proposition A.1. *Gradshteyn p.338*

$$\int_{-\infty}^{\infty} e^{ax} e^{-e^x} dx = \Gamma(a), \quad Re\ a > 0$$

Proposition A.2. *Gradshteyn p.573, 577*

$$\begin{aligned} \int_0^{\infty} x^n e^{-kx} (\ln x) dx &= \frac{n!}{k^{n+1}} \left[1 + \frac{1}{2} + \frac{1}{3} + \cdots \frac{1}{n} - \gamma - \ln k \right], \quad Re(k) > 0 \\ \int_0^{\infty} x^{v-1} (\ln x)^n e^{-ux} dx &= \frac{\partial^n}{\partial v^n} \left[u^{-v} \Gamma(v) \right], \quad v \geq 1, u > 0, n = 0, 1, 2, \cdots \end{aligned}$$

where γ denotes Euler constant and $Re(k)$ denotes The real part of the complex number k

Proposition A.3. *Gradshteyn p.340*

$$\int_0^{\infty} x^n e^{-ax} dx = n! a^{-n-1}, \quad Re\ a > 0$$

Proposition A.4. *Gradshteyn p.558*

$$\int_0^1 x^{u-1} \ln(1-x) dx = \frac{1}{u} \left[\psi(1+u) + \gamma \right], \quad Re\ u > -1$$

where $\psi(x) = \frac{d}{dx} \ln \Gamma(x)$

Proposition A.5. *Gradshteyn p.337*

$$\int_0^{\infty} x^m e^{-\beta x^n} dx = \frac{\Gamma(\gamma)}{n \beta^{\gamma}}, \quad \gamma = \frac{m+1}{n},\ m > 0,\ n > 0,\ \beta > 0$$

Proposition A.6. *Gradshteyn p.337*

$$\int_{-\infty}^{\infty} e^{-px^2 \pm qx} dx = \frac{\sqrt{\pi}}{p} e^{\frac{q^2}{4p^2}}, \quad Rep^2 > 0$$

Proposition A.7. *Gradshteyn p.337*

$$\int_0^\infty (x-a)e^{-\beta(x-b)^n}dx = \frac{\Gamma[\frac{2}{n},\beta(-b)^n]}{n\beta^{2/n}} - (a-b)\frac{\Gamma[\frac{1}{n},\beta(-b)^n]}{n\beta^{1/n}}, \quad n>0,\ \beta>0,\ |\arg b|<\pi$$

Proposition A.8. *(Gradshteyn p.573, 577)*

$$\begin{aligned}\int_0^\infty x^{v-1}e^{-ux}\ln x dx &= \frac{1}{u^v}\left[\Gamma'(v)-\Gamma(v)\log u\right], \quad v>0, u>0\\ \int_0^\infty x^{v-1}e^{-ux}(\ln x)^n dx &= \frac{\partial^n}{\partial v^n}\left[u^{-v}\Gamma(v)\right], \quad n=0,1,2,\cdots,\end{aligned}$$

Proposition A.9. *(Gradshteyn p.335)*

$$\begin{aligned}\int_{-\infty}^\infty \frac{e^{-\mu x}}{(1+e^{-x})^\nu}dx &= B\left(\mu,\nu-\mu\right), \quad 0<Re(\mu)<Re(\nu)\\ \int_{-\infty}^\infty \frac{e^{-\mu x}}{(e^{\beta/\gamma}+e^{-x/\gamma})^\nu}dx &= \gamma e^{\beta\left(\mu-\frac{\nu}{\gamma}\right)}B\left(\gamma\mu,\nu-\gamma\mu\right),\ Re\left(\frac{\nu}{\gamma}\right)>Re(\mu)>0,\ |Im(\beta)|<\pi Re(\gamma)\end{aligned}$$

where $B(x,y)=\int_0^1 t^{x-1}(1-t)^{y-1}dt$ *is Beta function.*

Proposition A.10. *(Gradshteyn p.325)*
For $a>0,\ ac>b^2$

$$\int_{-\infty}^\infty \frac{1}{(ax^2+2bx+c)^n}dx = \frac{(2n-3)!!\pi a^{n-1}}{(2n-2)!!(ac-b^2)^{n-0.5}}$$

where $(2n)!!=2\cdot 4\cdots(2n)$, $(2n+1)!!=1\cdot 3\cdots(2n+1)$, $0!!=1$ *and* $(-1)!!=1$.

Appendix B

Bibliography

[1] Aitchison, J., and Dunsmore, I. R., *Statistical Prediction Analysis*, Cambridge Univ. Press, London, New York, Melbourne *(1975)*

[2] Al-Aboud, Fahad M., *Bayesian estimations for the extreme value distribution using progressive censored data and asymmetric loss*, International Mathematical Forum, *4, no. 33, 1603–622 (2009)*

[3] Alexandersson, H., *A statistical analysis of wind, wind profiles and gust ratios at Granby, Uppsala*, Univ. of Uppsala, *Report No 55 (1979)*

[4] Anderson, T. W. and Sclove, S.L. *Introductory Statistical Analysis*, Houghton Mifflin Co., Boston *(1974)*

[5] Antle, C. E. and Rademaker, F., *An upper confidence limit on the maximum of future observations from a Type 1 extreme-value distribution*, Biometrika, *59, 475–477 (1972)*

[6] Avramidis, Athanassios N., and Wilson, James R. , *Correlation-induction techniques for estimating quantiles in simulation experiments*, Operations Research, *Vol. 46, No. 4, 574-591 (1998)*

[7] Bai, C., and Olshen, R. , *Discussion of "Theoretical comparison of bootstrap confidence intervals" by Peter Hall*, Ann. Statist., *16, 953-956(1988).*

[8] Barndorff-Nielsen, O.E. and Cox, David R., *Prediction and asymptotics*, Bernoulli, *Vol 2, Number 4, 319–340 (1996)*

[9] Beran, R., *Calibrating prediction regions*, Journal of the American Statistical Association, *Vol. 85, No. 411, 715–723(1990)*

[10] Bergstrom, H., *Distribution of extreme wind speed*, Wind energy report WE 92:2, Dep. of Meteorology, Univ. of Uppsala *(1992)*

[11] Borel, Emil, *Probability and Life, Dover, translated from the original, Les Probabilite et la Vie, 1943*, Presses Universitaire de France (1962)

[12] Casella,G. and Berger, R. L. *Statistical Inference*, Duxbury, 2nd Edition *(2002)*

[13] Chen, T. and Hoppe, F. M. *Simultaneous Confidence Intervals*, Encyclopedia of Biostatistics *(2005)*. doi: 10.1002/0470011815.b2a15147

[14] Chen, Zhenmin, *A simple method for estimating parameters of the location scale distribution family*, Journal of Statistical Computation and Simulation, *Vol. 81, No. 1, 49–58 (2011)*

[15] Cheng, R.C.H., and Amin, N.A.K., *Estimating parameters in continuous univariate distributions with a shifted origin*, Journal of the Royal Statistical Society, Series B, *45 (3), 394-403 (1983)*

[16] Cheng, R. C. H. and Traylor, L., *Non-regular maximum likelihood problems (with discussions)*, J. Roy. Statist. Soc. Ser. B, *57, 3–44 (1995)*

[17] Chernick, M.R., *Bootstrap Methods: A Guide for Practitioners and Researchers*, 2nd Edition. Wiley, New York *(2007)*

[18] Choi, J. and Srivastava, H.M., *Sums associated with the zeta function*, J. Math. Anal. Appl. *206, 103–120 (1997)*

[19] Choi, J. and Srivastava, H. M., *Evaluation of higher-order derivatives of the gamma function*, Ser. Mat.,*11, 9–18(2000)*

[20] Corsini, G., Gini, F., Greco, M.V. and Verrazzani, L., *Cramer-Rao bounds and estimation of the parameters of the Gumbel distribution* Aerospace and Electronic Systems, IEEE Transactions on, *Vol 31, Issue 3, 1202–1204 (1995)*

[21] Cramer,H.,*Mathematical Method of Statistics*, Princeton University Press, Princeton *(1958)*

[22] Davenport, A.G., *The dependence of wind loads on meteorological parameters*, Proc. Int. Res. Seminar, Wind Effects on Buildings and Structures, Ottawa, Univ. of Toronto Press, *19–82 (1967)*

[23] David, H. A., and Nagaraja, H. N., *Order Statistics* , Wiley, New Jersey, 3rd Edition *(2003)*

[24] Debashis,K,, *Construction of prediction intervals in location–scale families*, Biometrical Journal, *Vol 38, Issue 7, 865–878 (1996)*

[25] DeGroot, M. H., *Optimal Statistical Decisions*, McGraw-Hill Book Company, NewYork *(1970)*

[26] Downton, F., *Linear estimates of parameters in the extreme value distribution*, Technometrics, *Vol. 8, No. 1, 3–17 (1966)*

[27] Dudewicz , E. J., *The generalized bootstrap . In Bootstrapping and Related Techniques, Proceedings, Trier, FRG. (K. H. Jockel , G , Rothe , and W. Sendler , editors)*, Lecture Notes in Economics and Mathematical Systems , *Vol. 376, 31–37, Springer–Verlag , Berlin(1992)*

[28] Dunsmore, I. R., *The Bayesian predictive distribution in life testing models*, Technometrics, *vol.16, 455–460 (1974)*

[29] Efron, B., *Bootstrap methods: another look at the jackknife*, Ann. Statist. *7, 126 (1979)*

[30] Engelhardt, M., and Bain L. J., *Some results on point estimation for two-parameter Weibull or extreme-value distributions*, Technometrics, *16, 49–56 (1974)*

[31] Engelhardt, M., and Bain L. J., *Simplified statistical procedures for the Weibull or extreme-value distribution*, Technometrics, *19, 323–331 (1977)*

[32] Engelhardt, M., and Bain L. J., *Prediction limits and two-sample problems with complete censored Weibull data*, Technometrics, *21, 233–237 (1979)*

[33] Engelund, S. and Rackwitz, R., *On Predictive distribution functions for the three asymptotic extreme value distributions*, Structural Safety, *Vol. 11, 255–258 (1992)*

[34] Escobar, L.A. and Meeker,W.Q., *Statistical prediction based on censored life data*, Technometrics, *41, 113-124 (1999)*

[35] Fang, L., *Comparison of Prediction Intervals for the Gumbel Distribution*, M.Sc. thesis, McMaster University (2007)

[36] Fertig, K. W., Mayer, M. and Mann, N. R., *On constructing prediction intervals for samples from a Weibull or extreme value distribution*, Technometrics, *22, 567–573 (1980)*

[37] Fink, D., *A Compendium of Conjugate Priors*, Technical report, Environmental Statistical group, Department of Biology, Montana State University, USA *(1997)*

[38] Fiorentino, M. and Gabriele, S., *A correction for the bias of maximum-likelihood estimators of Gumbel parameters*, Journal of Hydrology, *Volume 73, Issue 1, p. 39–49(1984)*

[39] Fonseca, G., Giummole, F. and Vidoni, P., *Calibrating predictive distributions*, Journal of Statistical Computation and Simulation, *iFirst 2012, 1-11, DOI:10.1080/00949655.2012.709518*

[40] Fraser, D. A. S., *On fiducial inference*, Annals of Mathematical Statistics, *32, 661–676 (1961)*

[41] Geisser, Seymour, *Predictive Inference: An Introduction*, CRC Press (1993)

[42] Glaeser, H., *GRS method for uncertainty and sensitivity evaluation of code results and applications*, Sci. Tech. Nucl. Install. Article ID 798901, 7 pages, doi:10.1155/2008/798901(2008).

[43] Goodman, L.A., *A simple method for improving some estimators*, Ann. Math. Statist., *Vol 24, 114–117 (1953)*

[44] Gradshteyn, I.S. and Ryzhik, I.M., *Table of Integrals, Series, and Products*, Academic Press, Boston, 7nd Edition *(2007)*

[45] Greenwood, J.A., Landwehr, J.M., Matalas, N.C. and Wallis, J.R., *Probability-weighted moments: definition and relation to parameters of several distributions expressable in inverse form*, Water Resources Research, *15, 1049-1054(1979)*

[46] Grubbs, F., *Procedures for detecting outlying observations in samples*, Technometrics, *11(1), 1-21 (1969)*

[47] Gumbel, E. J., *On the independence of extremes in a sample*, Ann. Mathem. Statist., *17 No.1, 78–81.(1946)*

[48] Gumbel, E. J., *Statistics of Extremes*, Columbia Univ. Press,New York *(1958)*

[49] Guo, S. L., *A discussion on unbiased plotting positions for generalized extreme value distribution*, J. Hydrol., *121, 33-44 (1990)*

[50] Hahn, G.J., *Factors for Calculating Two-Sided Prediction Intervals for Samples from a Normal Distribution*, Journal of the American Statistical Association, *Volume 64, Issue 327 (1969)*

[51] Hahn, G.J., *Additional Factors for Calculating Prediction Intervals for Samples from a Normal Distribution*, Journal of the American Statistical Association, *Volume 65, Issue 332 (1970)*

[52] Hahn, G.J., and Meeker, W.Q., *Statistical Intervals: A Guide for Practitioners*, Wiley, NewYork *(1991)*

[53] Hall, P., Peng, L. and Tajvidi, N., *On prediction intervals based on predictive likelihood or bootstrap methods*, Biometrika, *86, pp. 871-880 (1999)*

[54] Hassanein, KM, *Estimation of the parameters of the extreme–value distribution by use of two or three order statistics*, Biometrika, *56, 2, 429–436 (1969)*

[55] Hinkley, D., *On estimating a symmetric distribution*, Biometrika, *3, 680–681 (1976)*

[56] Hosking, J. R. M., J. R. Wallis, and Wood, E. F., *Estimation of the generalized extreme-value distribution by the method of probability weighted moments*, Technometrics, *27(3), 251-261(1985)*

[57] Hosking, J. R. M., *L-moments: analysis and estimation of distributions using linear combinations of order statistics*, Journal of the Royal Statistical Society, *Series B, 52, 105–124(1990)*

[58] Hoppe, F. M. and Fang, L., *Bayesian prediction for the Gumbel distribution applied to feeder pipe thicknesses*, Proceedings of the 16th International Conference on Nuclear Engineering. Orlando FL, *May 11–15 (2008)*

[59] Hsieh, H. K., *Prediction intervals for Weibull order statistics*, Statistica Sinica *7: 1039–1051 (1997)*

[60] Jeffrey, A. and Dai, H.H., *Handbook of Mathematical Formulas and Integrals*, Academic Press, Boston, 4nd Edition *(2008)*

[61] Jenkinson, A.F., *The frequency distribution of the annual maximum (or minimum) values of meteorological elements*, Quarterly Journal of the Royal Meteorological Society, *81, 158–171(1955)*

[62] Johnson R., *Applied Multivariate Statistical Analysis*, Prentice Hall *(1992)*

[63] Jones, M. C., Marron, J. S., and Sheather, S. J., *A brief survey of bandwidth selection*, Journal of the American Statistical Association, *91, 401-407 (1996)*

[64] Jowitt, P.W., *The extreme-value type-1 distribution and the principle of maximum entropy*, Journal of Hydrology, *Vol 42, Issues 1–2, 23–38(1979)*

[65] Karian, Z.A., and Dudewicz, E.J., *Fitting Statistical Distributions: The Generalized Lambda Distribution and Generalized Bootstrap Methods*, Chapman and Hall/CRC *(2000)*

[66] Kim, H-J, *A Monte Carlo method for estimating prediction limits for the arithmetic mean of lognormal samples*, Communications in Statistics: Theory and Methods, *Volume 36, Number 11, 2159-2167(9) (2007)*

[67] Korchevsky V.M., Petrov V.V., *On the strong law of large numbers for sequences of dependent random variables*, Vestnik St.Petersburg University, Mathematics, *43, N 3, 143–147 (2010)*

[68] Kroese, A. H., *Distributional Inference: A Loss Function Approach*, Ph.D. Thesis, University of Groningen, The Netherlands *(1994)*

[69] Krishnamoorthy, K.,Lin, and Y.,Xia,Y., *Confidence limits and prediction limits for a Weibull distribution based on the generalized variable approach*, Journal of Statistical Planning and Inference, *139,2675–2684(2009)*

[70] Kushary, D., *Construction of Prediction Intervals in Location-Scale Families*, Biometrical Journal, *Volume 38, Issue 7, pages 865-878 (1996)*

[71] Landwehr, J., Matalas, N. and Wallis, J., *Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles*, Water Resour. Res., *15, 1055-1064(1979)*

[72] Lawless, J. F., *On the estimation of safe life when the underlying life distribution is Weibull*, Technometrics, *15, 857–865 (1973)*

[73] Lawless, J.F., and Fredette,M., *Frequentist prediction intervals and predictive distributions*, Biometrika, *92, 529–542(2005)*

[74] Lieblein, J. and Zelen, M., *Statistical investigation of the fatigue life of deep groove ball bearings*, J. Res. Nat. Bur. Stand., *Vol. 57, 273-316 (1956)*

[75] Lin, Y., *Asymptotics of bootstrapping mean on some smoothed empirical distributions*, Statistics and Decisions, *15, 301–306 (1997)*

[76] Lindley, D. V., *Approximate Bayesian methods*, Trabajos Estadistica, *Vol 31, 223–237 (1980)*

[77] Lloyd, E. H., *Least-squares estimation of location and scale parameters using order statistics*, Biometrika, *Vol. 39, No. 1/2, 88–95 (1952)*

[78] Longin F., *The asymptotic distribution of extreme stock market returns*, Journal of Business, *63, 383–408 (1996)*

[79] Mann, N. R., *Optimum estimators for linear functions of location and scale parameters*, Ann. Math. Statist., *Vol 40, 2149–2155 (1969).*

[80] Mann, N. R. Schafer, R. E. and Singpurwalla, N. D., *Methods for Statistical Analysis of Reliability and Life Data*, John Wiley and Sons, New York *(1974)*

[81] Mann N. R., *Warranty period for production lots based on fatigue-test data*, Engrg. Fracture Mech. *8, 123–130 (1976)*

[82] Mann, N. R. and Fertig, K. W., *Efficient unbiased quantile estimators for moderate size complete samples from extreme-value and Weibull distributions: confidence bounds and tolerance and prediction intervals*, Technometrics, *Vol. 19, No. 1, 87–93 (1977)*

[83] Martucci, G., Carniel, S., Chiggiato, J., Sclavo, M., Lionello, P., and Galati, M. B., *Statistical trend analysis and extreme distribution of significant wave height from 1958 to 1999 - an application to the Italian Seas*, Ocean Science Discussions, *Volume 6, Issue 3, 2005–2036 (2009)*

[84] Morrison J.E. and Smith J.A., *Stochastic modeling of flood peaks using the generalized extreme value distribution*, Water Resources Research, *38, 1305 (2002)*, doi:10.1029/2001WR000502.

[85] Nelson, W., *Weibull prediction of a future number of failures*, Quality and Reliability Engineering International, *16,23–26 (2000)*

[86] Nordman, D.J., and Meeker,W.Q., *Weibull prediction intervals for a future number of failures*, Technometrics, *44, 15–23(2002)*

[87] Oliveira, J. Tiago de *Decision results for the parameters of the extreme value (Gumbel) distribution based on the mean and the standard deviation*, Trabajos de Estadustica y de Investigacion Operativa, *Vol 14, pp.61–81 (1963)*

[88] Otten, A. and van Montfort, M. A. J., *Maximum-likelihood estimation of the general extreme-value distribution parameters*, J. Hydrol., *47(1/2), 187–192 (1980)*

[89] Pandey, M.D., Lu, D. and Komljenovic, D., *The impact of probabilistic modelling in life cycle management of nuclear piping systems*, ASME J. of Engineering for Gas, Turbine and Power, *133(1), 12901:1-7 (2011)*

[90] Patel, J. K., *Prediction intervals - a review*, Communications in Statistics–Theory and Methods, *13, 2393-2465 (1989)*

[91] Persson, K. and Ryden, J. *Exponentiated Gumbel Distribution for Estimation of Return Levels of Significant Wave Height*, Department of Mathematics, Uppsala University *(2007)*

[92] Phien, H.N., *A review of methods of parameter estimation for the extreme value type-1 distribution*, Journal of Hydrology *Vol 90, Issues 3–4, 251–268 (1987)*

[93] Pitman, E. J. G., *The estimation of the location and scale parameters of a continuous population of any given form*, Biometrika, *30, 391-421(1939)*

[94] Popov, N. K. Best Estimate and Uncertainty Analysis for CANDU Reactors, Lecture 33A. *Seminar and Training on Scaling, Uncertainty and 3D Coupled Calculations in Nuclear Technology (3D SUNCOP) Oct 08 - Oct 26, 2007, Hamilton, Ontario, Canada (2007)*

[95] Prescott, P. and Walden, A.T., *Maximum likelihood estimation of the parameters of the generalized extreme-value distribution*, Biometrika, *67, 723–724 (1980)*

[96] Prescott, P., and Walden, A. T., *Maximum-likelihood estimation of the parameters of the three-parameter generalized extreme-value distribution from censored samples*, J. Stat. Comput. Simul., *6, 241-250 (1983)*

[97] Raiffa, H., and Schlaifer, R., *Applied Statistical Decision Theory, Division of Research*, Graduate School of Business Administration, Harvard University, Boston *(1961)*

[98] Ranneby, Bo, *The maximum spacing method. An estimation method related to the maximum likelihood method*, Scandinavian Journal of Statistics, *11 (2), 93-112(1984)*

[99] Ren, C., Sun, D and Dey, Dipak K., *Comparison of Bayesian and frequentist estimation and prediction for a normal population*, Sankhya: The Indian Journal of Statistics, *Vol. 66, No. 4, 678–706 (2004)*

[100] Robert, C. P., *Simulation of truncated normal variables*, Statistics and Computing ,*Volume 5, Issue 2, 121–125 (1995)*

[101] Shao, J., *Mathematical Statistics: Exercises and Solutions*, Springer *(2005)*

[102] Silverman, B. W., *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London *(1986)*

[103] Smith, R. L., *Maximum likelihood estimation in a class of nonregular cases*, Biometrika, *72, 67-92 (1985)*

[104] Smith, R. L., *Bayesian and Frequentist Approaches to Parametric Predictive Inference*, Bayesian Statistics 6, Bernardo, J. M., Berger, J. O., Dawid, A. P., Smith, A. F. M. (eds , Richard L. Smith), Oxford Unversity Press, Oxford, *589–612 (1999)*

[105] Smith, R. M. and Bain, L. J., *An exponential power life-testing distribution*, Communications in Statistics - Theory and Methods, *4, 469-481(1975)*

[106] Sun , L. , and Muller-Schwarze , D., *Statistical resampling methods in biology: A case study of beaver dispersal patterns*, Am. J. Math. Manage. Sci. *16 , 463–502 (1996)*

[107] Sutcliffe, J.V., *Methods of Flood Estimation, A Guide to Flood Studies Report*, Report No.49, Institute of Hydrology,UK *(1978)*

[108] Swartz, G. B. and Murthy, V. K., *Minimum risk estimators with applications*, Journal of the Royal Statistical Society. Series B (Methodological), *Vol. 34, No. 3, 410–416 (1972)*

[109] Thomas, P., *2012 SERC NOTES: Order statistics*, extracted from http://www.cmsintl.org/cms_lecture_notes/2012_chap5.pdf.

[110] Tierney, T., and Kadane, J.B., *Accurate approximations for posterior moments and marginal densities*, Journal of the American Statistical Association, *Vol. 81, No. 393, 82–86(1986)*

[111] Tomlinson, A.I., *The frequency of high intensity rainfalls in New Zealand*, Water and Soil Technical Publication No. 19, Ministry of Works and Development, Wellington, New Zealand, *36 p (1980)*

[112] Ueki, M., and Fueda, K., *Adjusting estimative prediction limits*, Biometrika, *94, 509-511(2007)*

[113] Vidoni, P., *A note on modified estimative prediction limits and distributions*, Biometrika, *85, 949-953 (1998)*

[114] Vidoni, P., *Improved prediction intervals and distribution functions*, Scandinavian Journal of Statistics *36, 735-748 (2009)*

[115]

[116] Walsh, B., *Markov Chain Monte Carlo and Gibbs Sampling*, Lecture Notes, 2004, retrieved from http://web.mit.edu/ wingated/www/introductions/mcmc-gibbs-intro.pdf

[117] Wang, Q. J., *Direct sample estimators of L moments*, Water Resources Research, *32(12), 3617-3619(1996)*

[118] Weba, M., *Representation for bias and variance of the sample standard deviation*, Statistical Papers, *34, 369–375 (1993)*

[119] Wilks, S. S. *Determination of sample sizes for setting tolerance limits.* Ann. Mathem. Statist., *12, No.1, 91–96 (1941)*

[120] Wilson, E.B., *Probable inference, the law of succession, and statistical inference*, Journal of the American Statistical Association, *22 , 209-212 (1927)*

[121] Wong, T. S. T. and Li, W. K., *A note on the estimation of extreme value distributions using maximum product of spacings*, Time Series and Related Topics, *Vol. 52, 272–283 (2006)*

[122] Young, G. A., *Bootstrap: More than a stab in the dark? (with discussion)*, Statist. Sci., *9 , 382-415 (1994)*