Western Graduate&PostdoctoralStudies

Western University
Scholarship@Western

Electronic Thesis and Dissertation Repository

1-30-2018 10:00 AM

Advances in the Modeling of Heavy-tailed Distributions

Sang Jin Kang, *The University of Western Ontario*

Supervisor: Provost, Serge B., *The University of Western Ontario*
Co-Supervisor: Ren, Jiandong, *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Statistics and Actuarial Sciences
© Sang Jin Kang 2018

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Applied Statistics Commons, and the Statistical Methodology Commons

Recommended Citation

Kang, Sang Jin, "Advances in the Modeling of Heavy-tailed Distributions" (2018). *Electronic Thesis and Dissertation Repository*. 6003.
https://ir.lib.uwo.ca/etd/6003

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlswadmin@uwo.ca.

Abstract

Several advances are made in connection with the approximation and estimation of heavy-tailed distributions, some of which also befit other types of distributions. It is first explained that on initially applying the Esscher transform to heavy-tailed density functions such as the Pareto, Student-t and Cauchy, said densities can be approximated by employing a certain moment-based methodology. Alternatively, density approximants can be obtained by appropriately truncating such distributions or mapping them onto finite supports. These techniques are then extended to the context of density estimation, their validity being demonstrated by means of simulation studies. As well, illustrative actuarial examples are presented. Novel approaches involving the use of the Box-Cox transform in conjunction with empirical saddlepoint density estimates and generalized beta density functions are introduced for determining the empirical endpoints associated with various types of distributions. Additionally, an iterative algorithm and an approximation expressed in terms of a linear combination of Bernstein polynomials are proposed for securing smooth *bona fide* density functions. Finally, a certain polynomial adjustment is applied to bivariate empirical saddlepoint density estimates in order to accurately model bivariate data sets. The implementation of the proposed methodologies such as the constrained estimation of the four parameters of the generalized beta distribution and the adjusted bivariate empirical saddlepoint density estimation technique in the symbolic computing package *Mathematica* also represents a notable contribution of this dissertation.

Keywords: Density approximation, heavy-tailed distributions, *bona fide* density estimates, empirical distributional endpoints, empirical saddlepoint estimates, bivariate density estimation, symbolic computing.

It is the glory of God to conceal a matter:
to search out a matter is the glory of kings.
(Proverbs 25:2)

Acknowledgements

I would like to thank God for giving me the strength and wisdom to complete this thesis. I wish to express my sincere gratitude and appreciation to my supervisors, Professors Serge B. Provost and Jiandong Ren, for their inspired guidance and generous support. I am also very grateful to the Department of Statistical and Actuarial Sciences and the School of Graduate and Postdoctoral Studies for their financial support. I also would like to express my thanks to the thesis examiners, Drs. Song Cai, Yun-Hee Choi, David Stanford and Ricardas Zitikis for their valuable suggestions and comments. As well thanks are due to the faculty, staff and fellow graduate students in the Department of Statistical and Actuarial Sciences. Finally, I am indebted to my wife Grace and my three children Joseph, Ruth and Hannah for their love and support.

Contents

Abstract **i**

Dedication **ii**

Acknowledgements **iii**

List of Figures **vii**

List of Tables **x**

1 Introduction **1**

1.1 Overview of the proposed scholarly contributions 1

1.2 Characterizations of heavy-tailed distributions 2

1.3 Various types of moment-based density function approximations 4

1.4 Techniques used for modeling heavy-tailed distributions 5

1.4.1 Polynomially adjusted density approximation 5

1.4.2 The empirical saddlepoint density estimation approach 6

2 Density Approximation Techniques as Applied to Heavy-tailed Distributions **11**

2.1 Introduction 11

2.2 Density approximation via exponential tilting 11

2.2.1 Exponentially tilted distributions 11

2.2.2 The general algorithm 12

2.3 Truncating the support of a distribution for approximation purposes 13

2.4 Density approximation via the transformation of variables technique 14

2.5 Applications 15

2.5.1 Type II Pareto distribution 15

Approximation via exponential tilting 16

Approximating density functions from truncated distributions 21

Approximation via transformation of variables 21

2.5.2 The Student-*t* distribution 24
Approximation via exponential tilting 24
Approximating density functions from truncated distributions 28
Approximation via transformation of variables 28
2.5.3 Estimating densities from data by means of exponential tilting 29
Simulated data from a type II Pareto distribution 31
Simulated data from a Student-*t* distribution 33
Automobile insurance claims data set 34
The Danish fire data set 35

3 Novel Approaches for Estimating Distributional Endpoints 42
3.1 Introduction 42
3.2 Computational issues related to the empirical saddlepoint approach 43
3.3 Combining the Box-Cox transform and the empirical saddlepoint estimation method 44
3.4 Endpoints estimates based on the 4-parameter generalized beta distribution . . . 46
3.4.1 Penalized maximum likelihood estimation (PMLE) method 47
3.4.2 Feasibility-constrained moment matching (FCMM) technique 48
3.5 Simulation studies 50
3.5.1 A beta distribution 50
3.5.2 A gamma distribution 51
3.5.3 A generalized Pareto distribution 52

4 On Securing Bona Fide Density Functions 58
4.1 Introduction 58
4.2 General algorithm of obtaining smooth *bona fide* density functions 60
4.3 Numerical example 61
4.3.1 Comparing Glad's and Gajek's approaches 63
4.4 Smooth density estimates expressed in terms of Bernstein polynomials 64

5 Adjusted Empirical Bivariate Saddlepoint Estimates 71
5.1 Introduction 71
5.2 Empirical bivariate saddlepoint estimation 72
5.2.1 Computational issues related to the empirical saddlepoint estimate . . . 72
5.2.2 Normalized likelihood-like density function 72
5.3 Adjusted empirical bivariate saddlepoint estimates 74
5.4 Numerical examples 75

5.4.1 Simulated data from a bivariate normal distribution 75
5.4.2 Flood data set 75
5.4.3 Maximum speed data set 76
5.4.4 Simulated data from a mixture of a Dirichlet and bivariate beta density functions 78

6 Concluding Remarks and Further Research Directions 84
6.1 Concluding remarks 84
6.2 Further research 85

Appendix A Mathematica code 86

Curriculum Vitae 109

List of Figures

2.1 Plot of the tilted Lomax density function ($\alpha = 3.5, \beta = 3.5, \theta = 1.5$) 17

2.2 The exact (solid line) and approximated densities (dashed line) of the Lomax distribution for different polynomial degrees and tilting parameters 18

2.3 Plots of the log of the integrated Square error (ISE) with respect to θ (Top left panel: d=12 / Top right panel: d=18 / Bottom left panel: d=24 / Bottom right panel: d=32) 18

2.4 Exact density (solid line) and $f_{X_{d=32,\theta=1.12}}(x)$ (dashed line). $ISE = 1.45698 \times 10^{-10}$ 20

2.5 Exact density (solid line) and the approximant obtained without applying exponential tilting (dashed line) 20

2.6 Exact density (solid line) and the approximants obtained using different truncation points (dashed line) 22

2.7 The difference between exact density and the approximants obtained using different truncation points 22

2.8 Exact density (solid line) and $f_{X_{d=9,\delta=3.5}}(x)$ (dashed line). $ISE = 1.30606 \times 10^{-12}$ 23

2.9 Plots of the log of integrated square error (ISE) with respect to δ (Top left panel: d=7 / Top right panel: d=8 / Bottom left panel: d=9 / Bottom right panel: d=10) 24

2.10 Plot of the log of $ISE_{d,\theta}$ with $\theta = 1$ 27

2.11 Exact density (green solid line) and $f_{X_{d=37,\theta=1.12}}(x)$ (red dashed line). $ISE = 3.10482 \times 10^{-6}$ 27

2.12 Plots of the exact (solid line) and approximant (dashed line) of Student-t density for various degrees of freedom, ν (Top left panel: ν=1 / Top right panel: ν=2 / Bottom left panel: ν=3 / Bottom right panel: ν=4) 28

2.13 Plots of the exact (solid line) and approximant (dashed line) of the truncated Student-t density by applying symmetrization technique (Top left panel: ν=1 / Top right panel: ν=2 / Bottom left panel: ν=3 / Bottom right panel: ν=4) . . . 29

2.14 Plots of the exact (solid line) and approximants (dashed line) of Student-t densities obtained by applying the transformation of variables techniques (Top left panel: ν=1 / Top right panel: ν=2 / Bottom left panel: ν=3 / Bottom right panel: ν=4) 30

2.15 Plots of $AD_{d,\theta}$ vs θ leading to the selection of the optimal d and θ 32
2.16 Histogram of the simulated data from a Pareto II $(3.5, 35)$ distribution and plots of the exact density (solid line) and the density estimate (dashed line) 32
2.17 Histogram of the simulated data from a Student-t distribution on 3 degrees of freedom, the assumed density (solid line) and the density estimate (dashed line) 34
2.18 Density estimates and rescaled histogram for the auto insurance claims data set 35
2.19 Density estimates and histogram of the original data for the Danish fire data set 39

3.1 Histograms of the data simulated from a standard normal distribution and plots of the logarithm of the empirical saddlepoint estimates obtained from the rescaled data, the red dots indicating the minimum and maximum values of the rescaled data 45
3.2 Histograms of the data simulated from a standard normal distribution and density estimates obtained by fitting a generalized beta distribution (solid line: underlying density; dotted line: PMLE method; dashed line: FCMM technique) . 49
3.3 Histograms of the data simulated from a Beta $(2, 5)$ distribution and empirical density estimates based on Algorithm 3.1. (red dots: sample minima and maxima) 51
3.4 Histograms of the data simulated from a Beta$(2, 5)$ distribution and density estimates obtained by fitting a generalized beta distribution (solid line: underlying density; dotted line: PMLE method; dashed line: FCMM technique) 51
3.5 Histograms of the data simulated from a Gamma$(3, 2.5)$ distribution and empirical density estimates based on Algorithm 3.1. (red dots: sample minima and maxima) 53
3.6 Histograms of the data simulated from a Gamma $(3, 2.5)$ distribution and density estimates obtained by fitting generalized beta distributions (solid line: underlying density; dotted line: PMLE method; dashed line: FCMM technique) . 53
3.7 Histograms of the data simulated from a GP$(0, 1, 1)$ distribution and empirical density estimates based on Algorithm 3.1. (red dots: sample minima and maxima) 54
3.8 Histograms of the data simulated from a GP$(0, 1, 1)$ distribution and density estimates obtained by fitting generalized beta distributions (solid line: underlying density; dotted line: PMLE method; dashed line: FCMM technique) 55

4.1 Histogram of a simulated sample and plot of the underlying density function simulated from an equal mixture of Beta$(30, 5)$ and Beta$(6, 25)$ 62

4.2 Plots of $f_{d=14}(x)$ on various subranges of the interval (solid line: underlying density / dotted line: $f_{d_0=14}(x)$) 63
4.3 Plots of $f_{d_0=14}(x)$ and $f_{d_1=19}(x)$ on various subranges of the interval (dotted red: $f_{d_0=14}(x)$ / dot-dashed blue: $f_{d_1=19}(x)$) 64
4.4 Plots of $f_{d_0=14}(x)$, $f_{d_1=19}(x)$, and $f_{d_2=19}(x)$ on various subranges of the interval (dotted red: $f_{d_0=14}(x)$ / dot-dashed blue: $f_{d_1=19}(x)$) / dashed cyan: $f_{d_2=19}(x)$) . . . 65
4.5 Plots of $f_{d_0=14}(x)$, $f_{d_1=19}(x)$, $f_{d_2=19}(x)$, and $f_{FIN}(x)$ on various subranges of the interval (dotted red: $f_{d_0=14}(x)$ / dot-dashed blue: $f_{d_1=19}(x)$) / dashed cyan: $f_{d_2=19}(x)$) / solid purple: $f_{FIN}(x)$) 66
4.6 Plots of $f_{d_0=14}(x)$ and $f_{GL}(x)$ on various subranges of the interval (dashed line: $f_{d_0=14}(x)$ / solid line: $f_{GL}(x)$) 67
4.7 Plots of $f_{d_0=14}(x)$, $f_{1,GA}(x)$, $f_{2,GA}(x)$, and $f_{FIN,GA}(x)$ on various subranges of the interval (dotted red: $f_{d_0=14}(x)$ / dot-dashed blue: $f_{1,GA}(x)$) / dashed cyan: $f_{1,GA}(x)$) / solid purple: $f_{FIN,GA}(x)$) 68
4.8 Plots of the normalized truncated $f_{d_0=14}(x)$ (dotted lines) and the Bernstein polynomial estimates of various degrees (solid lines) 69
4.9 $B_{1000}(f;x)$ (solid line) and its spline approximant (dashed line) 69

5.1 Histogram of the rescaled data simulated from a bivariate normal distribution and the normalized exponential term of the empirical saddlepoint density . . . 76
5.2 Adjusted density estimate, $f_{2,2}(x,y)$, and a kernel density estimate of the original data simulated from a bivariate normal distribution 76
5.3 Histogram of the rescaled flood data and the normalized exponential term of the empirical saddlepoint estimate 78
5.4 Histogram of the original flood data, the adjusted density estimate, $f_{1,1}(x,y)$ and a kernel density estimate 78
5.5 Histogram of the rescaled maximum speed data and the normalized exponential term of the empirical saddlepoint estimate 79
5.6 Histogram of the original maximum speed data, the adjusted density estimate, $f_{2,1}(x,y)$, and a kernel density estimate 79
5.7 The pdf of the mixture of Dirichlet and beta density functions 80
5.8 Histogram of the original data simulated from a mixture of a Dirichlet and a bivariate beta density functions and contour plot 81
5.9 Adjusted density estimate, $f_{6,4}(x,y)$, and a kernel density estimate of the simulated data from a mixture of a Dirichlet and a bivariate beta density functions . 81

List of Tables

1.1 Classification of tail behaviors for certain distributions 3

2.1 $IS\,E_{d,\theta}$'s for the Pareto II ($\alpha = 3.5, \beta = 3.5$) for various values of θ (left column) and d (at the top) 19

2.2 VaR_κ for X and $X_{d,\theta}$ 21

2.3 TVaR_κ for X and $X_{d,\theta}$ 21

2.4 VaR_κ for X and $X_{d,\theta}$ 23

2.5 VaR_κ from the sample, exact distribution and density estimates 33

2.6 TVaR_κ from the sample, exact distribution and density estimates 34

2.7 Summary of the rescaled automobile insurance claims data set (Original standard deviation: 2,646.91) 35

2.8 VaR_κ based on the auto insurance claims data 36

2.9 TVaR_κ based on the auto insurance claims data 36

2.10 Summary of the rescaled Danish fire data (Original standard deviation: 8.52743) 37

2.11 Summary of the log-scaled Danish fire data 38

2.12 VaR_κ values obtained from the sample and the density estimates 38

3.1 Summary of endpoints and the extreme order statistics based on the empirical saddlepoint estimate (Data simulated from a standard normal distribution) . . . 44

3.2 Summary of endpoints and the extreme order statistics based on fitting a generalized beta distribution with the PMLE method (Data simulated from a standard normal distribution) 48

3.3 Summary of endpoints and the extreme order statistics based on fitting a generalized beta distribution with the FCMM technique (Data simulated from a standard normal distribution) 49

3.4 Summary of endpoints and the extreme order statistics based on the empirical saddlepoint estimates (Data simualted from a Beta $(2, 5)$ distribution) 50

3.5 Summary of endpoints and extreme order statistics based on fitting a generalized beta distribution with the PMLE method and FCMM technique (Data simulated from a Beta $(2,5)$ distribution) 52
3.6 Summary of endpoints and extreme order statistics based on the empirical saddlepoint estimate (Data simulated from a Gamma $(3,2.5)$ distribution) 52
3.7 Summary of endpoints and extreme order statistics based on fitting generalized beta distributions with the PMLE method and FCMM technique (Data simulated from a Gamma $(3,2.5)$ distribution) 54
3.8 Summary of endpoints and extreme order statistics based on the empirical saddlepoint estimates (Data simulated from a GP $(0,1,1)$ distribution) 55
3.9 Summary of endpoints and the extreme order statistics based on fitting a generalized beta distribution with the PMLE method and FCMM technique (Data simulated from a GP $(0,1,1)$ distribution) 55

4.1 Support at each iteration and final step 63

5.1 $SSD_{a,b}$'s for the rescaled bivariate normal simulated data with the values of a in the left column and those of b at the top 77
5.2 $SSD_{a,b}$'s for the rescaled flood data with the values of a in the left and those of b at the top 77
5.3 $SSD_{a,b}$'s for the rescaled maximum speed data with the values of a in the left column and those of b at the top 80

Chapter 1

Introduction

1.1 Overview of the proposed scholarly contributions

It is not uncommon to encounter heavy-tailed distributions in actuarial or econometric applications. A drawback associated with many of those distributions is that they only possess a finite number of moments, which curtails the applicability of moment-based density approximation or estimation techniques. In Chapter 2, we propose three approaches to modify such distributions so that all the moments of the resulting distributions exist. In particular, it is explained that on initially applying the Esscher transform to density functions such as the Pareto, Student-t and Cauchy, one can utilize a certain moment-based technique whereby the tilted density functions are expressed as the product of a base density function and a polynomial adjustment. Alternatively, density approximants can be secured by appropriately truncating the distributions or mapping them onto finite supports. The validity of these approaches is corroborated by making use of simulated data. Extensions to the context of density estimation, in which case sample moments are employed in lieu of exact moments, are discussed and illustrative examples involving actuarial data sets are presented. As well, applications hinging on type-II Pareto and Student-t distributions are provided. Novel approaches involving making use of the Box-Cox transform in conjunction with empirical saddlepoint density estimates and generalized beta density functions are introduced in Chapter 3 for determining the empirical endpoints of a distribution. It should be emphasized that the 'end points' obtained from the first approach are in fact points beyond which the distribution is of no practical significance. Such empirical points ought to be useful for instance in actuarial applications, in which case the distributions do not actually have an infinite support. It may happen that an approximant or an estimate of a density function could be slightly negative on the support of the distribution. Techniques such as those proposed by Gajek (1986) and Glad *et al.* (2003) have been suggested to circumvent

this shortcoming. However, the resulting density functions are not differentiable everywhere. We are proposing two new approaches in Chapter 4 namely, an iterative algorithm and a technique relying on approximating a function by means of Bernstein polynomials, for obtaining *smooth bona fide* density functions. A polynomial adjustment is applied to a bivariate empirical saddlepoint density estimate which is based on a sample estimate of the cumulant generating function in Chapter 5, which includes several numerical examples. As it turns out, the density functions obtained by making use of this new approach are more accurate than those determined from the widely utilized kernel density estimation approach. Finally, some concluding remarks are included in Chapter 6 where avenues for future developments are suggested.

Certain characterizations of distributional tail behavior are provided in the next subsection. Some of the main moment-based density approximation techniques are described in Section 1.3. Polynomially adjusted density approximants and empirical saddlepoint estimates are discussed in Section 1.4.

1.2 Characterizations of heavy-tailed distributions

Heavy-tailed distributions are of interest in distribution theory, data analysis, and various actuarial applications. Klugman *et al.* (2012) provided several classification categories between light- and heavy-tailed distribution which are based, for instance, on moments, the hazard rate function and the mean excess loss function. Parzen (1979) examined the limiting behavior of density quantile functions which can be expressed as

$$f\left(Q(u)\right) \sim \begin{cases} (1-u)^{\alpha} & \text{for } \alpha > 0 \text{ and } \alpha \neq 1 \\ (1-u)\left(\log \frac{1}{1-u}\right)^{1-\beta} & \text{for } \alpha = 1 \text{ and } 0 \leq \beta \leq 1 \end{cases} \tag{1.1}$$

where f and Q represent the density and quantile function, respectively, and $f_1(u) \sim f_2(u)$ denotes that the ratio $f_1(u)/f_2(u)$ converges to a positive finite constant as $u \to 1$. The parameter α determines three types of tail behavior: short tails, medium tails and long tails which correspond to $\alpha < 1$, $\alpha = 1$ and $\alpha > 1$, respectively.

In order to refine the tail classification advocated by Parzen (1979), Schuster (1984) relied on two quantities, namely,

$$\alpha = \lim_{u \to 1^-} -\frac{1-u}{f\left(Q(u)\right)} \frac{\partial \log\left[f\left(Q(u)\right)\right]}{\partial u} \tag{1.2}$$

and

$$c = \lim_{u \to 1^-} (1-u)/f\left(Q(u)\right) = \lim_{u \to 1^-} 1/h\left(Q(u)\right), \tag{1.3}$$

where f, Q, and h represent the density, quantile and hazard function, respectively, to obtain five categories of tail behavior:

$$\begin{array}{ccc} \text{Short} & 0 < \alpha < 1 & \\ \text{Medium-Short} & \alpha = 1 & c = 0 \\ \text{Medium-Medium} & \alpha = 1 & 0 < c < \infty \\ \text{Medium-Long} & \alpha = 1 & c = \infty \\ \text{Long} & \alpha > 1. & \end{array}$$

The latter criterion has a theoretical connection with the limiting size of extreme spacings.

The reader may refer to Rojo (1996), whose classification, which involves more categories, is based on the residual lifetime distributions. Rojo (2010) reviewed the aforementioned classifications of the tail behavior, carried out simulation studies and assigned categories to sample of observations. Table 1.1 provides a classification for some commonly used distributions as specified in Parzen (1979), Schuster (1984) and Rojo (1996). Heavy-tailed distributions belong to the medium-long and long tail categories in Schuster's classification.

Distribution	Parzen	Schuster	Rojo
Uniform	Short	Short	Super-Short
Beta	Short	Short	Super-Short
Exponential	Medium	Medium-Medium	Medium
Logistic	Medium	Medium-Medium	Medium
Extreme value	Medium	Medium-Short	Moderately-Short
Normal	Medium	Medium-Short	Weakly-Short
Lognormal	Medium	Medium-Long	Weakly-Long
Weibull ($k < 1$)	Medium	Medium-Long	Weakly-Long
Weibull ($k = 1$)	Medium	Medium-Medium	Medium
Weibull ($k > 1$)	Medium	Medium-Short	Weakly-Short
Cauchy	Long	Long	Weakly-Long
Pareto ($k > 1$)	Long	Long	Weakly-Long
Pareto ($k = 1$)	Long	Long	Moderately-Long
Pareto ($k < 1$)	Long	Long	Super-Long

Table 1.1: Classification of tail behaviors for certain distributions

1.3 Various types of moment-based density function approximations

Various density approximation techniques that are associated with the moments or the cumulants of a distribution have been proposed in the statistical literature. For instance, Pearson's curves, which are discussed in Solomon and Stephens (1978), rely only on the first few moments of a distribution. In that case, the density function, $f(x)$, is assumed to satisfy the following differential equation:

$$\frac{1}{f(x)}\frac{df(x)}{dx} = \frac{-(C_1 + x)}{C_0 + C_1 x + C_2 x^2}. \tag{1.4}$$

where C_0, C_1 and C_2 are constants.

Johnson curves, which are described in Elderton and Johnson (1969), make use of the first four moments of a distribution to approximate a density on the basis of a system of frequency curves whose support can be finite, infinite or semi-infinite. Edgeworth expansions as proposed by Edgeworth (1905) and further developed by Fisher and Cornish (1960) and Hill and Davis (1968), aim at increasing the accuracy of an approximation by making use of Taylor series expansions and appealing to the central limit theorem. The resulting approximants are expressed in terms of Hermite polynomials whose coefficients are determined from the cumulants of the target distribution.

The saddlepoint approximation, which was pioneered by Daniels (1954), has been extensively investigated for several decades. Goutis and Casella (1995) discussed the motivation for applying this technique. Jensen (1995) extensively covered the application of this method to various types of random variables, including i.i.d. sums, compound sums, Markov chains, and a sum of independent but not necessarily identically distributed variables. In this instance, the density approximant is expressed as

$$f(x) = \frac{1}{\sqrt{2\pi K''(\hat{s})}} \exp\left[K(\hat{s}) - \hat{s}x\right] \tag{1.5}$$

where $K(t)$ is the cumulant generating function, that is, $K(t) = \ln E\left(e^{tX}\right)$ and $\hat{s}$ is the solution to $K'(\hat{s}) = x$. The main advantage of the saddlepoint approximation is its accuracy in the tails of the target density.

1.4 Techniques used for modeling heavy-tailed distributions

1.4.1 Polynomially adjusted density approximation

Provost (2005) proposed a unified methodology for approximating density functions whereby the approximant is expressed as the product of a base density function and a polynomial adjustment. This methodology applies to various types of distributions, including those that are multimodal.

Let $f_X(x)$ be the density function of a continuous random variable X defined on the interval (a,b) and $E(X^\ell) \equiv \mu_X(\ell)$. Letting $Y = (X-u)/s$, $u \in \mathbb{R}$, $s \in \mathbb{R}^+$, the density function of Y, $f_Y(y) = s f_X(u+sy)$, is defined on the interval (a_0,b_0), where $a_0 = \frac{a-u}{s}$ and $b_0 = \frac{b-u}{s}$. The moments of Y can be evaluated as follows:

$$\mu_Y(\ell) = E\left(\frac{X-u}{s}\right)^\ell = \frac{1}{s^\ell}\sum_{k=0}^{\ell}\binom{\ell}{k}\mu_X(k)\,(-u)^{\ell-k}. \tag{1.6}$$

Let $\psi_Y(y)$ denote a base density function, that is, an initial approximation to $f_Y(y)$ and $m_Y(\ell)$ be its associated ℓ^{th} moment, that is, $\int_{a_0}^{b_0} y^\ell \psi_Y(y)\,\mathrm{d}y$. Assuming that $\mu_Y(i)$, $i = 1,2,\ldots$, and $m_Y(j)$, $j = 1,2,\ldots$ exist and are uniquely defined, the approximant to $f_Y(y)$ can be expressed as

$$f_{Y_d}(y) = \psi_Y(y)\sum_{k=0}^{d}\xi_k\, y^k \tag{1.7}$$

where d is a suitable degree for the polynomial adjustment and the ξ_k's are its coefficients.

The choice of the base density function, $\psi_Y(y)$, depends on the support of the random variable Y and the features of the distribution, as indicated by a histogram of the data. Such approximants are mathematically equivalent to those that are adjusted by certain linear combinations of orthogonal polynomials. For example, when beta, gamma and Gaussian distributions are taken as base densities, the adjustments will respectively involve Jacobi, Laguerre and Hermite polynomials, which can be generated from the associated weight functions, as discussed in Provost and Jiang (2012) and Ha and Provost (2008). The r parameters of $\psi_Y(y)$ can be estimated by making use of the method of moments, that is, by equating $\mu_Y(\ell)$ to $m_Y(\ell)$, $\ell = 1,2,\ldots,r$.

The coefficients, ξ_k, of the polynomial degree k are determined by solving the system of equations resulting from equating the hth moment of the target distribution to the hth moment associated with the density approximant for $h = 0, 1, \ldots, d$, that is,

$$\int_{a_0}^{b_0} y^h\,\psi_Y(y)\sum_{k=0}^{d}\xi_k\, y^k\,\mathrm{d}y = \int_{a_0}^{b_0} y^h f_Y(y)\,\mathrm{d}y, \qquad h = 0, 1, \ldots, d. \tag{1.8}$$

This leads to a system of linear equations which yields the following solution in matrix notation:

$$\begin{pmatrix} \xi_0 \\ \xi_1 \\ \vdots \\ \xi_d \end{pmatrix} = \begin{pmatrix} m_Y(0) & m_Y(1) & \cdots & m_Y(d-1) & m_Y(d) \\ m_Y(1) & m_Y(2) & \cdots & m_Y(d) & m_Y(d+1) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ m_Y(d) & m_Y(d+1) & \cdots & m_Y(2d-1) & m_Y(2d) \end{pmatrix}^{-1} \begin{pmatrix} \mu_Y(0) \\ \mu_Y(1) \\ \vdots \\ \mu_Y(d) \end{pmatrix} \tag{1.9}$$

The degree of the polynomial adjustment is chosen to be the degree at which the integrated squared difference (ISD) as defined in Equation (1.10) or the integrated squared error (ISE) as specified by Equation (1.11) stop decreasing noticeably or satisfy a predetermined tolerance level. The ISD between approximants of successive degrees is defined as

$$ISD_d{}^{\Delta} = \int_{a_0}^{b_0} (f_{Y_d}(x) - f_{Y_{d+1}}(x))^2 \mathrm{d}x, \; d = r+1, r+2, \ldots \tag{1.10}$$

If the target density is known, the ISE between the exact and the approximated densities is evaluated as follows:

$$ISE_d = \int_{a_0}^{b_0} (f_{Y_d}(x) - f_Y(x))^2 \mathrm{d}x, \; d = r+1, r+2, \ldots \tag{1.11}$$

The minimum degree of the polynomial adjustment is set to be $r+1$ since the base density already matches the first r moments.

Finally, the approximate density for X obtained by applying the inverse transformation is given by

$$f_{X_d}(x) = \psi_Y\left(\frac{x-u}{s}\right) \sum_{k=0}^{d} \frac{\xi_k}{s} \left(\frac{x-u}{s}\right)^k. \tag{1.12}$$

Generally, the approximants can be regarded as nearly *bona fide* density functions. However, it should be noted that the resulting density approximants may occasionally happen to be slightly negative on certain sub-ranges of their supports. In order to obtain a truly *bona fide* density function, one can set the density to be zero wherever it becomes negative and normalize the resulting function so that it integrates to one. Alternatively, an algorithm proposed by Gajek (1986) can be applied. This aspect is investigated in Chapter 4.

1.4.2 The empirical saddlepoint density estimation approach

Suppose that the analytical form of the cumulant generating function is unknown. Given the sample of observations $x_1, x_2, \ldots, x_n$, the empirical cumulant generating function, as introduced by Davison and Hinkley (1988), is estimated by

$$\hat{K}_n(s) = \log\left(\frac{1}{n}\sum_{i=1}^{n} e^{sx_i}\right) \tag{1.13}$$

where n is the sample size. The estimates of the first and second derivative of $\hat{K}_n(s)$ are

$$\hat{K}_n{}'(s) = \frac{\sum_{i=1}^{n} x_i\, e^{sx_i}}{\sum_{i=1}^{n} e^{sx_i}} \tag{1.14}$$

and

$$\hat{K}_n{}''(s) = \frac{\sum_{i=1}^{n} x_i^2\, e^{sx_i}}{\sum_{i=1}^{n} e^{sx_i}} - \hat{K}_n{}'(s)^2. \tag{1.15}$$

The empirical saddlepoint estimation, denoted by $\hat{f}_n(x)$, is expressed as follows:

$$\hat{f}_n(x) = \left(2\pi\, \hat{K}_n{}''(\hat{s})\right)^{-\frac{1}{2}} e^{\hat{K}_n(\hat{s})-\hat{s}x} \tag{1.16}$$

where $\hat{s}$ is such that $\hat{K}_n{}'(\hat{s}) = x$.

Feuerverger (1989) discussed the properties of the empirical saddlepoint approximation, as well as those of the empirical moment and cumulant generating functions. Monti and Ronchetti (1993) investigated the relationship between the empirical log-likelihood and the empirical saddlepoint approximations to obtain nonparametric density estimates and confidence regions for a multivariate M-estimator.

The following propositions will be used in this dissertation.

Proposition 1.4.1 *The boundary of $\hat{K}_n{}'(s)$ lies in the interval* $(\min x_i, \max x_i)$.

Proof Let $x_{(i)}$ be the ordered observations from a sample where $x_{(1)} < x_{(2)} < \cdots < x_{(n)}$ and n is the sample size. Making use of Equation (1.14),

$$\begin{aligned}\lim_{s\to\infty} \hat{K}_n{}'(s) &= \lim_{s\to\infty} \frac{x_{(n)}\, e^{sx_{(n)}} + \sum_{i=1}^{n-1} x_{(i)}\, e^{sx_{(i)}}}{e^{sx_{(n)}} + \sum_{i=1}^{n-1} e^{sx_{(i)}}} \\ &= \lim_{s\to\infty} \frac{x_{(n)} + \sum_{i=1}^{n-1} x_{(i)}\, e^{s(x_{(i)}-x_{(n)})}}{1 + \sum_{i=1}^{n-1} e^{s(x_{(i)}-x_{(n)})}}.\end{aligned} \tag{1.17}$$

Since $x_{(i)} - x_{(n)} < 0$ for any $x_{(i)}$, the limit of the sums in both the numerator and the denominator is zero. Therefore $\lim_{s\to\infty} \hat{K}_n{}'(s) = x_{(n)}$.

In the same way,

$$\begin{aligned}\lim_{s\to-\infty} \hat{K}_n{}'(s) &= \lim_{s\to-\infty} \frac{x_{(1)}\, e^{sx_{(1)}} + \sum_{i=2}^{n} x_{(i)}\, e^{sx_{(i)}}}{e^{sx_{(1)}} + \sum_{i=2}^{n} e^{sx_{(i)}}} \\ &= \lim_{s\to-\infty} \frac{x_{(1)} + \sum_{i=2}^{n-1} x_{(i)}\, e^{s(x_{(i)}-x_{(1)})}}{1 + \sum_{i=2}^{n-1} e^{s(x_{(i)}-x_{(1)})}}\end{aligned} \tag{1.18}$$

As $s(x_{(i)} - x_{(1)}) < 0$ for any $x_{(i)}$ and $s \to -\infty$, the limit of sums in both the numerator and the denominator is zero. Therefore, $\lim_{s\to-\infty} \hat{K}_n{}'(s) = x_{(1)}$. Hence, the result holds.

Proposition 1.4.2 *$\hat{K}_n(s)$ is a strictly convex function.*

Proof Let X be a convex set in a real vector space and let $f : X \to \mathbb{R}$ be a function. The function f is strictly convex if

$$\forall x_1 \neq x_2 \in X, \forall t \in (0,1): \quad f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2). \tag{1.19}$$

Replacing $f(x)$ with $\hat{K}_n(s)$ on the left side in Equation (1.19) and making use of Hölder's inequality, the empirical cumulant generating function can be expressed as

$$\begin{aligned}
\hat{K}_n\left(ta + (1-t)b\right) &= \log\left(\frac{1}{n}\sum_{i=1}^{n} e^{(ta+(1-t)b)x_i}\right) \\
&= \log\left(\frac{1}{n}\sum_{i=1}^{n} e^{(ax_i)t} e^{(bx_i)(1-t)}\right) \\
&\leq \log\left(\left(\frac{1}{n}\sum_{i=1}^{n} e^{ax_i}\right)^{t}\left(\frac{1}{n}\sum_{i=1}^{n} e^{bx_i}\right)^{1-t}\right) \\
&= t\,\hat{K}_n(a) + (1-t)\,\hat{K}_n(b)
\end{aligned}$$

which establishes the strict convexity. Replacing $\hat{K}_n{}'(s)$ with $\hat{K}_n{}''(s)$ in Equation (1.17) and (1.18) confirms that both $\lim_{s\to\infty} \hat{K}_n{}''(s)$ and $\lim_{s\to-\infty} \hat{K}_n{}''(s)$ are zero, and then the condition of strict convexity is being met.

Bibliography

[1] Daniels, H. E. (1954). Saddlepoint approximations in statistics. *The Annals of Mathematical Statistics* **25**(4), 631–650.

[2] Davison, A. C. and Hinkley, D. V. (1988). Saddlepoint approximations in resampling methods. *Biometrika* **75**(3), 417–431.

[3] Edgeworth, F. Y. (1905). The law of error I. *Proceedings of the Cambridge Philosophical Society* **20**, 36–65.

[4] Elderton, W. P. and Johnson N. L. (1969). *Systems of frequency curves*. Cambridge University Press, New York.

[5] Feuerverger, A. (1989). On the empirical saddlepoint approximation. *Biometrika* **75.5**(3), 457–464.

[6] Fisher R. A. and Cornish E. A. (1960). The percentile points of distributions having known cumulants. *Technometrics* **2**(2), 209–225.

[7] Gajek, L. (1986). On improving density estimators which are not bona fide functions. *The Annals of Statistics* **14**(4), 1612–1618.

[8] Glad, I. K., Hjort, N. L., and Ushakov, N. G. (2003). Correction of density estimators that are not densities. *Scandinavian Journal of Statistics* **30**(2), 415–427.

[9] Goutis C. and Casella G. (1999). Explaining the saddlepoint approximation. *The American Statistician* **53**(3), 216–224.

[10] Ha, H. and Provost, S. B. (2008). On the inversion of certain moment matrices. *Linear Algebra and its Applications* **430**(10), 2650–2658.

[11] Hill, G. W. and Davis, A. W. (1968). Generalized Asymptotic Expansions of Cornish–Fisher Type. *Annals of Mathematical Statistics* **39**(4), 1264–1273.

[12] Jensen, J. L. (1995). *Saddlepoint Approximations*. Clarendon Press: Oxford Statistical Science Series 16, Oxford.

[13] Klugman, S. A. and Panjer, H. H. and Willmot, G. E. (2012). *Loss models: from data to decisions, 4th Ed.* John Wiley & Sons: Hoboken, New Jersey.

[14] Monti, A. C. and Ronchetti, E. (1993). On the relationship between empirical likelihood and empirical saddlepoint approximation for multivariate M-estimators. *Biometrika* **80**(2), 329–338.

[15] Parzen, E. (1979). Nonparametric statistical data modeling. *Journal of the American statistical association* **74**(365), 105–121.

[16] Provost, S. B. (2005). Moment–based density approximants. *The Mathematica Journal* **9**, 727–756.

[17] Provost, S. B. and Jiang, M. (2012). Orthogonal polynomial density estimates: Alternative representation and degree selection. *International Journal of Computational and Mathematical Sciences* **6**, 12–29.

[18] Rojo, J. (1996). On tail categorization of probability laws. *Journal of the American Statistical Association* **91**(433), 378–384.

[19] Rojo, J. and Ott, R. C. (2010). Testing for tail behavior using extreme spacings. *arXiv preprint arXiv:1011.6458*

[20] Schuster, E. F. (1984). Classification of probability laws by tail behavior. *Journal of the American Statistical Association* **79**(388), 936–939.

[21] Solomon, H. and Stephens M. (1978). Approximations to density functions using Pearson curves. *Journal of the American Statistical Association* **73**(361), 153–160.

Chapter 2

Density Approximation Techniques as Applied to Heavy-tailed Distributions

2.1 Introduction

Novel density estimation and approximation techniques applying to heavy-tailed distribution are introduced in this chapter.

In Section 2.2, we propose to initially apply to the exponential tilting technique (the Esscher transform) before approximating the densities of heavy-tailed distributions. Then, a so-called polynomially adjusted density approximation technique is utilized. As explained in Provost (2005), Ha and Provost (2007) and Chapter 1 of this thesis, the resulting approximant is expressed as the product of a base density and a polynomial adjustment. The following two sections cover the approximation of such distributions by forming a compact support through truncation and by applying the transformation of variables technique, respectively. Several illustrative examples involving heavy-tailed distributions such as the Type II Pareto (Lomax), Cauchy, and Student-t are presented in the last section. Special consideration is also paid to symmetric distributions.

2.2 Density approximation via exponential tilting

2.2.1 Exponentially tilted distributions

An exponentially tilted distribution is obtained by transforming the target probability density function, $f_X(x)$, into the following probability density function,

$$f_{X_\theta}(x) = \frac{e^{-\theta x} f_X(x)}{\mathcal{L}_{f_X(x)}(\theta)}, \ \theta \in \mathbb{R} \tag{2.1}$$

where $\mathcal{L}_{f_X(x)}(\theta) = \int_{-\infty}^{\infty} e^{-\theta x} f_X(x)\,\mathrm{d}x$ is the bilateral Laplace transform of $f_X(x)$. The original definition of the Laplace transform, that is, the unilateral Laplace transform, is utilized when the support of X is the positive half-line.

This technique has previously been utilized in actuarial science and financial mathematics. Esscher (1932) initially used this approach to approximate the distribution of aggregate claims at a point of interest, x_0, the tilting parameter θ being such that the mean of the transformed distribution equals x_0. Gerber and Shiu (1994) and Elliot *et al.* (2005) utilized the exponential tilting method to value derivative securities whose prices were governed by certain stochastic processes with stationary and independent increments. Bühlmann (1980) introduced the Esscher premium principle as a risk measure within the framework of utility theory and risk exchange. Cox *et al.* (2005) demonstrated how to utilize multivariate exponential tilting for pricing certain securities. This technique is also used in importance sampling, as discussed for instance in Siegmund (1976).

In order to apply the polynomially adjusted density approximation methodology described in Section 1.4.1, all the moments associated with the target density and the base density must be finite. Accordingly, this approximation approach cannot be directly applied to heavy-tailed distributions, which often have a limited number of moments. Examples of such distributions include the type II Pareto and the Cauchy distributions. Moreover, as was shown in Provost (2005), the resulting approximant may prove inaccurate when the tail behavior of a distribution under consideration does not match that of the assumed base density.

To mitigate those issues, we propose to exponentially tilt the target density function before attempting to determine a polynomially adjusted approximant.

2.2.2 The general algorithm

The following algorithm is utilized to approximate density functions that are transformed via the exponential tilting technique on the basis of their theoretical moments.

1. Given the density function, $f_X(x)$, we obtain the corresponding exponentially tilted density function, $f_{X_\theta}(x)$, as specified by Equation (2.1).

2. The ℓth moment associated with the tilted density is determined as follows:

$$\mu_{X_\theta}(\ell) = \int_{-\infty}^{\infty} x^\ell f_{X_\theta}(x)\mathrm{d}x. \tag{2.2}$$

3. The approximant of the tilted density, $\tilde{f}_{X_{d,\theta}}(x)$, is expressed as the product of the base density, $\psi_{X_\theta}(x)$, and a polynomial adjustment of degree d:

$$\tilde{f}_{X_{d,\theta}}(x) = \psi_{X_\theta}(x) \sum_{k=0}^{d} \xi_k x^k. \tag{2.3}$$

4. The methodology described in Section 1.4.1 is applied to estimate the parameter(s) of the base density and to determine the coefficients of the polynomial adjustment, which are obtained by equating $\int_{-\infty}^{\infty} x^h f_{X_\theta}(x)\,\mathrm{d}x$ to $\int_{-\infty}^{\infty} x^h \tilde{f}_{X_{d,\theta}}(x)\,\mathrm{d}x$ for $h = 0, 1, \ldots, d$.

5. The approximant of the target density is then recovered by applying the inverse tilting transformation, that is,

$$f_{X_{d,\theta}}(x) = \exp(\theta x) \cdot \mathcal{L}_{f_X(x)}(\theta) \cdot \tilde{f}_{X_{d,\theta}}(x). \tag{2.4}$$

6. Using the integrated squared error (ISE) formula,

$$ISE_{d,\theta} = \int_{\infty}^{\infty} \left(f_{X_{d,\theta}}(x) - f_X(x)\right)^2 \mathrm{d}x, \quad \theta \in \mathbb{R},\ d = 3, 4, \ldots, \tag{2.5}$$

we select the '*optimal*' polynomial degree d as well as the appropriate tilting parameter, θ, according to the criteria specified in Section 1.4.1.

2.3 Truncating the support of a distribution for approximation purposes

Let X be a random variable having a finite number of moments, whose support is the interval $(-\infty, \infty)$. Let $g_{X_{(a,b)}}(x)$ denote the corresponding truncated distribution on the interval (a, b), whose density function is given by

$$g_{X_{(a,b)}}(x) = \frac{f_X(x)}{\int_a^b f_X(x)\mathrm{d}x}, \quad a < x < b \tag{2.6}$$

where $f_X(x)$ is the density function of X. For the case of a target density function defined on the interval $[0, \infty)$, one can set the right-truncation points b as a certain percentile of the target density, leaving $a = 0$ since the right-tail behaviour is of interest. If both tails are present and the distribution is symmetric around zero, one can choose $a = -b$ as the truncation points so that a and b represent p^{th} and $(100 - p)^{\text{th}}$ percentile, respectively.

As the support of $X_{(a,b)}$ is now compact, all the moments of the truncated random variable exists. And thus, the approximation methodology described in Section 1.4.1 can be applied. Accordingly, the target density, $f_X(x)$, can be approximated as follows:

1. The ℓth moment associated with $g_{X_{(a,b)}}(x)$ as specified in Equation (2.6) is given by

$$\mu_{X_{(a,b)}}(\ell) = \frac{\int_a^b x^\ell f_X(x)\mathrm{d}x}{\int_a^b f_X(x)\mathrm{d}x}. \tag{2.7}$$

2. A uniform density on the interval (a, b) is used as base density and the approximant is then

$$\tilde{f}_{X_{d,(a,b)}}(x) = \frac{1}{b-a}\sum_{k=0}^{d} \xi_k x^k. \tag{2.8}$$

3. The coefficients of the polynomial adjustments, ξ_k, are determined by solving the linear system resulting from equating $\int_a^b x^h g_X(x)\,\mathrm{d}x$ to $\int_a^b x^h \tilde{f}_{X_{d,(a,b)}}(x)\,\mathrm{d}x$ for $h = 0, 1, \ldots, d$.

4. The approximant of the target density on the interval (a, b) is then

$$f_{X_{d,(a,b)}}(x) = \left(\int_a^b f_{X_\theta}(x)\mathrm{d}x\right) \cdot \tilde{f}_{X_{d,(a,b)}}(x). \tag{2.9}$$

5. The '*optimal*' polynomial degree d is selected on the basis of the *ISE* between $f_{X_{d,(a,b)}}(x)$ and $f_X(x)$, denoted by $ISE_{X_{d,(a,b)}}$.

2.4 Density approximation via the transformation of variables technique

Consider a random variable X defined on the interval $[0, \infty)$ and the transformed variable, $Y = \frac{X}{X+\delta}$, whose support is the interval $[0, 1)$. Since all the moments of Y exist, the density approximation methodology discussed in Section1.4.1 can be implemented.

More specifically, this methodology involves the following steps:

1. Given $f_X(x)$, the density function of X, the density function of $Y = \frac{X}{X+\delta}$ is

$$f_Y(y) = \frac{\delta}{(1-y)^2}\, f_X\left(\frac{y\delta}{1-y}\right) \tag{2.10}$$

   and the required associated moments are evaluated.

2. The approximant of the density of Y, $f_{Y_d}(x)$, is expressed as a polynomial as in Equation (2.8), with $a = 0$ and $b = 1$.

3. The coefficients of this polynomial, denoted by ξ_k, are determined by solving the linear systems resulting from equating $\int_0^1 y^h f_Y(y)\,\mathrm{d}y$ to $\int_0^1 y^h f_{Y_d}(y)\,\mathrm{d}y$ for $h = 0, 1, \ldots, d$.

4. The approximant of the target density of X is then obtained by applying the inverse transformation:

$$f_{X_{d,\delta}}(x) = \frac{\delta}{(x+\delta)^2}\, f_{Y_d}\left(\frac{x}{x+\delta}\right). \tag{2.11}$$

5. The '*optimal*' polynomial degree, d, and the parameter, δ, is selected on the basis of the *ISE* between $f_{X_{d,\delta}}(x)$ and $f_X(x)$, denoted by $ISE_{d,\delta}$.

2.5 Applications

In this section, we apply the three proposed density approximation methodologies to certain heavy-tailed distributions. More specifically, we will approximate the density functions of the type II Pareto, Cauchy, and Student-t distributions. Additionally, the approximation technique relying on exponential tilting is applied in the context of density estimation on the basis of a sample of observed values. Applications involving heavy-tailed actuarial data sets are presented.

2.5.1 Type II Pareto distribution

In the actuarial context, Pareto distributed individual loss random variables are utilized to model extreme loss and risky types of insurance coverage. Since this distribution is not exponentially bounded, it results in very large values having positive probabilities.

Let X follow a type II Pareto (also called Lomax) distribution, denoted Pareto II (α, β), with density function

$$f_X(x) = \frac{\alpha\beta^\alpha}{(x+\beta)^{\alpha+1}}, \quad \text{for } x \geq 0 \text{ and } \alpha, \beta > 0. \tag{2.12}$$

The ℓth moment of X is given by

$$\mu_X(\ell) = \begin{cases} \beta^\ell\, \Gamma(\ell+1)\, \Gamma(\alpha-\ell)\, /\, \Gamma(\alpha) & \text{if } \ell < \alpha \\ \infty & \text{if } \ell \geq \alpha \end{cases}, \tag{2.13}$$

the number of finite moments depending on the value of the shape parameter α.

Approximation via exponential tilting

According to Nadarajah (2006), the analytical solution of the Laplace transform can be expressed either in terms of an incomplete gamma function or the Whittaker function. The solution involving the incomplete gamma function is

$$\mathcal{L}_{f_X(x)}(\theta) = \alpha(\beta\theta)^{\alpha} \cdot e^{\beta\theta} \cdot \Gamma(-\alpha, \beta\theta) \tag{2.14}$$

where $\Gamma(a, x) = \int_x^{\infty} t^{a-1}\, e^{-t} dt$, whereas the solution given in terms of Whittaker's function is

$$\mathcal{L}_{f_X(x)}(\theta) = \alpha(\beta\theta)^{\frac{\alpha-1}{2}} \cdot e^{\frac{\beta\theta}{2}} \cdot \mathcal{W}_{-\frac{\alpha+1}{2},-\frac{\alpha}{2}}(\beta\theta), \tag{2.15}$$

where $\mathcal{W}_{c,d}(z)$ denotes the Whittaker function, which can be expressed as

$$\mathcal{W}_{c,d}(z) = \frac{e^{-\frac{z}{2}}\, z^{d+\frac{1}{2}}}{\Gamma(d-c+\frac{1}{2})} \int_0^{\infty} t^{d-c-\frac{1}{2}}\, (1+t)^{d+c-\frac{1}{2}}\, e^{-zt}\, \mathrm{d}t, \tag{2.16}$$

where $\Gamma(z) = \int_0^{\infty} t^{z-1}\, e^{-t} \mathrm{d}t$.

We make use of Equation (2.15) to obtain the corresponding tilted distribution as Equation (2.14) can occasionally produce complex numbers.

Consider $X \sim$ Pareto II (α, β), where the symbol " $\sim$ " means "is distributed as". We obtained a new representation of the ℓth moment of the tilted Type II Pareto distribution, which is

$$\mu_{X_\theta}(\ell) = \left(\frac{\beta}{\theta}\right)^{\ell/2} \Gamma(\ell+1) \frac{\mathcal{W}_{-\frac{\alpha+\ell+1}{2},-\frac{\alpha-\ell}{2}}(\beta\theta)}{\mathcal{W}_{-\frac{\alpha+1}{2},-\frac{\alpha}{2}}(\beta\theta)}. \tag{2.17}$$

The tilted density is then approximated by making use of Equation (2.3).

No matter what positive real value is being used for the tilting parameter, the tilted density takes on a finite value at $x = 0$. Figure 2.1 shows a plot of the tilted density function of the Pareto II $(3.5, 3.5)$ where the tilting parameter is $\theta = 1.5$. This suggests taking the exponential density function as base density, that is, $\psi_X(x) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}}$, $x > 0$, where parameter λ is estimated by $\mu_{X_\theta}(1)$.

Other steps that are described in the algorithm included in Section 2.2, such as determining the coefficients of the polynomial adjustments and recovering the approximant of the target density, are then implemented. The resulting function is finally normalized in order to obtain a *bona fide* density function.

We select the degree of the polynomial adjustment, d, and the tilting parameter, θ, such that the *ISE* with respect to d and θ, as defined in Equation (2.5), is minimized.

Specifying the value of the tilting parameter can be challenging as it can theoretically take on any positive real values. As suggested by Figure 2.2, the approximant appears to be robust with respect to the choice of the tilting parameter.

To illustrate how the tilting parameter is chosen, let θ be discretized to two decimals. Figure 2.3 illustrates how the $IS\,E_{d,\theta}$ behaves with respect to given values of θ and d, the polynomial adjustment degree. Generally, up to polynomial degree 32, the higher the polynomial degree, the lower the $IS\,E_{d,\theta}$ becomes. For any polynomial degree, d, there is a certain local minimum of $IS\,E_{d,\theta}$ around a tilting value in the neighbourhood of $\theta = \frac{\alpha}{\beta}$. However, as $\arg\min_{\theta} IS\,E_{d,\theta}$ gradually increases as d increases, it is difficult to select one particular numerical value of θ as the '*optimal*' tilting parameter. Furthermore, when the polynomial degree is higher, the approximant pattern gradually becomes irregular and exhibits outliers mainly due to the presence of potentially singular points occurring when $IS\,E_{d,\theta}$ is evaluated numerically.

In light of these considerations, we take $\theta = \frac{\alpha}{\beta} = 1$ as the reference point for the tilting parameter. The resulting $IS\,E_{d,\theta}$ values shown in Table 2.1 indicate that the local minimum of $IS\,E_{d,\theta}$ occurs at $d = 32$ and $\theta = 1.12$.

After applying the inverse tilting transformation, the approximant obtained with $d = 32$ and $\theta = 1.12$ is seen to be in close agreement with the target density in Figure 2.4.

Figure 2.5 where a gamma density is used as base density illustrates how inaccurate the resulting approximant can be when one resorts to the polynomially adjusted approximation methodology without exponential tilting.

Related to actuarial applications, risk measures are important tools in pricing, loss reserv-

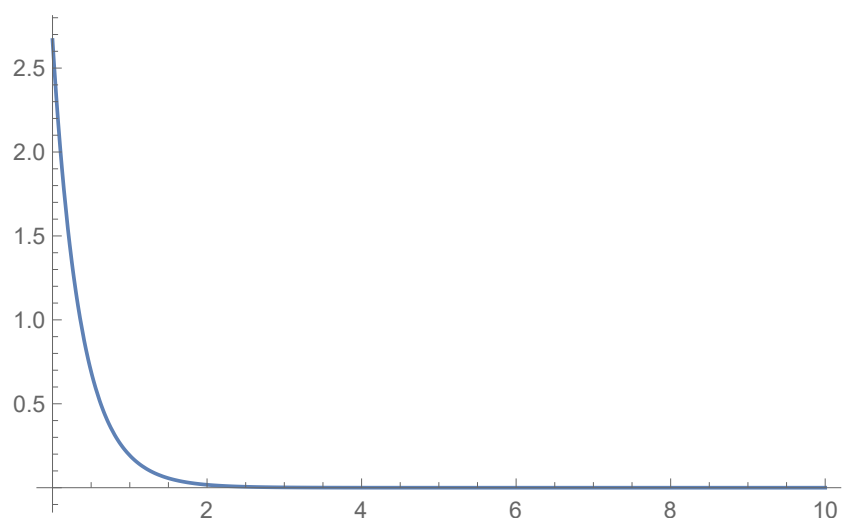
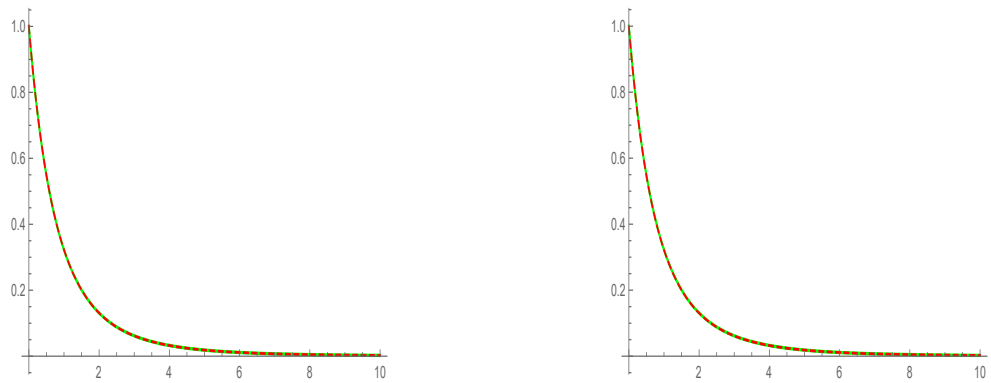


Figure 2.1: Plot of the tilted Lomax density function ($\alpha = 3.5, \beta = 3.5, \theta = 1.5$)



(a) $d = 14$ and $\theta = 3$ (b) $d = 16$ and $\theta = 0.8$

Figure 2.2: The exact (solid line) and approximated densities (dashed line) of the Lomax distribution for different polynomial degrees and tilting parameters

ing and quantitative risk management. Formally, given a random variable X with distribution function $F_X(x)$ and a confidence level $\kappa \in (0, 1)$, the Value-at-Risk (VaR) and Tail Value-at-Risk (TVaR) are respectively defined as

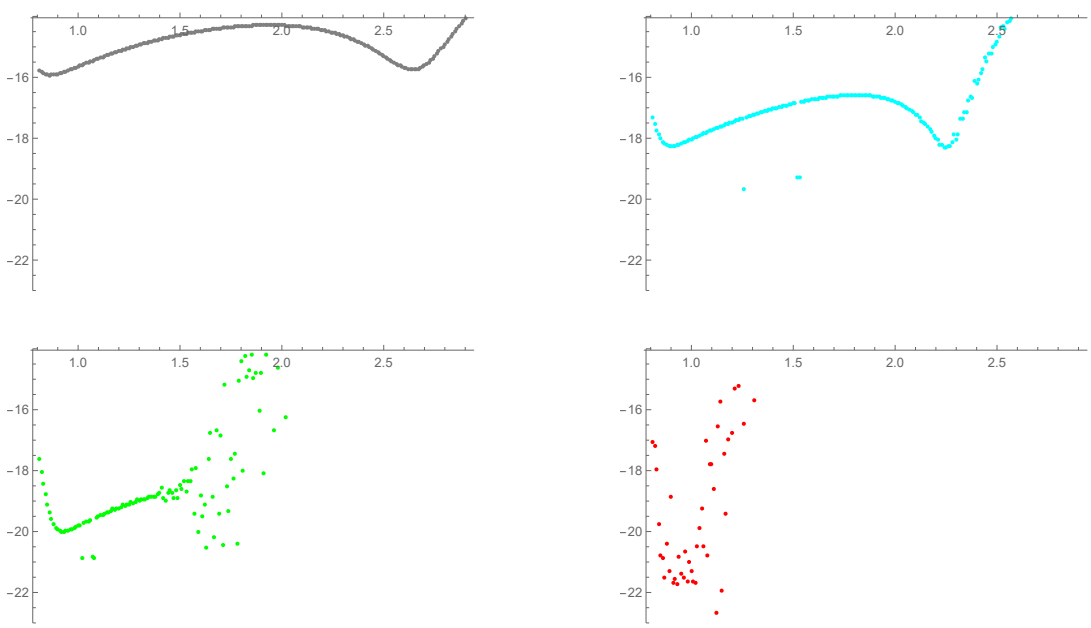


Figure 2.3: Plots of the log of the integrated Square error (ISE) with respect to θ
(Top left panel: d=12 / Top right panel: d=18 /
Bottom left panel: d=24 / Bottom right panel: d=32)

$$\mathrm{VaR}_\kappa(X) = \inf\{\ell \in \mathbb{R} : F_X(\ell) \geq \kappa\} \tag{2.18}$$

and

$$\mathrm{TVaR}_\kappa(X) = E(X|X > \mathrm{VaR}_\kappa(X)) = \frac{1}{1-\kappa}\int_\kappa^1 \mathrm{VaR}_t(X)\,\mathrm{d}t\,. \tag{2.19}$$

In the actuarial literature, those two risk measures are known as the quantile premium principle and conditional tail expectation, respectively.

Let $X_{d,\theta}$ be a random variable associated with the density approximant obtained from applying the exponential tilting technique, $f_{X_{d,\theta}}(x)$. Letting $X \sim$ Pareto II$(3.5, 3.5)$ as in the previous example, Tables 2.2 and 2.3 provide specific values of VaR and TVaR for the exact and several approximated distributions. We observe that VaR and TVaR values obtained from the approximated distributions are generally quite close to those determined from the exact distribution.

	26	27	28	29	30	31	32	33	34	35
1	1.49061E-09	1.08512E-09	9.23611E-10	7.52682E-10	5.953E-10	4.57864E-10	5.66844E-10	6.09409E-10	6.84966E-09	2.3436E-08
1.01	1.53571E-09	1.22344E-09	9.49539E-10	7.77049E-10	6.32834E-10	4.34349E-10	3.9275E-10	5.20791E-10	2.26372E-09	**1.9667E-09**
1.02	1.58918E-09	1.26012E-09	9.82395E-10	8.30968E-10	6.14509E-10	5.5261E-10	3.83891E-10	4.99977E-10	**2.75015E-10**	1.2383E-08
1.03	1.63385E-09	1.29692E-09	9.72445E-10	9.54928E-10	4.72387E-10	1.45007E-09	1.29628E-09	1.18213E-08	4.39682E-08	1.2267E-06
1.04	1.70458E-09	1.30206E-09	1.08318E-09	8.00901E-10	8.29508E-10	3.30702E-10	2.26703E-09	5.95117E-09	8.85908E-08	2.5198E-07
1.05	1.74003E-09	1.41848E-09	1.00344E-09	1.13852E-09	3.62603E-10	2.93586E-09	4.31604E-09	3.40082E-08	1.15705E-07	2.2295E-06
1.06	1.81514E-09	1.41895E-09	1.1304E-09	9.41529E-10	5.72097E-10	1.40564E-09	1.27666E-09	2.62748E-08	1.52015E-07	2.5022E-06
1.07	1.87428E-09	1.44098E-09	1.26039E-09	6.53731E-10	2.11622E-09	2.07634E-09	4.02014E-08	2.23593E-07	2.03221E-06	7.205E-06
1.08	1.94318E-09	1.4902E-09	1.2294E-09	9.17847E-10	8.37077E-10	5.38434E-10	9.27253E-10	4.3055E-10	1.69338E-08	3.6961E-08
1.09	1.94292E-09	1.66023E-09	1.04757E-09	1.53416E-09	2.37589E-10	6.61543E-09	1.84787E-08	1.79549E-07	1.05305E-06	1.2324E-05
1.1	2.04903E-09	1.59794E-09	1.32109E-09	8.63558E-10	1.45751E-09	3.45979E-10	1.86874E-08	1.16536E-07	1.24969E-06	5.8666E-06
1.11	2.1717E-09	1.52186E-09	1.66558E-09	5.24648E-10	2.86501E-09	9.1443E-10	8.22249E-09	1.39329E-09	3.18481E-07	3.4514E-06
1.12	2.23233E-09	1.60062E-09	1.55925E-09	8.36026E-10	1.17872E-09	7.42794E-10	**1.45698E-10**	1.002E-08	1.08483E-08	8.0852E-08
1.13	2.29127E-09	1.64248E-09	1.69424E-09	5.95484E-10	3.30866E-09	3.43545E-09	6.47341E-08	4.33466E-07	4.12174E-06	2.4156E-05
1.14	2.25715E-09	1.96971E-09	**1.02891E-09**	2.85321E-09	1.24387E-09	3.40148E-08	1.49259E-07	1.09429E-06	6.44014E-06	5.7735E-05
1.15	2.39616E-09	1.82269E-09	1.53145E-09	1.19609E-09	7.52665E-10	1.56486E-09	2.90138E-10	1.27062E-08	1.14797E-06	2.2223E-05
1.16	2.41926E-09	1.98698E-09	1.346E-09	1.55629E-09	7.66298E-10	**1.79652E-10**	2.63977E-08	2.76974E-07	2.92208E-06	1.4001E-05

Table 2.1: $IS\,E_{d,\theta}$'s for the Pareto II ($\alpha = 3.5, \beta = 3.5$) for various values of θ (left column) and d (at the top)

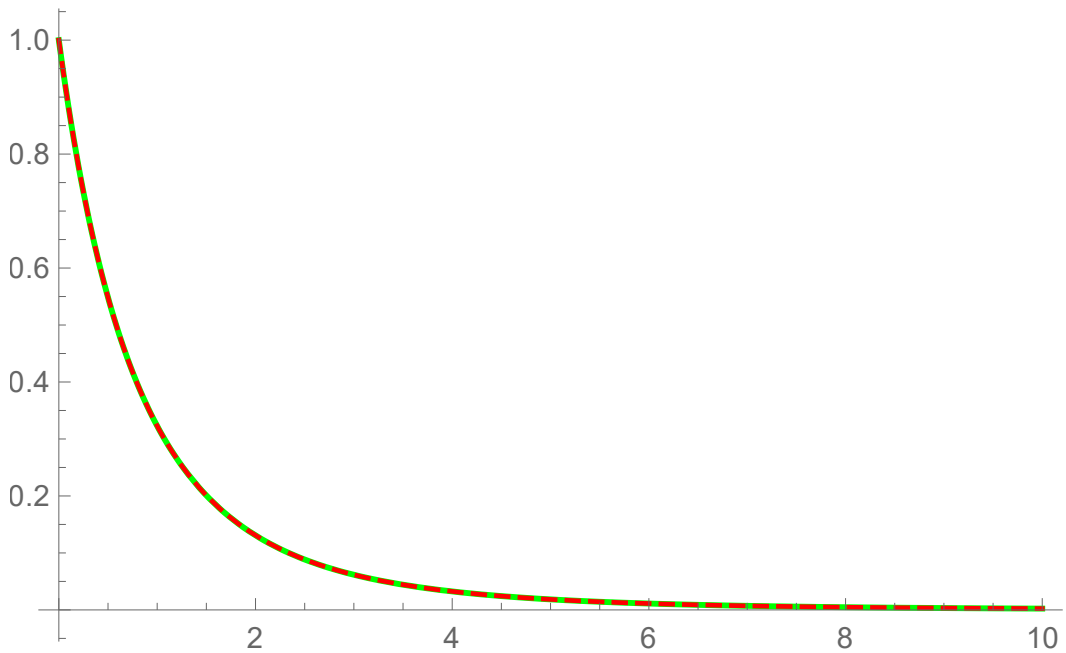


Figure 2.4: Exact density (solid line) and $f_{X_{d=32,\theta=1.12}}(x)$ (dashed line).
$ISE = 1.45698 \times 10^{-10}$

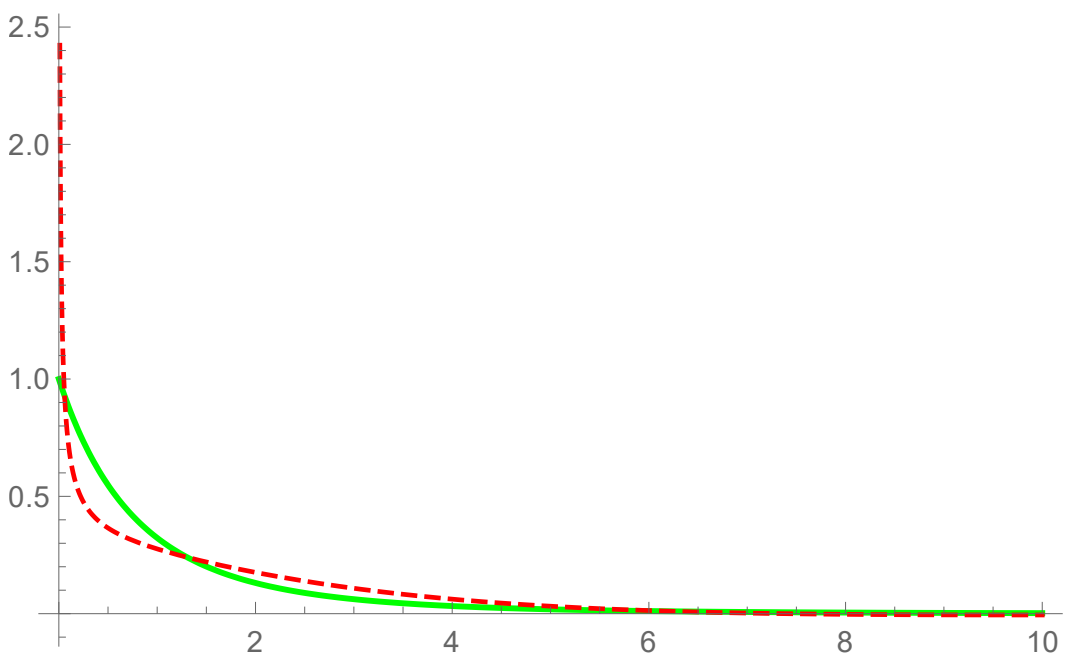


Figure 2.5: Exact density (solid line) and the approximant obtained without applying exponential tilting (dashed line)

κ	$\mathrm{VaR}_\kappa(X)$	$\mathrm{VaR}_\kappa(X_{d=30,\theta=0.88})$	$\mathrm{VaR}_\kappa(X_{d=31,\theta=1})$	$\mathrm{VaR}_\kappa(X_{d=32,\theta=1.12})$
0.8	2.04337	2.04318	2.04317	2.04335
0.9	3.25744	3.25691	3.2569	3.25739
0.99	9.54658	9.53524	9.53499	9.54544
0.999	21.689	21.4872	21.4679	21.6657
ISE	N/A	7.94042×10^{-10}	4.57864×10^{-10}	1.45698×10^{-10}

Table 2.2: VaR_κ for X and $X_{d,\theta}$

Approximating density functions from truncated distributions

We now follow the algorithm presented in Section 2.3 to obtain an approximant for a type II Pareto density function with parameters ($\alpha = 3.5, \beta = 3.5$) on the intervals [0, 95th percentile] and [0, 99th percentile]. Figure 2.6 shows that the exact density superimposed on each of the two approximants, which consists of polynomials of degrees 25.

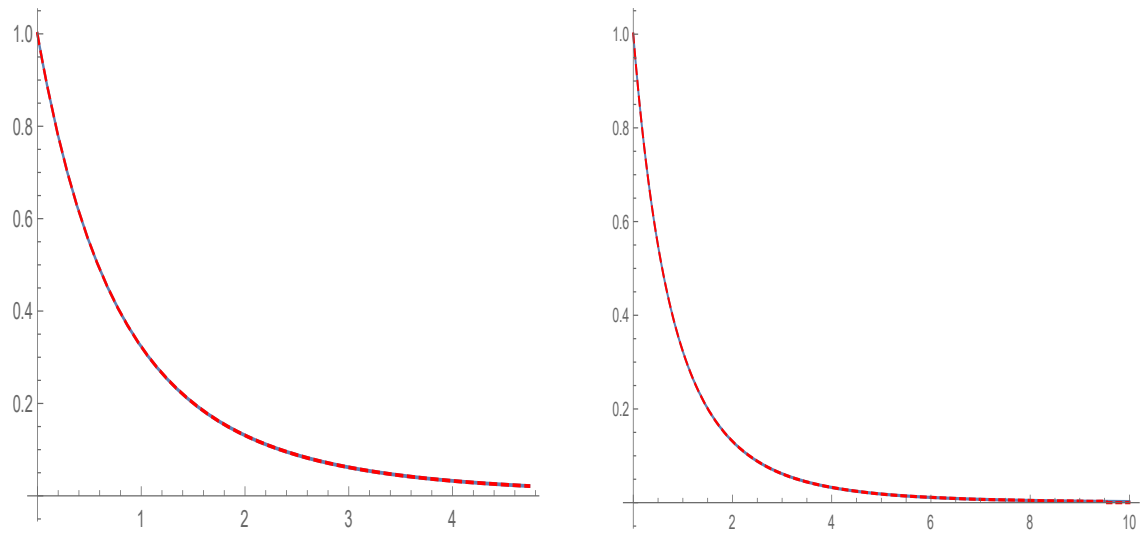
The differences between the exact density and its approximants are plotted in Figure 2.7. We observe that as the support becomes narrower, the error between the exact and approximated density functions and the *ISE* become smaller. In addition, the magnitude of error term increases steeply in the extreme right tail.

Approximation via transformation of variables

Consider the type II Pareto density function, $f_X(x) = \frac{\alpha\beta^\alpha}{(x+\beta)^{\alpha+1}}$, and the density function obtained by applying the transformation $Y = \frac{X}{X+\delta}$ with $\delta = \beta$, which turns out to be a beta density function, that is given by $f_Y(y) = \alpha(1-y)^{\alpha-1}, \ 0 \le y \le 1$. For illustrative purposes, we are making use of a Pareto II $(3.5, 3.5)$ distribution. On applying the transformation with $\delta = 3.5$, the resulting 9th degree polynomial approximant is seen to be in close agreement with the target density in Figure 2.8.

κ	$\mathrm{TVaR}_\kappa(X)$	$\mathrm{TVaR}_\kappa(X_{d=30,\theta=0.88})$	$\mathrm{TVaR}_\kappa(X_{d=31,\theta=1})$	$\mathrm{TVaR}_\kappa(X_{d=32,\theta=1.12})$
0.8	4.26072	4.24757	4.24679	4.25501
0.9	5.96042	5.93443	5.93289	5.94903
0.99	14.7652	14.5228	14.5089	14.6532
0.999	31.7646	29.6872	29.5781	30.6841
ISE	N/A	7.94042×10^{-10}	4.57864×10^{-10}	1.45698×10^{-10}

Table 2.3: TVaR_κ for X and $X_{d,\theta}$

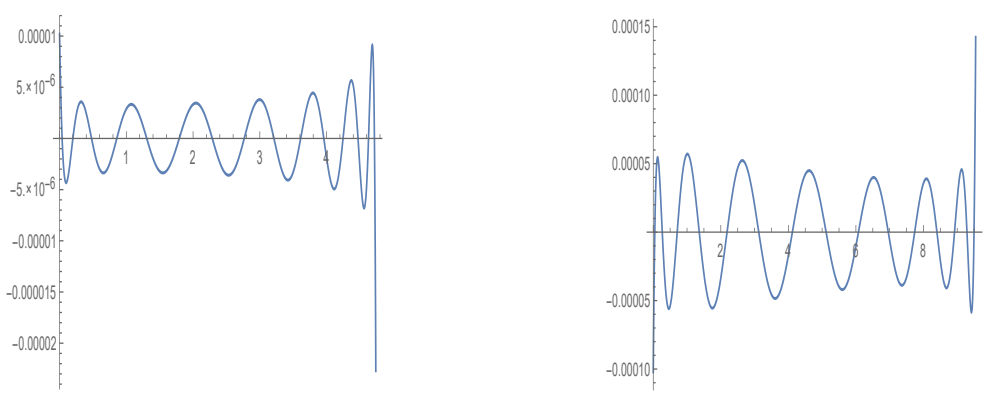


(a) Approximant on [0, 95th percentile]
$ISE = 4.49857 \times 10^{-11}$

(b) Approximant on [0, 99th percentile]
$ISE = 1.40519 \times 10^{-7}$

Figure 2.6: Exact density (solid line) and the approximants obtained using different truncation points (dashed line)

Generally, the '*optimal*' approximant is obtained when the $ISE_{d,\delta}$ is minimized with respect to the transformation parameter δ and the polynomial degree, d. Figure 2.9 illustrates the patterns followed by the log of $ISE_{d,\delta}$ with respect changes in δ discretized by two decimal points, while keeping certain polynomial degrees, d, fixed. The discontinuities of $ISE_{d,\delta}$'s are observed due to issues associated with the numerical integration. The $ISE_{d,\delta}$ is minimized at



(a) Approximant on [0, 95th percentile]

(b) Approximant on [0, 99th percentile]

Figure 2.7: The difference between exact density and the approximants obtained using different truncation points

$(d, \delta) = (5.2, 9)$, the correspond numerical value being 1.31067×10^{-14}. As in Figure 2.8, the resulting approximant is in close agreement with the target density. Additionally, Table 2.4 provides specific values of the VaR associated the exact density and the approximant obtained via exponential tilting and transformation of variables.

The TVaR associated with approximant via transformation of variables cannot be evaluated as Equation (2.11) becomes

$$f_{X_{d,\delta}}(x) = \sum_{k=0}^{d} \xi_k \frac{\delta \cdot x^k}{(x+\delta)^{k+2}}, \tag{2.20}$$

in which case even the first moment associated with the approximant is unavailable.

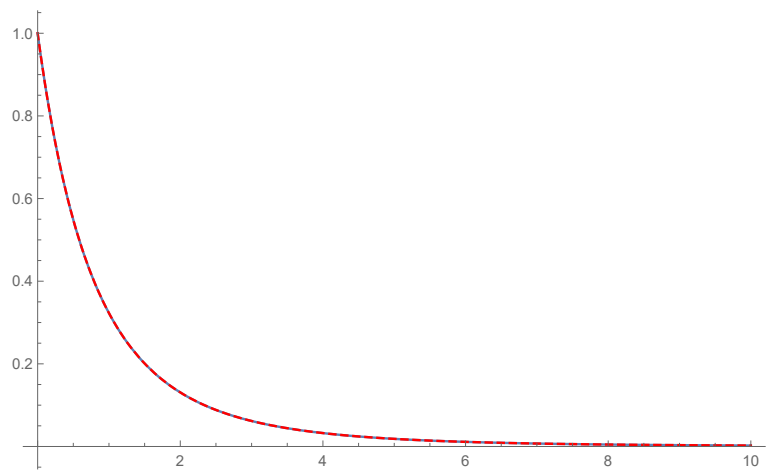


Figure 2.8: Exact density (solid line) and $f_{X_{d=9,\delta=3.5}}(x)$ (dashed line).
$ISE = 1.30606 \times 10^{-12}$

κ	$\mathrm{VaR}_\kappa(X)$	$\mathrm{VaR}_\kappa(X_{d=32,\theta=1.12})$	$\mathrm{VaR}_\kappa(X_{d=9,\delta=5.2})$
0.8	2.04337	2.04335	2.04337
0.9	3.25744	3.25739	3.25744
0.99	9.54658	9.54546	9.54659
0.999	21.689	21.662	21.6888
ISE	N/A	1.45698×10^{-10}	1.31067×10^{-14}

Table 2.4: VaR_κ for X and $X_{d,\theta}$

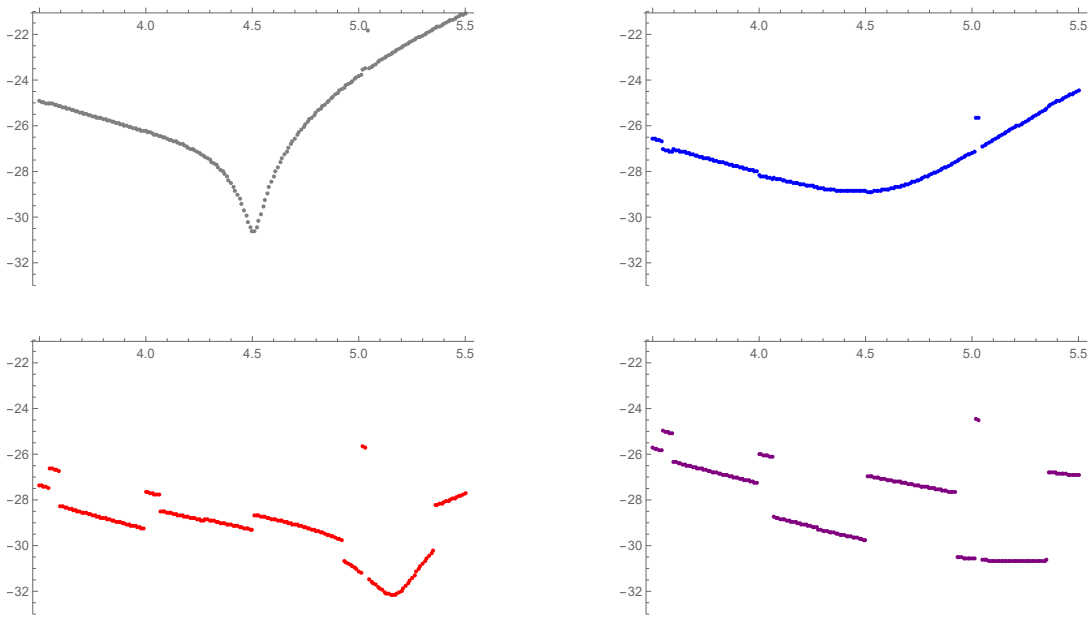


Figure 2.9: Plots of the log of integrated square error (ISE) with respect to δ
(Top left panel: d=7 / Top right panel: d=8 /
Bottom left panel: d=9 / Bottom right panel: d=10)

2.5.2 The Student-t distribution

Let Y be a Student-t random variable with degrees of freedom ν, whose density function is given by

$$f_Y(y;\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\,\pi}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{y^2}{\nu}\right)^{-\frac{\nu+1}{2}},\ -\infty < y < \infty,\ \nu > 0. \tag{2.21}$$

For $\nu > 1$, the ℓth moment of Y is given by

$$\mu_X(\ell) = \begin{cases} 0 & \ell \text{ odd when } 0 < \ell < \nu \\ \frac{1}{\sqrt{\pi}\Gamma(\frac{\nu}{2})}\left[\Gamma(\frac{\ell+1}{2})\Gamma(\frac{\nu-\ell}{2})\nu^{\frac{\ell}{2}}\right] & \ell \text{ even when } 0 < \ell < \nu \end{cases}. \tag{2.22}$$

Approximation via exponential tilting

As the Laplace transform of the Student-t distribution on ν degrees of freedom does not exist and the number of available moment which depends on ν is finite, we consider the half-Student-t distribution and approximate its distribution by utilizing a symmetrization technique, which was discussed in Ha (2007).

As the desity function of Y is symmetric about zero, the transformation $X = |Y|$ yields the half-Student-t density function: $f_X(x) = 2f_Y(x),\ x > 0$. For instance, when $\nu = 1$ and 3, the analytic solutions of the Laplace transform of the half-Student-t distribution are respectively

$$\mathcal{L}_{f_Y(y;\nu=1)}(\theta) = \frac{2\mathrm{Ci}(\theta)\sin(\theta) + (\pi - 2\mathrm{Si}(\theta))\cos(\theta)}{\pi} \tag{2.23}$$

and

$$\mathcal{L}_{f_Y(y;\nu=3)}(\theta) = \frac{2 \times G^{3,1}_{1,3}\left(\frac{3}{4}\theta^2 \,\middle|\, \begin{matrix} \frac{1}{2} \\ 0, \frac{1}{2}, \frac{3}{2} \end{matrix}\right)}{\pi^{\frac{3}{2}}} \tag{2.24}$$

where $\mathrm{Ci}(\theta) = -\int_\theta^\infty \frac{\cos(t)}{t}\,\mathrm{d}t$, and $\mathrm{Si}(\theta) = \int_0^\theta \frac{\sin(t)}{t}\,\mathrm{d}t$ represent the cosine integral and the sine integral respectively, and

$$G^{m,n}_{p,q}\left(z \,\middle|\, \begin{matrix} a_1, \ldots, a_p \\ b_1, \ldots, b_q \end{matrix}\right) = \frac{1}{2\pi i}\int_{\mathcal{L}} \frac{\left(\prod_{j=1}^m \Gamma\left(b_j - s\right)\right)\left(\prod_{j=1}^n \Gamma\left(1 - a_j + s\right)\right)}{\left(\prod_{j=n+1}^p \Gamma\left(a_j - s\right)\right)\left(\prod_{j=m+1}^q \Gamma\left(1 - b_j + s\right)\right)} z^s\,\mathrm{d}s \tag{2.25}$$

denotes Meijer's G-function, where m, n, p, and q are integers with $0 \le m \le q$ and $0 \le n \le p$, $i = \sqrt{-1}$ and the infinite contour of integration $\mathcal{L}$, which represents a suitable closed contour in the complex plane, separates the poles of $\Gamma\left(b_j - s\right)$, $j = 1, 2, \ldots, m$ from the poles of $\Gamma\left(1 - a_j + s\right)$, $j = 1, 2, \ldots, n$.

Here are new representations of the ℓth moment of the tilted half-Student-t distribution. When $\nu = 1$ (also referred to as the half-Cauchy distribution), one has

$$\mu_{X_\theta}(\ell;\theta;\nu = 1) = \begin{cases} 1 & \text{if } \ell = 0 \\ \frac{(\pi - 2\mathrm{Si}(\theta))\sin(\theta) - 2\mathrm{Ci}(\theta)\cos(\theta)}{2\mathrm{Ci}(\theta)\sin(\theta) + (\pi - 2\mathrm{Si}(\theta))\cos(\theta)} & \text{if } \ell = 1 \\ \frac{-2\mathrm{Ci}(\theta)\sin(\theta) + 2\mathrm{Si}(\theta)\cos(\theta) + \frac{2}{\theta} - \pi\cos(\theta)}{2\mathrm{Ci}(\theta)\sin(\theta) + (\pi - 2\mathrm{Si}(\theta))\cos(\theta)} & \text{if } \ell = 2 \\ \frac{G^{3,1}_{1,3}\left(\frac{\theta^2}{4} \,\middle|\, \begin{matrix} \frac{1-\ell}{2} \\ \frac{1-\ell}{2}, 0, \frac{1}{2} \end{matrix}\right)}{2\mathrm{Ci}(\theta)\sin(\theta) + (\pi - 2\mathrm{Si}(\theta))\cos(\theta)} \frac{1}{\sqrt{\pi}} & \text{if } \ell = 3, 4, \ldots, \end{cases} \tag{2.26}$$

and when $\nu = 3$, one has

$$\mu_{X_\theta}(\ell;\theta;\nu=3)=\begin{cases} 1 & \text{if } \ell=0 \\ \dfrac{\sqrt{3}^{\ell}\times G^{3,1}_{1,3}\left(\frac{3}{4}\theta^2 \,\middle|\, \begin{matrix}\frac{1-\ell}{2}\\ 0,\frac{1}{2},\frac{3-\ell}{2}\end{matrix}\right)}{G^{3,1}_{1,3}\left(\frac{3}{4}\theta^2 \,\middle|\, \begin{matrix}\frac{1}{2}\\ 0,\frac{1}{2},\frac{3}{2}\end{matrix}\right)} & \text{if } \ell=1,2 \\ \dfrac{\sqrt{3}^{\ell}\times G^{3,1}_{1,3}\left(\frac{3}{4}\theta^2 \,\middle|\, \begin{matrix}\frac{1-\ell}{2}\\ \frac{3-\ell}{2},0,\frac{1}{2}\end{matrix}\right)}{G^{3,1}_{1,3}\left(\frac{3}{4}\theta^2 \,\middle|\, \begin{matrix}\frac{1}{2}\\ 0,\frac{1}{2},\frac{3}{2}\end{matrix}\right)} & \text{if } \ell=3,4,\ldots \end{cases} \tag{2.27}$$

If the degree of freedom is odd and greater than or equal to three, the ℓth tilted moment is represented by a ratio of Meijer's G-function. However, when the degrees of freedom are even, an explicit solution does not exist so that each tilted moment needs to be evaluated via numerical integration.

As the plot of the tilted half-Student-t density function is decreasing and takes a finite value at $x = 0$, the exponential density function is used as base density, its parameter being estimated from $\mu_{X_\theta}(1)$.

On observing the behavior of $IS\,E_{d,\theta}$ with respect to the tilting parameter, θ, we determined that fixing $\theta = \frac{\nu+1}{\nu+2}$, as the reference point for the tilting parameter generally yields accurate approximants. Exceptionally, when we attempt to approximate the half-Cauchy distribution, for which $\nu = 1$, setting the tilting parameter θ equal to one significantly lowers the $IS\,E_{d,\theta}$. The alternating behavior of the log of $IS\,E_{d,\theta}$ with $\theta = 1, \nu = 1$ is presented in Figure 2.10. Once the approximant of half-Cauchy density, $f_{X_{d,\theta}}(x)$ as specified in Equation (2.4), is obtained after applying the inverse tilting transformation, the approximant of the Cauchy density, which is symmetric, is obtained as follows:

$$f_{Y_{d,\theta}}(x) = \frac{1}{2} f_{X_{d,\theta}}(-x)\ \mathbb{1}_{(-\infty,0)}(x) + \frac{1}{2} f_{X_{d,\theta}}(x)\ \mathbb{1}_{(0,\infty)}(x) \tag{2.28}$$

where $\mathbb{1}_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$.

Figure 2.11 shows the target density superimposed on the approximant consisting of a 37th degree polynomial.

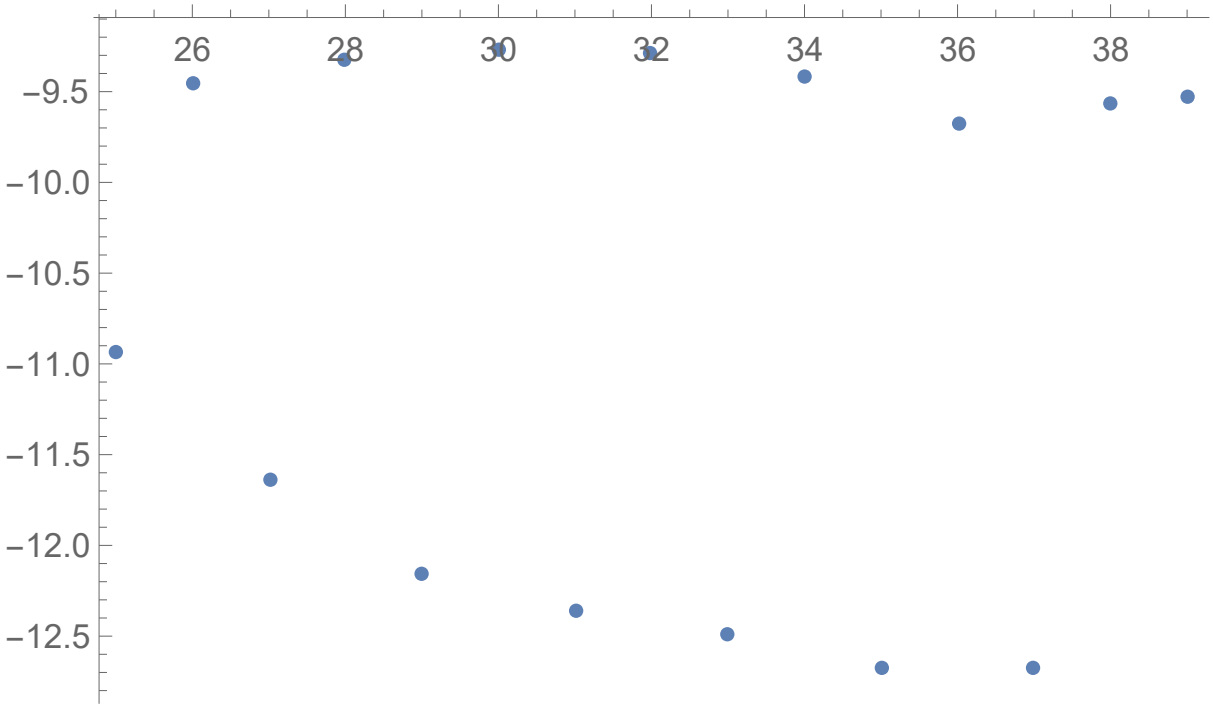
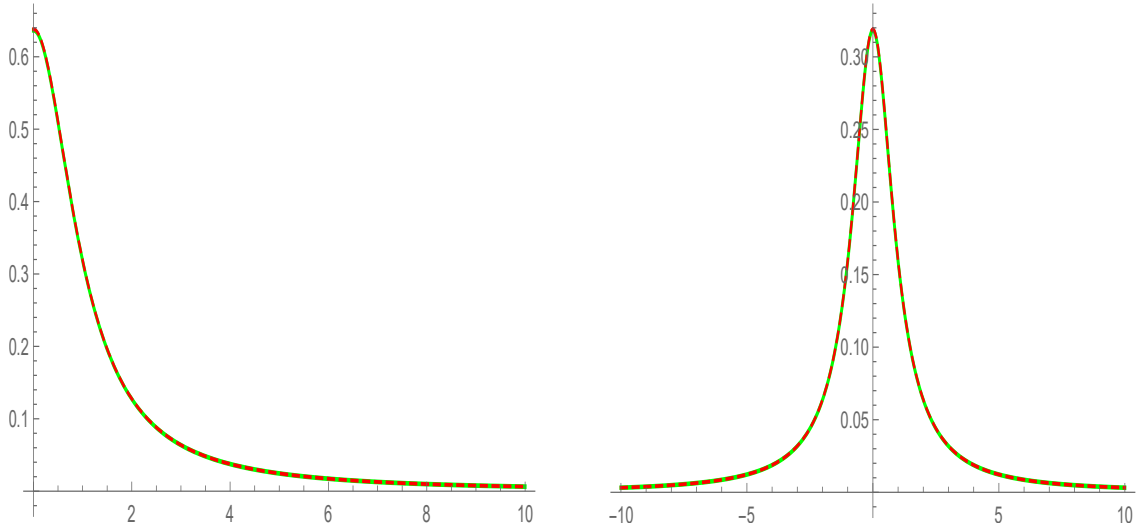


Figure 2.10: Plot of the log of $IS\,E_{d,\theta}$ with $\theta = 1$



(a) Half-Cauchy density

(b) Cauchy density

Figure 2.11: Exact density (green solid line) and $f_{X_{d=37,\theta=1.12}}(x)$ (red dashed line).
$IS\,E = 3.10482 \times 10^{-6}$

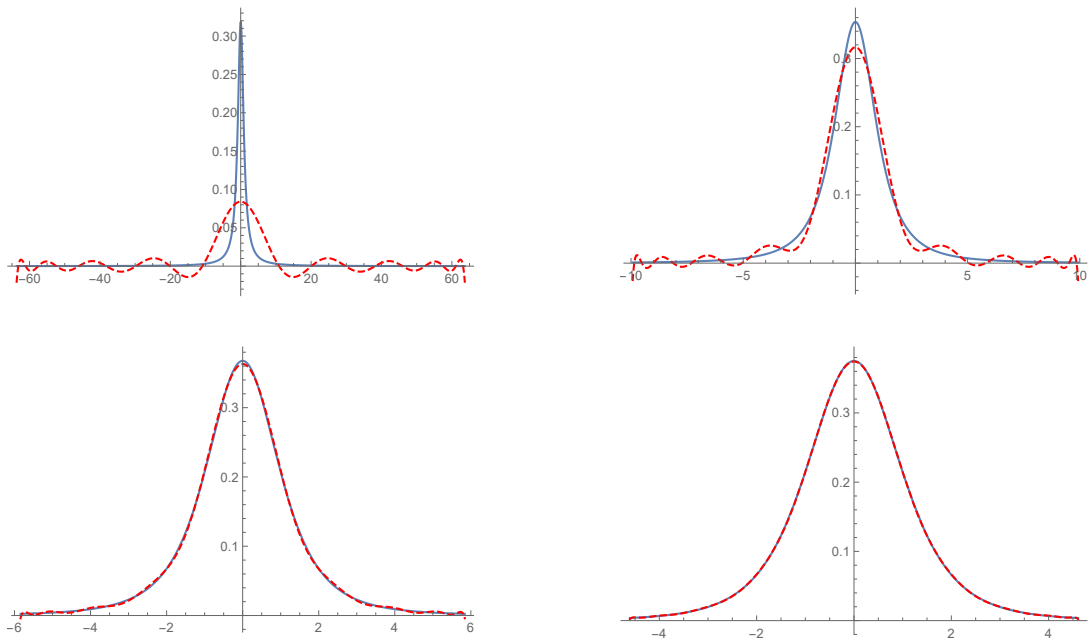


Figure 2.12: Plots of the exact (solid line) and approximant (dashed line) of Student-t density for various degrees of freedom, ν (Top left panel: ν=1 / Top right panel: ν=2 / Bottom left panel: ν=3 / Bottom right panel: ν=4)

Approximating density functions from truncated distributions

We now attempt to approximate the truncated density function of a Student-t random variable by applying the methodology discussed in Section 2.3. Fixing the polynomial degree to 19, plots of the resulting approximants of the Student-t density functions with degrees of freedom 1, 2, 3, and 4 within two-sided 99% probability intervals are presented in Figure 2.12. For low numbers of degrees of freedom, the approximants show oscillating patterns, which is not the case (except when $\nu = 1$) when the symmetrization technique is applied as can be seen from Figure 2.13.

Approximation via transformation of variables

Once the symmetrization technique is applied, the half-Student-t density functions with any degree of freedom, ν, are approximated by making use of the methodology in Section 2.4. The approximants of the Student-t density functions are recovered by using Equation 2.28.

The approximants of the Student-t density functions with degree of freedom 1, 2, 3, and 4, along with $ISE_{d,\delta}$'s are included in Figure 2.14.

2.5.3 Estimating densities from data by means of exponential tilting

This section explains how the density approximation methodology that is based on exponential tilting, which is discussed in Section 2.2, is applied when a sample of observations is available. As the counterpart of dentisy approximants, which make use of the exact tilted moments, the density estimates rely on the sample tilted moments.

Consider an observation vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. In the context of density estimation in conjunction with the exponential tilting, the following points ought to be considered:

1. Before the density is estimated, the sample should be rescaled so that the sample standard deviation be equal to one. Heavy-tailed features are detected from the extreme quantiles of the sample or from the positive-sloped mean excess plot, as discussed by Ghosh and Resnick (2010). To justify the application of the exponential tilting technique, the resulting estimate needs to be compared to that obtained without tilting.

2. The ℓth sample tilted moment with tilting parameter θ, denoted by $m_\theta(\ell)$, is given by

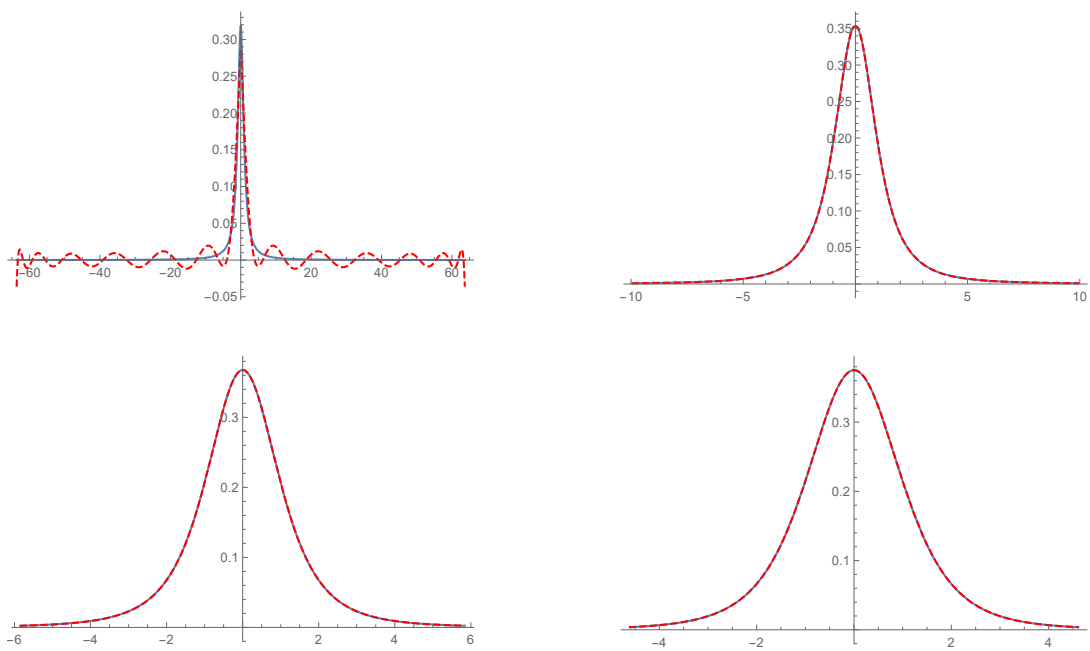
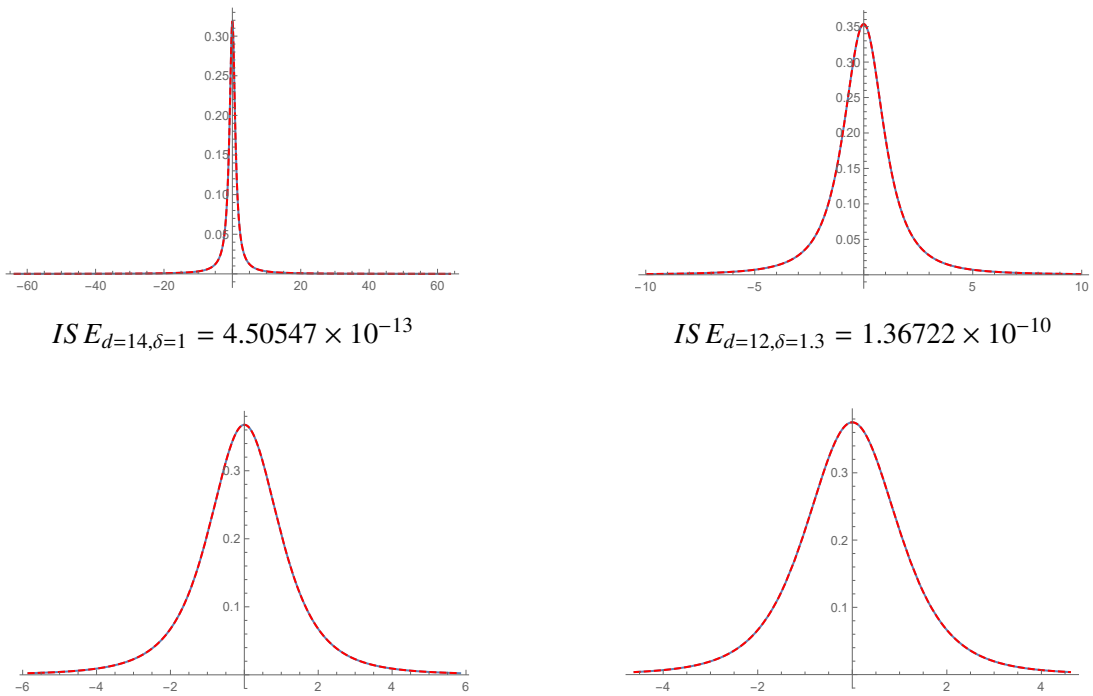


Figure 2.13: Plots of the exact (solid line) and approximant (dashed line) of the truncated Student-t density by applying symmetrization technique (Top left panel: ν=1 / Top right panel: ν=2 / Bottom left panel: ν=3 / Bottom right panel: ν=4)



$IS\,E_{d=14,\delta=1} = 4.50547 \times 10^{-13}$ $IS\,E_{d=12,\delta=1.3} = 1.36722 \times 10^{-10}$

$IS\,E_{d=12,\delta=1.5} = 3.70535 \times 10^{-10}$ $IS\,E_{d=17,\delta=2} = 2.78199 \times 10^{-13}$

Figure 2.14: Plots of the exact (solid line) and approximants (dashed line) of Student-t densities obtained by applying the transformation of variables techniques (Top left panel: ν=1 / Top right panel: ν=2 / Bottom left panel: ν=3 / Bottom right panel: ν=4)

$$m_\theta(\ell) = \frac{\sum_{j=1}^{n} \exp\left(-\theta x_j\right) x_j^{\ell}}{\sum_{j=1}^{n} \exp\left(-\theta x_j\right)}. \tag{2.29}$$

3. A histogram of the data can be utilized to identify an appropriate distribution to be used as base density function. For non-negative samples, gamma or exponential density functions are usually suitable as base densities.

   For the gamma base density function, that is,

$$\psi_X(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, \; x \geq 0, \tag{2.30}$$

   the parameters α and β are estimated by

$$\alpha = \frac{m_\theta(1)^2}{m_\theta(2) - m_\theta(1)^2} \;\text{ and }\; \beta = \frac{m_\theta(2)}{m_\theta(1)} - m_\theta(1). \tag{2.31}$$

If the histogram is more or less symmetric, the symmetrization technique discussed in Section 2.5.2 is applied to the half-distribution; then a gamma or exponential density function can be utilized as base density function, with a shift of location if necessary.

4. The choice of the tilting parameter, θ, as well as the degree of the polynomial adjustment, d, is based on the Anderson-Darling test. Let $AD_{d,\theta}$ denote the Anderson-Darling test statistic evaluated at the resulting cumulative distribution function (CDF) whose associated tilting parameter and polynomial adjustment degree are respectively θ and d, that is,

$$AD_{d,\theta} = -n - \sum_{k=1}^{n} \frac{2k-1}{n} \left[\log\left(1 - F_{d,\theta}(x_{(n-k+1)})\right) + \log(F_{d,\theta}(x_{(k)})\right] \tag{2.32}$$

where n is the sample size, $F_{d,\theta}(x)$ denotes the CDF determined from the density estimate, $f_{d,\theta}(x)$, and $x_{(1)} < x_{(2)} < \cdots < x_{(n)}$ is the ordered sample.

Such a discrepancy measure can be used even when $f_{d,\theta}(x)$ is slightly negative at some sample points. However, the log-likelihood statistics or related discrepancy measures, such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC), are inappropriate when the adjusted density estimates take on negative values.

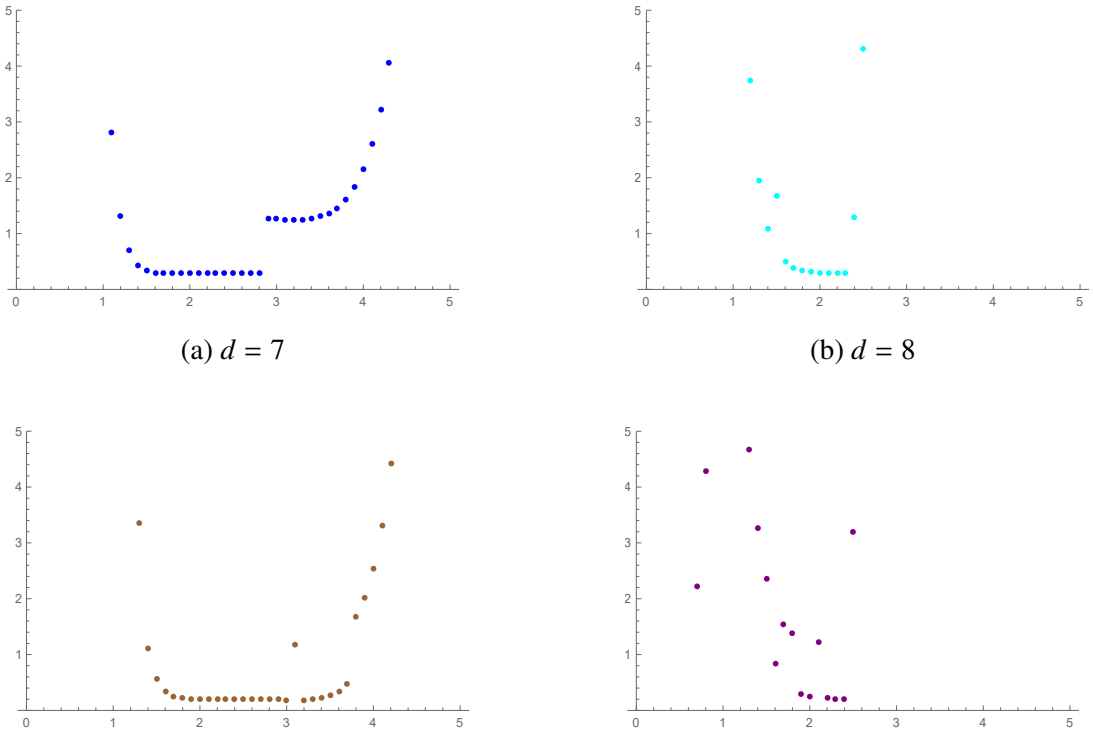
5. From the several candidates of $f_{d,\theta}(x)$ having low $AD_{d,\theta}$ values, it is desirable to choose one that has no root. If, however, we cannot find such a density estimate, then we choose the one having an even number of roots in the right tail. The resulting density estimate can be made *bona fide* by setting the density to be zero wherever it becomes negative and normalizing it.

Simulated data from a type II Pareto distribution

Consider a simulated sample of size 5,000 from a Pareto II $(3.5, 35)$ distribution. Once the values are rescaled by dividing them by their standard deviation, the methodology described in Section 2.2 is applied, except that the exact tilted moments are replaced with the sample tilted moments. In this case, an exponential density function is used as base density function.

Plots of $AD_{d,\theta}$ values for polynomial adjustments of various degrees are shown in Figure 2.15. When taking parsimony into account, the graphs indicate that one can select $d = 9$ and $\theta = 3$, as suitable polynomial degree and tilting parameter, respectively.

The density estimate is obtained after applying the inverse tilting and scale transformations. The resulting density estimate is plotted in Figure 2.16.



(a) $d = 7$ (b) $d = 8$

(c) $d = 9$ (d) $d = 10$

Figure 2.15: Plots of $AD_{d,\theta}$ vs θ leading to the selection of the optimal d and θ

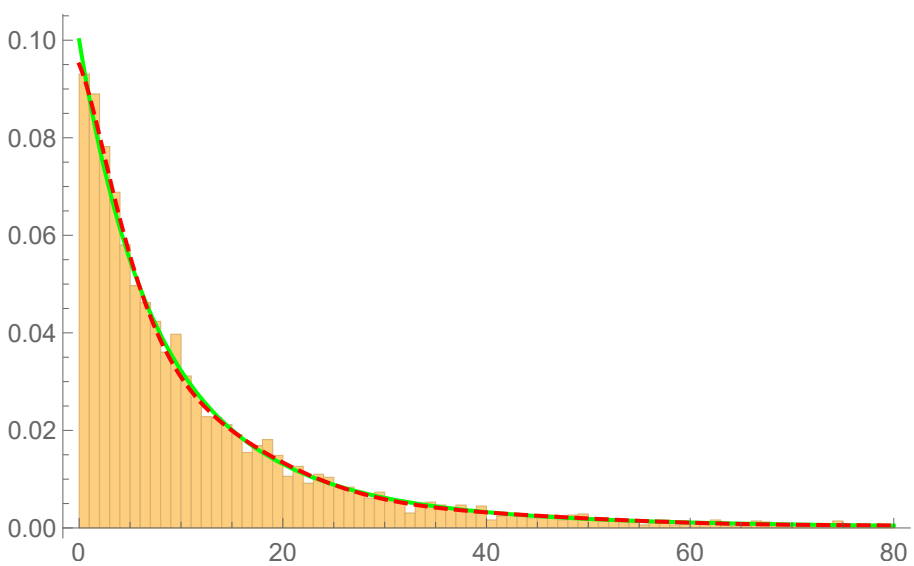


Figure 2.16: Histogram of the simulated data from a Pareto II $(3.5, 35)$ distribution and plots of the exact density (solid line) and the density estimate (dashed line)

κ	Empirical	Pareto (3.5,35)	TEP ($d = 9$ and $\theta = 3$)	BCP ($d = 4$ and $\lambda = 0.1$)
0.8	20.3432	20.4337	20.4421	20.2569
0.9	33.0825	32.5774	32.7712	33.0464
0.95	48.7676	47.3741	48.2939	49.1294
0.99	100.004	95.4658	97.6907	104.338

Table 2.5: VaR_κ from the sample, exact distribution and density estimates

Before comparing the risk measures associated with the density estimates, we introduce the following notations:

1. Tilted Exponential Polynomial (TEP) as the density estimate resulting from the exponential tilting technique, where the exponential density function is used as the base density (d and θ represent the degree of the polynomial adjustment and tilting parameter, respectively).

2. Tilted Gamma polynomial (TGP) as the density estimate resulting from applying the exponential tilting technique, where the gamma density function is used as the base density (d and θ represent the degree of the polynomial adjustment and tilting parameter, respectively).

3. Box-Cox Polynomial (BCP) as the density estimate obtained by applying the Box-Cox transformation in conjunction with a polynomial adjustment. This transformation, introduced by Box and Cox (1964), is defined as

$$y(\lambda) = \begin{cases} \frac{y^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \log y & \text{if } \lambda = 0 \end{cases} . \tag{2.33}$$

   Once the sample is transformed to a more or less symmetric shape by applying Box-Cox transform, the normal density function which is used as base density, is estimated and a polynomial adjustment is applied, as discussed in Section 1.4.1 (d and λ represent the degree of the polynomial adjustment and transformation parameter, respectively).

Once those estimates are back-transformed, we can compare the risk measures, which is done in Tables 2.5 and 2.6 for VaR and TVaR, respectively.

Simulated data from a Student-t distribution

Consider a simulated sample of size 10,000 from a Student-t distribution on 3 degrees of freedom. After rescaling, symmetrizing and applying the exponential tilting technique along with

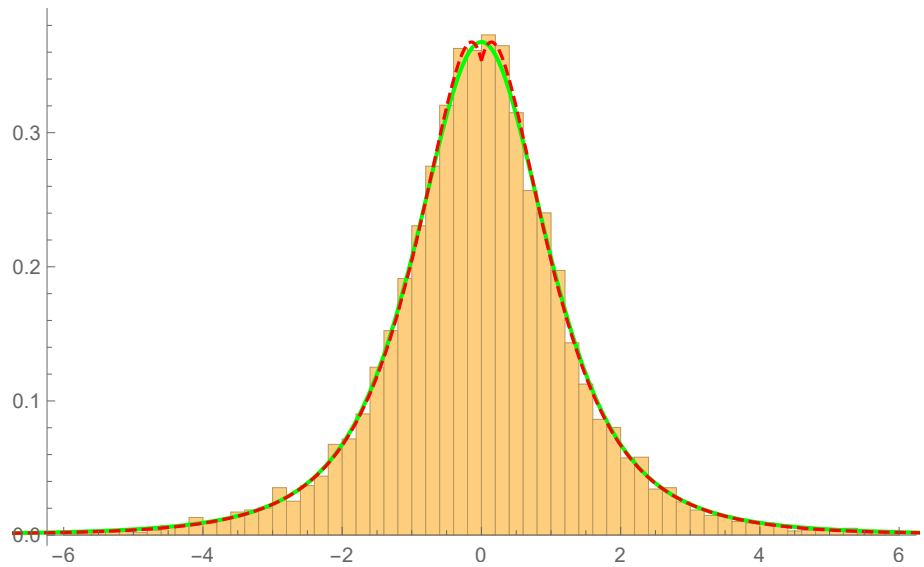


Figure 2.17: Histogram of the simulated data from a Student-t distribution on 3 degrees of freedom, the assumed density (solid line) and the density estimate (dashed line)

making use of an exponential density function as base density for estimating the 'half Student-t' density function, one can select $d = 7$ and $\theta = 0.9$ as the degree of the polynomial adjustment and tilting parameter. The plot of the density estimate superimposed on a histogram of the sample data are included in Figure 2.17.

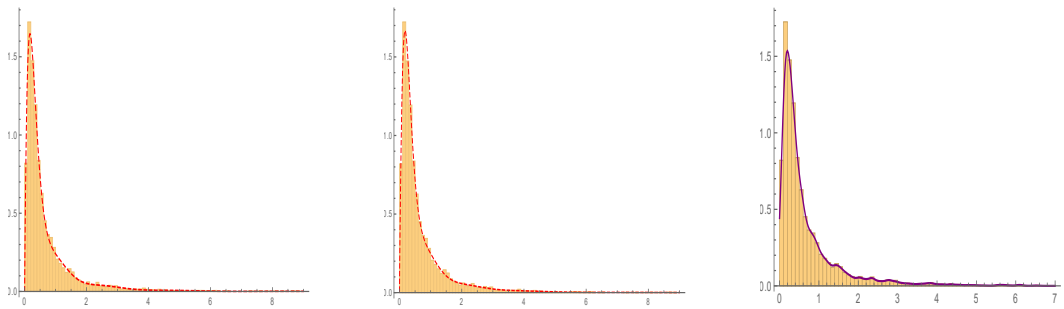
Automobile insurance claims data set

This data set, which was introduced by Frees (2011), consists of 6,773 individual claim amounts for private passenger automobile insurance issued by a large US mid-western property and casualty insurance company. This data set was originally used to examine claim distributions for several risk classification categories.

The summary statistics for the rescaled data whose standard deviation is equal to one are

κ	Empirical	Pareto (3.5,35)	TEP ($d = 9$ and $\theta = 3$)	BCP ($d = 4$ and $\lambda = 0.1$)
0.8	44.9521	42.6072	42.6498	44.6046
0.9	64.2477	59.6042	59.7103	63.5354
0.95	89.0128	80.3238	79.9156	87.2058
0.99	182.639	147.652	137.345	166.829

Table 2.6: TVaR_κ from the sample, exact distribution and density estimates



(a) TEP ($d = 11, \theta = 1.1$) $AD = 0.592737$

(b) TGP ($d = 9, \theta = 1.5$) $AD = 0.409576$

(c) Kernel density estimate $AD = 24.6244$

Figure 2.18: Density estimates and rescaled histogram for the auto insurance claims data set

presented in Table 2.7. The heavy-tailed nature of the sample is clearly suggested by the high kurtosis measure as well as an extreme maximum observed value.

Interestingly, after the proposed density estimation methodology is applied including the exponential tilting and several $AD_{d,\theta}$ values are compared, both a TEP ($d = 11, \theta = 1.1$) estimate and a TGP ($d = 9, \theta = 1.5$) estimate provide suitable models fits as can be seen from Figure 2.18. We note that the exponential tilting is also useful in the context of density estimation since otherwise the tail of the resulting density estimate presents undesirable fluctuations with subintervals where the estimates are negative. A kernel estimate is also included in this figure for comparison purposes, indicating that it is outperformed by the proposed estimates.

Additionally, neither density estimate has a root in the right tail. VaR and TVaR are compared in Tables 2.8 and 2.9, respectively.

The Danish fire data set

The Danish fire data set is often used in applications involving extreme value theory; see for instance McNeil (1997) and Resnick (1997). Collected at Copenhagen Reinsurance from 1980

Sample size	6773	Min	0.70008
Mean	0.70008	1st quartile	0.19787
Skewness	6.23567	2nd quartile	0.37844
Kurtosis	87.2775	3rd quartile	0.80751
		Max	22.6679

Table 2.7: Summary of the rescaled automobile insurance claims data set (Original standard deviation: 2,646.91)

to 1990, the data consists of inflation-adjusted total fire losses, expressed in millions of Danish Krone. Although the sample size of the full data set is 2,492, we confine our attention to 2,156 losses exceeding one million Krone, taking into account the insurance deductible.

As various density estimate candidates are suggested, we describe each one, along with the determination of the parameters, such as the degree of polynomial adjustment (d), the tilting parameter (θ) and the Box-Cox transform parameter (λ).

1. **Density I: TEP from the rescaled data**

   Table 2.10 provides summary statistics for the rescaled data, which are clearly indicative of a heavy-tailed distribution. After comparing the density estimates obtained with and without exponential tilting, along with the choice of a gamma or exponential density functions as base densities, a TEP ($d = 9, \theta = 1.15$) estimate which has no root in the right tail is a possible candidate. However, the corresponding Anderson-Darling test statistic, which is equal to 44.29 is large due to the fact that the deductible was not taken into consideration. As Figure 2.19 (a) reveals, the back-transformed density estimate does not provide a good fit around the mode of the distribution.

2. **Density II: Location-shifted TEP from the rescaled data**

   This shortcoming is circumvented by making use of a shifted density estimate. Given a deductible level, d_0, and denoting by σ the standard deviation of the sample $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, where $x_i > d_0$ for $i = 1, 2, \ldots, n$, the linear transformation $\mathbf{y} = \frac{\mathbf{x}-d_0}{\sigma}$

κ	Empirical	TEP ($d = 11, \theta = 1.1$)	TGP ($d = 9, \theta = 1.5$)	BCP ($d = 7, \lambda = 0$)
0.8	2545	2591.97	2576.89	2534.5
0.9	4171.01	4159.98	4185.58	4224.53
0.95	6356.9	6468.07	6356.05	6358.44
0.99	12091.5	12413.1	11971.5	12668.8

Table 2.8: VaR_κ based on the auto insurance claims data

κ	Empirical	TEP ($d = 11, \theta = 1.1$)	TGP ($d = 9$ and $\theta = 1.5$)	BCP ($d = 7$ and $\lambda = 0$)
0.8	5518.88	5452.73	5585.01	5521.61
0.9	7772.38	7641.58	8028.22	7800.34
0.95	10415.8	10156.1	10609.9	10465
0.99	18263.7	17027.9	19404.8	18070.7

Table 2.9: TVaR_κ based on the auto insurance claims data

Sample size	2156	Min	0.117608
Mean	0.398392	1st quartile	0.156057
Skewness	18.7093	2nd quartile	0.208751
Kurtosis	483.462	3rd quartile	0.34858
		Max	30.781

Table 2.10: Summary of the rescaled Danish fire data
(Original standard deviation: 8.52743)

ensures that the rescaled sample vector has a minimum value that is close to zero. The minimizer of the Anderson-Darling test statistic is determined and no roots are present in the right tail of the resulting density estimate that was obtained with exponential tilting from an exponential density function as the base density, its parameters being $\hat{d} = 5$ and $\hat{\theta} = 4.8$, which yields $AD_{\hat{d},\hat{\theta}} = 11.2854$.

3. **Density III: Location-shifted TGP from the rescaled data**

   When the gamma density function is utilized as base density for determining a density estimate with exponential tilting, the fit of the distribution is further improved as the $AD_{\hat{d},\hat{\theta}}$ value is then 0.730527 with $\hat{d} = 9$ and $\hat{\theta} = 3.6$.

4. **Density IV: BCP from the rescaled data**

   Once the methodology for estimating the BCP that is described in Section 6.3.1 is applied, in which case the Box-Cox transform parameter takes on the value -1 and a degree 5 for the polynomial adjustment is suitable.

5. **Density V: Exponential polynomial from the log-scaled data**

   As the deductible level, d_0, is one, other possible density estimates result from the log-scaled data. The summary statistics for the log-scaled data that are included in Table 2.11 suggest a light-tailed distribution, in which case exponential tilting may not be necessary. Using the exponential density function as base density in conjunction with a polynomial adjustment of degree 9 results in two roots in the right tail. Although the Anderson-Darling test statistics are slightly smaller with adjustments of larger degrees, the number of roots in the tail then increases, which would preclude choosing a higher polynomial adjustment degree. *Bona fide* density estimates can be obtained by following the methodology described in Chapter 4.

6. **Density VI: TEP from the log-scaled data**

Sample size	2156	Min	0.00288882
Mean	0.790965	1st quartile	0.285752
S.D.	0.716333	2nd quartile	0.576672
Skewness	1.76413	3rd quartile	1.0894
Kurtosis	7.1908	Max	5.57311

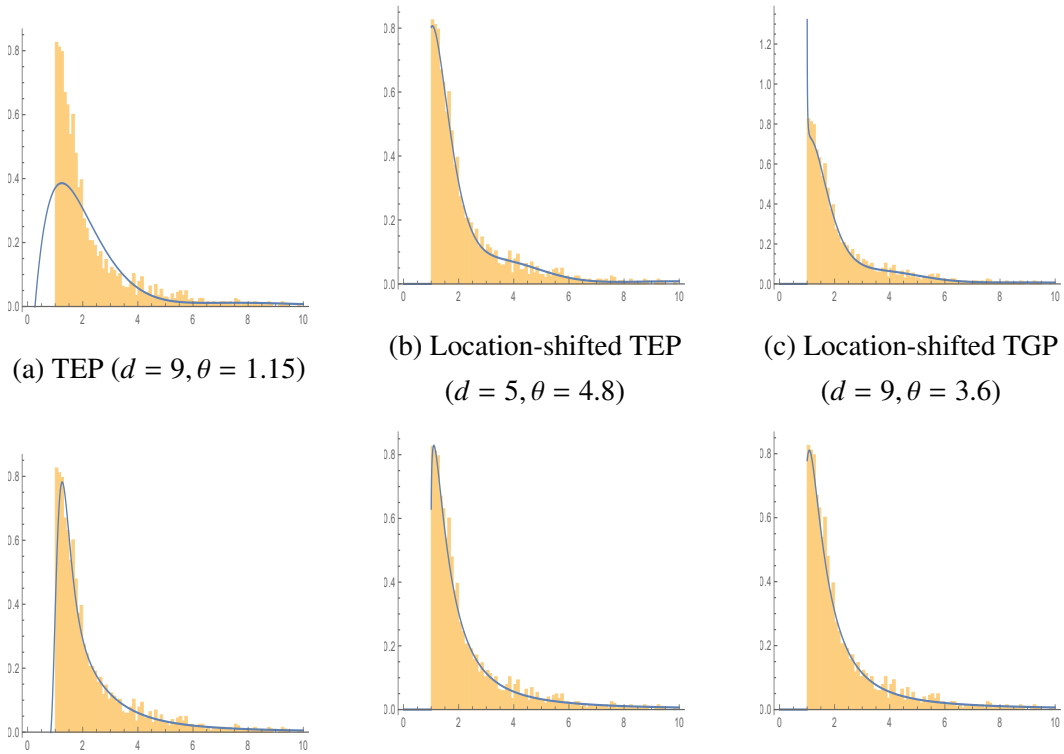
Table 2.11: Summary of the log-scaled Danish fire data

A density estimate with exponential tilting for which $d = 7$ and $\theta = 0.7$ is considered and its VaR is compared to those associated with aforementioned density candidates. Table 2.12 presents specific VaR values associated the various density estimates. Figure 2.19 includes plots of the density estimates superimposed of a histogram of the original data.

In light of the Anderson-Darling test statistics, the VaR values appearing in Table 2.11 and parsimony considerations, density estimate VI could be viewed as being the most suitable.

κ	Empirical	Density II	DensityIII	Density IV	Density V	Density VI
0.8	3.48145	3.35407	3.51668	3.32754	3.4303	3.32754
0.9	5.56174	4.87409	5.35254	5.11591	5.70866	5.73878
0.95	10.0111	7.20204	9.32785	7.94532	9.50778	9.7081
0.99	26.2146	12.9562	22.1245	26.3172	27.7974	27.7927

Table 2.12: VaR_κ values obtained from the sample and the density estimates



(a) TEP ($d = 9, \theta = 1.15$)

(b) Location-shifted TEP ($d = 5, \theta = 4.8$)

(c) Location-shifted TGP ($d = 9, \theta = 3.6$)

(d) BCP ($d = 7, \lambda = -1$)

(e) Exponential polynomial ($d = 9$) from log-scaled data

(f) TEP ($d = 7, \delta = 0.7$) from log-scaled data

Figure 2.19: Density estimates and histogram of the original data for the Danish fire data set

Bibliography

[1] Box, G. E. P. and Cox, D. R. An analysis of transformations. (1964) *Journal of the Royal Statistical Society. Series B (Methodological)* **26**(2), 211–252.

[2] Bühlmann, H. (1980). An economic premium principle. *ASTIN Bulletin: The Journal of the IAA* **11**(1), 52–60.

[3] Cox, S. H., Lin, Y., and Wang, S. (2006). Multivariate exponential tilting and pricing implications for mortality securitization. *Journal of Risk and Insurance* **73**(4), 719–736.

[4] Elliott, R. J., Chan, L., and Siu, T. K. (2005). Option pricing and Esscher transform under regime switching. *Annals of Finance* **1**(4), 423–432.

[5] Esscher, F. (1932). On the probability function in the collective theory of risk. *Skandinavisk Aktuarietidskrift* **15**(3), 175–195.

[6] Frees, E. W. (2011). *Regression modeling with actuarial and financial applications*. Cambridge University Press, New York.

[7] Gerber, H. U. and Shiu, E. S. W. (1994). Option pricing by Esscher transforms. *Transactions of the Society of Actuaries* **46**, 99–191.

[8] Ghosh, S. and Resnick, S. (2010) A discussion on mean excess plots. *Stochastic Processes and their Applications* **120**(8), 1492–1517.

[9] Ha, H. (2007). Moment–based density approximation algorithm for symmetric distributions. *The Korean Communications in Statistics* **14**(3), 583–592.

[10] Ha, H. and Provost, S. B. (2007). A viable alternative to resorting to statistical tables. *Communications in Statistics Simulation and Computation* **36**, 1135–1151.

[11] McNeil, A. (1997). Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin: The Journal of the IAA* **27**(1), 117–137.

[12] Nadarajah, S. and Kotz, S. (2006). On the Laplace transform of the Pareto distribution. *Queueing Systems* **54**, 243–244.

[13] Provost, S. B. (2005). Moment–based density approximants. *The Mathematica Journal* **9**, 727–756.

[14] Resnick, S. I. (1997). Discussion of the Danish data on large fire insurance losses. *Astin Bulletin* **27**(1), 139–151.

[15] Siegmund, D. (1976). Importance sampling in the Monte Carlo study of sequential tests. *The Annals of Statistics* **4**(4), 673–684.

Chapter 3

Novel Approaches for Estimating Distributional Endpoints

3.1 Introduction

Given a continuous distribution function F, its left and right endpoints are defined as $L := \inf\{x : F(x) > 0\}$ and $U := \sup\{x : F(x) < 1\}$. Various statistical methods have been proposed for estimating endpoints, particularly $U < +\infty$. Hall (1982) considered the maximum likelihood estimate which is expressed in terms of the largest order statistics. Hall and Wang (1999) proposed minimum-distance estimates, which include Greenwood's statistic. Estimates obtained from maximizing a semiparametric likelihood in a Bayesian context were explored by Hall and Wang (2005). Estimates involving high-order moments and normal measurement errors were discussed by Girard *et al.* (2012) and Leng *et al.* (2018). Confidence intervals for endpoint estimates were investigated by Loh (1984) and Li *et al.* (2011).

The estimation of endpoints is also closely linked to extreme value theory. In fact, an endpoint is finite if the generalized extreme value distribution belongs to the reverse Weibull domain. Caers and Maes (1998) determined endpoint estimates by making use of a tail-weighted quantile function. Alves and Nevas (2014) proposed a technique for estimating a finite endpoint in the Gumbel domain. As the behavior of the tails has to be taken account, endpoint estimates can be obtained as a compromise between Hill's estimator and a linear combination of order statistics; this is extensively discussed in Resnick (2007).

Such estimates have been utilized in various applications, including the maximum life span of human beings discussed by Aarssen and De Haan (1994) and Alves *et al.* (2017), records in athletics investigated by Einmahl and Magnus (2008) and Hall (1999), and production frontiers analyzed by Cazals *et al.* (2002).

The aforementioned methods of estimating endpoints are based on certain underlying distributional assumptions, such as expressing a density function $f(x)$ as $x^k L(x)$ where k is real number and $L(x)$ is slowly varying function, or presuming a specific type of the extreme value distribution as underlying density function.

As the underlying distribution is usually unknown in the case of a sample, it can prove challenging to determine endpoints, especially in the case of heavy-tailed distributions. This chapter investigates alternative approaches for estimating endpoints, which are based on the entire sample of observations and its moments.

We propose to determine empirical endpoints by making use of the empirical saddlepoint approximation method, which is described in Section 1.4.2. Some computational issues encountered in evaluating the empirical saddlepoint estimates are discussed in Section 3.2. As explained in Section 3.3, this method can be utilized in conjunction with the Box-Cox transform for samples consisting of non-negative observations. As well, endpoint estimates based on the 4-parameter generalized beta distribution are proposed in Section 3.4. In Section 3.5, we conduct simulation studies and determine endpoints by making use of each method. We initially generated samples from known distributions to elicit procedures for determining endpoints.

3.2 Computational issues related to the empirical saddlepoint approach

One of the computational challenges associated with empirical saddlepoint estimation resides in determining a solution $\hat{s}$ as specified in Section 1.4.2, for values of x lying outside the range of the sample of observations. In connection with this problem, Fasiolo *et al.* (2018) proposed extended empirical saddlepoint approximations that are expressed as an adjusted weighted average of the empirical saddlepoint approximation and the theoretical cumulant generating function of a multivariate normal distribution. However, this method is valid only if the normality assumption, which can be inferred to some extent from the summary statistics, holds.

We make use of a numerical solution that relies on numerous iterations of the Newton-Raphson method to determine the empirical saddlepoint such that $\hat{K}_n{}'(\hat{s}) = x$ where $\hat{K}_n(s)$ denotes the empirical cumulant generating function defined in Equation (1.13). Another issue associated with this estimate stems from the fact that the Jacobian of the second derivative of the cumulant generating function can be singular at some points located outside the range of the sample. Although $\hat{K}_n{}''(s)$ cannot be equal to zero theoretically, it can be zero numerically for some points lying outside of the range of the sample, which makes it impossible to evaluate

Sample size	L	$x_{(1)}$	$x_{(n)}$	U
500	−3.08013	−2.52098	2.80998	3.98349
2,500	−3.92801	−3.33604	3.54335	4.81975
10,000	−5.22163	−4.22639	3.79761	5.04246

Table 3.1: Summary of endpoints and the extreme order statistics based on the empirical saddlepoint estimate (Data simulated from a standard normal distribution)

the empirical saddlepoint estimates at such points.

The procedure for estimating endpoints can be described as follows: First, the original sample values are rescaled by dividing them by the sample standard deviation. When the standard deviation of the original data set is large, many missing points are encountered while evaluating empirical saddlepoint estimates. To mitigate this issue, rescaling is advisable before determining empirical saddlepoint estimates for points lying outside the range of the sample.

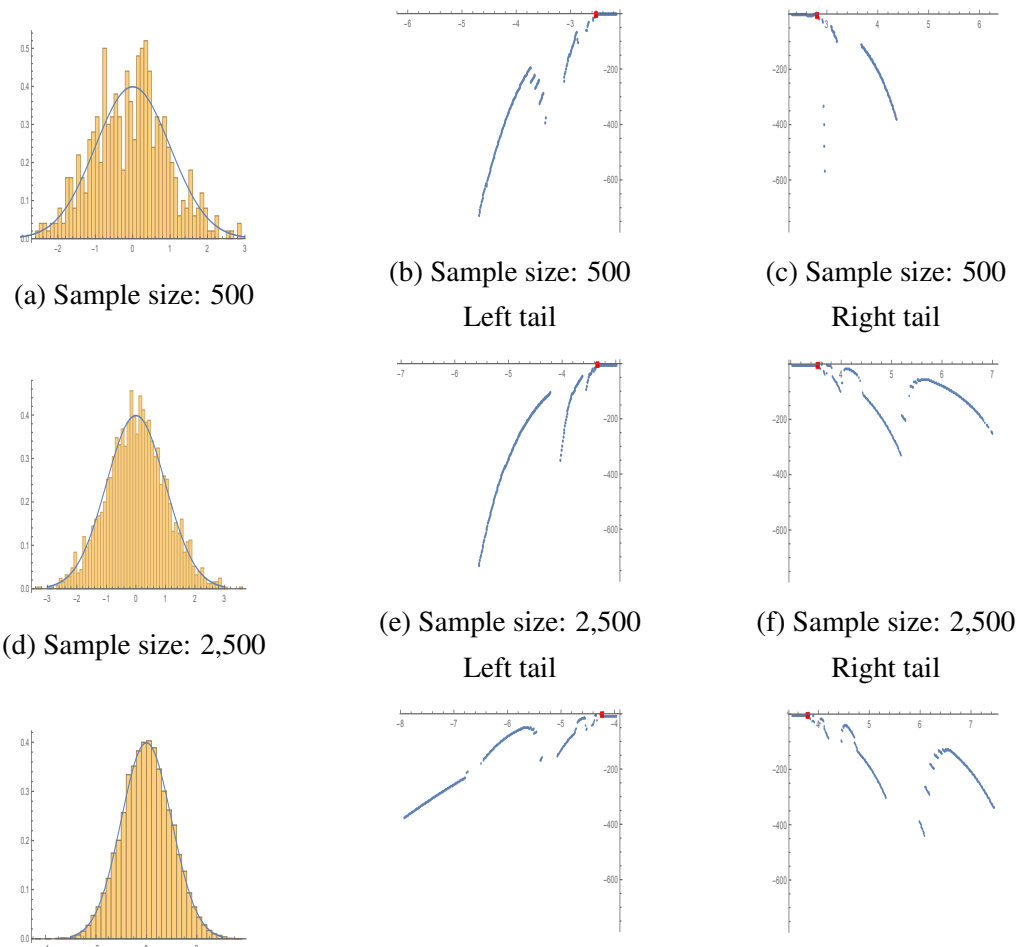
Notwithstanding these remarks, the patterns of the logarithm of the empirical saddlepoint estimates determined for points lying outside of the range of the sample of observations provide an indication for obtaining empirical endpoints. For illustrative purposes, consider samples of sizes 500, 2,500, and 10,000, that are simulated from a standard normal distribution.

Once the sample is rescaled, we investigate the behavior of the logarithm of the empirical saddlepoint estimates. Figure 3.1 presents the plots of those estimates which are transformed back to the original scale, along with the minimum and maximum values of the sample, which are indicated by red dots. The empirical endpoint is chosen to be the value of the support at which the density estimate as determined from the first discernible arc of the empirical saddlepoint estimates (interpolated or extrapolated values thereof) are of the order $e^{-200} \approx 10^{-87}$. The same tolerance level will be used in all the applications. Table 3.1 includes the empirical endpoint estimates, L and U, along with the extreme order statistics.

3.3 Combining the Box-Cox transform and the empirical saddlepoint estimation method

The Box-Cox transform, as defined in Equation (2.33), is a useful tool for estimating the empirical endpoints since the transformed data and the plot of the logarithm of the estimate are then more symmetrical.

The following procedure is implemented for estimating empirical endpoints:



(a) Sample size: 500

(b) Sample size: 500
Left tail

(c) Sample size: 500
Right tail

(d) Sample size: 2,500

(e) Sample size: 2,500
Left tail

(f) Sample size: 2,500
Right tail

(g) Sample size: 10,000

(h) Sample size: 10,000
Left tail

(i) Sample size: 10,000
Right tail

Figure 3.1: Histograms of the data simulated from a standard normal distribution and plots of the logarithm of the empirical saddlepoint estimates obtained from the rescaled data, the red dots indicating the minimum and maximum values of the rescaled data

Algorithm 3.1

1. Apply the Box-Cox transform as specified by Equation (2.33). The parameter, λ, is such that the following profile log likelihood function, $l(\lambda)$, is maximized:

$$\ell(\lambda) = -\frac{n}{2}\log\left(2\pi\,\sigma^2(\lambda)\right) - \frac{1}{2}\sum_{i=1}^{n}\frac{(x_i(\lambda)-\mu(\lambda))^2}{\sigma^2(\lambda)} + (\lambda-1)\sum_{i=1}^{n}\log x_i \tag{3.1}$$

where x_i, $x_i(\lambda)$, $\mu(\lambda)$, $\sigma^2(\lambda)$ represent the ith observed sample value, the corresponding transformed value, the mean of the transformed sample and the variance of the trans-

formed sample, respectively.

2. Rescale the transformed data by dividing it by $\sigma(\lambda)$.

3. Evaluate the logarithm of the empirical saddlepoint estimates lying outside of the range of the transformed data resulting from applying Steps 1 and 2.

4. Choose the empirical endpoints such that the estimates in Step 3 meet a tolerance level of e^{-200}, while taking into account that the transformed support is

$$(0,\infty) \rightarrow \begin{cases} \left(-\frac{1}{\lambda\sigma^2(\lambda)}, \infty\right) & \text{if } \lambda > 0 \\ (-\infty, \infty) & \text{if } \lambda = 0 \\ \left(-\infty, -\frac{1}{\lambda\sigma^2(\lambda)}\right) & \text{if } \lambda < 0 \end{cases} . \tag{3.2}$$

   If an empirical endpoint obtained from the Box-Cox-rescaled data lies beyond a lower or upper bound as specified by Equation (3.2), which is indicative of a significant concentration of the data in the neighborhood of the minimum or maximum value of the sample, an empirical endpoint that is solely based on the rescaled data (using the same tolerance) is sought.

5. Apply the back-transformation, based on the intervals specified in Equation (3.2), so that the empirical endpoints can be identified on the original scale.

   It should be pointed out that the resulting value, which is referred to as an empirical endpoint, is not an endpoint per say but a point beyond which the distribution is of no practical significance. The further an empirical endpoint lies in the tail a rescaled distribution, the heavier the tail.

   It should also be noted that e^{-200} is a tolerance that proves suitable for the various distributions so far considered. Of course, one may set a lower or higher tolerance that might better suit one's own purpose.

Samples simulated from various types of distributions, including a heavy-tailed one, are used to illustrate this methodology in Section 3.5.

3.4 Endpoints estimates based on the 4-parameter generalized beta distribution

Consider the 4-parameter generalized beta distribution, denoted by GB(α, β, L, U), whose density function is given by

$$f(x) = \frac{1}{\mathrm{B}(\alpha,\beta)} \frac{(x-L)^{\alpha-1}\,(U-x)^{\beta-1}}{(U-L)^{\alpha+\beta-1}} \quad \text{for } L \le x \le U, \tag{3.3}$$

where $\mathrm{B}(\alpha,\beta) = \frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}$, the shape parameters α and β being positive, and for instance $\Gamma(v) = \int_0^\infty t^{v-1}\,e^{-t}\,\mathrm{d}t$. This approach provides estimated endpoints specified by the generalized beta distribution which is utilized as a distributional model as opposed to the empirical endpoints obtained from the saddlepoint approach. As a result, particularly large values are obtained as estimated endpoints in the case of heavy-tailed distributions whose tails converge very slowly to zero.

There exist two main methods for estimating the parameters of this distribution: the penalized maximum likelihood estimation (PMLE) method and the feasibility-constrained moment matching (FCMM) technique. When the distribution is very skewed, it is advisable to apply the Box-Cox transform prior to estimating the parameters.

3.4.1 Penalized maximum likelihood estimation (PMLE) method

Letting the ordered observations be $x_{(1)} < x_{(2)} < \cdots < x_{(n)}$ where n is the sample size, Wang (2005) added penalization components to the likelihood function so that $L \le x_{(1)}$, $U \ge x_{(n)}$, and α and β remain positive. Given the $x_{(i)}$'s, $i = 1,\ldots,n$, the original and penalized likelihood functions, denoted by l_o and l_p, respectively, are

$$\begin{aligned} l_o(\alpha,\beta,L,U) = &-n\ \log(\mathrm{B}(\alpha,\beta)) - n\,(\alpha+\beta-1)\log(U-L) \\ &+ (a-1)\left[\sum_{i=1}^{n}\log(x_{(i)}-L)\right] + (b-1)\left[\sum_{i=1}^{n}\log(U-x_{(i)})\right] \end{aligned} \tag{3.4}$$

and

$$l_p(\alpha,\beta,L,U) = l_o(\alpha,\beta,L,U) + \log\left(\frac{x_{(1)}-L}{x_{(2)}-L}\right) + \log\left(\frac{U-x_{(n)}}{U-x_{(n-1)}}\right). \tag{3.5}$$

The parameters are estimated by solving the constrained maximization problem:

$$\begin{aligned} &\max_{\alpha,\beta,L,U} l_p(\alpha,\beta,L,U) \\ &\text{subject to } \alpha > 0, \beta > 0, L \le x_{(1)}, \text{and } U \ge x_{(n)}. \end{aligned} \tag{3.6}$$

Making use of the previously simulated samples of sizes 500, 2,500, and 10,000 from the standard normal distribution, the estimated values of the endpoints obtained with the PMLE method are included in Table 3.2.

Sample size	L	$x_{(1)}$	$x_{(n)}$	U
500	−4.12643	−2.52098	2.80998	5.16627
2,500	−15.5561	−3.33604	3.54335	19.1293
10,000	−72.4937	−4.22639	3.79761	19.2211

Table 3.2: Summary of endpoints and the extreme order statistics based on fitting a generalized beta distribution with the PMLE method (Data simulated from a standard normal distribution)

3.4.2 Feasibility-constrained moment matching (FCMM) technique

AbouRizk *et al.* (1994) proposed several methods for estimating the parameters of the generalized beta distribution. One of them referred to as the FCMM technique provides computational efficiency as well as a close match to the central moments.

Consider a random variable X that is distributed as $\mathrm{GB}(\alpha, \beta, L, U)$. Then the mean, variance, skewness, and kurtosis of X can be expressed as follows:

$$\mu = E(X) = L + (U - L)\frac{\alpha}{\alpha + \beta} \tag{3.7}$$

$$\sigma^2 = Var(X) = (U - L)^2 \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \tag{3.8}$$

$$\gamma_1 = E\left[\left(\frac{X - \mu}{\sigma}\right)^3\right] = \frac{2(\beta - \alpha)\sqrt{\alpha + \beta + 1}}{(\alpha + \beta + 1)\sqrt{\alpha\beta}} \tag{3.9}$$

$$\gamma_2 = E\left[\left(\frac{X - \mu}{\sigma}\right)^4\right] = \frac{3(\alpha + \beta + 1)[2(\alpha + \beta)^2 + \alpha\beta(\alpha + \beta - 6)]}{\alpha\beta(\alpha + \beta + 2)(\alpha + \beta + 3)} \tag{3.10}$$

Let $\bar{x}, s^2, b_1, b_2$ denote the sample mean, variance, skewness and kurtosis, which are given by

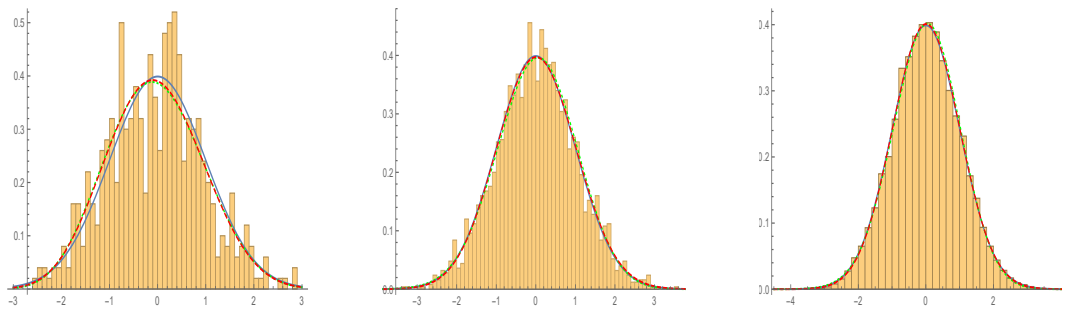
$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3.11}$$

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{3.12}$$

$$b_1 = \frac{1}{n}\sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s}\right)^2 \tag{3.13}$$

Sample size	L	$x_{(1)}$	$x_{(n)}$	U
500	−4.82484	−2.52098	2.80998	6.03919
2,500	−15.576	−3.33604	3.54335	19.1591
10,000	−16.8676	−4.22639	3.79761	17.9655

Table 3.3: Summary of endpoints and the extreme order statistics based on fitting a generalized beta distribution with the FCMM technique (Data simulated from a standard normal distribution)



(a) Sample size: 500 (b) Sample size: 2,500 (c) Sample size: 10,000

Figure 3.2: Histograms of the data simulated from a standard normal distribution and density estimates obtained by fitting a generalized beta distribution (solid line: underlying density; dotted line: PMLE method; dashed line: FCMM technique)

$$b_2 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^4 \tag{3.14}$$

where the x_i's and n denote the observations and the sample size, respectively.

As the moments in Equation (3.7)-(3.10) are functions of the four parameters, we estimate those parameters by solving the constrained non-linear optimization problem:

$$\min_{\alpha,\beta,L,U} (\mu(\alpha,\beta,L,U) - \bar{x})^2 + \left(\sigma^2(\alpha,\beta,L,U) - s^2\right)^2 + (\gamma_1(\alpha,\beta) - b_1)^2 + (\gamma_2(\alpha,\beta) - b_2)^2$$
$$\text{subject to } \alpha > 0,\ \beta > 0,\ L \le x_{(1)}, \text{ and } U \ge x_{(n)}. \tag{3.15}$$

The estimated values of the endpoints obtained by applying the FCMM technique to the data simulated from the standard normal distribution are included in Table 3.3. The plots of the generalized beta density estimates are shown in Figure 3.2, superimposed on a histogram of the data.

3.5 Simulation studies

We generate samples from known distributions and apply the methodologies herein advocated to obtain empirical and estimated endpoints.

3.5.1 A beta distribution

Let Beta (α, β) denote a beta distribution whose density function is given by

$$f(x) = \frac{x^{\alpha-1}\,(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}, \quad 0 \le x \le 1, \tag{3.16}$$

where both α and β are shape parameters and $\mathrm{B}(\alpha,\beta) = \frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.

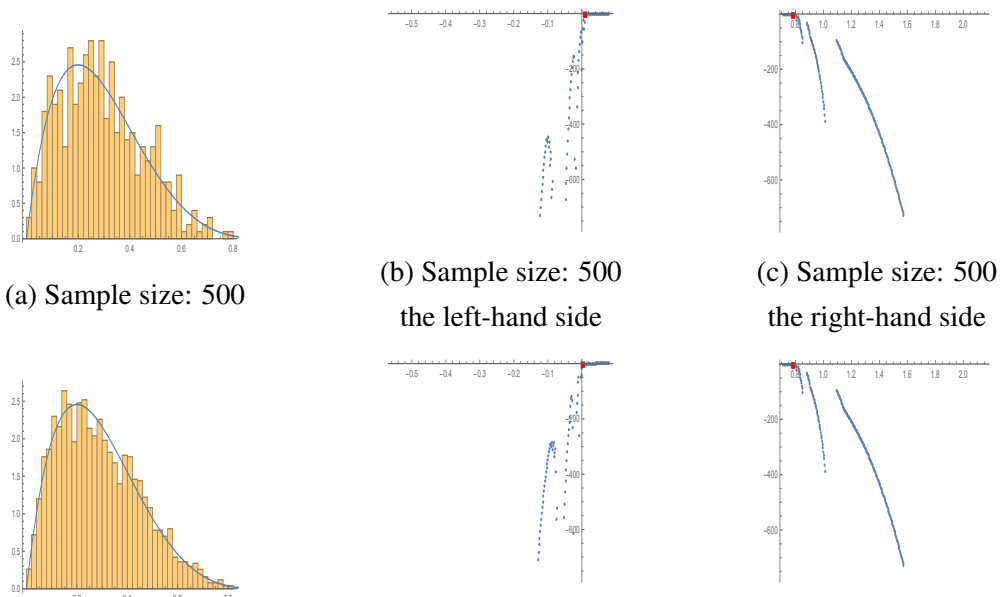
Consider samples of size 500 and 2,500, that are simulated from a Beta$(2, 5)$ distribution. Once the Box-Cox transform is applied to the data which is then rescaled, the behavior of the logarithm of empirical saddlepoint estimates is investigated.

Histograms of these samples along with the plots of the logarithm of the estimates lying outside the range of the transformed sample are shown in Figure 3.3. Since the empirical endpoint falls outside the transformed support in the left tail, only the rescaled data is being utilized in accordance with Algorithm 3.1, the resulting density estimates being included in Figure 3.3 (b) and (e). Table 3.4 includes the empirical endpoint estimates corresponding to a tolerance of e^{-200} applied to the empirical saddlepoint density estimates, along with the extreme order statistics.

Figure 3.4 shows the plots of the density estimates based on the generalized beta distribution whose parameters are determined by making use of the two methods described in Section 3.4, which are superimposed on a histogram of the data. The plot based on the PMLE method is not included for the sample of size 2,500 as the fit of the resulting density estimate leaves something to be desired. Table 3.5 provides endpoint estimates based on the fitted generalized beta distribution along with the extreme order statistics. Based on prior knowledge, negative values could of course be set equal to zero.

Sample size	L	$x_{(1)}$	$x_{(n)}$	U
500	−0.011126	0.009831	0.786359	0.965942
2,500	−0.007959	0.003414	0.863242	0.978048

Table 3.4: Summary of endpoints and the extreme order statistics based on the empirical saddlepoint estimates (Data simualted from a Beta $(2, 5)$ distribution)



(a) Sample size: 500

(b) Sample size: 500
the left-hand side

(c) Sample size: 500
the right-hand side

(d) Sample size: 2,500

(e) Sample size: 2,500
the left-hand side

(f) Sample size: 2,500
the right-hand side

Figure 3.3: Histograms of the data simulated from a Beta $(2,5)$ distribution and empirical density estimates based on Algorithm 3.1.
(red dots: sample minima and maxima)

(a) Sample size: 500

(b) Sample size: 2,500

Figure 3.4: Histograms of the data simulated from a Beta$(2,5)$ distribution and density estimates obtained by fitting a generalized beta distribution
(solid line: underlying density; dotted line: PMLE method; dashed line: FCMM technique)

3.5.2 A gamma distribution

Let Gamma (α,β) denote a gamma distribution whose density function is given by

Method	Sample size	L	$x_{(1)}$	$x_{(n)}$	U
PMLE	500	−0.111147	0.009831	0.786359	0.851286
FCMM	500	−0.011930	0.009831	0.786359	0.920395
FCMM	2,500	−0.004612	0.003414	0.863242	1.02274

Table 3.5: Summary of endpoints and extreme order statistics based on fitting a generalized beta distribution with the PMLE method and FCMM technique (Data simulated from a Beta $(2, 5)$ distribution)

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, \quad x \geq 0, \tag{3.17}$$

where α and β are the positive shape and scale parameters, respectively. Consider samples of size 500 and 2,500 simulated from the Gamma $(3, 2.5)$ distribution.

We follow the same procedure, that is, applying the Box-Cox transform, rescaling the resulting values so that the standard deviation becomes one, tracking the patterns of the empirical saddlepoint estimates, and determining the empirical endpoints on the original scale. Figure 3.5 shows histograms of the samples along with the plots of the logarithm of the density estimates lying outside of the range from the transformed sample. The resulting empirical endpoint estimates along with the sample minima and maxima are included in Table 3.6.

The density plots of the fitted generalized beta distributions are superimposed on histograms of the samples in Figure 3.6. The resulting endpoints are specified in Table 3.7. The estimate based on the PMLE approach applied to the sample of size 2,500 is not included as an upper bound estimate (U) since it is not consistent with the other upper bound estimates.

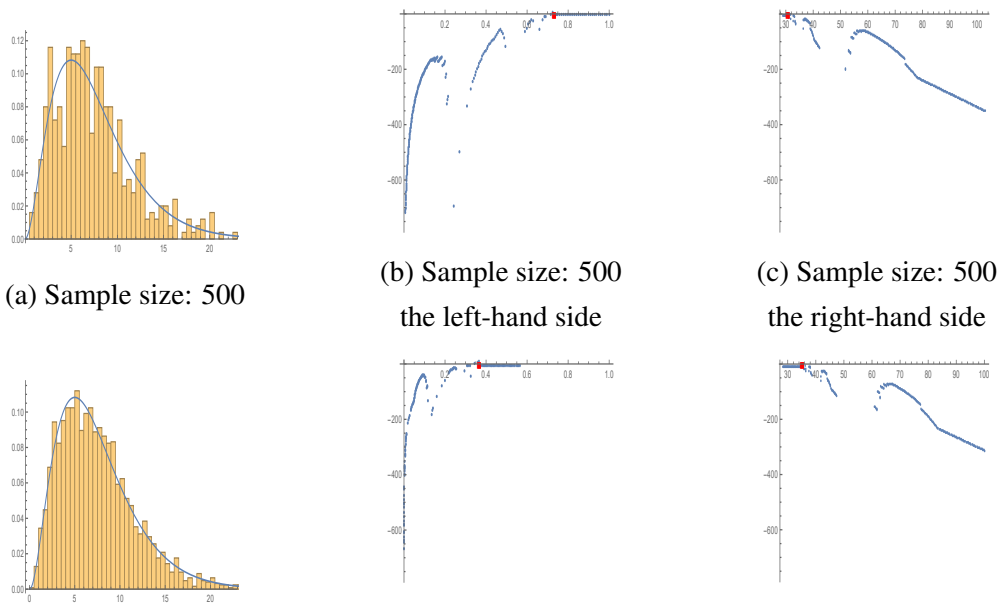
3.5.3 A generalized Pareto distribution

Let GP (m, α, s) denote a generalized Pareto distribution whose density function is given by

$$f(x) = \frac{1}{s}\left(1 + \frac{x-m}{\alpha s}\right)^{-\alpha-1}, \quad x \geq m, \tag{3.18}$$

Sample size	L	$x_{(1)}$	$x_{(n)}$	U
500	0.355658	0.73347	30.823	75.5683
2,500	0.129403	0.129403	35.2076	80.8711

Table 3.6: Summary of endpoints and extreme order statistics based on the empirical saddlepoint estimate (Data simulated from a Gamma $(3, 2.5)$ distribution)



(a) Sample size: 500

(b) Sample size: 500
the left-hand side

(c) Sample size: 500
the right-hand side

(d) Sample size: 2,500

(e) Sample size: 2,500
the left-hand side

(f) Sample size: 2,500
the right-hand side

Figure 3.5: Histograms of the data simulated from a Gamma$(3, 2.5)$ distribution and empirical density estimates based on Algorithm 3.1.
(red dots: sample minima and maxima)

(a) Sample size: 500

(b) Sample size: 2,500

Figure 3.6: Histograms of the data simulated from a Gamma $(3, 2.5)$ distribution and density estimates obtained by fitting generalized beta distributions
(solid line: underlying density; dotted line: PMLE method; dashed line: FCMM technique)

where m, α and s are the location, shape and scale parameters, respectively. This distribution is utilized in conjunction with the peak-over-threshold method in the context of extreme value theory, as discussed by Hosking and Wallis (1987) and Falk and Guillou (2008).

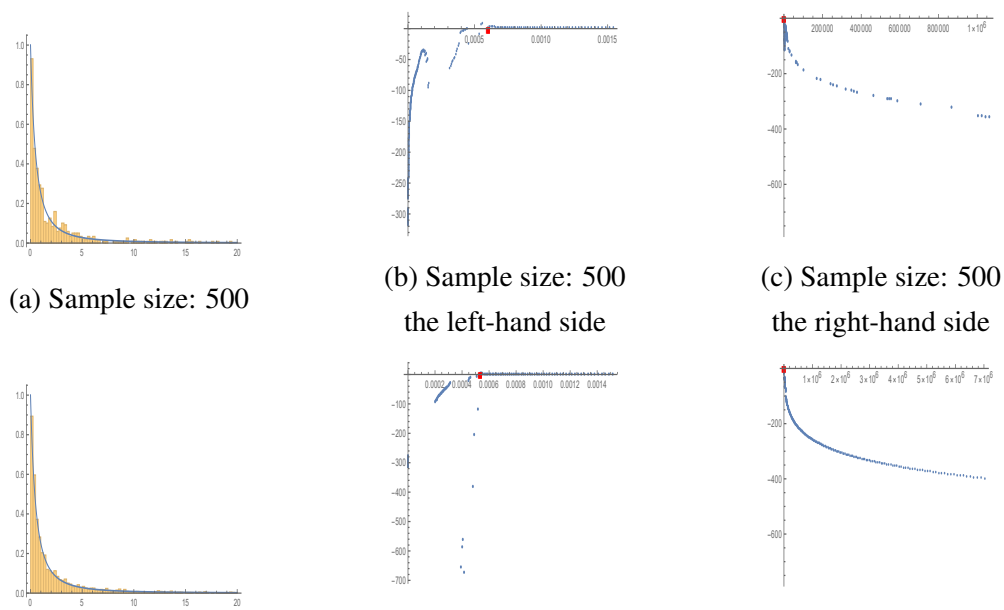
Samples of sizes 500 and 2,500 are simulated from the GP $(0, 1, 1)$ distribution. Histograms

Method	Sample size	L	$x_{(1)}$	$x_{(n)}$	U
PMLE	500	0.27828	0.73347	30.823	6.62003×10^6
FCMM	500	0.73347	0.73347	30.823	1.67244×10^8
FCMM	2,500	0.246125	0.366762	35.2076	4.67695×10^5

Table 3.7: Summary of endpoints and extreme order statistics based on fitting generalized beta distributions with the PMLE method and FCMM technique
(Data simulated from a Gamma $(3, 2.5)$ distribution)

of the samples along with the plots of the logarithm of the estimates lying outside the range of the transformed data are shown in Figure 3.7. Table 3.8 includes empirical endpoint estimates based on the empirical saddlepoint approach along with extreme order statistics.

The fit based on the generalized beta distribution is not adequate when the original data is utilized. Thus, we fit this distribution to the Box-Cox transformed data. The resulting density estimates obtained by applying the inverse transform provide a suitable fit, as shown in Figure 3.8. The density estimate obtained by applying the FCMM technique on the sample of size 500



(a) Sample size: 500

(b) Sample size: 500
the left-hand side

(c) Sample size: 500
the right-hand side

(d) Sample size: 2,500

(e) Sample size: 2,500
the left-hand side

(f) Sample size: 2,500
the right-hand side

Figure 3.7: Histograms of the data simulated from a GP(0, 1, 1) distribution and empirical density estimates based on Algorithm 3.1.
(red dots: sample minima and maxima)

Sample size	L	$x_{(1)}$	$x_{(n)}$	U
500	4.71009×10^{-6}	6.03906×10^{-4}	1,374.79	128,836
2,500	9.56331×10^{-5}	5.45281×10^{-4}	9129.48	403,837

Table 3.8: Summary of endpoints and extreme order statistics based on the empirical saddlepoint estimates (Data simulated from a GP (0, 1, 1) distribution)

(a) Sample size: 500 (b) Sample size: 2,500

Figure 3.8: Histograms of the data simulated from a GP(0, 1, 1) distribution and density estimates obtained by fitting generalized beta distributions (solid line: underlying density; dotted line: PMLE method; dashed line: FCMM technique)

is not included since the upper endpoint estimate coincides with the maximum of the sample. Table 3.9 includes the endpoint estimates based on the fitted generalized beta distributions.

Method	Sample size	L	$x_{(1)}$	$x_{(n)}$	U
PMLE	500	7.0872×10^{-64}	6.03906×10^{-4}	1,374.79	1.49087×10^{13}
PMLE	2,500	2.61842×10^{-46}	5.45281×10^{-4}	9129.48	3.03384×10^{17}
FCMM	2,500	5.45281×10^{-4}	5.45281×10^{-4}	9129.48	$6.96802 \times 10^{948,484}$

Table 3.9: Summary of endpoints and the extreme order statistics based on fitting a generalized beta distribution with the PMLE method and FCMM technique (Data simulated from a GP (0, 1, 1) distribution)

Bibliography

[1] Aarssen, K. and De Haan, L. (1994). On the maximal life span of humans. *Mathematical Population Studies* **4**(4), 259–281.

[2] AbouRizk, S. M., Halpin, D. W., and Wilson, J. R. (1994). Fitting beta distributions based on sample data. *Journal of Construction Engineering and Management* **120**(2), 288–305.

[3] Alves, I. F. and Neves, C. (2014). Estimation of the finite right endpoint in the Gumbel domain. *Statistica Sinica* **24**, 1811–1835.

[4] Alves, I. F., Neves, C., and Rosário, P. (2017). A general estimator for the right endpoint with an application to supercentenarian womens records. *Extremes* **20**(1), 199–237.

[5] Caers, J. and Maes, M. A. (1998). Identifying tails, bounds and end-points of random variables. *Structural Safety* **20**(1), 1–23.

[6] Cazals, C., Florens, J. P., and Simar, L. (2002). Nonparametric frontier estimation: a robust approach. *Journal of econometrics* **106**(1), 1–25.

[7] Einmahl, J. H. and Magnus, J. R. (2008). Records in athletics through extreme-value theory. *Journal of the American Statistical Association* **103**(484), 1382–1391.

[8] Falk, M. and Guillou, A. (2008). Peaks-over-threshold stability of multivariate generalized Pareto distributions. *Journal of Multivariate Analysis*, **99**(4), 715–734.

[9] Fasiolo, M., Wood, S. N., Hartig, F., and Bravington, M. V. (2018). An extended empirical saddlepoint approximation for intractable likelihoods. *Electronic Journal of Statistics*, **12**(1), 1544–1578.

[10] Girard, S., Guillou, A., and Stupfler, G. (2012). Estimating an endpoint with high-order moments. *Test* **21**(4), 697–729.

[11] Hall, P. (1982). On estimating the endpoint of a distribution. *The Annals of Statistics* **11**(1), 556–568.

[12] Hall, P. and Wang, J. Z. (1999). Estimating the end-point of a probability distribution using minimum-distance methods. *Bernoulli* **5**(1), 177–189.

[13] Hall, P. and Wang, J. Z. (2005). Bayesian likelihood methods for estimating the end point of a distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**(5), 717–729.

[14] Hosking, J. R.and Wallis, J. R. (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, **29**(3), 339–349.

[15] Leng, X., Peng, L., Wang, X., and Zhou, C. (2018). Endpoint estimation for observations with normal measurement errors. *Extremes*, https://doi.org/10.1007/s10687-018-0332-x

[16] Li, D., Peng, L., and Qi, Y. (2011). Empirical likelihood confidence intervals for the endpoint of a distribution function. *Test* **20**(2), 353–366.

[17] Loh, W. Y. (1984). Estimating an endpoint of a distribution with resampling methods. *The Annals of Statistics* **12**(4), 1543–1550.

[18] Resnick, S. I. (2007). *Heavy-tail Phenomena: Probabilistic and Statistical Modeling*. Springer Science and Business Media, New York.

[19] Wang, J. Z. (2005). A note on estimation in the four-parameter beta distribution. *Communications in Statistics-Simulation and Computation* **34**(3), 495–501.

Chapter 4

On Securing Bona Fide Density Functions

4.1 Introduction

Some classes of efficient density estimation techniques may not necessarily produce *bona fide* density functions. For instance, they may produce estimators that take slightly negative values or do not integrate to one. Such density estimates can create serious problems for practitioners, as pointed out by Silverman (1986, p. 69), and several methods have been proposed for modifying them with a view to obtain *bona fide* density estimates. Hall and Murison (1993), Devroye and Györfi (1985), and Kałuszka (1998) mentioned that a *bona fide* density estimator, $\check{f}(x)$, can be generated by normalizing the positive part of the initial estimate of the density denoted by $f(x)$:

$$\check{f}(x) = \frac{\max(f(x), 0)}{\int \max(f(x), 0)}. \tag{4.1}$$

Gajek (1986) proposed the following iterative scheme which makes use of a certain weight function whose tail should be weighted strongly enough to make a density estimate *bona fide*. Given the initial estimator, $\hat{f}(x)$ and weight function, $w(x)$, following the nth iteration, one has:

$$f_{n,GA}(x) = \max(f_{n-1,GA}(x), 0) - \frac{C_{n-1} - 1}{w(x) \int \frac{1}{w(x)} \, \mathrm{d}x} \quad \text{for} \quad n = 1, 2, \ldots \tag{4.2}$$

with $f_{0,GA}(x) = f(x)$ and $C_{n-1} = \int \max\left(f_{n-1,GA}(x), 0\right) \, \mathrm{d}x$.

As the iteration process always guarantees that the integral of $f_{n,GA}(x)$ equals one, the only condition to obtain a *bona fide* density function is that $f_{n,GA}(x) \geq 0$ for any x.

Glad *et al.* (2003) proposed a uniform vertical shift of the density in conjunction with the removal of the negative part, resulting in a *bona fide* density estimator. Typically, the integral

of the initial estimator which removes the negative part is greater than one, and the *bona fide* estimator, $f_{GL}(x)$, is obtained as follows:

$$f_{GL}(x) = max(f(x) - \xi, 0). \tag{4.3}$$

where ξ is chosen such that $\int f_{GL}(x)\,\mathrm{d}x = 1$.

All three methodologies aim at removing negative parts. However, such corrections present some drawbacks. First, there is no possibility of generating simulated values on the subintervals where the function is set equal to zero. As the negative parts of the initial density estimator occur mostly in the tail, one might confine the distribution to a compact support, resulting in generating the simulated value only within that support. Secondly, the resulting *bona fide* density estimators may not be everywhere differentiable. This leads to complicated closed forms for the cumulative density function (CDF) and possible computational inefficiencies in the calculation of the CDF's.

In order to address the aforementioned issues, we propose a novel methodology for generating smooth *bona fide* density approximants having a compact support. This approach is based on the moments of the normalized non-negative parts of the initial approximant. Unlike other *bona fide* density estimators, this estimator is not only *bona fide* density but also differentiable within the support of the density.

In Chapter 1, we introduced a polynomially-adjusted density approximation (PADA) methodology for distributions having a compact support. As discussed in Provost (2005), the resulting approximant which has a polynomial form is equivalent to those obtained by means of linear combinations of Legende polynomials. This methodology is a cornerstone of the subsequent sections both with respect to density approximation and density estimation, which respectively make use of exact and sample moments.

Section 4.2 discusses the process of obtaining a smooth *bona fide* density estimator. We propose to make use of an iterative scheme in conjunction with the PADA methodology. The final adjustment involves a vertical shift of the estimator at the last iteration, which is then normalized.

Section 4.3 presents a numerical example and compares the density estimates resulting from applying our methodology with those obtained with Glad's and Gajek's approaches. In Section 4.4, we modify the initial density estimate by removing the negative parts and then by approximating the resulting function by means of Bernstein polynomials.

4.2 General algorithm of obtaining smooth *bona fide* density functions

Once properly normalized, the approximants or the estimates are often *bona fide* density functions. However, it may happen that they be slightly negative on certain sub-ranges of their support. In order to obtain differentiable bona fide density functions on their entire support, that is, *smooth bona fide* density functions, the following procedures is implemented.

1. An initial density estimate, $f_{d_0}(x)$, is obtained by applying the PADA methodology. The polynomial degree d_0 is the minimizer of the sum of squared difference (SSD_{d_0}), that is,

$$SSD_{d_0} = \sum_{i=1}^{n} \left(F_e(x_i) - F_{d_0}(x_i)\right)^2 \tag{4.4}$$

   where $F_e(x_i)$ denotes the empirical cumulative distribution function (CDF) and $F_{d_0}(x_i)$ the CDF associated with $f_{d_0}(x)$ evaluated at the sample point x_i.

2. When an estimate defined on the interval (a, b) is negative on subranges of the intervals (a, a_0) and (b_0, b) where $f_d(a_0) = 0$, and $f_d(b_0) = 0$, the initial estimate is obtained from truncating its support to the interval $[a_0, b_0]$ and removing any remaining negative parts within $[a_0, b_0]$. Such a correction which is also discussed in Efromovich (1999), is expressed as

$$f_{0,d_0}(x) = \frac{\max(f_{X_{d_0}}(x), 0)}{\int_{a_0}^{b_0} \max(f_{X_{d_0}}(x), 0)\, \mathrm{d}x}\ ,\ a_0 < x < b_0. \tag{4.5}$$

   As $\max(f_{X_{d_0}}(x), 0) = \frac{1}{2}(f_{X_{d_0}}(x) + \sqrt{f_{X_{d_0}}(x)^2})$, we reexpress Equation (4.5) as

$$f_{0,d_0}(x) = \frac{(f_{X_{d_0}}(x) + \sqrt{f_{X_{d_0}}(x)^2})}{\int_{a_0}^{b_0} (f_{X_{d_0}}(x) + \sqrt{f_{X_{d_0}}(x)^2})\, \mathrm{d}x} \tag{4.6}$$

   to improve computational efficiency.

3. Set $i = 1$.

4. The ith iteration of the approximant of degree d_i, $f_{d_i}(x)$, is obtained by approximating $f_{i-1,d_{i-1}}(x)$ with the PADA methodology, that is,

$$f_{d_i}(x) = \frac{1}{b-a} \sum_{k=0}^{d_i} \xi_k \, y^k. \tag{4.7}$$

Note that we use the initial support interval $[a, b]$. Once the new truncation points (a_i, b_i) are determined where $f_{d_i}(a_i) = 0$ and $f_{d_i}(b_i) = 0$, the new corrected estimate, $f_{i,d_i}(x)$, is expressed as

$$f_{i,d_i}(x) = \begin{cases} f_{d_i}(x) & \text{for } a_i < x < b_i \\ 0 & \text{for } a < x < a_i \text{ and } b_i < x < b \end{cases}. \tag{4.8}$$

If $\min f_{i,d_i}(x) \geq 0$, we make use of a multiplicative normalizing constant to obtain a *smooth bona fide* density function. For the case where the minimum value of f_{i,d_i} takes slightly negative value, that is, for example $-10^{-2} \leq \min f_{i,d_i}(x) \leq 0$, we vertically shift $f_{i,d_i}(x)$ and adjust the resulting function with a normalizing constant, that is,

$$f_{FIN}(x) = \frac{\left(f_{i,d_i}(x) + \left|\min f_{i,d_i}(x)\right|\right)}{\int_\alpha^\beta \left(f_{i,d_i}(x) + \left|\min f_{i,d_i}(x)\right|\right) \mathrm{d}x}, \tag{4.9}$$

where α and β are points of intersection with the abscissa that delimit the final support of the approximant, which is the desired *smooth bona fide* density function, and stop. Otherwise,

5. We replace $f_{i,d_i}(x)$ with

$$f_{i,d_i}(x) = \frac{(f_{X_{d_i}}(x) + \sqrt{f_{X_{d_i}}(x)^2})}{\int_{a_i}^{b_i} (f_{X_{d_i}}(x) + \sqrt{f_{X_{d_i}}(x)^2}) \, \mathrm{d}x}; \tag{4.10}$$

6. Set $i = i + 1$ and return to Step 4.

4.3 Numerical example

Let Beta(α, β) denote a beta distribution with shape parameters α and β. Consider a sample of 200 observations simulated from a mixture of two equally weighted beta density functions with parameters $(\alpha_1 = 30, \beta_1 = 5)$ and $(\alpha_1 = 6, \beta_1 = 25)$. The assumed density function is expressed as

$$f_X(x) = \frac{1}{2} \frac{x^{29}(1-x)^4}{B(30,5)} + \frac{1}{2} \frac{x^5(1-x)^{24}}{B(6,25)} \tag{4.11}$$

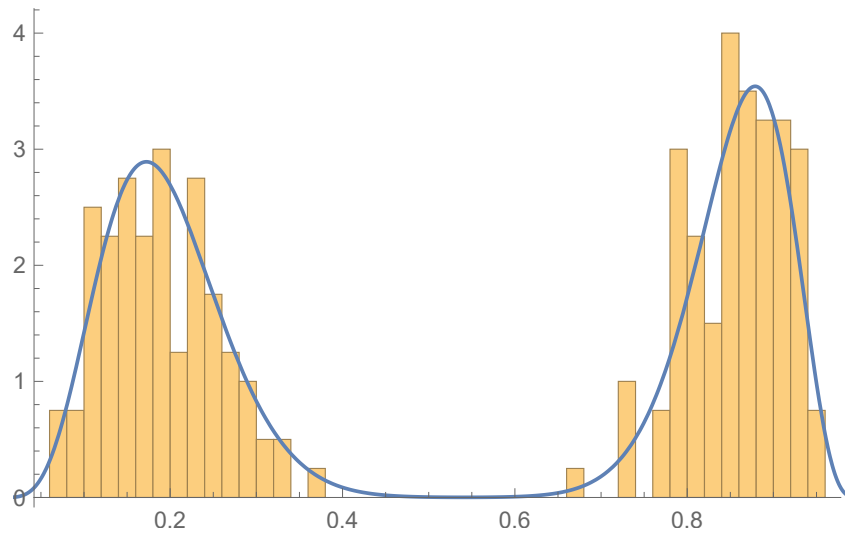


Figure 4.1: Histogram of a simulated sample and plot of the underlying density function simulated from an equal mixture of Beta(30, 5) and Beta(6, 25)

where $\mathrm{B}(\alpha,\beta) = \frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}$. The histogram of the sample along with the plot of the assumed density shown in Figure 4.1 are clearly indicative of a separated bimodal distribution. Without loss of generality, the initial support is set to be that of the underlying distribution, that is, [0, 1].

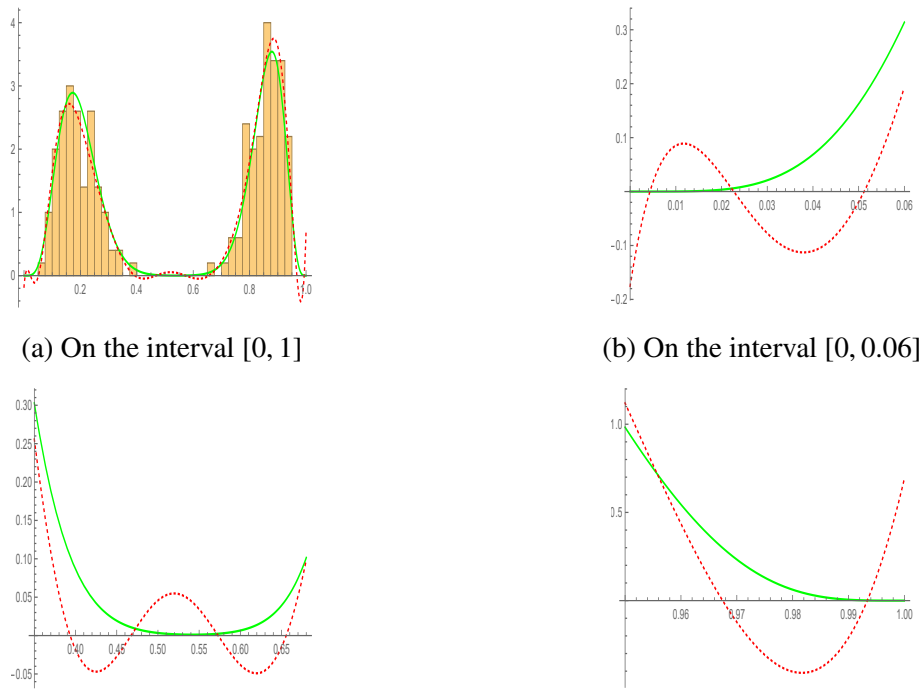
The initial polynomial estimate is $f_{d_0=14}(x)$. Its plot as well as the underlying density on entire interval is superimposed on a histogram of the sample in Figure 4.2 and shown along with the underlying density on some subranges of the support.

Once the truncation points are set and the associated polynomial density is determined, the first iteration is completed. The plots of $f_{d_0=14}(x)$ and $f_{d_1=19}(x)$ are compared on three subranges within the initial interval in Figure 4.3.

Similarly, plots $f_{d_2=19}(x)$, resulting from the second iteration, along with those of $f_{d_0=14}(x)$ and $f_{d_1=19}(x)$, are shown in Figure 4.4.

As the density approximant on the third iteration does not show a significant improvement compared with $f_{d_2=19}(x)$ and the local minimum takes a negative value within the truncation point evaluated on the second iteration, the iteration procedure is stopped. Then a vertical shift of $f_{d_2=19}(x)$ that makes it nonnegative and the normalizing of the resulting function leads to a *smooth bona fide* density function denoted by $f_{FIN}(x)$. Density estimates at each iteration stage and the final density function are plotted in Figure 4.5.

Notice that, with each iteration step, the support, (a_i, b_i), becomes gradually wider, as Table 4.1 indicates.



(a) On the interval $[0, 1]$

(b) On the interval $[0, 0.06]$

(c) On the interval $[0.35, 0.068]$

(d) On the interval $[0.95, 1]$

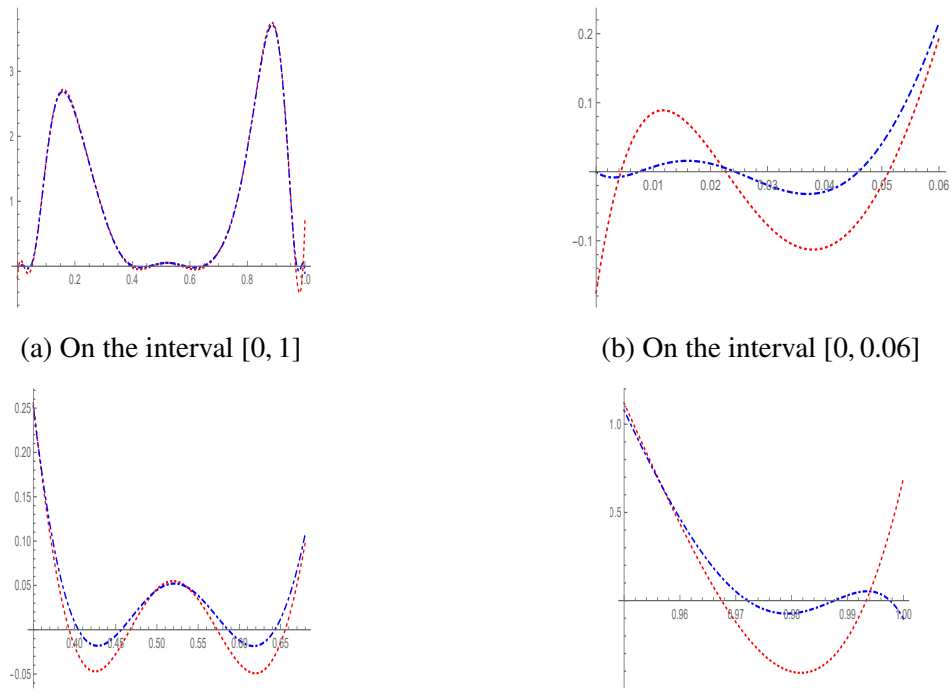
Figure 4.2: Plots of $f_{d=14}(x)$ on various subranges of the interval (solid line: underlying density / dotted line: $f_{d_0=14}(x)$)

4.3.1 Comparing Glad's and Gajek's approaches

Once the initial estimate, $f_{d_0=14}(x)$, is obtained, applying Equation (4.3) with a suitable value of ξ leads to the estimate obtained from Glad's approach, $f_{GL}(x)$. Figure 4.6 shows that the corrected density function, $f_{GL}(x)$, cannot eliminate several variations found in both tails and the subregion located between the two modes. Even if the tail areas can be corrected by truncating at (a_0, b_0), the *bona fide* densities obtained with this method may not be everywhere differentiable.

Number of iteration step (i)	a_i	b_i
0	0.0512034	0.967463
1	0.0459074	0.972093
2	0.0438354	0.973245
Final	0.0421001	0.973735

Table 4.1: Support at each iteration and final step



(a) On the interval $[0, 1]$

(b) On the interval $[0, 0.06]$

(c) On the interval $[0.35, 0.068]$

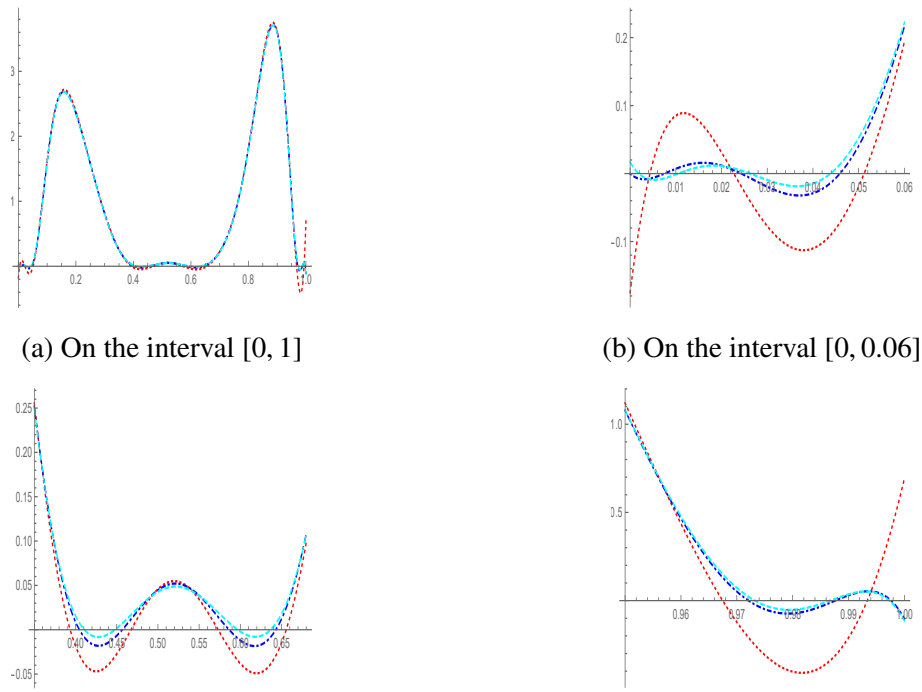
(d) On the interval $[0.95, 1]$

Figure 4.3: Plots of $f_{d_0=14}(x)$ and $f_{d_1=19}(x)$ on various subranges of the interval (dotted red: $f_{d_0=14}(x)$ / dot-dashed blue: $f_{d_1=19}(x)$)

Gajek's approach assumes strong weights in the tails as weight function, $w(x)$; however, Equation (4.2) entails a converse adjustment in the tails. Using the reciprocal of Epanechnikov's density function as weight function, that is, $w(x) = \frac{1}{6x(1-x)}$, and applying repeatedly the iterative scheme relying on Equation (4.2) produces several density function corrections at each iteration stage, denoted by $f_{n,GL}(x)$, where n represents the iteration number. The plot of the final density estimate, denoted by $f_{FIN,GA}(x)$, along with corrections at each iteration are included in Figure 4.7, indicating minimal adjustment in the tails. Even if the estimate truncated at (a_0, b_0) is used as the initial estimate, a smooth correction between the modes of the sample cannot be obtained.

4.4 Smooth density estimates expressed in terms of Bernstein polynomials

Let f be a continuous real-valued function of the interval $[0, 1]$. The sequence of Bernstein polynomials, denoted by $B_n(f; x)$, is expressed as



(a) On the interval [0, 1]

(b) On the interval [0, 0.06]

(c) On the interval [0.35, 0.068]

(d) On the interval [0.95, 1]

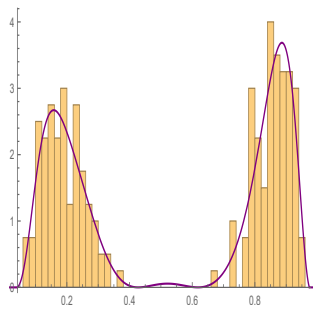
Figure 4.4: Plots of $f_{d_0=14}(x)$, $f_{d_1=19}(x)$, and $f_{d_2=19}(x)$ on various subranges of the interval (dotted red: $f_{d_0=14}(x)$ / dot-dashed blue: $f_{d_1=19}(x)$) / dashed cyan: $f_{d_2=19}(x)$)

$$B_n(f;x) = \sum_{j=0}^{n} f\left(\frac{j}{n}\right) \binom{n}{j} x^j \, (1-x)^{n-j} \tag{4.12}$$

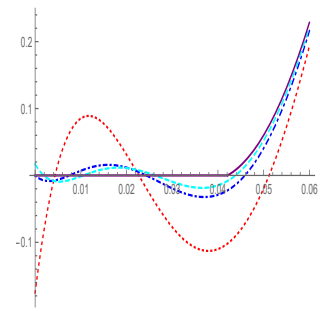
The reader may refer to Farouki (2012), which provided a historical perspective on the use of the Bernstein polynomials and their applications.

Consider the initial estimate, $f_{d_0=14}(x)$, obtained in Section 4.3, which is truncated at $a_0 = 0.0512$ and $b_0 = 0.9675$ as specified in Table 4.1. Once the estimate is normalized, $B_n(f;x)$ as defined in Equation (4.12) is estimated, which guarantees that the resulting approximants is continuous and differentiable at any point in the interval [0, 1]. As shown in Figure 4.8, the higher the degree n, the more accurate the approximant is. Then the resulting smooth function can be further approximated with a spline. For example, a second-degree spline approximant of $B_{1000}(f;x)$ which is based on 40 equidistant points of the interval [0, 1], is plotted in Figure 4.9 along with the approximated function.

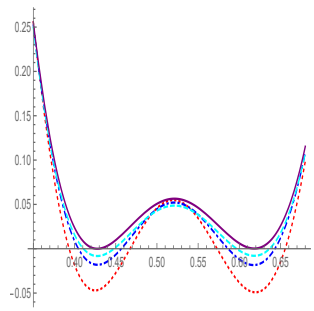
In cases where the function to be approximated points to a clearly separated bimodal or multimodal distribution, which can be assessed by inspecting a preliminary representation of the distribution such as a histogram, it is advisable to apply the proposed methodologies on



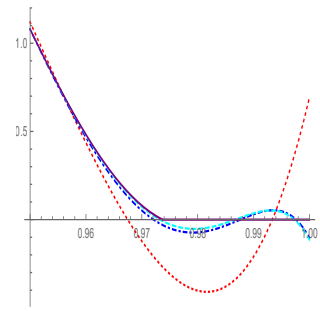
(a) On the interval $[0, 1]$



(b) On the interval $[0, 0.06]$



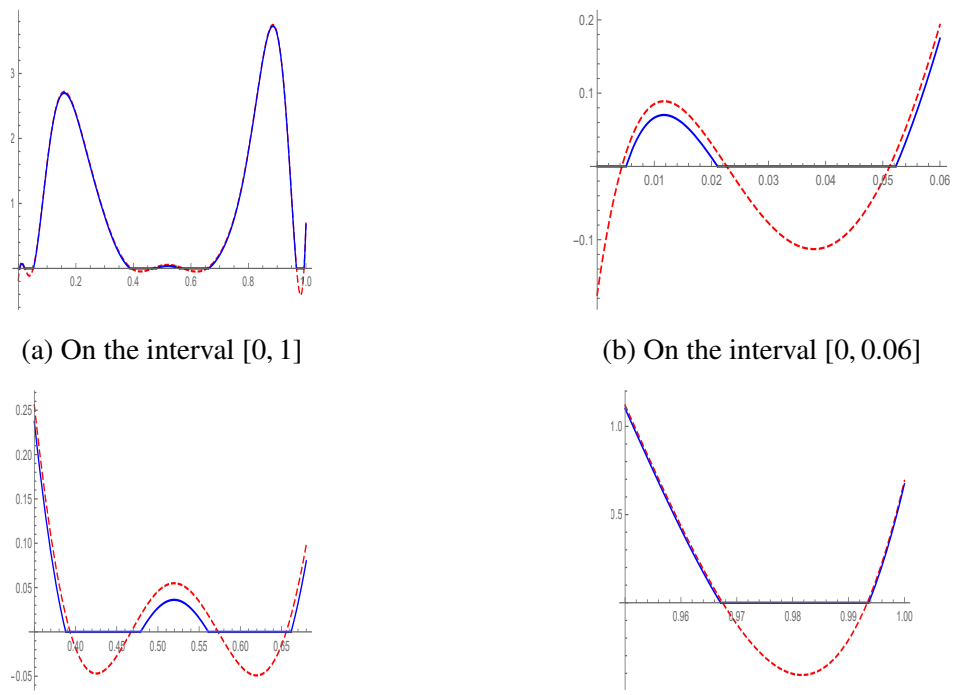
(c) On the interval $[0.35, 0.068]$



(d) On the interval $[0.95, 1]$

Figure 4.5: Plots of $f_{d_0=14}(x)$, $f_{d_1=19}(x)$, $f_{d_2=19}(x)$, and $f_{FIN}(x)$ on various subranges of the interval (dotted red: $f_{d_0=14}(x)$ / dot-dashed blue: $f_{d_1=19}(x)$) / dashed cyan: $f_{d_2=19}(x)$) / solid purple: $f_{FIN}(x)$)

subintervals wherein the data is only sparse in the neighborhoods of their respective endpoints and then combine the appropriately weighted density estimates so obtained in a single one.



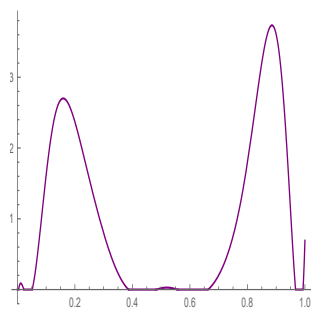
(a) On the interval $[0, 1]$

(b) On the interval $[0, 0.06]$

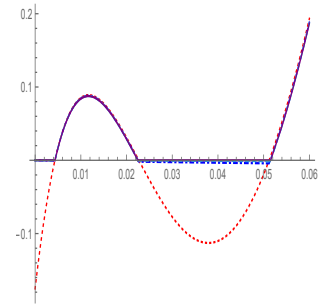
(c) On the interval $[0.35, 0.068]$

(d) On the interval $[0.95, 1]$

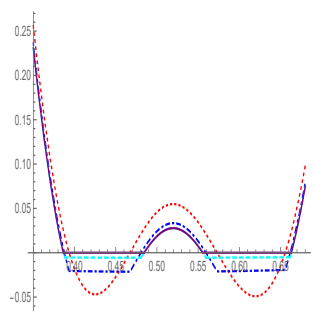
Figure 4.6: Plots of $f_{d_0=14}(x)$ and $f_{GL}(x)$ on various subranges of the interval (dashed line: $f_{d_0=14}(x)$ / solid line: $f_{GL}(x)$)



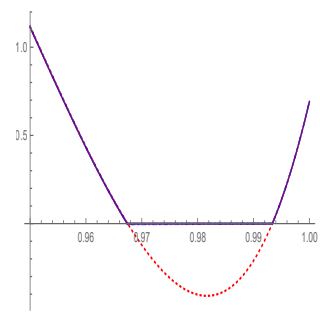
(a) On the interval $[0, 1]$



(b) On the interval $[0, 0.06]$

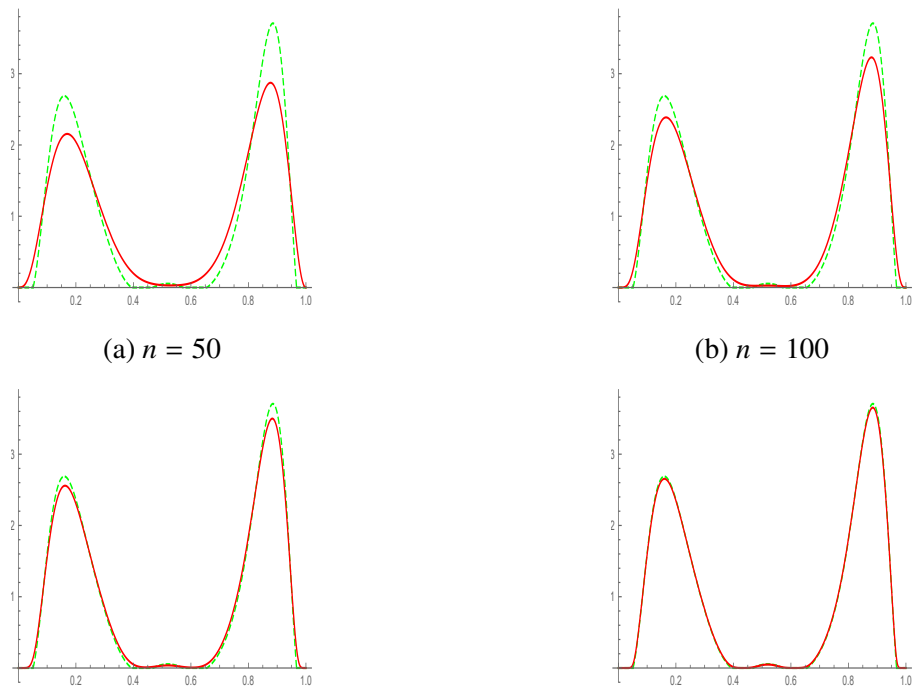


(c) On the interval $[0.35, 0.068]$



(d) On the interval $[0.95, 1]$

Figure 4.7: Plots of $f_{d_0=14}(x)$, $f_{1,GA}(x)$, $f_{2,GA}(x)$, and $f_{FIN,GA}(x)$ on various subranges of the interval (dotted red: $f_{d_0=14}(x)$ / dot-dashed blue: $f_{1,GA}(x)$) / dashed cyan: $f_{1,GA}(x)$) / solid purple: $f_{FIN,GA}(x)$)



(a) $n = 50$ (b) $n = 100$

(c) $n = 250$ (d) $n = 1,000$

Figure 4.8: Plots of the normalized truncated $f_{d_0=14}(x)$ (dotted lines) and the Bernstein polynomial estimates of various degrees (solid lines)

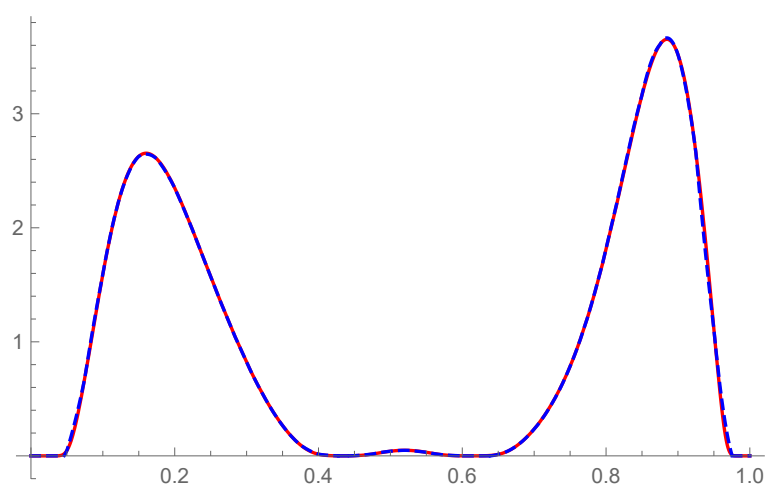


Figure 4.9: $B_{1000}(f;x)$ (solid line) and its spline approximant (dashed line)

Bibliography

[1] Devroye L. and Gyöfri L. (1985). *Nonparametric Density Estimation. The L_1 View*. Wiley, New York.

[2] Efromovich, S. (1999). *Nonparametric Curve Estimation: Methods, Theory, and Applications*. Springer Science and Business Media, New York.

[3] Farouki, R. T. (2012). The Bernstein polynomial basis: A centennial retrospective. *Computer Aided Geometric Design* **29**(6), 379–419.

[4] Gajek, L. (1986). On improving density estimators which are not bona fide functions. *The Annals of Statistics* **14**(4), 1612–1618.

[5] Glad, I. K., Hjort, N. L., and Ushakov, N. G. (2003). Correction of density estimators that are not densities. *Scandinavian Journal of Statistics* **30**(2), 415–427.

[6] Hall, P. and Murison, R. D. (1993). Correcting the negativity of high-order kernel density estimators. *Journal of Multivariate Analysis* **47**(1), 103–122.

[7] Kałuszka, M. (1998). On the Devroye-Györfi methods of correcting density estimators. *Statistics and probability letters* **37**(3), 249–257.

[8] Provost, S. B. (2005). Moment–based density approximants. *The Mathematica Journal* **9**, 727–756.

[9] Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

Chapter 5

Adjusted Empirical Bivariate Saddlepoint Estimates

5.1 Introduction

The saddlepoint appproximation is one of the most efficient statistical technique for obtaining very accurate tail probabilities. Since the pioneering work of Daniels (1954), further developments have been described for instances in Butler (2007), Jensen (1995), Kolassa (1994) and Goutis and Casella (1999). The multivariate saddlepoint density is expressed as

$$f(\mathbf{x}) = (2\pi)^{-\frac{d}{2}}\left|K''(\hat{\mathbf{s}})\right|^{-\frac{1}{2}} \exp\left[K(\hat{\mathbf{s}}) - \hat{\mathbf{s}}^T\mathbf{x}\right] \tag{5.1}$$

where d is the dimension and $K(\mathbf{s})$ is the cumulant generating function, that is, $K(\mathbf{s}) = \ln E\left(e^{\mathbf{s}^T\mathbf{x}}\right)$ and $\hat{\mathbf{s}}$ is the solution to $K'(\hat{\mathbf{s}}) = \mathbf{x}$.

When a sample of observations is available, empirical saddlepoint estimates are used to estimate density functions. In this chapter, we investigate an improved method for estimating bivariate density functions by combining an empirical saddlepoint estimate and a bivariate polynomial adjustment. After introducing the empirical bivariate saddlepoint estimate in Section 5.2, we propose to use the normalized exponential term of this estimate as an initial estimate. Statistically, this term is connected to likelihood ratio statistics. The proposed adjusted bivariate density estimate is expressed as the product of a spline representation of the normalized exponential term and a bivariate polynomial adjustment, which is discussed in Section 5.3. Section 5.4 presents several examples to illustrate how the methodology is applied.

5.2 Empirical bivariate saddlepoint estimation

The univariate empirical saddlepoint estimate defined in Equation (1.16) can be extended to the bivariate case. Suppose that the analytical form of the cumulant generating function is unknown. Given a bivariate sample of observations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ where $\mathbf{x}_j = \left(x_{1j}, x_{2j}\right)^T$, the empirical cumulant generating function is given by

$$\hat{K}_n(\boldsymbol{\theta}) = \log\left(\frac{1}{n}\sum_{i=1}^{n} e^{\boldsymbol{\theta}^T \mathbf{x}_i}\right) \tag{5.2}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2)^T$ and n is the sample size. The estimates of the derivatives of the function $\hat{K}_n(\boldsymbol{\theta})$ are

$$\hat{K}_n{}'(\boldsymbol{\theta}) = \frac{\sum_{i=1}^{n} \mathbf{x}_i \, e^{\boldsymbol{\theta}^T \mathbf{x}_i}}{\sum_{i=1}^{n} e^{\boldsymbol{\theta}^T \mathbf{x}_i}} \tag{5.3}$$

and

$$\hat{K}_n{}''(\boldsymbol{\theta}) = \frac{\sum_{i=1}^{n} \mathbf{x}_i \, \mathbf{x}_i{}^T \, e^{\boldsymbol{\theta}^T \mathbf{x}_i}}{\sum_{i=1}^{n} e^{\boldsymbol{\theta}^T \mathbf{x}_i}} - \hat{K}_n{}'(\boldsymbol{\theta}) \, \hat{K}_n{}'(\boldsymbol{\theta})^T. \tag{5.4}$$

The empirical saddlepoint estimation, denoted by $\hat{f}_n(\mathbf{x})$, is expressed as follows:

$$\hat{f}_n(\mathbf{x}) = (2\pi)^{-1} \left|\hat{K}_n{}''(\hat{\boldsymbol{\theta}})\right|^{-\frac{1}{2}} e^{\hat{K}_n(\hat{\boldsymbol{\theta}}) - \hat{\boldsymbol{\theta}}^T \mathbf{x}} \tag{5.5}$$

where $\hat{\boldsymbol{\theta}}$ is the saddlepoint solution of the system of equation of

$$\hat{K}_n{}'(\hat{\boldsymbol{\theta}}) = \mathbf{x}. \tag{5.6}$$

5.2.1 Computational issues related to the empirical saddlepoint estimate

Some of the computational issues encountered in connection with making use of the empirical saddlepoint estimate are described in Section 3.2. Another difficulty in estimating the empirical saddlepoint resides in the tail area as the convergence rate is then slower. This may lead to overestimating the density around the tail, which is mainly due to underestimating $\hat{K}_n{}''(\hat{\boldsymbol{\theta}})$, even if the saddlepoint solution, $\hat{\boldsymbol{\theta}}$, exists. The reader may refer to Feuerverger (1989) and Fasiolo *et al.* (2018) for a discussion of the asymptotics of the empirical saddlepoint estimate.

5.2.2 Normalized likelihood-like density function

Theoretically, the moment generating function, $M(s)$, of random variable X can be expressed as follows in terms of the cumulant generation function, $K(s)$:

$$e^{K(s)} = M(s) = \int_{-\infty}^{\infty} e^{sx} f(x)\,\mathrm{d}x, \tag{5.7}$$

which can be reexpressed as

$$1 = \int_{-\infty}^{\infty} e^{sx-K(s)} f(x)\,\mathrm{d}x. \tag{5.8}$$

The integrand is a density indexed by the parameter s, which shall be denoted

$$f(x; s) = e^{sx-K(s)} f(x). \tag{5.9}$$

This is referred to the Esscher transform (or exponentially tilted density function) of X. On rearranging Equation (5.9), one has

$$e^{K(s)-sx} = \frac{f(x)}{f(x; s)}. \tag{5.10}$$

Butler (2007) discussed the statistical interpretation of Equation (5.10) in connection with testing hypotheses on s and with likelihood ratio statistics. If the observed value of X is x, the maximum likelihood estimate of s maximizes $f(x; s)$, which coincides with the saddlepoint solution, $K'(s) = x$. The application of likelihood statistics in conjunction with the saddlepoint approximation is discussed by Barndorff-Nielsen (1983), Barndorff-Nielsen and Cox (1984), Fraser (1988), and Jensen (1992).

The following proposition states the connection between likelihood ratio statistics and the exponential term of the empirical bivariate saddlepoint estimate.

Proposition 5.2.1 *Given* $\mathbf{x}$*, whose associated saddlepoint solution is given in Equation (5.6), the local maximum of* $e^{\hat{K}_n(\boldsymbol{\theta})-\boldsymbol{\theta}^T\mathbf{x}}$ *is obtained at* $\mathbf{x} = \bar{\mathbf{x}}$ *where* $\bar{\mathbf{x}}$ *is the sample mean vector.*

Proof As $\hat{K}_n{}'(\hat{\boldsymbol{\theta}})$ is the function of $\mathbf{x}$ used to obtain the saddlepoint, we let $g(\mathbf{x}) = e^{\hat{K}_n(\boldsymbol{\theta}(\mathbf{x}))-\boldsymbol{\theta}(\mathbf{x})^T\mathbf{x}}$.

The partial derivative of $g(\mathbf{x})$ with respect to $\mathbf{x}$ is expressed as

$$\begin{aligned}\frac{\partial g}{\partial \mathbf{x}} &= \left(\frac{\partial \hat{K}_n}{\partial \boldsymbol{\theta}}\frac{\partial \boldsymbol{\theta}}{\partial \mathbf{x}} - \frac{\partial \boldsymbol{\theta}}{\partial \mathbf{x}}\mathbf{x} - \boldsymbol{\theta}(\mathbf{x})\right) g(\mathbf{x}) \\ &= \left[\left(\frac{\partial \hat{K}_n}{\partial \boldsymbol{\theta}} - \mathbf{x}\right)\frac{\partial \boldsymbol{\theta}}{\partial \mathbf{x}} - \boldsymbol{\theta}(\mathbf{x})\right] g(\mathbf{x}).\end{aligned} \tag{5.11}$$

Since $\frac{\partial \hat{K}_n}{\partial \boldsymbol{\theta}} = \mathbf{x}$, the only condition for $\frac{\partial g}{\partial \mathbf{x}} = \mathbf{0}$ to hold is $\boldsymbol{\theta}(\mathbf{x}) = \mathbf{0}$. In light of Proposition 1.4.2, the local maximum of $g(\mathbf{x})$ is found at $\hat{K}_n(\mathbf{0}) = \bar{\mathbf{x}}$. Hence, the result.

On normalizing $g(\mathbf{x})$, the resulting function,

$$g^*(\mathbf{x}) = \frac{g(\mathbf{x})}{\int_{\mathcal{R}^2} g(\mathbf{x})\,\mathrm{d}x}, \tag{5.12}$$

becomes a density function, $g^*(\mathbf{x})$ denoting the normalized likelihood-like density function.

5.3 Adjusted empirical bivariate saddlepoint estimates

Zareamoghaddam *et al.* (2017) proposed a bivariate density estimation methodology that makes use of a bivariate polynomial adjustment. Similarly, the proposed density estimate is expressed as the product of a normalized likelihood-like density function, $g^*(x, y)$, which is used as base density, and a bivariate polynomial adjustment, that is,

$$f_{p,q}(x_1,\ x_2) = g^*(x_1,\ x_2) \sum_{i=0}^{p} \sum_{j=0}^{q} \xi_{i,j}\, {x_1}^i\, {x_2}^j \tag{5.13}$$

where p and q denotes the polynomial degree associated with x_1 and x_2, respectively.

Let the $(k, \ell)^{\text{th}}$ joint sample moment be denoted by $\mu(k, \ell) = \frac{1}{n}\sum_{j=1}^{n} {x_{1j}}^k {x_{2j}}^\ell$ where $(x_{1j},\ x_{2j})$ is the jth bivariate sample observation, and let the $(k, \ell)^{\text{th}}$ joint moment associated with the base density, $g^*(x_1,\ x_2)$, be denoted by $m(k, \ell) = \int\int_{\mathcal{R}^2} {x_1}^k\, {x_2}^\ell\, g^*(x_1,\ x_2)\,\mathrm{d}x_2\,\mathrm{d}x_1$.

To determine the coefficients $\xi_{i,j}$ of the polynomial adjustment, the joint sample moments are equated to the joint moments associated with $f_{p,q}(x_1,\ x_2)$:

$$\begin{aligned}
\mu(k, \ell) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^k y^\ell f_{p,q}(x,\ y)\,\mathrm{d}x_2\,\mathrm{d}x_1 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^k y^\ell\, g^*(x_1,\ x_2) \sum_{i=0}^{p}\sum_{j=0}^{q} \xi_{i,j}\, {x_1}^i\, {x_2}^j\,\mathrm{d}x_2\,\mathrm{d}x_1 \\
&= \sum_{i=0}^{p}\sum_{j=0}^{q} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \xi_{i,j}\, {x_1}^{k+i}\, {x_2}^{\ell+j}\, g^*(x_1,\ x_2)\,\mathrm{d}x_2\,\mathrm{d}x_1,
\end{aligned}$$

for $k = 0,\ \ldots,\ p$ and $\ell = 0,\ \ldots,\ q$, which yields the following $(p+1)\times(q+1)$ linear equations:

$$\mu(k, \ell) = \sum_{i=0}^{p}\sum_{j=0}^{q} \xi_{i,j}\, m(k+i,\ \ell+j),\ \ k = 0,\ 1,\ 2, \ldots,\ p,\ \ \ell = 0,\ 1,\ 2, \ldots,\ q. \tag{5.14}$$

Therefore, the $\xi_{i,j}$'s can be obtained by solving the linear system $\mathbb{M}\xi = \boldsymbol{\mu}$ where ξ and $\boldsymbol{\mu}$ are vectors of dimensions $(p+1)\times(q+1)$ whose $(i(q+1)+(j+1))^{\text{th}}$ component, $\xi_{i,j}$ and $\mu(i, j)$,

appear in the same order for $i = 0, 1, \ldots, p$ and $j = 0, 1, \ldots, q$. Increasing p and q should theoretically result in greater accuracy.

The polynomial degrees p and q are such that the sum of squared differences ($SSD_{p,q}$)

$$SSD_{p,q} = \sum_{i=1}^{n} \left(F_e(x_{1i}, x_{2i}) - F_{p,q}(x_{1i}, x_{2i}) \right)^2 \tag{5.15}$$

is minimized, $F_e(x_{1i}, x_{2i})$ denoting the empirical bivariate cumulative distribution function (CDF) and $F_{p,q}(x_{1i}, x_{2i})$, the CDF associated with the density estimate $f_{p,q}(x_1, x_2)$ evaluated at the sample point (x_{1i}, x_{2i}).

5.4 Numerical examples

5.4.1 Simulated data from a bivariate normal distribution

Let $\mathcal{N}_d(\boldsymbol{\mu}, \mathbb{V})$ denote a d-dimensional multivariate normal distribution where $\boldsymbol{\mu}$ and $\mathbb{V}$ denote the mean vector and covariance matrix. Consider a sample of 500 observations simulated from a $\mathcal{N}_2\left(\begin{pmatrix}3\\5\end{pmatrix}, \begin{pmatrix}16 & 4\\4 & 4\end{pmatrix}\right)$ distribution.

Once the original data set is transformed so that the covariance matrix is converted to a correlation matrix, the exponential term of the empirical saddlepoint estimate is evaluated locally at multiple points, including some points outside of the boundary of the sample. Then, a second order bivariate interpolating spline is fitted to these points, representing a likelihood-like statistics. This spline is normalized on the support, and the resulting surface serves as an initial approximate density function, namely the base density function. The density estimate, $f_{\hat{a},\hat{b}}(x, y)$, is obtained once the initial estimate is adjusted by an $(\hat{a} + 1) \times (\hat{b} + 1)$-term bivariate polynomial, where $\hat{a}$ and $\hat{b}$ are such that $SSD_{\hat{a},\hat{b}}$ is minimized.

Figure 5.1 shows a histogram of the rescaled data and the normalized exponential term of the empirical saddlepoint density estimate. After applying the inverse transformation, the resulting density estimate is plotted along with a kernel density estimate (KDE) in Figure 5.2. Table 5.1 includes the values of $SSD_{a,b}$ for various polynomial degrees a and b.

5.4.2 Flood data set

The flood data was previously analyzed by Yue (2001) to model a five-parameter bivariate gamma distribution. Making use of daily stream-flow values, this data set consists of flood peaks, volumes, and duration in the Madawaska basin, Quebec, Canada, covering an area of 2,690 km^2, from 1919 to 1995. In this case, only the peaks and the volumes are considered.

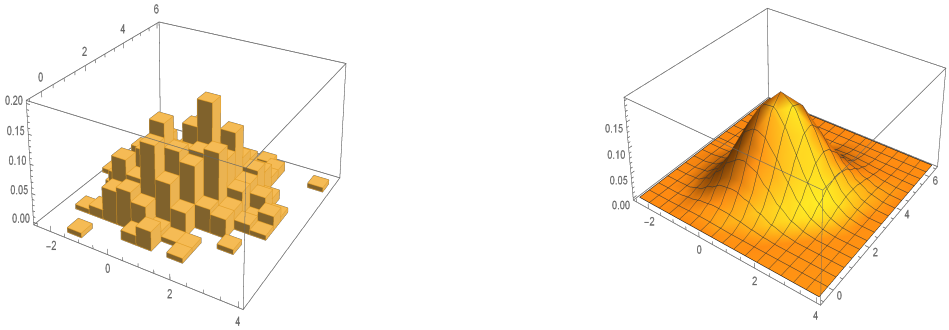
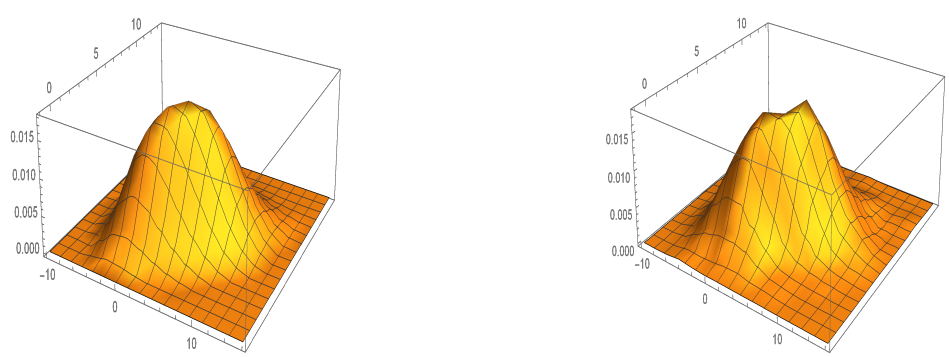


Figure 5.1: Histogram of the rescaled data simulated from a bivariate normal distribution and the normalized exponential term of the empirical saddlepoint density

The rescaled histogram and the plot of the normalized exponential term of the empirical saddlepoint density estimate are shown in Figure 5.3. The plot of the density estimate resulting from applying the inverse transformation and that of the KDE are included in Figure 5.4. Table 5.2 provides the $SSD_{a,b}$ values for various polynomial degrees a and b, the minimum of $SSD_{a,b}$ being found at $(a, b) = (1, 1)$.

5.4.3 Maximum speed data set

The maximum speed data was utilized by Castillo *et al.* (2005) to estimate the parameters of the generalized extreme value distribution. This data set consists of the maximum car speed recorded over 200 dry weekends at registered points on a highway and on a mountain. The maximum speed was recorded for the first 1,000 cars passing through the given locations.



$SSD_{2,2} = 0.0338222$ $SSD_{KDE} = 0.059261$

Figure 5.2: Adjusted density estimate, $f_{2,2}(x, y)$, and a kernel density estimate of the original data simulated from a bivariate normal distribution

	0	1	2	3	4	5
0	0.693781	0.379499	0.175481	0.173828	0.22172	0.221641
1	0.482737	0.15753	0.0726933	0.0746046	0.0869459	
2	0.296449	0.0590618	**0.0338222**	0.0364008		
3	0.323535	0.0574198	0.0353862			
4	0.349345	0.0609876				
5	0.3492					

Table 5.1: $SSD_{a,b}$'s for the rescaled bivariate normal simulated data with the values of a in the left column and those of b at the top

	0	1	2	3
0	0.690451	0.223898	0.0667203	0.060882
1	0.413571	**0.0492777**	0.0711331	
2	0.142395	0.0613828		
3	0.142368			

Table 5.2: $SSD_{a,b}$'s for the rescaled flood data with the values of a in the left and those of b at the top

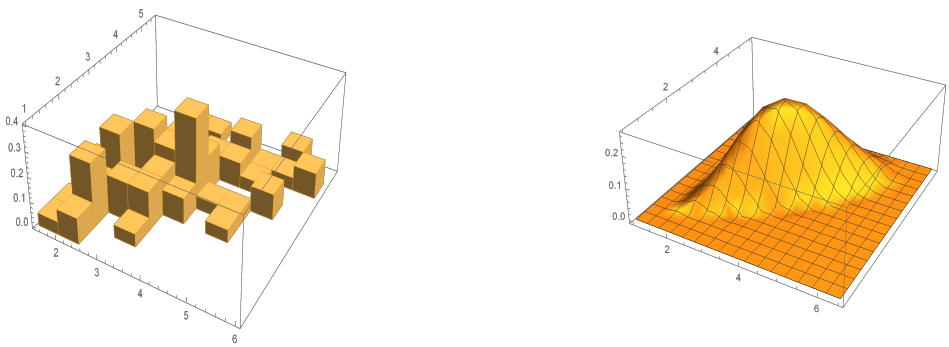
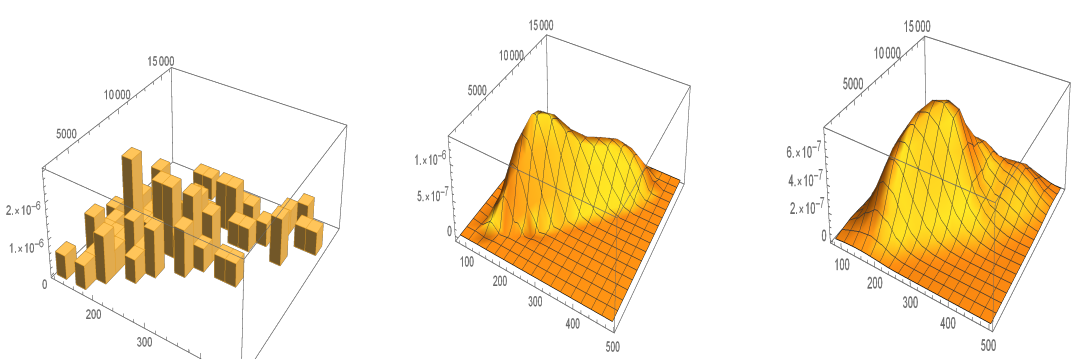


Figure 5.3: Histogram of the rescaled flood data
and the normalized exponential term of the empirical saddlepoint estimate

Figure 5.5 displays the rescaled histogram and the plot of the normalized exponential term of the empirical saddlepoint density estimate. Figure 5.6 shows the plot of the density estimate after applying the inverse transformation and that of a KDE. Table 5.3 presents the values of $SSD_{a,b}$ for various polynomial degrees a and b, the minimum $SSD_{a,b}$ being found at $(a, b) = (2, 1)$.

5.4.4 Simulated data from a mixture of a Dirichlet and bivariate beta density functions

The proposed methodology is also applied to a bimodal distribution. Let Dirichlet(a, b, c) denote the bivariate Dirichlet distribution, whose density function is given by



$SSD_{1,1} = 0.0492777$ $SSD_{KDE} = 0.0782292$

Figure 5.4: Histogram of the original flood data,
the adjusted density estimate, $f_{1,1}(x, y)$ and a kernel density estimate

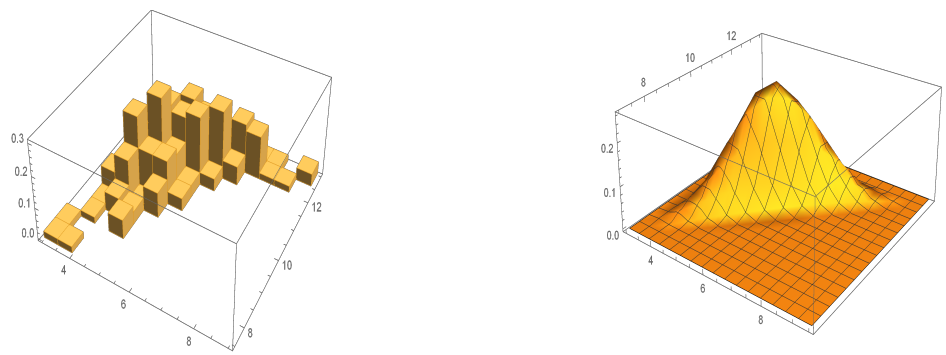


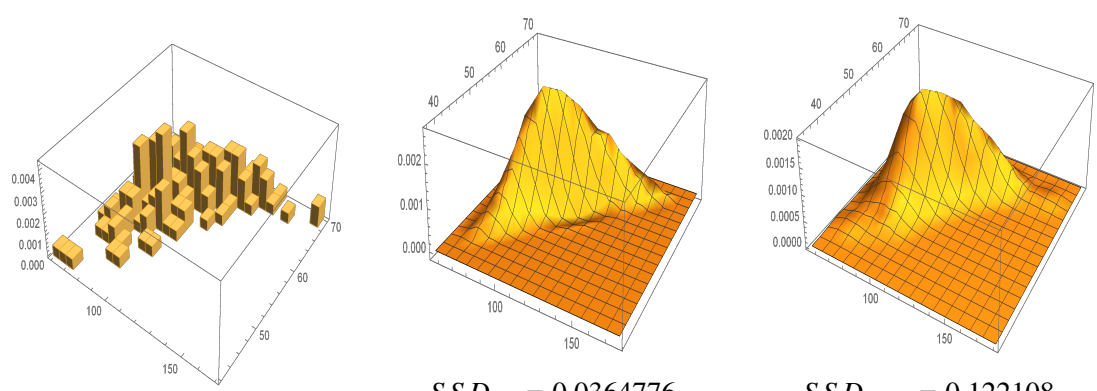
Figure 5.5: Histogram of the rescaled maximum speed data
and the normalized exponential term of the empirical saddlepoint estimate

$$f(x,y) = \frac{x^{a-1}\, y^{b-1}\, (1-x-y)^{c-1}\, \Gamma(a+b+c)}{\Gamma(a)\, \Gamma(b)\, \Gamma(c)} \quad \text{for } x > 0,\ y > 0,\ \text{and } x + y < 1. \tag{5.16}$$

Let Beta $((a1, b1), (a2, b2))$ denote the product of independent univariate beta density functions, which is given by

$$f(x,y) = \frac{x^{a1-1}\, (1-x)^{b1-1}}{\mathrm{B}(a1, b1)} \cdot \frac{y^{a2-1}\, (1-y)^{b2-1}}{\mathrm{B}(a2, b2)} \quad \text{for } 0 < x < 1 \text{ and } 0 < y < 1. \tag{5.17}$$

Consider 2,500 values simulated from an equal mixture of Beta $((4, 2), (5, 3))$ and Dirichlet (2,3,2) density functions. The plot of the resulting density is shown in Figure 5.7. A histogram of the simulated data set and the corresponding contour plot are included in Figure 5.8. The



$SSD_{2,1} = 0.0364776$ $\qquad$ $SSD_{KDE} = 0.122108$

Figure 5.6: Histogram of the original maximum speed data,
the adjusted density estimate, $f_{2,1}(x, y)$, and a kernel density estimate

	0	1	2	3	4
0	0.350409	0.728009	0.768645	0.758816	0.665429
1	0.20402	0.044134	0.0411254	0.0547479	
2	0.177573	**0.0364776**	0.0470069		
3	0.190414	0.0457248			
4	0.187497				

Table 5.3: $SSD_{a,b}$'s for the rescaled maximum speed data with the values of a in the left column and those of b at the top

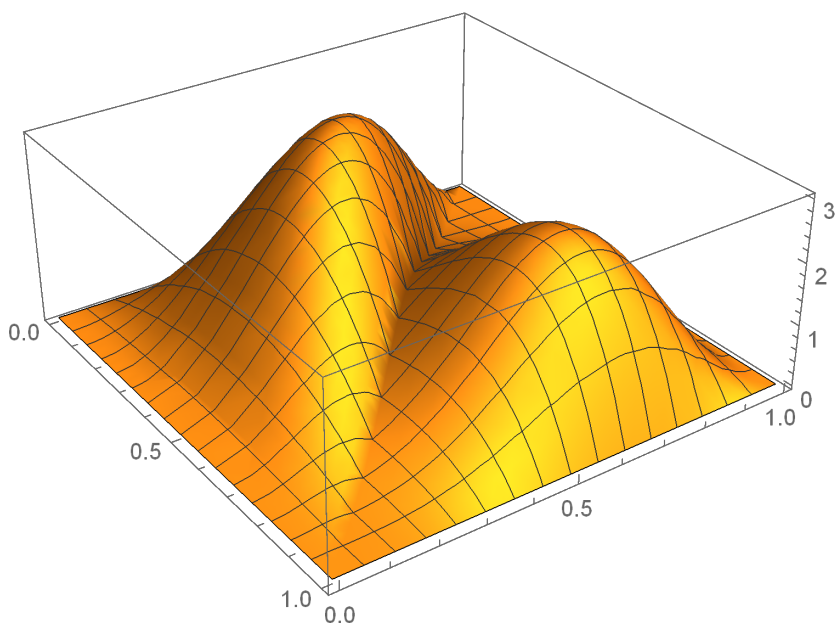


Figure 5.7: The pdf of the mixture of Dirichlet and beta density functions

adjusted density estimate is plotted in Figure 5.9, along with a KDE. This estimate involves higher orders for the polynomial adjustment in order to conform to the bimodality of the distribution. Additionally, the resulting density estimate exhibits a better fit since its associated SSD (0.0536) is lower than that of the KDE (0.0750).

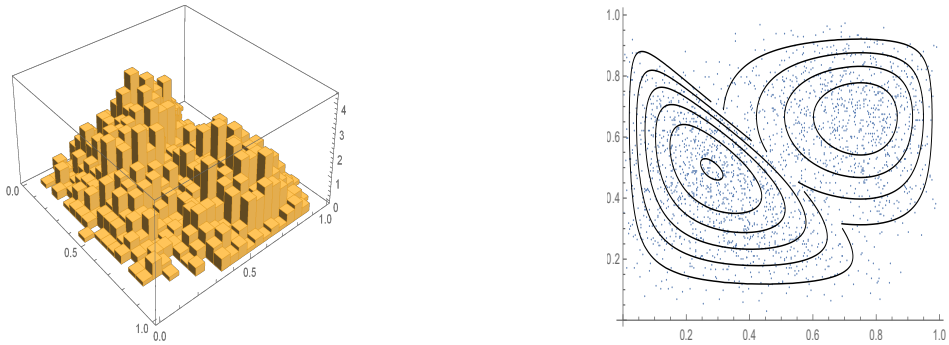
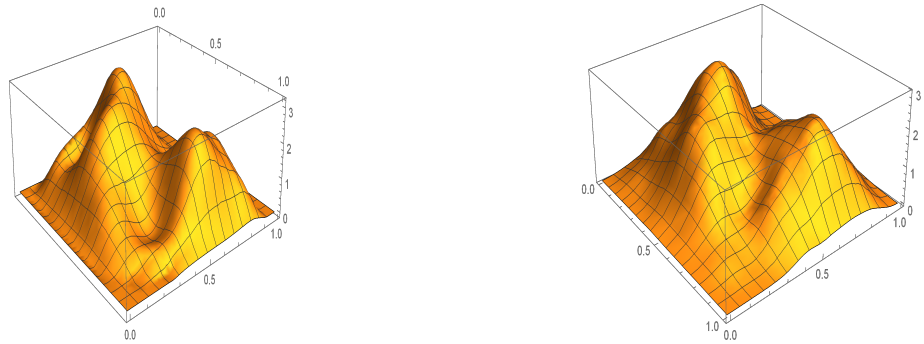


Figure 5.8: Histogram of the original data simulated from a mixture of a Dirichlet and a bivariate beta density functions and contour plot



$SSD_{6,4} = 0.0536251$ $SSD_{KDE} = 0.075049$

Figure 5.9: Adjusted density estimate, $f_{6,4}(x, y)$, and a kernel density estimate of the simulated data from a mixture of a Dirichlet and a bivariate beta density functions

Bibliography

[1] Barndorff-Nielsen O. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70**(2), 343–365.

[2] Barndorff-Nielsen O. and Cox D. R. (1984). Bartlett adjustments to the likelihood ratio statistic and the distribution of the maximum likelihood estimator. *Journal of the Royal Statistical Society. Series B (Methodological)* **46**(3), 483–495.

[3] Butler, R. W. (2007). *Saddlepoint Approximations with Applications (Vol 22)*. Cambridge University Press, Cambridge.

[4] Castillo, E., Hadi, A. S., Balakrishnan, N., and Sarabia, J. M. (2005). *Extreme Value and Related Models with Applications in Engineering and Science*. Wiley, Hoboken, NJ.

[5] Daniels, H. E. (1954). Saddlepoint approximations in statistics. *The Annals of Mathematical Statistics* **25**(4), 631–650.

[6] Fasiolo, M., Wood, S. N., Hartig, F., and Bravington, M. V. (2018). An extended empirical saddlepoint approximation for intractable likelihoods. *Electronic Journal of Statistics* **12**(1), 1544-1578.

[7] Feuerverger, A. (1989). On the empirical saddlepoint approximation. *Biometrika* **76**(3), 457–464.

[8] Fraser, D. A. S. (1988). Normed likelihood as saddlepoint approximation. *Journal of Multivariate Analysis* **27**(1), 181–193.

[9] Goutis C. and Casella G. (1999). Explaining the saddlepoint approximation. *The American Statistician* **53**(3), 216–224.

[10] Jensen, J. L. (1992). The modified signed likelihood statistic and saddlepoint approximations. *Biometrika* **79**(4), 693–703.

[11] Jensen, J. L. (1995). *Saddlepoint Approximations*. Clarendon Press: Oxford Statistical Science Series 16, Oxford.

[12] Kolassa, J. E. (1994). *Series Approximation Methods in Statistics*. Springer-Verlag, New Yori.

[13] Yue, S. (2001). A bivariate gamma distribution for use in multivariate flood frequency analysis. *Hydrological Processes* **15**(6), 1033–1045.

[14] Zareamoghaddam, H., Provost, S. B. and Ahmed, S. E. (2017). A moment-based bivariate density estimation methodology applicable to big data modeling. *Journal of Probability and Statistical Science* **15**(2), 135–152.

Chapter 6

Concluding Remarks and Further Research Directions

6.1 Concluding remarks

Several developments are introduced in connection with the approximation and estimation of heavy-tailed density functions, some of which also apply to other types of distributions. It is explained in Chapter 2 that on initially applying the exponential tilting technique to heavy-tailed density functions such as the Pareto, Student-t and Cauchy, one can approximate the resulting densities by employing the polynomially adjusted denisty approximation technique described in Section 1.4. Alternatively, density approximants can be obtained by truncation. Density approximations can then be secured between the points of truncation. Incidentally, in the context of density estimation, this method is related to the application of the peak-over-threshold model, a Pareto distribution being used for modeling purposes beyond the truncation point. The third methodology involves transforming the distributions so that their supports are finite. These techniques are then extended to the context of density estimation and their validity is corroborated by means of simulation studies. As well, illustrative actuarial applications are presented.

Novel approaches whereby the Box-Cox transform and the empirical saddlepoint density estimation technique are utilized for determining the empirical endpoints associated with various types of distributions in Chapter 3. An alternative technique based on the four-parameter generalized beta density is also proposed.

Additionally, an iterative algorithm and an approximation expressed in terms of a linear combination of Bernstein polynomials were introduced in Chapter 4 for obtaining smooth *bona fide* density functions.

A certain polynomial adjustment was applied to bivariate empirical saddlepoint density estimates in Chapter 5 in order to accurately model bivariate data sets.

The implementation of the proposed methodologies such as the constrained estimation of the four parameters of the generalized beta distribution and the adjusted bivariate empirical saddlepoint density estimation technique in the symbolic computing package *Mathematica* also represents a notable contribution of this dissertation. The *Mathematica* code needed for implementing the main examples in each chapter, which constitutes another significant contribution of this thesis, is included in the Appendix.

6.2 Further research

For instance, one could consider applying the Esscher transform or the transformation of variables technique to multivariate heavy-tailed distributions so that the multivariate counterpart of the polynomially adjusted density estimation methodology could be applied.

Similarly, the univariate techniques proposed to secure *bona fide* density functions could be extended to the multivariate distributions; the iterative approach is directly applicable whereas the alternative methodology would rely on multivariate Bernstein polynomials.

As well, empirical multivariate saddlepoint density estimates could be utilized for estimating the endpoints of a multivariate distribution. This approach could also be used for obtaining multivariate density estimates with polynomial adjustments.

Moreover, products of generalized beta densities could be employed for determining endpoints in each direction after standardization; then on taking the inverse transformation, one could observe where these points are mapped with respect to the original distribution.

Appendix A

Mathematica code

The Mathematica code utilized for implementing the main numerical examples presented in this dissertation is included in this appendix. The evaluation of the moments which are utilized for approximating or estimating density functions is carried out with rational numbers so as to prevent any loss of precision.

A.1 Code for Chapter 2

A.1.1 The density approximants of a type-II Pareto (3.5,3.5) distribution obtained via exponential tilting (Section 2.5.1)

The target density

```
f[x_] := a b^a / (b + x)^(a+1)
a = 3.5; b = 3.5;
Plot[f[x], {x, 0, 10}, PlotRange → All]
```

Comparing the exact and numerically evaluated tilted moments given the tilting parameter $\theta = \frac{a}{b}$

```
nc[s_] := a (b * s)^((a-1)/2) e^((b*s)/2) WhittakerW[-(a + 1)/2, -a/2, b * s];
fstar[x_, s_] := e^(-s x) f[x] / nc[s]
μNInt[h_, s_] := μNInt[h, s] = NIntegrate[x^h fstar[x, s], {x, 0, ∞}] // Chop
m[h_, s_] :=
 m[h, s] = Rationalize[(b/s)^(h/2) Gamma[h + 1] WhittakerW[(-a-h-1)/2, (-a+h)/2, b s] / WhittakerW[(-a-1)/2, -a/2, b s], 10^-300]
Table[μNInt[n, a/b], {n, 0, 25}]
Table[N[m[h, a/b]], {h, 0, 25}]
```

- **The approximant of the tilted density**

```
α = 1;
β[s_] := m[1, s];
ψ[y_, s_] := ψ[y, s] = PDF[GammaDistribution[α, β[s]], y];
mgfbase[t_, s_] :=
 mgfbase[t, s] = MomentGeneratingFunction[GammaDistribution[α, β[s]], t]
mmbase[n_, s_] := mmbase[n, s] = D[mgfbase[t, s], {t, n}] /. t -> 0;
IMb[x_, s_, n_] := IMb[x, s, n] =
  (Inverse[Table[mmbase[q + r, s], {q, 0, n}, {r, 0, n}]]. Table[m[j, s], {j, 0, n}]).
   Table[x^j, {j, 0, n}];
fappbefore[x_, s_, n_] := fappbefore[x, s, n] = ψ[x, s] IMb[x, s, n]
Simplify[N[fappbefore[x, 3, 16]]]
Show[Plot[fstar[x, 3], {x, 0, 10}, PlotRange → All, PlotStyle → {Green, Thick}],
 Plot[Evaluate[fappbefore[x, 3, 16]], {x, 0, 10},
  PlotRange → All, PlotStyle → {Red, Dashed}]]
```

- **The approximant of the target density via the inverse transformation**

```
fapp[x_, s_, n_] :=
 fapp[x, s, n] = e^(s x) * Rationalize[nc[s], 10^-30] * fappbefore[x, s, n]
nc1[s_, n_] := nc1[s, n] = Rationalize[Integrate[fapp[x, s, n], {x, 0, ∞}], 10^-30];
fappfinal[x_, s_, n_] := fappfinal[x, s, n] = fapp[x, s, n] / nc1[s, n]
Simplify[N[fappfinal[x, 1.12, 32]]]
Off[Reduce::"ratnz"]
N[Reduce[fappfinal[x, 1.12, 32] == 0, x]]
Show[Plot[f[x], {x, 0, 10}, PlotRange → All, PlotStyle → {Green, Thick}],
 Plot[Evaluate[fappfinal[x, 1.12, 32]],
  {x, 0, 10}, PlotRange → All, PlotStyle → {Red, Dashed}]]
```

- **Comparing the integrated squared errors for several approximants**

```
ISE[s_, h_] := ISE[s, h] = Integrate[(fappfinal[x, s, h] - f[x])^2, {x, 0, ∞}]
ListPlot[Table[{s, Log[ISE[s, 12]]}, {s, 0.81, 2.9, 0.01}],
 PlotRange → {-23, -14}, PlotStyle → Gray]
ListPlot[Table[{s, Log[ISE[s, 18]]}, {s, 0.81, 2.9, 0.01}],
 PlotRange → {-23, -14}, PlotStyle → Cyan]
ListPlot[Table[{s, Log[ISE[s, 24]]}, {s, 0.81, 2.9, 0.01}],
 PlotRange → {-23, -14}, PlotStyle → Green]
ListPlot[Table[{s, Log[ISE[s, 32]]}, {s, 0.81, 2.9, 0.01}],
 PlotRange → {-23, -14}, PlotStyle → Red]
```

- **Comparing VaR and TVaR**

```
U[k_?NumericQ] := Integrate[fappfinal[x, 0.88, 30], {x, 0, k}]
V[k_?NumericQ] := Integrate[fappfinal[x, 1, 31], {x, 0, k}]
W[k_?NumericQ] := Integrate[fappfinal[x, 1.12, 32], {x, 0, k}]
q = 0.99;
N[Rationalize[b * ((1 - q)^(-1/a) - 1), 10^-100]]
q1 = N[Rationalize[FindRoot[U[j] == q, {j, start}][[1]][[2]], 10^-100]]
q2 = N[Rationalize[FindRoot[V[j] == q, {j, start}][[1]][[2]], 10^-100]]
q3 = N[Rationalize[FindRoot[W[j] == q, {j, start}][[1]][[2]], 10^-100]]
```

$$\frac{1}{1-q}\int_q^1 b * \left((1-u)^{-\frac{1}{a}} - 1\right) du$$

```
Integrate[x fappfinal[x, 0.88, 30], {x, q1, ∞}]
 Integrate[fappfinal[x, 0.88, 30], {x, q1, ∞}]
Integrate[x fappfinal[x, 1, 31], {x, q2, ∞}]
 Integrate[fappfinal[x, 1, 31], {x, q2, ∞}]
Integrate[x fappfinal[x, 1.12, 32], {x, q3, ∞}]
 Integrate[fappfinal[x, 1.12, 32], {x, q3, ∞}]
```

A.1.2 The density approximants of a type-II Pareto (3.5,3.5) distribution obtained by truncation (Section 2.5.1)

The approximant of the truncated density function on the interval [0, 95th percentile]

```
f[x_] := a b^a / (b + x)^(a+1)
q[x_] := b * ((1 - x)^(-1/a) - 1);
a = 3.5; b = 3.5;
v = 0.95;
m1[h_] := m1[h] = Rationalize[Integrate[x^h f[x], {x, 0, q[v]}], 10^-500]
N[Table[m1[h], {h, 0, 30}]]
m1trunc[h_] := m1trunc[h] = m1[h] / m1[0]
m1base[h_] := m1base[h] = Rationalize[q[v]^h / (1 + h), 10^-500]
IMb1[x_, n_] := IMb1[x, n] =
   (Inverse[Table[m1base[r + c], {r, 0, n}, {c, 0, n}]]. Table[m1trunc[j], {j, 0, n}]).
    Table[x^j, {j, 0, n}];
fapp1[x_, n_] := fapp1[x, n] = If[0 ≤ x ≤ q[v],
    m1[0] Rationalize[1 / q[v], 10^-500] IMb1[x, n], 0];
Show[Plot[f[x], {x, 0, q[v]}, PlotRange → All, AxesOrigin → {0, 0}], Plot[
  Evaluate[fapp1[x, 25]], {x, 0, q[v]}, PlotStyle → {Red, Dashed}, PlotRange → All]]
```

A.1.3 Density approximants of a type-II Pareto (3.5,3.5) distribution obtained via transformation of variables (Section 2.5.1)

```
f[x_] := a b^a / (b + x)^(a+1)
g[y_, θ_] := θ/(1 - y)^2 f[θ y/(1 - y)];
a = 3.5; b = 3.5;
Plot[f[x], {x, 0, 10}, PlotRange → All]
Plot[g[y, b], {y, 0, 1}, PlotRange → All]
m[h_, θ_] := Rationalize[Integrate[x^h g[x, θ], {x, 0, 1}], 10^-500]
mbase[h_] := 1/(h + 1)
IMb[x_, θ_, n_] :=
  (Inverse[Table[mbase[r + c], {r, 0, n}, {c, 0, n}]]. Table[m[j, θ], {j, 0, n}]).
   Table[x^j, {j, 0, n}];
gapp[x_, θ_, n_] := IMb[x, θ, n];
Show[Plot[g[x, 2], {x, 0, 1}, PlotRange → All, AxesOrigin → {0, 0}], Plot[
  Evaluate[gapp[x, 2, 9]], {x, 0, 1}, PlotStyle → {Red, Dashed}, PlotRange → All]]
fapp[x_, θ_, n_] := θ/(x + θ)^2 gapp[x/(x + θ), θ, n]
Show[Plot[f[x], {x, 0, 10}, PlotRange → All, AxesOrigin → {0, 0}],
 Plot[fapp[x, 3.5, 9], {x, 0, 10}, PlotStyle → {Red, Dashed}, PlotRange → All]]
N[Reduce[gapp[x/(x + 52/10), 52/10, 9] == 0, x]]
```

A.1.4 The density approximants of a Cauchy distribution obtained via exponential tilting (Section 2.5.2)

Approximated density of a half-Cauchy distribution

```
f[x_] := 2/(π (1 + x^2))
Plot[f[x], {x, 0, 6}, PlotRange → All]
nc[s_] :=
  nc[s] = Rationalize[(2 CosIntegral[s] Sin[s] + Cos[s] (π - 2 SinIntegral[s]))/π, 10^-50];
fstar[x_, s_] = fstar[x, s] = e^(-s x) f[x]/nc[s];
Plot[Evaluate[fstar[x, 1]], {x, 0, 6}, PlotRange → All]
m[s_, h_] := Which[h == 0, (2 CosIntegral[s] Sin[s] + Cos[s] (π - 2 SinIntegral[s]))/(2 CosIntegral[s] Sin[s] + Cos[s] (π - 2 SinIntegral[s])),
  h == 1, (-2 Cos[s] CosIntegral[s] + Sin[s] (π - 2 SinIntegral[s]))/(2 CosIntegral[s] Sin[s] + Cos[s] (π - 2 SinIntegral[s])), h == 2,
   (2/s - π Cos[s] - 2 CosIntegral[s] Sin[s] + 2 Cos[s] SinIntegral[s]) /
    (2 CosIntegral[s] Sin[s] + Cos[s] (π - 2 SinIntegral[s])),
  h ≥ 3, MeijerG[{{1/2 - h/2}, {}}, {{1/2 - h/2, 0, 1/2}, {}}, s^2/4] /
    ((2 CosIntegral[s] Sin[s] + Cos[s] (π - 2 SinIntegral[s])) π^(1/2))]
ψ[y_, s_] := ψ[y, s] = PDF[GammaDistribution[α, β[s]], y];
mmbase[s_, n_] := mmbase[s, n] = β[s]^n Factorial[n]
IMb[x_, s_, n_] :=
  IMb[x, s, n] = (Inverse[Table[mmbase[s, r + t], {r, 0, n}, {t, 0, n}]].
      Table[Rationalize[m[s, j], 10^-50], {j, 0, n}]).Table[x^j, {j, 0, n}];
fappbefore[x_, s_, n_] := fappbefore[x, s, n] = ψ[x, s] IMb[x, s, n]
Show[Plot[Evaluate[fstar[x, 1]], {x, 0, 60}, PlotRange → All,
  PlotStyle → {Green, Thick}], Plot[Evaluate[fappbefore[x, 1, 9]],
  {x, 0, 60}, PlotRange → All, PlotStyle → {Red, Dashed}]]
fapp[x_, s_, h_] := e^(s x) * nc[s] * fappbefore[x, s, h]
nc1[s_, h_] := Rationalize[Integrate[fapp[x, s, h], {x, 0, ∞}], 10^-30];
fappfinal[x_, s_, h_] := fapp[x, s, h] / nc1[s, h]
Off[Reduce::"ratnz"]
N[Reduce[fappfinal[x, 1, 37] == 0, x]]
Show[Plot[f[x], {x, 0, 10}, PlotRange → All, PlotStyle → {Green, Thick}],
 Plot[Evaluate[fappfinal[x, 1, 37]], {x, 0, 10},
  PlotRange → All, PlotStyle → {Red, Dashed}]]
```

- **Obtaining an approximant of a Cauchy density function by applying the symmetrization technique**

```
g[x] = 1/(π (1 + x^2));
Show[
 Plot[Evaluate[g[x]], {x, -10, 10}, PlotRange → All, PlotStyle → {Green, Thick}],
 Plot[Evaluate[1/2 fappfinal[x, 1, 37] + 1/2 fappfinal[-x, 1, 37]],
  {x, -10, 10}, PlotRange → All, PlotStyle → {Red, Dashed}]]
```

A.1.5 The density estimate obtained via exponential tilting for data simulated from a type-II Pareto (3.5,35) distribution

- **Using an exponential density as base density**

```
SeedRandom[1437]
RNGUnif = RandomReal[1, 5 * 10^3];
α = 3.5; β = 35;
RNGPareto = β * ((1 - RNGUnif)^(-1/α) - 1);
Show[Histogram[RNGPareto, {0, 60, 1}, "PDF"],
 Plot[PDF[ParetoDistribution[b, a, 0], x], {x, 0, 60}, PlotRange → All]]
n1 = Length[RNGPareto];
μ = 1/n1 Total[RNGPareto];
σ = Sqrt[1/(n1 - 1) Total[(RNGPareto - μ)^2]];
moddata = RNGPareto / σ;
LaplaceMoment[h_, θ_] :=
 LaplaceMoment[h, θ] = Rationalize[Total[e^(-θ moddata) moddata^h]/n1, 10^-100]
μ1[h_, θ_] := μ1[h, θ] = LaplaceMoment[h, θ]/LaplaceMoment[0, θ];
a2 = 1;
b2[θ_] := b2[θ] = μ1[1, θ];
ψ2[y_, θ_] := PDF[GammaDistribution[a2, b2[θ]], y];
mgfbase2[t_, θ_] :=
  MomentGeneratingFunction[GammaDistribution[a2, b2[θ]], t];
mm2[n_, θ_] := D[mgfbase2[t, θ], {t, n}] /. t -> 0;
IMb2[x_, θ_, n_] :=
  (Inverse[Rationalize[Table[mm2[r + s, θ], {r, 0, n}, {s, 0, n}], 10^-300]].
     Table[μ1[j, θ], {j, 0, n}]).Table[x^j, {j, 0, n}];
fappbefore2[x_, θ_, n_] := ψ2[x, θ] IMb2[x, θ, n];
fappafter2[x_, θ_, n_] :=
 fappafter2[x, θ, n] = (e^(θ x) * LaplaceMoment[0, θ] * fappbefore2[x, θ, n]) /
   Integrate[e^(θ t) * LaplaceMoment[0, θ] * fappbefore2[t, θ, n], {t, 0, ∞}]
Show[Histogram[moddata, {0, 6, 0.1}, "PDF"], Plot[Evaluate[fappafter2[x, 3.5, 8]],
  {x, 0, 10}, PlotRange → All, PlotStyle → {Red, Dashed}]]
N[Reduce[IMb2[x, 3, 9] == 0, x, Reals]]
```

- **Anderson-Darling test with various tilting parameters and polynomial adjustment degrees**

```
𝒟[θ_, h_] := ProbabilityDistribution[fappafter2[x, θ, h], {x, 0, ∞}]
𝒫[θ_, h_] := 𝒫[θ, h] = AndersonDarlingTest[moddata, 𝒟[θ, h], "TestStatistic"]
ListPlot[Table[{θ, 𝒫[θ, 3]}, {θ, 0.5, 5, 0.1}], PlotStyle → Black, PlotRange → {0, 5}]
ListPlot[Table[{θ, 𝒫[θ, 4]}, {θ, 0.5, 5, 0.1}], PlotStyle → Pink, PlotRange → {0, 5}]
ListPlot[Table[{θ, 𝒫[θ, 5]}, {θ, 0.5, 5, 0.1}], PlotStyle → Red, PlotRange → {0, 5}]
ListPlot[Table[{θ, 𝒫[θ, 6]}, {θ, 0.5, 5, 0.1}],
 PlotStyle → Orange, PlotRange → {0, 5}]
ListPlot[Table[{θ, 𝒫[θ, 7]}, {θ, 0.5, 5, 0.1}], PlotStyle → Blue, PlotRange → {0, 5}]
ListPlot[Table[{θ, 𝒫[θ, 8]}, {θ, 0.5, 5, 0.1}], PlotStyle → Cyan, PlotRange → {0, 5}]
ListPlot[Table[{θ, 𝒫[θ, 9]}, {θ, 0.5, 5, 0.1}], PlotStyle → Brown, PlotRange → {0, 5}]
ListPlot[Table[{θ, 𝒫[θ, 10]}, {θ, 0.5, 5, 0.1}],
 PlotStyle → Purple, PlotRange → {0, 5}]
```

A.1.6 The density estimate obtained via exponential tilting for the Automobile insurance claim data set

Using a gamma density function as base density

```
AutoClaim = Import["C:\\Users\\owner\\Documents\\John\\Research\\Moment-based
      density\\AutoClaim.xlsx", {"Data", 1}];
Transpose[AutoClaim];
data = %[[1]];
n = Length[data]
μ = 1/n Total[data];
σ = Sqrt[1/(n - 1) Total[(data - μ)^2]];
Histogram[data, 50, "PDF"]
moddata = data / σ;
LaplaceMoment[h_, θ_] :=
 LaplaceMoment[h, θ] = Rationalize[Total[e^(-θ moddata) moddata^h]/n, 10^-100]
μ1[h_, θ_] := μ1[h, θ] = LaplaceMoment[h, θ]/LaplaceMoment[0, θ];
a3[θ_] := a3[θ] = μ1[1, θ]^2/(μ1[2, θ] - μ1[1, θ]^2);
b3[θ_] := b3[θ] = (μ1[2, θ] - μ1[1, θ]^2)/μ1[1, θ];
ψ3[y_, θ_] := ψ3[y, θ] = PDF[GammaDistribution[a3[θ], b3[θ]], y];
mgfbase3[t_, θ_] :=
  mgfbase3[t, θ] = MomentGeneratingFunction[GammaDistribution[a3[θ], b3[θ]], t];
mm3[n_, θ_] := mm3[n, θ] = D[mgfbase3[t, θ], {t, n}] /. t -> 0;
IMb3[x_, θ_, n_] := IMb3[x, θ, n] =
   (Inverse[Rationalize[Table[mm3[r + s, θ], {r, 0, n}, {s, 0, n}], 10^-300]].
      Table[μ1[j, θ], {j, 0, n}]).Table[x^j, {j, 0, n}];
fappbefore3[x_, θ_, n_] := fappbefore3[x, θ, n] = ψ3[x, θ] IMb3[x, θ, n];
fappafter3[x_, θ_, n_] :=
 fappafter3[x, θ, n] = (e^(θ x) * LaplaceMoment[0, θ] * fappbefore3[x, θ, n]) /
   Integrate[e^(θ t) * LaplaceMoment[0, θ] * fappbefore3[t, θ, n], {t, 0, ∞}]
Show[Histogram[moddata, {0, 9, 0.1}, "PDF"], Plot[Evaluate[fappafter3[x, 1.5, 9]],
  {x, 0, 9}, PlotRange → All, PlotStyle → {Red, Dashed}]]
N[Reduce[IMb3[x, 1.5, 9] == 0, x, Reals]]
```

A.2 Code for Chapter 3

A.2.1 Procedure for determining endpoints by making use of the Box-Cox transform and the empirical saddlepoint estimates

- **Data simulated from a GP (0,1,1) distribution (sample size: 2,500)**

```
SeedRandom[5832]
RNGUnif = RandomVariate[UniformDistribution[], 2500];
m = 0;
s = 1;
α = 1;
RNGPareto = m + (s * (RNGUnif^(-α) - 1))/α;
n = Length[RNGUnif];
f[x_] := 1/s (1 + (x - m)/(α s))^(-α-1)
Show[Histogram[RNGPareto, {0, 20, 0.25}, "PDF"],
 Plot[f[x], {x, 0, 20}, PlotRange → All]]
{Min[RNGPareto], Max[RNGPareto]}
```

- **Applying the Box-Cox transform and rescaling it with the resulting standard deviation**

```
autoboxcox[λ_] := If[λ == 0, Log[RNGPareto], (RNGPareto^λ - 1)/λ]
prolikelihood[λ_] :=
 Total[Log[PDF[NormalDistribution[Mean[autoboxcox[λ]], StandardDeviation[
       autoboxcox[λ]]], autoboxcox[λ]]]] + (λ - 1) * Total[Log[RNGPareto]]
ListPlot[Table[{λ, prolikelihood[λ]}, {λ, -1.5, 0.5, 0.1}]]
λ1 = 0;
Histogram[autoboxcox[λ1]]
{Min[autoboxcox[λ1]], Max[autoboxcox[λ1]]}
σ1 = StandardDeviation[autoboxcox[λ1]]
standdata = autoboxcox[λ1]/σ1;
Histogram[standdata, 50, "PDF"]
{Min[standdata], Max[standdata]}
```

- **The logarithm of the empirical saddlepoint estimates determined for points lying outside of the range of the sample**

```
data = Rationalize[standdata, 10^-100];
k[t_] := k[t] = Log[Sum[E^(data[[j]] t) / n, {j, 1, n}]];
k1[t_] := k1[t] = k'[t];
k2[t_] := k2[t] = k''[t];
ξ[q_] := ξ[q] = FindRoot[k1[ξ] == q, {ξ, 0}][[1]][[2]];
g1[q_] := (1/(2 π k2[ξ[q]]))^(1/2) e^(k[ξ[q]]-ξ[q] q);

w[q_] := w[q] = Sign[ξ[q]] * Sqrt[2 * (q * ξ[q] - k[ξ[q]])];
v[q_] := v[q] = ξ[q] * Sqrt[k2[ξ[q]]];
FSaddle[q_] :=
  FSaddle[q] = CDF[NormalDistribution[0, 1], w[q] + (1 / w[q]) Log[(v[q] / w[q])]];
Off[FindRoot::jsing, FindRoot::lstol]
tf1 = Table[{e^((-7.5+.01 i)*σ1), Log[1/e^((-7.5+.01 i)*σ1) g1[-7.5 + 0.01 i]]}, {i, 0, 400}]
Show[ListPlot[{{0, 0}, {e^(-3.5*σ1), -750}}, PlotStyle → None],
 ListPlot[tf1, PlotRange → All],
 ListPlot[{{e^(Min[data]*σ1), 0}}, PlotStyle → Red, PlotMarkers → {"■", 15}]]
tf2 = Table[{e^((4.5+.01 i)*σ1), Log[1/e^((4.5+.01 i)*σ1) g1[4.5 + 0.01 i]]}, {i, 0, 400}]
Show[ListPlot[{{e^(4.5*σ1), 0}, {e^(8.5*σ1), -750}}, PlotStyle → None],
 ListPlot[tf2, PlotRange → All],
 ListPlot[{{e^(Max[data]*σ1), 0}}, PlotStyle → Red, PlotMarkers → {"■", 15}]]
```

A.2.1 Obtaining estimated endpoints based on the generalized beta distribution

- **Penalized maximum likelihood estimate method**

```
likelihood[x_] :=
 -Log[Beta[a, b]] - (a + b - 1) Log[u - l] + (a - 1) * Log[x - l] + (b - 1) * Log[u - x]
opt = NMaximize[{Total[likelihood[autoboxcox[λ1]]] +
     Log[(Sort[autoboxcox[λ1]][[1]] - l)/(Sort[autoboxcox[λ1]][[2]] - l)] + Log[(u - Sort[autoboxcox[λ1]][[n]])/(u - Sort[autoboxcox[λ1]][[n - 1]])],
    a > 0, b > 0, l < Sort[autoboxcox[λ1]][[1]], u > Sort[autoboxcox[λ1]][[n]]},
   {a, b, l, u}, MaxIterations → 1000]
a1 = opt[[2]][[1]][[2]]
b1 = opt[[2]][[2]][[2]]
l1 = opt[[2]][[3]][[2]]
u1 = opt[[2]][[4]][[2]]
```

- **Feasibility constrained moment matching technique**

```
opt1 = NMinimize[{(l + (u - l) a/(a + b) - Mean[autoboxcox[λ1]])^2 +
    ((u - l)^2 (a * b)/((a + b)^2 (a + b + 1)) - StandardDeviation[autoboxcox[λ1]]^2)^2 +
    (2 (b - a) Sqrt[a + b + 1]/((a + b + 2) Sqrt[a * b]) - Skewness[autoboxcox[λ1]])^2 +
    (3 (a + b + 1) (2 (a + b)^2 + a * b * (a + b - 6))/(a * b * (a + b + 2) (a + b + 3)) - Kurtosis[autoboxcox[λ1]])^2, a > 0,
   b > 0, l < Sort[autoboxcox[λ1]][[1]], u > Sort[autoboxcox[λ1]][[n]]}, {a, b, l, u}]
a2 = opt1[[2]][[1]][[2]]
b2 = opt1[[2]][[2]][[2]]
l2 = opt1[[2]][[3]][[2]]
u2 = opt1[[2]][[4]][[2]]
```

- **Fitting the density estimates by applying the inverse transform**

```
f1[x_] := If[l1 ≤ x ≤ u1, 1/Beta[a1, b1] (x - l1)^(a1-1) (u1 - x)^(b1-1)/(u1 - l1)^(a1+b1-1), 0]
f2[x_] := If[l2 ≤ x ≤ u2, 1/Beta[a2, b2] (x - l2)^(a2-1) (u2 - x)^(b2-1)/(u2 - l2)^(a2+b2-1), 0]
Show[Histogram[RNGPareto, {0, 20, 0.25}, "PDF"],
 Plot[f[x], {x, 0, 20}, PlotRange → All],
 Plot[1/x f1[Log[x]], {x, 0, 20}, PlotStyle → {Green, Dotted}, PlotRange → All],
 Plot[1/x f2[Log[x]], {x, 0, 20}, PlotStyle → {Red, Dashed}, PlotRange → All]]
```

A.3 Code for Chapter 4

A.3.1 Procedure for obtaining a smooth *bona fide* density function

Simulated data from a mixture of equally weighted beta density functions

```
SeedRandom[8267]
RNGUnif = RandomReal[1, 200];
partition01 = Select[RNGUnif, # > 0.5 &];
partition02 = Select[RNGUnif, # ≤ 0.5 &];
n1 = Length[partition01];
n2 = Length[partition02];
RNGBeta1 = RandomVariate[BetaDistribution[30, 5], n1];
RNGBeta2 = RandomVariate[BetaDistribution[6, 25], n2];
RNGMixture = Join[RNGBeta1, RNGBeta2];
f[x_] := 1/2 PDF[BetaDistribution[30, 5], x] + 1/2 PDF[BetaDistribution[6, 25], x];
Show[Histogram[RNGMixture, 50, "PDF"], Plot[f[x], {x, 0, 1}, PlotRange → All]]
```

Initial density estimate

```
n = Length[RNGMixture];
m[h_] := m[h] = Rationalize[Total[RNGMixture^h]/n, 10^-20];
ψ[y_] := PDF[UniformDistribution[], y];
mm[h_] := 1/(h + 1);
IMb[x_, n_] :=
  (Inverse[Table[mm[r + c], {r, 0, n}, {c, 0, n}]]. Table[m[j], {j, 0, n}]).
   Table[x^j, {j, 0, n}];
fapp[x_, n_] := ψ[x] IMb[x, n];
N[Reduce[fapp[x, 14] == 0, x]]
Show[Histogram[RNGMixture, {0, 1, 0.025}, "PDF"],
 Plot[f[x], {x, 0, 1}, PlotRange → All, PlotStyle → Green],
 Plot[Evaluate[fapp[x, 14]], {x, 0, 1}, PlotRange → All, PlotStyle → {Red, Dotted}]]
```

- **Density estimate obtained from the first iteration**

```
N[Reduce[fapp[x, 14] == 0, x]]
v1 = Rationalize[Reduce[fapp[x, 14] == 0, x][[4]][[2]], 10^-100];
v2 = Rationalize[Reduce[fapp[x, 14] == 0, x][[9]][[2]], 10^-100];
N[{v1, v2}]
μ11[j_] := μ11[j] =
  Rationalize[NIntegrate[x^j fapp[x, 14], {x, v1, v2}, WorkingPrecision → 20], 10^-300]
μ12[j_] := μ12[j] = Rationalize[NIntegrate[x^j Sqrt[fapp[x, 14]^2],
    {x, v1, v2}, WorkingPrecision → 20], 10^-300]
μ1mod[j_] := μ1mod[j] = (1/2 * (μ11[j] + μ12[j]))/(1/2 * (μ11[0] + μ12[0]))
ψ1[y_] := PDF[UniformDistribution[], y];
mm1[h_] := 1/(h + 1); IMb1[x_, j_] := IMb1[x, j] =
  (Inverse[Table[mm1[r + c], {r, 0, j}, {c, 0, j}]]. Table[μ1mod[h], {h, 0, j}]).
   Table[x^h, {h, 0, j}]; fapp1[x_, j_] := fapp1[x, j] = ψ1[x] IMb1[x, j];
Show[Plot[Evaluate[fapp[x, 14]], {x, 0, 1}, PlotRange → All,
  PlotStyle → {Red, Dotted}], Plot[Evaluate[fapp1[x, 19]],
  {x, 0, 1}, PlotStyle → {Blue, DotDashed}, PlotRange → All]]
```

- **Density estimate obtained from the second iteration**

```
N[Reduce[fapp1[x, 19] == 0, x]]
v3 = Rationalize[Reduce[fapp1[x, 19] == 0, x][[4]][[2]], 10^-100];
v4 = Rationalize[Reduce[fapp1[x, 19] == 0, x][[9]][[2]], 10^-100];
N[{v3, v4}]
μ21[j_] := μ21[j] = Rationalize[
   NIntegrate[x^j fapp1[x, 19], {x, v3, v4}, WorkingPrecision → 20], 10^-300]
μ22[j_] := μ22[j] = Rationalize[NIntegrate[x^j Sqrt[fapp1[x, 19]^2],
    {x, v3, v4}, WorkingPrecision → 20], 10^-300]
μ2mod[j_] := μ2mod[j] = (1/2 * (μ21[j] + μ22[j]))/(1/2 * (μ21[0] + μ22[0]))
ψ2[y_] := PDF[UniformDistribution[], y];
mm2[h_] := 1/(h + 1); IMb2[x_, j_] := IMb2[x, j] =
  (Inverse[Table[mm2[r + c], {r, 0, j}, {c, 0, j}]]. Table[μ2mod[h], {h, 0, j}]).
   Table[x^h, {h, 0, j}]; fapp2[x_, j_] := fapp2[x, j] = ψ2[x] IMb2[x, j];
Show[Plot[Evaluate[fapp[x, 14]], {x, 0, 1}, PlotRange → All,
  PlotStyle → {Red, Dotted}], Plot[Evaluate[fapp1[x, 19]], {x, 0, 1},
  PlotRange → All, PlotStyle → {Blue, DotDashed}], Plot[Evaluate[fapp2[x, 19]],
  {x, 0, 1}, PlotStyle → {Cyan, Dashed}, PlotRange → All]]
```

- **Final *bona fide* density estimate**

```
N[Reduce[fapp2[x, 19] == 0]]
v5 = Rationalize[Reduce[fapp2[x, 19] == 0, x][[5]][[2]], 10^-100];
v6 = Rationalize[Reduce[fapp2[x, 19] == 0, x][[10]][[2]], 10^-100];
N[Reduce[D[fapp2[x, 19], x] == 0]]
a1 = Reduce[D[fapp2[x, 19], x] == 0][[5]][[2]];
b1 = Reduce[D[fapp2[x, 19], x] == 0][[7]][[2]];
N[fapp2[a1, 19]]
N[fapp2[b1, 19]]
mn = Rationalize[fapp2[a1, 19], 10^-100];
fab[x_] := fapp2[x, 19] + Abs[mn]
N[Reduce[fab[x] == 0]]
v7 = Reduce[fab[x] == 0][[4]][[2]];
v8 = Reduce[fab[x] == 0][[5]][[2]];
norcst = Rationalize[NIntegrate[fab[x], {x, v7, v8}], 10^-100];
fn[x_] := fn[x] = fab[x]/norcst
Show[Histogram[RNGMixture, 50, "PDF"],
 Plot[Which[v7 ≤ x ≤ v8, fn[x], x < v7, 0, x > v8, 0],
  {x, 0, 1}, PlotRange → All, PlotStyle → Purple]]
```

A.3.2 Density estimate obtained from Glad's approach

```
S[k_?NumericQ] :=
  Integrate[1/2 ((fapp[x, 14] - k) + Sqrt[(fapp[x, 14] - k)^2]), {x, 0, 1}];
c1 = Rationalize[FindRoot[S[k] == 1, {k, 0.01}][[1]][[2]], 10^-300];
fglad[x_] := 1/2 ((fapp[x, 14] - c2) + Sqrt[(fapp[x, 14] - c1)^2]);
Show[Plot[fapp[x, 14], {x, 0, 1}, PlotRange → All, PlotStyle → {Red, Dashed}],
 Plot[fglad[x], {x, 0, 1}, PlotRange → All, PlotStyle → Blue]]
```

A.3.3 Density estimate obtained from Gajek's approach

Density estimate obtained from the first iteration

```
μp11 = Rationalize[NIntegrate[fapp[x, 14], {x, 0, 1}, WorkingPrecision → 20], 10^-300];
μp12 = Rationalize[
   NIntegrate[Sqrt[fapp[x, 14]^2], {x, 0, 1}, WorkingPrecision → 20], 10^-300];
c11 = 1/2 (μp11 + μp12);
h[x_] := 1/(6 x (1 - x));
fgajek1[x_] := 1/2 * (fapp[x, 14] + Sqrt[fapp[x, 14]^2]) - (c11 - 1)/h[x];
Show[Plot[Evaluate[fapp[x, 14]],
  {x, 0, 1}, PlotRange → All, PlotStyle → {Red, Dotted}],
 Plot[fgajek1[x] , {x, 0, 1}, PlotStyle → {Blue, DotDashed}, PlotRange → All]]
```

Density estimate obtained from the second iteration

```
μp21 = Rationalize[NIntegrate[fgajek1[x], {x, 0, 1}, WorkingPrecision → 20], 10^-300];
μp22 = Rationalize[
   NIntegrate[Sqrt[fgajek1[x]^2], {x, 0, 1}, WorkingPrecision → 20], 10^-300];
c12 = 1/2 (μp21 + μp22);
fgajek2[x_] := 1/2 (fgajek1[x] + Sqrt[fgajek1[x]^2]) - (c12 - 1)/h[x];
Show[Plot[Evaluate[fapp[x, 14]],
  {x, 0, 1}, PlotRange → All, PlotStyle → {Red, Dotted}],
 Plot[fgajek1[x] , {x, 0, 1}, PlotStyle → {Blue, DotDashed}, PlotRange → All],
 Plot[fgajek2[x] , {x, 0, 1}, PlotStyle → {Cyan, Dashed}, PlotRange → All]]
```

A.3.4 Smooth density estimate expressed in terms of Bernstein polynomials

```
v1 = Rationalize[Reduce[fapp[x, 14] == 0, x][[4]][[2]], 10^-100];
v2 = Rationalize[Reduce[fapp[x, 14] == 0, x][[9]][[2]], 10^-100];
N[{v1, v2}]
fapptrunc[x_, h_] := If[v1 ≤ x < v2, 1/2 (fapp[x, h] + Sqrt[fapp[x, h]^2]), 0];
nc = Integrate[fapptrunc[x, 14], {x, 0, 1}];
fappbern2[x_, k_] :=
 fappbern2[x, k] = Sum[(fapptrunc[v/k, 14]/nc BernsteinBasis[k, v, x]), {v, 0, k}]
Show[Plot[Evaluate[fapptrunc[x, 14]/nc], {x, 0, 1}, PlotStyle → {Green, Dashed}],
 Plot[Evaluate[fappbern2[x, 1000]], {x, 0, 1}, PlotStyle → Red]]
```

A.4 Code for Chapter 5

A.4.1 Procedure for obtaining an adjusted empirical saddlepoint estimate

Simulated data from a mixture of Dirichlet and bivariate beta density functions

```
SeedRandom[2283]
RNGUnif = RandomReal[1, 2500];
partition01 = Select[RNGUnif, # > 0.5 &];
partition02 = Select[RNGUnif, # ≤ 0.5 &];
n1 = Length[partition01];
n2 = Length[partition02];
RNGDirichlet = RandomVariate[DirichletDistribution[{2, 3, 2}], n1];
RNGBetaX = RandomVariate[BetaDistribution[4, 2], n2];
RNGBetaY = RandomVariate[BetaDistribution[5, 3], n2];
RNGBeta = Transpose[{RNGBetaX, RNGBetaY}];
data = Join[RNGDirichlet, RNGBeta];
n = Length[data];
Plot3D[1/2 PDF[DirichletDistribution[{2, 3, 2}], {x, y}] +
  1/2 PDF[BetaDistribution[4, 2], x] * PDF[BetaDistribution[5, 3], y],
 {x, 0, 1}, {y, 0, 1}, Exclusions → None]
Histogram3D[data, 20, "PDF"]
Show[ListPlot[data], ContourPlot[1/2 PDF[DirichletDistribution[{2, 3, 2}], {x, y}] +
    1/2 PDF[BetaDistribution[4, 2], x] * PDF[BetaDistribution[5, 3], y],
   {x, 0, 1}, {y, 0, 1}, ContourShading → None]]
Σ = Covariance[data]
rescaled = data.MatrixPower[DiagonalMatrix[Diagonal[Σ]], -0.5];
Histogram3D[rescaled, 20, "PDF"]
```

- **Fitting a second order bivariate interpolating spline to the locally evaluated exponential terms of the empirical saddlepoint estimates**

```
data1 = Rationalize[rescaled, 10^-100];
k[t1_, t2_] := k[t1, t2] = Log[Sum[E^(data1[[j]][[1]]*t1+data1[[j]][[2]]*t2) / n, {j, 1, n}]];
k11[t1_, t2_] := k11[t1, t2] = D[k[x1, x2], {x1, 1}, {x2, 0}] /. {x1 → t1, x2 → t2};
k12[t1_, t2_] := k12[t1, t2] = D[k[x1, x2], {x1, 0}, {x2, 1}] /. {x1 → t1, x2 → t2};
k21[t1_, t2_] := k21[t1, t2] = D[k[x1, x2], {x1, 2}, {x2, 0}] /. {x1 → t1, x2 → t2};
k22[t1_, t2_] := k22[t1, t2] = D[k[x1, x2], {x1, 1}, {x2, 1}] /. {x1 → t1, x2 → t2};
k23[t1_, t2_] := k23[t1, t2] = D[k[x1, x2], {x1, 0}, {x2, 2}] /. {x1 → t1, x2 → t2};
ξ[x1_, x2_] :=
 ξ[x1, x2] = FindRoot[{k11[t1, t2] == x1 && k12[t1, t2] == x2}, {{t1, 0}, {t2, 0}}]
r1[t1_, t2_] := ⅇ^(k[ξ[t1,t2][[1]][[2]],ξ[t1,t2][[2]][[2]]]-{ξ[t1,t2][[1]][[2]],ξ[t1,t2][[2]][[2]]}.{t1,t2})
ListPlot3D[Flatten[Table[{x, y, r1[x, y]}, {x, 0, 4.5, 0.25}, {y, 0, 5.5, 0.25}], 1],
 PlotRange → All]
gs = Interpolation[Flatten[Table[{x, y, r1[x, y]},
    {x, 0, 4.5, 0.25}, {y, 0, 5.5, 0.25}], 1], InterpolationOrder → 2]
nc = NIntegrate[gs[x, y], {x, 0, 4.5}, {y, 0, 5.5}]
Plot3D[gs[x, y] / nc, {x, 0, 4.5}, {y, 0, 5.5}, PlotRange → All]
```

- **Applying a bivariate polynomial adjustment**

```
X = Transpose[rescaled][[1]];
Y = Transpose[rescaled][[2]];
jm[a_, b_] := jm[a, b] = Rationalize[Total[X^a Y^b]/n, 10^-100]
mm[i_, j_] :=
 mm[i, j] = Rationalize[NIntegrate[x^i y^j gs[x, y], {x, 0, 4.5}, {y, 0, 5.5}], 10^-100]
mm1[i_, j_] := mm1[i, j] = mm[i, j]/mm[0, 0]
Off[Inner::"normal"]
f3[L1_List, L2_List] := Inner[Plus, L1, L2, List];
L3[δ1_, δ2_] := L3[δ1, δ2] = Flatten[Table[{i, j}, {i, 0, δ1}, {j, 0, δ2}], 1];
P3[δ1_, δ2_] :=
 P3[δ1, δ2] = Table[f3[L3[δ1, δ2][[i]], L3[δ1, δ2][[j]]], {i, 1, Length[L3[δ1, δ2]]},
   {j, 1, Length[L3[δ1, δ2]]}]; MM[δ1_, δ2_] := MM[δ1, δ2] =
  Rationalize[Table[mm1[P3[δ1, δ2][[i, j]][[1]], P3[δ1, δ2][[i, j]][[2]]],
    {i, Length[P3[δ1, δ2]]}, {j, Length[P3[δ1, δ2]]}], 10^-100];
μvect[δ1_, δ2_] := μvect[δ1, δ2] = Table[jm[L3[δ1, δ2][[i, 1]], L3[δ1, δ2][[i, 2]]],
    {i, Dimensions[L3[δ1, δ2]][[1]]}];
ϵ[δ1_, δ2_] := ϵ[δ1, δ2] = Inverse[MM[δ1, δ2]].μvect[δ1, δ2];
polfun[x_, y_, δ1_, δ2_] :=
  polfun[x, y, δ1, δ2] = Flatten[Table[x^i y^j, {i, 0, δ1}, {j, 0, δ2}], 1];
fapp1[x_, y_, δ1_, δ2_] := fapp1[x, y, δ1, δ2] =
  gs[x, y]/nc ϵ[δ1, δ2].polfun[x, y, δ1, δ2]
Plot3D[Evaluate[fapp1[x, y, 6, 4]], {x, 0, 4.5}, {y, 0, 5.5}, PlotRange → All]
Plot3D[Det[MatrixPower[DiagonalMatrix[Diagonal[Σ]], -0.5]]
  Evaluate[fapp1[MatrixPower[DiagonalMatrix[Diagonal[Σ]], -0.5][[1]][[1]] * x,
    y * MatrixPower[DiagonalMatrix[Diagonal[Σ]], -0.5][[2]][[2]], 6, 4]],
 {x, 0, 1}, {y, 0, 1}, PlotRange → All]
```

- **Comparing the sums of squared differences**

```
edis = EmpiricalDistribution[rescaled];
SSD[i_, j_] :=
 SSD[i, j] = Sum[(NIntegrate[fapp1[x, y, i, j], {x, 0, rescaled[[k]][[1]]},
       {y, 0, rescaled[[k]][[2]]}] - CDF[edis, rescaled[[k]]])^2, {k, 1, n}]
SSD[0, 0]
SSD[6, 4]
```

Curriculum Vitae

Education

Ph.D. Candidate in Statistics 2014 - Present
The University of Western Ontario, London, ON

- Supervisors: Dr. Serge B. Provost and Dr. Jiandong Ren
- Research interests: Distribution theory, Density function approximation and estimation, Extreme value theory, Credibility theory, Econometrics
- Thesis title: ”Advances in the Modeling of Heavy-tailed Distributions”

Master of Science in Statistics 2012 - 2014
The University of Western Ontario, London, ON

- Supervisor: Dr. Bruce L. Jones
- Specialization: Actuarial Science
- Project: ”Applications and Issues of Hierarchical Credibility Theory in Group Health Benefit”

Master of Finance 2009 - 2010
Schulich School of Business, York University, Toronto, ON

- Member of Schulich Finance Association
- Participated in Toronto CFA Society Investment Research Challenge

Bachelor of Arts in Economics 2005 - 2007
The University of Western Ontario, London, ON

Bachelor of Business Administration 1998 - 2004
Chung-Ang University, Seoul, Republic of Korea

- Specialization: Finance

Teaching and working experience

Teaching and research assistant 2013 - 2018
The University of Western Ontario

- Facilitated tutorial sessions for several courses
- Managed and scheduled the Statistics Help Center and the proctoring of exams
- Conducted teaching evaluations

Cashier 2007 - 2009
Happy Days Mini Mart, London, ON

Private mathematics tutor 2003 - 2004
Seoul, Republic of Korea

Soldier and non-commission officer 2000 - 2002
552nd Military Police Company, Pusan, Korea
(One of U.S. army units, responsible for policing the armed force and serving combat support in the Korean peninsula)

Conferences

- The 53rd Actuarial Research Conference, Presenter 2018
  University of Western Ontario, ON, Canada
- 2017 IMS-FIPS Workshop, Presenter 2017
  University of Maryland, Baltimore County, MD, USA
- The 50th Actuarial Research Conference, Participant 2015
  University of Toronto, ON, Canada

Computer skills

Statistical Software: Mathematica, R, Excel with macro, SPSS, STATA
Operating Systems: Windows
Other software: MS Office, LaTeX, Bloomberg Terminal

Honours and awards

Faculty of Science Graduate Student Teaching Award 2018
The University of Western Ontario

Western Graduate Research Scholarship 2013 - 2018
The University of Western Ontario

Dean's Honour List 2004
Chung-Ang University

Chung-Su Scholarship 1999 - 2004
Chung-Ang University

Working papers

- Moment-Based Density Approximation Techniques as Applied to Heavy-tailed Distributions. To be submitted to the *International Journal of Statistics and Probability*.
- Novel Approaches for Estimating Distributional Endpoints.
- On Securing Bona Fide Density Functions.
- Adjusted Empirical Bivariate Saddlepoint Estimates.