
Table of Contents

<i>Abstract</i>	iii
<i>Acknowledgements</i>	v
<i>Table of Contents</i>	vi
<i>List of Tables</i>	xi
<i>List of Figures</i>	xiii
<i>List of Abbreviations</i>	xv
1. Introduction	1
1.1 Nonparametric Density Estimation	1
1.2 Nonparametric Density Estimation under Shape Restrictions	2
1.3 Motivation	3
1.4 Contributions	4
1.5 Outline of the Thesis	5
2. Literature Review	7
2.1 Shape-restricted Density Estimation	7
2.2 Nonparametric Estimation of a Unimodal Density Function	9
2.2.1 Kernel-based estimation	9
2.2.2 Spline-based estimation	12
2.3 Nonparametric Estimation of a Log-concave Density Function	15
2.3.1 Log-concave density	15

2.3.2	Nonparametric maximum likelihood estimation of a log-concave density function	17
2.3.3	Smooth nonparametric estimation of a log-concave density function	18
2.3.4	Algorithms for log-concave density estimation	18
2.4	Nonparametric Estimation of a Unimodal and Heavy-tailed Distribution	20
2.4.1	Heavy-tailed distributions	20
2.4.2	Parametric models for heavy-tailed distributions	21
2.4.3	Nonparametric models for heavy-tailed distributions	23
2.5	Nonparametric Mixtures	25
3.	<i>A Fast Algorithm for Log-concave Density Estimation</i>	29
3.1	Introduction	29
3.2	Characterization of the Nonparametric Maximum Likelihood Estimate	32
3.2.1	Parametrizations and log-likelihood function	33
3.2.2	Characterization	34
3.2.3	Theoretical properties	34
3.3	Computation	38
3.3.1	Main idea	38
3.3.2	Updating masses	39
3.3.3	Algorithm	41
3.4	Convergence	42
3.5	Numerical Studies	45
3.5.1	Setup	45
3.5.2	Simulation studies	46
3.5.3	Real-world data	46
3.6	Summary	51
4.	<i>Smooth Log-Concave Density Estimation</i>	55
4.1	Introduction	55
4.1.1	Smooth estimation of a unimodal density function	56
4.1.2	Smooth estimation of a log-concave density function	57
4.1.3	Overview	58
4.2	Smoothness Assumption	60
4.2.1	φ'' decreasing	60
4.2.2	φ'' increasing	62

4.2.3	φ'' increasing and then decreasing	63
4.2.4	φ'' decreasing and then increasing	64
4.2.5	Log-concavity properties of commonly-used distributions	65
4.2.6	Theoretical properties	67
4.3	Computation	67
4.3.1	Derivatives of the log-likelihood function	68
4.3.2	Updating coefficients	68
4.3.3	Expanding and reducing knot sets	70
4.3.4	Algorithm	71
4.4	Assessing Log-concavity	72
4.4.1	Bootstrapping	72
4.4.2	Test statistics	73
4.5	Simulation Studies	73
4.5.1	Between piecewise quadratic estimators	73
4.5.2	Against other estimators	75
4.6	Real-world Data	79
4.6.1	Reliability data	84
4.6.2	Log-return data	87
4.6.3	Timings	91
4.7	Summary	93
5.	<i>An Application of Log-concave Density Estimation: ROC Curve Estimation</i>	97
5.1	Introduction	97
5.2	ROC Curve Estimation Based on Log-concave Density Estimates	99
5.3	Simulation Studies	100
5.4	An Example	101
5.5	Summary	111
6.	<i>Nonparametric Estimation for Heavy-tailed Distributions under Shape Restrictions</i>	115
6.1	Empirical Motivation	115
6.2	Introduction	116
6.2.1	Nonparametric estimation of heavy-tailed distributions	117
6.2.2	Nonparametric estimation of unimodal distributions	118
6.2.3	Overview	119

6.3	Maximum Likelihood Estimation for Unimodal Heavy-tailed Distributions	119
6.3.1	The log-convex-concave-convex density estimator	119
6.3.2	Maximum likelihood estimation and its characterization	122
6.3.3	Relations to parametric distributions	123
6.4	Computation	124
6.4.1	Derivatives of the log-likelihood function	124
6.4.2	Updating coefficients	125
6.4.3	Expanding and reducing knot sets	126
6.4.4	Algorithm	127
6.5	Bootstrap Test	127
6.6	Simulation Studies	129
6.7	Financial Data	132
6.7.1	Setup	135
6.7.2	Testing for financial data	138
6.7.3	Density estimation	139
6.8	Summary	147
7.	<i>Heavy Tails and Value at Risk Estimation</i>	151
7.1	Introduction	151
7.2	Estimators for Comparison	152
7.3	Heavy Tails Analysis	153
7.4	VaR estimation	155
7.5	Summary	159
8.	<i>Summary and Future Works</i>	161
8.1	Summary	161
8.2	Future Works	163

List of Tables

3.1	Settings for the algorithms.	45
3.2	Performance of algorithms in simulation studies	47
3.3	Results of the real-world data sets.	54
4.1	Two distributions of Type 1.	74
4.2	Simulation results for two distributions in terms of MISE and MHD.	75
4.3	Four distributions of different types.	77
4.4	Simulation results for four scenarios in terms of MISE and MHD.	79
4.5	Descriptive statistics of reliability data.	84
4.6	p -values from bootstrap test for reliability data.	84
4.7	Results based on the three estimators for reliability data.	87
4.8	Descriptive statistics of log returns.	88
4.9	p -values from the bootstrap test for log-return data.	88
4.10	Results based on the three estimators for six-month log-return data.	93
4.11	Results based on the three estimators for one-year log-return data.	94
4.12	Running times (s) of CNMLCDS for real-world data.	94
5.1	Two scenarios for ROC curve estimation.	101
5.2	Results based on ASE values in simulated studies.	103
5.3	Descriptive statistics of pancreases data.	103
5.4	p -values from bootstrap test for pancreases data.	104
5.5	Results based on whole pancreases data in terms of ASE and KS.	111
5.6	Results for ASE values for the pancreases data.	112
5.7	Results for MR (%) values for the pancreases data.	112

6.1	Intervals with corresponding gradient functions.	126
6.2	Two heavy-tailed distributions we simulated from.	130
6.3	Results for two heavy-tailed distributions, in terms of MISE and MHD.	131
6.4	Financial data (closing price).	134
6.5	Optimal number of mixture components for a UMBP estimate.	137
6.6	Descriptive statistics of the daily log-returns of financial datasets.	138
6.7	Test for the daily log-returns of financial datasets.	139
6.8	Model fitting for financial data sets.	143
6.9	Cross-validation results for MSFT data.	147
6.10	Cross-validation results for SPX data.	148
6.11	Cross-validation results for NZD/USD data.	149
7.1	Tail performance results in terms of MWISE.	155
7.2	VaR estimation with Kupiec's test results.	159

List of Figures

3.1	CPU times for Gamma(2, 1) with different sample sizes.	48
3.2	Histograms and log-concave density estimates for simulation data.	49
3.3	Density plots with gradient curve for reliability data.	50
3.4	Density plots with gradient curve for daily log-return of S&P 500.	52
3.5	Density plots with gradient curve for daily log volatilities.	53
4.1	Illustration of different types of log-concave functions.	59
4.2	Finding knots for the middle part.	71
4.3	Illustration plots for a truncated normal and a skew normal.	74
4.4	Boxplots for the ISE and HD for a truncated normal and a skew normal.	76
4.5	Illustration plots for exponential, gamma, beta and logistic distributions.	78
4.6	Boxplots for the ISE for exponential, gamma, beta and logistic distributions.	80
4.7	Boxplots for the HD for exponential, gamma, beta and logistic distributions.	81
4.8	Histograms and log-concave density estimates for exponential and gamma distributions.	82
4.9	Histograms and log-concave density estimates for beta and logistic distributions.	83
4.10	Log-density with gradient curve for reliability data computed by CNMLCDS.	85
4.11	Q-Q plots for reliability data form different estimators.	86
4.12	Histograms and log-concave density estimates for reliability data.	86
4.13	Log-density with gradient curve for log returns of S&P 500 index.	90
4.14	Q-Q plots for daily log-returns of S&P 500 index form different estimators.	91
4.15	Histograms and log-concave density estimates for daily log returns of S&P 500 index.	92
5.1	Density plots and ROC curves from simulated data.	102

5.2	Log-density with gradient curve for log-CA 19-9 measurements.	105
5.3	Log-density with gradient curve for log-CA 125 measurements.	106
5.4	Q-Q plots for the log-CA 19-9 measurements.	107
5.5	Q-Q plots for the log-CA 125 measurements.	108
5.6	Density plots and induced ROC curves from pancreases data.	110
6.1	The histogram of SPX daily log-returns.	116
6.2	Log-density function and the second derivative of a student- t distribution.	120
6.3	Illustration plots for two heavy-tailed distributions.	131
6.4	Boxplots for the ISE and HD for simulated data sets with sample size 100.	133
6.5	Boxplots for the ISE and HD for simulated data sets with sample size 1000.	134
6.6	Histograms and varying density estimates for typical log-normal and Student's t	135
6.7	Financial data overview.	136
6.8	Gradient curves computed by CNMLCCC for financial datasets.	140
6.9	Histograms and different estimates for financial datasets.	142
6.10	Q-Q plots from different estimates for MSFT data.	144
6.11	Q-Q plots from different estimates for SPX data.	145
6.12	Q-Q plots from different estimates for NZD/USD data.	146
7.1	Box plots of the WISE from LN $(0, 1)$	156
7.2	Box plots of the WISE from WB $(0.5, 1)$	157
7.3	Tail plots among different estimates from different scenarios.	158
7.4	VaR estimation for financial datasets.	160

List of Abbreviations

AD	Anderson-Darling
ASA	active set algorithm
ASE	average square error
CDF	cumulative distribution function
CN	conditional number
CNM	constrained Newton method
CNMLCCC	constrained Newton method for log-convex-concave-convex density
CNMLCD	constrained Newton method for log-concave density
CNMLCDS	constrained Newton method for smooth log-concave density
HD	Hellinger distance
ICMA	iterative convex minorant algorithm
ISE	integrated square error
KER	kernel
KDE	kernel density estimate
KKT	Karush Kuhn Tucher
KMCE	kernel density estimator based on modified champernowne distribution
KL	Kullback-Leibler
KS	Kolmogorov-Smirnov
LC	log concave
LCDS	smooth log-concave density estimate
LCCC	log-convex-concave-convex
MD	mean difference
MHD	mean Hellinger distance
MISE	mean integrated squared error
MLE	maximum likelihood estimator
MR	misclassification rate
NPMLE	npnparametric maximum likelihood estimate
PAVA	pool-adjacent-violators algorithm
PL	piecewise linear
POF	proportion of failure
PQ	piecewise quadratic

PQC	combined piecewise quadratic
ROC	operating characteristic curve
SN	skew normal
SNPMLE	smooth npnparametric maximum likelihood estimate
ST	skewed student-t
TN	truncated normal
UMBP	unimodal bernstein polynomials
UBRS	unimodal regression spline
VaR	Value-at-Risk
WISE	weighted integrated square error

Chapter 1

Introduction

1.1 Nonparametric Density Estimation

Density estimation is a fundamental problem in statistics and has attracted a lot of research interest in the last few decades. It is about finding an estimate of some underlying probability density function using observed data, based on much or little prior knowledge of the distribution. It has been applied in many fields, including finance, economics, climatology, genetics, hydrology and physiology; see [Sheather et al. \(2004\)](#) and the references therein. Density estimation can be either parametric, in the sense that the data is assumed to be drawn from one of a known family of distributions defined by using a finite number of parameters, or nonparametric where much less rigid assumptions are made about the distribution. Parametric approaches have the advantages of easy interpretation and fast fitting, but they also have the drawback of potential misspecification. In this thesis, we shall focus on nonparametric methods because of their generality and letting the data speak for themselves for determining the density estimate. Nonparametric density estimation can further serve as a building block for solving other statistical problems, such as regression, classification and clustering. It is truly a highly important statistical research topic.

The histogram is perhaps the earliest nonparametric estimator of a univariate density function. Alternatives to the histogram include the naive estimator, kernel estimator, adaptive kernel estimator and maximum penalized likelihood estimators. More detailed descriptions of these estimators were given by [Silverman \(1986\)](#); see also [Wand and Jones \(1994\)](#) and [Simonoff \(2012\)](#). Typically, these nonparametric estimators require some smoothing parameters, the choice of which can be hard in practice. There are also, of course, much literature that concerns the issue of how to choose

the degree of smoothness of the estimate (Terrell, 1990).

In many applications, one problem of many nonparametric methods is that they may fail to satisfy some practically known properties. For example, in state price density estimation, the call price should be a monotonically decreasing convex function of the strike price and the state price density must be nonnegative and integrate to one. Some nonparametric models introduce a tuning parameter to incorporate some of these characteristics and other implications of economic theory. However, these models require many observations on the strike prices in one day, but this can be an problem in the interest rate option market. It is also difficult to properly determine the value of a tuning parameter. Many practical functions are known to exhibit certain shapes, even though they have no known parametric forms, for examples, the utility, cost and profit functions in economics, dose-response relationships in medicine, the hazard rate and the mean residual life function in reliability and survival analysis. It is natural to make a good use of such prior knowledge about the underlying function that has been established or widely observed in a physical field. When such shape-related information is available, a good approach would be to estimate the curve under the corresponding restrictions.

1.2 Nonparametric Density Estimation under Shape Restrictions

Shape-restricted density estimation is useful for producing estimates having some desired qualitative features. The qualitative features that might be of interest include monotonicity, unimodality, or convexity. Estimation of a function under such shape restrictions offers many advantages, such as meeting the physical requirements in a context, giving an easy interpretation and may having higher estimation efficiency. The maximum likelihood and least squares are two popular criteria for nonparametric estimation of a function under shape restrictions. A class of weighted bootstrap techniques was applied to density estimation under constraints, see Hall and Presnell (1999). The nonparametric maximum likelihood estimator for a nonincreasing density was studied by Grenander (1956). Other researchers have since studied other shape-restricted estimation problems. We refer the reader to Bartoszyński et al. (1981) for the nonparametric maximum likelihood estimator (NPMLE) of a decreasing failure rate function, to Rufibach (2007), Dümbgen and Rufibach (2009) and Anderson-Bergman (2014) for log-concave density estimation, and to Groeneboom et al. (2001) for convex density estimation. Many methods have also been developed for such problems, but using the least squares approach; see Hildreth (1954) for concave function estimation, Barlow (1972) and Robertson et al. (1988) for monotone regression function estimation, Mammen (1991), Groeneboom et al. (2001) and Birke and Dette (2007) for convex or concave

function estimation, [Ait-Sahalia and Duarte \(2003\)](#) and [Yatchew and Härdle \(2006\)](#) for state price density estimation, and [Meyer \(2012\)](#) and [Turnbull and Ghosh \(2014\)](#) for smooth unimodal density estimation.

1.3 Motivation

Our main motivation for investigating nonparametric density estimation under shape restrictions comes from fitting an asset return distribution in finance. Practitioners often assume for convenience that the distribution of random variables under investigation is normal. Since the 1960s, however, empirical evidence has led many to reject this Gaussian assumption in favour of various heavy-tailed alternatives. The features of high peak and heavy-tailedness in financial asset returns are commonly accepted now. Unimodality of the underlying density function is also cited as a reasonable assumption in finance and econometrics. Such considerations motivate us to study shape-restricted density estimation, in particular its applications to financial problems.

Unimodality is viewed as a reasonable assumption in many problems. The construction of the [Grenander \(1956\)](#) estimator and its properties can be straightforwardly extended to the case of a unimodal density with a known mode. More precisely, it was to combine a nondecreasing Grenander's estimate to the left of the mode with a nonincreasing one to the right. However, [Woodroffe and Sun \(1993\)](#) pointed out that the estimator is inconsistent at the known mode. If the mode is unknown, the situation is getting harder. The likelihood function in this case is not bounded and hence the nonparametric maximum likelihood estimate (NPMLE) does not exist ([Birgé, 1997](#)). For starters, we turn to investigate the class of log-concave densities which is the subset of the class of unimodal densities. The allure of the log-concavity assumption is that the NPMLE of a log-concave density exists, is consistent and can be computed without any tuning parameter ([Walther, 2002, 2009](#); [Pal et al., 2007](#)). We shall propose a new algorithm for finding the NPMLE of a log-concave density and compared with other existing algorithms.

The NPMLE of a log-concave density function is piecewise linear and thus is not smooth. Discontinuities in the first order derivatives may be deemed unsatisfactory if a smoothness assumption is desired. The lack of smoothness is often a disadvantage in some applications. Taking classification as an example, the zero mass outside the convex hull of the data makes classification of a new observation undefined, even when it is apparently closer to the observations of one class. In such a situation, a smooth density estimate would be preferable. A smooth estimate has a more attractive visual appearance and usually offers significantly improved estimation accuracy, especially for small-sized samples ([Chen and Samworth, 2013](#)). Hence we will further investigate the smooth

nonparametric maximum likelihood estimator of a log-concave density function.

Log-concave density estimation can be readily applied to some financial data that is not highly heavy-tailed. It provides pretty good estimation even for small-sized samples. However, log-concavity is not a good assumption for the high-peaked and heavy-tailed data; see also [Meyer \(2012\)](#). We would like to further estimate highly heavy-tailed distributions, while preserving unimodality. The most difficult for modelling the whole unimodal and heavy-tailed distribution is how to accurately capture the tails without losing the important information in the main body of the density. We impose different shape restrictions on the tails and the main body of the density function, in particular the log-convexity restriction on the tails and the log-concavity restriction on the main body.

Specifically, we focus on the nonparametric density estimation under shape restrictions from the maximum likelihood point of view and are particularly interested in smooth estimators. While the main goal of this thesis is to propose new techniques that address the problems of nonparametric density estimation subject to shape constraints, we also need to find general, efficient algorithms to compute these estimates since these problems have no closed-form solutions.

1.4 Contributions

The crucial target of this thesis is to propose a novel approach to nonparametric density estimation under shape restrictions and develop new fast algorithms to solve these problems. The proposed framework is applicable to the case of log-concavity. Moreover, this framework extends in a nice manner for the heavy tailed, and possibly skewed distributions in order to estimate the return of an asset in financial fields. Monotone increasing and decreasing shape restrictions can also be achieved by adjusting our proposed approaches. The main contributions of this thesis are given as follows:

- Develop a new algorithm for computing the NPMLE of a log-concave density function.
- Propose a new family of smooth log-concave density functions.
- Propose a new estimator for unimodal heavy-tailed distributions.
- Compare different algorithms and estimators by using simulated and real-world data.
- Study some practical problems: the receiver operating characteristic curve estimation and value-at-risk estimation.

- Develop new tests for log-concavity and for detecting the functional form implied in the new estimator of unimodal heavy-tailed distributions.
- Implement the new algorithms in R and make them publicly available in R packages.

1.5 Outline of the Thesis

The thesis is organized as follows. In Chapter 2, we first describe some fundamental issues in nonparametric density estimation under shape restrictions. Then a literature review is given on the existing methods to estimating unimodal, log-concave and heavy-tailed density functions. In the end, nonparametric mixtures and their maximum likelihood estimation are reviewed, owing to their relations to the methods proposed in this thesis.

Chapter 3 first describes the nonparametric maximum likelihood estimator of a log-concave density function in a great detail. The computational difficulty with real observations and the new idea for solving this problem are also discussed. A new algorithm is presented for computing the NPMLE of a log-concave density. The properties of the NPMLE has been studied and the convergence of this new algorithm is also theoretically established. It is numerically compared with some other existing algorithms in terms of computing time.

As an extension of Chapter 3, Chapter 4 is concerned with the nonparametric estimation of a density function under both log-concavity and smoothness assumptions. We propose several new smooth nonparametric maximum likelihood estimators of a log-concave density function. New algorithms are developed for the proposed estimators. Comparisons among our new smooth methods with other smooth and non-smooth approaches are made through both simulated and real-world datasets. Tests for log-concavity are also studied based on the proposed algorithm for computing the NPMLE of a log-concave density function in Chapter 3.

Chapter 5 is an application of the methodology proposed in Chapter 4 to the receiver operating characteristic curve estimation. We compared different estimators via simulated and real-world data.

Chapter 6 focuses on the problem of nonparametric maximum likelihood estimation of a unimodal and heavy-tailed density function. By imposing different constraints on the main body (log-concavity) and tail domain (log-convexity), a very general hybrid is proposed. The algorithm for this estimator has also been presented. We report the results of numerical studies of simulation and real-world data sets that compare the performance of our shape-restricted estimator to other estimators in the literature. We also develop a new test for detecting the functional form implied by this new estimator.

Tail fitting is further studied in details in Chapter 7. We also apply the method proposed in Chapter 6 to the value at risk estimation. Numerical studies have also been conducted to compare different estimators.

Chapter 8 gives a summary that concludes the thesis. Some interesting aspects for future works related to the dissertation are discussed.

Chapter 2

Literature Review

2.1 Shape-restricted Density Estimation

Density estimation under shape restrictions has been studied and gradually improved upon in the last few decades. The allure is the prospect of obtaining fully automatic nonparametric estimators, without using any tuning parameter, the value of which is often difficult to choose. Furthermore, shape-restricted density estimators allow the modelling assumptions to be tailored to more closely match reality and ensure that estimates have the desired shape characteristics for all samples, not just on average or asymptotically (Wolters, 2012).

Shape-restricted density estimation essentially assumes that $f_0 \in \mathcal{F}$, where f_0 is the true density and \mathcal{F} contains all functions that follow certain shape constraints. Choosing a suitable family \mathcal{F} is important. It should be both “large” enough to include the true f_0 , or at the very least, f_0 can be well approximated by elements from \mathcal{F} , and “small” enough so that the LSE or the MLE is well-defined and consistent with respect to certain norms (Chen, 2013). The shape restrictions imposed on \mathcal{F} should correspond to real-world problems and their use in practice should be easily justifiable.

Once choosing a family \mathcal{F} , one can use either the least squares estimator (LSE) or the maximum likelihood estimator (MLE) to estimate f_0 . For a special family \mathcal{F} , the LSE is defined as

$$\hat{f} = \operatorname{argmin}_{f \in \mathcal{F}} \left\{ \int f^2(x) dx - 2 \int f(x) dF_n(x) \right\},$$

where F_n denotes the empirical distribution obtained from the sample. The MLE, on the other

hand, is defined as

$$\hat{f} = \operatorname{argmax}_{f \in \mathcal{F}} \int \log f(x) dF_n(x).$$

The qualitative constraints that have received the most attention in the literature are monotonicity, unimodality and convexity.

We note that throughout the thesis, for simplicity of notation, we typically use x_1, x_2, \dots, x_n for the ordered copy of a random sample, i.e., the order statistics, owing to the nature of shape-restricted estimation. For methods under no shape restrictions, a random sample does not have to be ordered and the formulae of these methods are often irrelevant to the ordering of a sample. Distinctions will be made, if any ambiguity arises.

Monotonicity

Monotonicity is the most basic shape constraint for a real-valued function on \mathbb{R} . In many situations, this constraint can be imposed on the data in a straightforward way. [Grenander \(1956\)](#) studied the nonparametric maximum likelihood estimation for a non-increasing density function on $[0, \infty)$ and showed that the estimator must be a step function with jumps only at some observations. The pool-adjacent-violators algorithm (PAVA) ([Ayer et al., 1955](#); [Robertson et al., 1988](#)) can be applied to compute the estimate. More details about PAVA will be described later. However, this estimator is inconsistent at the mode(zero here); see [Balabdaoui et al. \(2011\)](#).

Unimodality

Unimodality is cited as a reasonable assumption in many problems. Note that the class of unimodal densities includes monotone densities as a special case. The [Grenander \(1956\)](#) estimator can be straightforwardly extended to the case of a unimodal density with a known mode, but it does not directly adapt to the case with an unknown mode ([Rao, 1969](#); [Woodroffe and Sun, 1993](#)). [Woodroffe and Sun \(1993\)](#) proposed a consistent estimator by introducing a penalty term for the value at the mode based on maximum likelihood. [Bickel and Fan \(1996\)](#), who also used the PAVA, discussed some problems in unimodal density estimation and plugged in a consistent point estimate of the mode location. The linear spline is used by [Meyer and Woodroffe \(2004\)](#), who developed a consistent decreasing density estimator which is forced to be concave on an interval containing the mode. [Meyer \(2012\)](#) proposed a quadratic spline estimator for a decreasing density function. A unimodal density estimator is obtained by piecing together two isotonic density estimators at a known mode. A smooth log-concave density estimator was proposed by [Dümbgen and Rufibach](#)

(2009). Anderson-Bergman (2014) introduced a new, more flexible shape constraint, “inverse convex”, for survival analysis and other types of heavy-tailed data.

Convexity

Groeneboom et al. (2001) considered a piecewise linear estimator in the decreasing and convex density estimation, but this estimator has a tendency to spike at the mode. A support reduction algorithm was proposed by Groeneboom et al. (2008) to compute this estimator. A cubic spline estimator for decreasing and convex density function was developed by Meyer (2012).

Of course, there are also many other shape restrictions; see Bartoszynski et al. (1981) based on MLE for estimating the intensity function related nonstationary poisson process, Ait-Sahalia and Duarte (2003), Yatchew and Härdle (2006), Birke and Dette (2007) and Horowitz and Lee (2015) for convexity based on LSE.

2.2 Nonparametric Estimation of a Unimodal Density Function

A density f on the real line is said to be unimodal if there exists a point M such that f is nondecreasing on $(-\infty, M)$ and nonincreasing on (M, ∞) . Then M is known as the mode of the density. When the true density is unimodal, there are two good reasons to enforce unimodality. First, making use of the shape information would help improve estimation accuracy (Wolters, 2012). Second, incorporating the constraint will eliminate spurious modes that may reduce the effectiveness of the density estimate as an exploratory tool and communication aid (Wolters, 2012). Maximum likelihood estimation of a unimodal density with a known mode can be accomplished by using two decreasing estimators on either side of the mode. However, if the mode is unknown and has to be estimated as well, the maximum likelihood estimator does not exist because the likelihood is unbounded if the mode is allowed to vary (Birgé, 1997). Several smoothed unimodal estimators have been proposed using kernel ideas and spline methods.

2.2.1 Kernel-based estimation

The kernel method is the most popular and conceptually simplest nonparametric approach. It is of wide applicability, particularly in the univariate case, and is probably the method whose properties are best understood (Parzen, 1962; Fryer, 1977; Silverman, 1986). It is defined as the weighted average of kernel functions centred at the observed values. Given $x_i, i = 1, \dots, n$, the kernel

density estimator (KDE) is

$$\hat{f}(x; h) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right),$$

where $K(\cdot)$ is the kernel function satisfying $\int K(x) dx = 1$ and h is a positive number, known as the bandwidth. The behaviour of a KDE relies strongly on the choice of the value of the smoothing parameter h .

Many research efforts have been made in smooth estimation of a unimodal density function using kernels. [Silverman \(1981\)](#) proposed a bandwidth test for unimodality based on nonparametric density estimation. However, this test can not form the basis for a unimodal density estimator. In the case where the mode is known, [Fougères \(1997\)](#) proposed a unimodal estimator based on a unimodal rearrangement of the kernel estimator. [Cheng et al. \(1999\)](#) treated a general unimodal density as a transformation of some known unimodal template and then introduced a recursive method for estimating the transformation. A smoothing estimator was constructed during the algorithm by using the kernel technique. A kernel estimator was also considered as the derivative of the least concave majorant of the distribution by [Eggermont and LaRiccia \(2000\)](#). [Hall and Huang \(2002\)](#) proposed a method for rendering unimodal a standard kernel density estimator by minimizing the integrated squared distance between a conventional density estimator and its reweighted version. This estimator needs to remove spurious wiggles in the tails of the conventional density estimator which can result in a detrimental increase in the density estimator at other places, leading to poor mean squared error performance. It also commonly suffers from the difficulty for heavy-tailed distributions. The kernel method for estimating monotone, convex and log-concave densities was proposed by [Birke \(2009\)](#). [Dümbgen and Rufibach \(2009\)](#) proposed a smooth log-concave density estimate by convolving their nonparametric maximum likelihood density estimate with a Gaussian density, which preserves the log-concavity shape constraint. This estimate was further studied by [Chen and Samworth \(2013\)](#) who developed a new test of log-concavity and by [Rufibach \(2012\)](#) who developed a new smooth estimator of the ROC curve based on the log-concavity assumption of the constituent distributions.

It would be preferable if unimodality could be achieved by adding a conceptually simple modification to a standard nonparametric estimator. Data sharpening, as advanced by [Braun and Hall \(2001\)](#), is one approach that operates in this way and can improve upon the performance of numerous estimators. Data sharpening refers to methods for preprocessing data. Since the introduction of data sharpening methods by [Choi and Hall \(1999\)](#) and [Choi et al. \(2000\)](#), they become an attractive approach to achieve unimodality by perturbing the data and improve the performance of the stan-

standard kernel density estimator by adding a simple modification. Data sharpening involves altering the positions of data values, controlled by minimizing a measure of the total distance that data are moved subject to unimodality on the estimator. Given $\mathbf{x} = (x_1 \cdots x_n)^T$, and letting $\mathbf{y} = (y_1 \cdots y_n)^T$ be the new sharpened data vector, the original problem of unimodal density estimation can be set up as the sharpened KDE problem

$$\hat{f}_{\mathbf{y}}(x; h) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - y_i}{h}\right),$$

which is algebraically the same as the standard kernel density estimator, only a subscript \mathbf{y} added to \hat{f} , indicating which data vector is used to produce the estimate. The usual KDE is $\hat{f}_{\mathbf{x}}$, that is, when $\mathbf{y} = \mathbf{x}$. The best sharpened data vector \mathbf{y} can be defined as a solution to a constrained minimization problem:

$$\mathbf{y} = \underset{\mathbf{z}}{\operatorname{argmin}} \sum_{i=1}^n D(x_i, z_i),$$

subject to the unimodality that (with h fixed) there exists m such that $\hat{f}'_{\mathbf{y}}(x) \geq 0$ when $x \leq m$ and $\hat{f}'_{\mathbf{y}}(x) \leq 0$ when $x \geq m$. D is a nonnegative and symmetric distance function, e.g., the Euclidean distance.

For a data-sharpening method, it requires the choice of a distance function D . A natural choice, used by [Braun and Hall \(2001\)](#) and [Hall and Kang \(2005\)](#), is a norm of the difference $\mathbf{y} - \mathbf{x}$, defined as

$$L_{\alpha}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n |x_i - y_i|^{\alpha}, \quad 1 \leq \alpha \leq 2.$$

[Braun and Hall \(2001\)](#) successfully applied data sharpening to obtain unimodal estimates, without providing any theoretical support or clear guidance to the choice of α . [Hall and Kang \(2005\)](#) provided both theoretical and numerical properties of the data sharpening method based on L_1 distance function. They produced a smooth unimodal estimator with very good mean squared error performance. [Wolters \(2009\)](#) proposed a greedy algorithm for unimodal kernel density estimation following [Braun and Hall \(2001\)](#) in applying data sharpening to a KDE.

A common feature of kernel-based methods is that they introduce certain tuning parameters, such as the order of the kernel or the bandwidth. Researchers have made a great effort for choosing properly these parameters, which is far from trivial; see [Fryer \(1977\)](#), [Cao et al. \(1994\)](#) and [Jones et al. \(1996\)](#). An inappropriate bandwidth may breed the danger of under- or over-smoothing. It also usually involves minimizing a measure of global effectiveness of a curve estimate, such as the

Integrated Square Error (ISE), the Mean ISE (MISE) or other performance measures. However, these criteria are not good for capturing the unimodality of a density function. With the shape of unimodality known, spline estimators based on maximum likelihood are an alternative approach to the estimation of a unimodal density.

2.2.2 Spline-based estimation

A spline is a piecewise polynomial constructed in such a way that it is also continuous or even smooth at the points, called knots, at which two polynomials are pieced together. Splines can be used to approximate virtually any smooth function, at least if a sufficient large number of knots are used. It extends the advantages of polynomials to include greater flexibility, local effects of parameter value changes and the possibility of imposing shape constraints on the estimate.

Bickel and Fan (1996) proposed several methods for estimating a unimodal density based on the maximum likelihood method. They applied the pregrouping technique to the maximum likelihood method to reduce peaking problems and save computational cost. A plug-in MLE was firstly introduced, but it is discontinuous. Then they introduced a smoother estimate by finding the MLE satisfying the monotonicity restrictions among linear splines. However, this linear spline MLE gives zero mass outside the range of the observed values and does not produce a qualitatively different curve from the plug-in MLE itself. A smoothed curve was further obtained by solving an isotone cubic spline regression problem. Let \hat{f} be the plug-in MLE estimate of a unimodal density of Bickel and Fan (1996). Denote by z_1, \dots, z_N the midpoints of the \hat{f} histogram bins, and by y_1, \dots, y_N the corresponding heights. As shown in Bickel and Fan (1996), the number of bins should be $25 \sim 50$ depending on the number of data points. Setting $z_2, z_6, \dots, z_{4k+2}$ ($k = \lfloor (N-2)/4 \rfloor$) as initial knots θ , the $m = k + 4$ basis functions are

$$\begin{cases} \delta_1(x) = 1, & \delta_2(x) = x, & \delta_3(x) = x^2, & \delta_4(x) = x^3, \\ \delta_j(x) = (x - \theta_{j-4})_+^3 & \text{for } j = 5, \dots, m. \end{cases}$$

Then, letting $\log(f(x)) = \sum_{j=1}^m s_j \delta_j(x)$, use least squares to estimate the parameter $\mathbf{s} = (s_1, \dots, s_m)^T$ that minimizes

$$\sum_{i=1}^N \left[\log(y_i) - \sum_{j=1}^m s_j \delta_j(z_i) \right]^2 \omega_i,$$

where ω_i is the area of the histogram estimate for the i th bin. In each iteration, first run the regression step, and then delete the knot having the smallest absolute t -value (if $|t| < 3$, where t means t statistic). Repeat this process until no $|t|$ is smaller than 3. After the normalization process,

the smooth density estimate was finally obtained. Unfortunately, the resultant smooth curve is not necessarily unimodal.

Meyer and Woodroffe (2004) introduced a consistent version using a decreasing linear spline estimator with concave interval at the mode and determined its rate of convergence in the Hellinger metric. If the concavity assumption turns out to be valid over a given interval, then this estimator needs a user-defined penalty or smoothing parameter. Otherwise, the concavity interval can be used as a penalty device and is allowed to go to zero as the sample size increases. Meyer (2008) proposed an algorithm for the cubic monotone case and extended it to a convexity constraint, as well as some variants such as increasing concavity. Later, Meyer (2012) also obtained a smooth unimodal estimator by introducing quadratic splines with knots spaced in the support. It has been pointed out that the mode can be given or estimated using polynomial kernel density estimation in a sufficiently fast way; see Eddy (1980). With a known mode, and without loss of generality, assuming the mode is 0, it places k_1 interior knots to the left of the mode, and k_2 interior knots to the right of the mode. There are $m = k_1 + k_2 + 3$ knots in all by adding the mode itself and the exterior knots encompassing the domain of the function (d_l, d_u) . It starts with the basis functions for the decreasing part (the right of the mode). Letting $\delta_1, \dots, \delta_{k_2+2}$ be the right-hand basis functions defined on the knots $0 = \theta_0 < \theta_1, \dots, \theta_{k_2} < \theta_{k_2+1} = d_u$ and $\delta_{k_2+3}, \dots, \delta_{k_1+k_2+4}$ be the increasing basis functions defined on the k_1 interior knots. Let $m = k_1 + k_2 + 4$, a smooth unimodal density estimator can be written as

$$\hat{f}(x) = \sum_{j=1}^m s_j \delta_j,$$

subject to $s_j \geq 0$ for $j = 1, \dots, m$ and $\sum_{j=1}^{k_2+2} s_j = \sum_{j=k_2+3}^{k_1+k_2+4} s_j$. The parameters can be obtained by minimizing the integrated squared error

$$\int_{-\infty}^{+\infty} (\hat{f}(x) - f_n(x))^2 dx,$$

where f_n is the empirical density function. Data-driven compromises are introduced to choose the number and positions of knots in the case of density estimation with a known support. When support is unknown, exterior knots should better be chosen to span the data, and then apply data-driven compromises to the interior knots.

Turnbull and Ghosh (2014) proposed a unimodal density estimator using Bernstein polynomials. Consider a Bernstein polynomial of order $m - 1$ to estimate f which is a unimodal, continuous

density function on support $[a, b]$,

$$f_m(x, \boldsymbol{\omega}) = \frac{1}{b-a} \sum_{k=1}^m \omega_k f_b \left(\frac{x-a}{x-b}; k, m-k+1 \right),$$

where $\boldsymbol{\omega} = (\omega_1, \dots, \omega_m)^T \geq \mathbf{0}$ is a vector of weights, $\sum_{k=1}^m \omega_k = 1$ and $\exists k^* \in \{1, 2, \dots, m\}$ such that the ω_k 's are non-decreasing for $k \leq k^*$ and non-increasing for $k \geq k^*$; $f_b(\cdot)$ is the Beta density function with shape parameters k and $m-k+1$. When no information is known about the support of f , let $a = x_{(1)} - s/\sqrt{n}$ and $b = x_{(n)} + s/\sqrt{n}$, where $x_{(1)}$ and $x_{(n)}$ are the first and last order statistics of the data, and s is the sample standard deviation. These bounds are motivated by the fact that $Pr[X_{(1)} - \frac{s}{\sqrt{n}} \leq X_{n+1} \leq X_{(n)} + \frac{s}{\sqrt{n}}] \geq Pr[X_{(1)} \leq X_{n+1} \leq X_{(n)}] = \frac{n-1}{n+1}$. A novel procedure for selecting the number of weights (m) is provided and the mixing weights $\boldsymbol{\omega}$ are chosen through quadratic programming techniques subject to linear inequality constraints.

Sometimes, the measure of inaccuracy, such as the integrated squared error, does not reliably reflect qualitative fidelity (see, e.g., [Kooperberg and Stone \(1991\)](#)). In order to make a good use of the available shape information, one can apply the logspline model to smooth unimodal estimation. In the logspline method of density estimation, the logarithm of the unknown density function is approximated by a polynomial spline, the unknown coefficients of which are estimated by maximum likelihood. The logspline density estimation for the univariate data set can be found in [Kooperberg and Stone \(1992\)](#), [Stone et al. \(1997\)](#), [Koo et al. \(1999\)](#) and [Koo and Kooperberg \(2000\)](#).

Consider the logspline density estimation without binning, and let $\delta_1, \dots, \delta_m$ be a set of basis functions that spans a space of polynomial splines. The exponential family based on these basis functions has the form

$$f(x; \mathbf{s}) = \exp \{s_1 \delta_1(x) + \dots + s_m \delta_m(x) - C(\mathbf{s})\}$$

where $C(\mathbf{s}) = \log \left\{ \int \exp \left(\sum_{i=1}^m s_i \delta_i(x) \right) dx \right\}$ is the normalizing constant and \mathbf{s} is the coefficient vector. The corresponding log-likelihood function has the form given by

$$l(\mathbf{s}) = \sum_{i=1}^n \sum_{j=1}^m s_j \delta_j(x_i) - nC(\mathbf{s}).$$

The density estimator is derived by maximizing the log-likelihood function above. Similar to [Bickel and Fan \(1996\)](#), [Stone et al. \(1997\)](#) also used cubic splines to model a log-density function. The difference is to assume that the second and third derivatives are zero outside the range of the

observed values. For this reason, extrapolation outside the range of the x -values will be a linear function. The knots are selected based on a stepwise knot addition and a stepwise knot deletion depending on Wald statistics (Kooperberg and Stone, 1992). Final model selection is based on the BIC as shown in Stone et al. (1997). The algorithm is available in R package `logspline`. However, the resultant curve is still not necessarily unimodal.

As we mentioned earlier in Section 2.2, the nonparametric maximum likelihood estimate of a unimodal density function with a unknown mode does not exist. In this situation, the log-concavity for which the estimator has the advantage of not having to specify the mode might be a useful surrogate for unimodality. Log-concave density functions, as a special important case, are particularly appealing to us.

2.3 Nonparametric Estimation of a Log-concave Density Function

2.3.1 Log-concave density

A density function f on $\mathcal{X} \subset \mathbb{R}$ is log-concave if its logarithm is concave. That is, it is of the form

$$f(x) = e^{\varphi(x)},$$

for some concave function $\varphi : \mathcal{X} \rightarrow (-\infty, \infty)$ that satisfies $\int e^{\varphi(x)} dx = 1$. One prime example is the normal density, where $\varphi(x)$ is a quadratic in x . In fact, it includes many parametric distribution families, such as the gamma distribution $\text{Gamma}(r, \gamma)$ for shape parameter $r \geq 1$, the beta distribution (with shape parameter a) $\text{Beta}(a, b)$ for $a, b \geq 1$ and the logistic distribution (Bagnoli and Bergstrom, 2005; Balabdaoui et al., 2009). Log-concavity is also appealing for many other properties. For instance, if the density function f is log-concave on (a, b) and the corresponding cumulative distribution function F is also log-concave on (a, b) , then the left side integral of F ($\int_a^x F(t) dt$) is also log-concave on (a, b) (Bagnoli and Bergstrom, 2005). Furthermore, the class of log-concave densities is the subclass of unimodal densities (Walther, 2002). By imposing log-concavity, one can avoid the spiking problem in unimodal density estimation (Dümbgen and Rufibach, 2009). Although the nonparametric maximum likelihood estimate of a unimodal density does not exist, that of a log-concave density does. It also has the desirable consistency and rates of convergence properties, and is unique and fully automatic, in the sense that there is no need for a tuning parameter (Walther, 2002, 2009; Pal et al., 2007). Moreover, log-concave densities can adaptively capture skewness, subexponential tails and nondecreasing hazard rates (Karlin, 1968; Barlow and Proschan, 1975). They offer a flexible nonparametric alternative to purely parametric

models and form an important subfamily of unimodal distributions.

With so many attractive properties, log-concave distributions have been found to be very useful in many fields, such as economics, finance, sampling and nonparametric Bayesian analysis. Many interesting propositions in the economics of information are built on the log-concavity assumption. Log-concavity was applied in the analysis of auctions to characterize efficient auctions (Matthews, 1987). In the games of incomplete information, it is assumed that the principal does not know a relevant characteristic of an agent. From the principal's point of view, the agent's type is a random variable, with distribution function F . Laffont and Tirole (1988) assumed that F is log-concave. The powerful applications of ρ -concavity have been found by Caplin and Nalebuff (1991) to the theory of elections and income distribution. In the economics of uncertainty and information, An (1998) provided a complete characterization of log-concavity. More applications in economics have been summarized by Bagnoli and Bergstrom (2005). Chen (2014) proposed semiparametric time series models with log-concave innovations and discussed their consistency. An application to sampling can be found in Gilks and Wild (1992), who proposed a method for rejection sampling from any univariate log-concave probability density function. Furthermore, Mengersen and Tweedie (1996) showed that, for a log-concave density on the real line, the Metropolis-Hastings algorithm, under very general conditions, is geometrically ergodic which is an important class of a homogeneous Markov chain on a countable state space. Thus, the assumption of log-concavity may be desirable in terms of the convergence rate of the associated Markov chain; see also Brooks et al. (1998). Bobkov et al. (2011) provided a concentration property of the information in data which comes from a log-concave distribution. Moreover, log-concavity is also useful in reliability theory (Barlow and Proschan, 1975), and clustering, e.g., Walther (2009) and the references therein. The log-concave density estimator can also be used to improve accuracy in the estimation of the tail index of a generalized Pareto distributions (Müller and Rufibach, 2009). The estimation of a wider class of log-concave densities, named as ρ -concave densities, has been investigated by Koenker and Mizera (2010). Tests for log-concavity can be found in An (1996) and Walther (2002) for univariate densities and Hazelton (2011) for multivariate densities. More properties and results for log-concavity have been reviewed and formulated by Saumard and Wellner (2014).

To the author's best knowledge, limited large sample theory is available for the log-concave estimator thus far. For the univariate case, Pal et al. (2007) established the consistency of the log-concave nonparametric maximum likelihood estimator with respect to the Hellinger metric, while Dümbgen and Rufibach (2009) provided results on the uniform consistency on compact subsets of the interior of the support and showed that if $\log(f)$ belongs to a Hölder class with exponent $\gamma \in [1, 2]$, then the estimated density \hat{f} and true density on compact subsets of the interior of

$\{f > 0\}$ are uniformly consistent with rate $O_p((\log n/n)^{\gamma/(2\gamma+1)})$ which is at least $O_p((\log n/n)^{1/3})$ and typically $O_p((\log n/n)^{2/5})$. Balabdaoui et al. (2009) showed the limit distribution theory for maximum likelihood estimation of a log-concave density. The maximum likelihood estimation of a multivariate log-concave density has been discussed by Cule et al. (2010), who showed that multivariate log-concave distributions are a very well-behaved nonparametric class. Its theoretical properties and consistency can be found in Cule et al. (2010) and Schuhmacher and Dümbgen (2010). More theoretic work about the nonparametric estimation under shape constraints can be found in Groeneboom and Jongbloed (2014).

2.3.2 Nonparametric maximum likelihood estimation of a log-concave density function

One of our goals is to estimate f based on a random sample drawn from a log-concave univariate density function. Under the log-concavity constraint, the MLE has received much more attention than the LSE due to its computational attractiveness. For a log-concave density f on \mathbb{R} , the normalized log-likelihood function of f is given by

$$\int \log f dF_n = \int \varphi dF_n,$$

where F_n stands for the empirical distribution function obtained from the sample. Many efforts have been made to relax the constraint on f being a probability density and to get a criterion function to maximize over the convex set of all concave functions φ . The most popular way is to employ the standard trick of adding a Lagrange term to the above function, which leads to the functional

$$\Psi_n(\varphi) = \int \varphi dF_n - \int \exp \varphi(x) dx,$$

owing to Theorem 3.1 of Silverman (1982). Then the nonparametric maximum likelihood estimate of $\varphi = \log f$ is the maximizer of this functional over all concave functions,

$$\hat{\varphi}_n = \operatorname{argmax}_{\varphi \text{ concave}} \Psi_n(\varphi)$$

and $\hat{f}_n = \exp \hat{\varphi}_n$.

Let $x_1 \leq x_2 \leq \dots \leq x_n$ be the order statistics. Then the nonparametric maximum likelihood estimator $\hat{\varphi}_n$ is known to be continuous and piecewise linear on $[x_1, x_n]$ with knots contained in

$\{x_1, \dots, x_n\}$, and $\hat{\varphi}_n = -\infty$ on $\mathbb{R} \setminus [x_1, x_n]$. Refer to [Walther \(2002\)](#), [Pal et al. \(2007\)](#), [Rufibach \(2007\)](#) and [Dümbgen and Rufibach \(2009\)](#), for properties such as the uniqueness and shape of $\hat{\varphi}_n$, and other basic ones.

2.3.3 Smooth nonparametric estimation of a log-concave density function

A characteristic feature of the nonparametric maximum likelihood estimator of a log-concave density function is that it is not smooth. A smooth estimate has a more attractive visual appearance.

It is somewhat surprising to find that very few papers directly deal with smooth nonparametric estimation of a log-concave density function. Several approaches have been considered by [Dümbgen and Rufibach \(2009\)](#) to estimate the unknown density f . They found that the version by a convolution of the NPMLE of a log-concave density function with a Gaussian density not only provides smoothness but also preserves the log-concave shape. The convoluted estimator is given by

$$\hat{f}^*(x) = \int \phi_{\hat{\gamma}}(x - y) d\hat{F}(y),$$

where \hat{F} is the distribution function corresponding to the NPMLE \hat{f} and $\phi_{\hat{\gamma}}$ denotes the normal density of $\mathcal{N}(0, \hat{\gamma}^2)$. The bandwidth $\hat{\gamma}$, as suggested by [Dümbgen and Rufibach \(2009\)](#), is chosen based on the fact that the variance of \hat{f}^* should coincide with $\hat{\sigma}^2$ (an unbiased estimator of the variance of the distribution), hence

$$\hat{\gamma}^2 = \hat{\sigma}^2 - \text{Var}(\hat{F}).$$

Note that the discontinuities of \hat{f} at the first and last order statistics are smoothed out by $\hat{f}^*(x)$. However, \hat{f}^* highly depends on and is rather close to \hat{f} ([Chen and Samworth, 2013](#)).

2.3.4 Algorithms for log-concave density estimation

Since the nonparametric maximum likelihood estimate of a log-concave density function $\hat{\varphi}(x)$ must be a piecewise linear continuous function on $[x_1, x_n]$ with knots only at some of x_1, \dots, x_n ([Walther, 2002](#); [Rufibach, 2007](#)), computing this estimate is commonly formulated as a convex optimization problem and is shown to have an equivalent dual formulation as a constrained maximum Shannon entropy problem ([Dümbgen and Rufibach, 2009](#); [Koenker and Mizera, 2010](#)).

More precisely, the maximization of $\left\{ \sum_{i=1}^n \varphi(x_i) - n \int e^{\varphi(x)} dx \right\}$ over a piecewise linear, concave

φ can be transformed to maximize the function

$$n\varphi_1 + \sum_{i=2}^n (n-i+1)(x_i - x_{i-1})h_i - ne^{\varphi_1} \sum_{i=2}^n \frac{e^{\sum_{k=2}^i (x_k - x_{k-1})h_k} - e^{\sum_{k=2}^{i-1} (x_k - x_{k-1})h_k}}{h_i}$$

under the constraint that $\mathbf{h} = (h_1, h_2, \dots, h_n)^T$ belongs to the cone

$$\{\mathbf{h} \in \mathbf{R}^n : h_n \leq \dots \leq h_2\},$$

where $h_1 = \varphi_1 = \varphi(x_1)$ and $h_i = \{\varphi(x_i) - \varphi(x_{i-1})\} / (x_i - x_{i-1})$, $i = 2, \dots, n$, i.e., the slope between x_{i-1} and x_i . This equivalent transformation holds according to Theorem 3.1 of [Silverman \(1982\)](#).

There are several algorithms for solving the transformed problem. The log-barrier algorithm ([Terlaky and Vial, 1998](#)) which is proposed for the general optimization problem under linear constraints can be directly employed to solve this kind of problem. The iterative convex minorant algorithm (ICMA) ([Groeneboom and Wellner, 1992](#); [Jongbloed, 1998](#); [Walther, 2002](#); [Rufibach, 2007](#)) which is developed for convex optimization problem over a convex cone was implemented by [Rufibach \(2007\)](#) using the Hermite interpolation line search in the Newton procedure for this problem. However, the two algorithms are very slow, even for a small-sized sample. [Dümbgen et al. \(2007\)](#) makes use of a faster active set algorithm (ASA) ([Fletcher, 1987](#)) for computing the log-concave MLE. [Anderson-Bergman \(2014\)](#) provided a new method for finding the log-concave NPMLE. Named as the log concave NPMLE algorithm (LC), it combines the ideas of [Fletcher \(1987\)](#) and [Dümbgen et al. \(2007\)](#). The main problem in computing the NPMLE of a log-concave density is how to efficiently locate all relevant knots and update the changes of slope at the knots.

Note that the smooth version of [Dümbgen and Rufibach \(2009\)](#) needs to first find the NPMLE \hat{f} using, e.g., the ASA algorithm. Denoting by \hat{F} the corresponding CDF, the variance of \hat{F} is given by

$$\text{Var}(\hat{F}) = \int_{x_1}^{x_n} (x - \bar{X})\hat{f}(x) dx,$$

where \bar{X} is the mean of the sample. Hence the default bandwidth is the square root of $(\hat{\sigma}^2 - \text{Var}(\hat{F}))$.

2.4 Nonparametric Estimation of a Unimodal and Heavy-tailed Distribution

Not every unimodal density is log-concave. A good example is the Student's t -distribution, which (with finite degrees of freedom) has tails heavier than that of a log-concave density function. Log-concavity is not a good assumption for data which has a very high kurtosis; see also Meyer (2012). The distributions of financial data such as stock prices and bonds yields are typically heavy-tailed, high-peaked and skewed, as compared with the normal distribution. When estimating the distributions of such financial data, we need to relax the log-concavity constraint.

2.4.1 Heavy-tailed distributions

Skewness and Kurtosis

Skewness, which measures the symmetry of a distribution, and kurtosis, which measures the tail flatness, give us insight into the shape of the distribution. Skewness and kurtosis are the 3rd and 4th standardized central moments. More precisely, let X be a random variable with mean μ and standard deviation σ . The skewness and kurtosis of its distribution are given by, respectively,

$$\text{Sk} = \text{E} \left(\frac{(X - \mu)^3}{\sigma^3} \right) \quad \text{and} \quad \text{Kur} = \text{E} \left(\frac{(X - \mu)^4}{\sigma^4} \right).$$

A normal distribution has skewness 0 and kurtosis 3. Hence, we speak of a left (or right) skewed distribution if the skewness is negative (or positive) and a heavy-tailed, or leptokurtic, distribution if its kurtosis is greater than 3. In many financial applications it is the right tail of the distribution that is of interest, but a distribution may have a heavy left tail, or both tails may be heavy. Throughout the thesis, we may use fat, thick or long tails as synonyms of heavy tails.

Overview

Heavy-tailed distributions arise in many fields such as finance, economics, hydrology, geology and physics. In finance, for example, most of the classical financial models rely heavily on the assumption that the random variables under investigation follow normal distributions, such as Markowitz's Portfolio Theory (Markowitz, 1952), the Capital Asset Pricing Model of Share (Frencha, 2003), and Black-Scholes' formula (Black and Scholes, 1973). However, the Gaussian assumption was rejected by Mandelbrot (1963) and Fama (1965), who pointed out that the distribution of asset returns is not well approximately by the Gaussian. Since then, a lot of empirical researches

have been carried out which lead to some stylized facts that the distribution of asset returns is skewed and has a higher peak and heavier tails than the normal distribution; see [Rachev et al. \(2005\)](#) and [Stoyanov et al. \(2011\)](#). Data of this kind can also be found in economics ([McDonald, 1984](#); [Charpentier and Flachaire, 2014](#)), healthcare ([Manning et al., 2005](#)), non-life insurance ([Klugman et al., 2012](#)), World Wide Web traffic data ([Maiboroda and Markovich, 2004](#); [Markovitch and Krieger, 2000](#)) and survival analysis ([Anderson-Bergman, 2014](#)), to name a few. For more applications, see [Meerschaert and Scheffler \(2003\)](#) and the references therein.

With the development in computing power and the availability of large data sets, there are major advances in the study of heavy-tailed distributions and processes. Apart from a full distribution modelling, many approaches are developed based on the extreme value theory which deals with extreme deviations from the median of a probability distribution, and they are important tools for modelling heavy tails.

The extreme value theory has a long history, with many applications, such as extreme temperatures, floods and winds ([Stott et al., 2004](#); [Buishand et al., 2008](#); [Stoyanov et al., 2011](#)). [Embrechts et al. \(1997\)](#) provided applications to insurance and finance. [Gencay and Selcuk \(2006\)](#) investigated extreme value theory studies in finance literature and examined the dynamics of extreme values of overnight borrowing rates using generalized Pareto distributions. [Stoyanov et al. \(2011\)](#) proposed two extreme value theory-based models, the block of maxima method and the peaks-over-threshold method, for modelling financial returns. For more applications in finance, refer to [Novak \(2011\)](#). Extreme value theory is a powerful and fairly robust framework to examine the tail behaviour of a distribution. However, models based on extreme value theory are only for modelling the tail behaviour while largely ignoring the rest of the distribution. Sometimes, the research interest also lies in the central section or the main body of the distribution. In our thesis, we focus on full density estimation.

There are two broad classes of estimators that are used to estimate a full distribution. One is parametric models, and the other, nonparametric ones.

2.4.2 Parametric models for heavy-tailed distributions

Parametric models, such as the Student's t , Stable, Pareto, Weibull, Lévy, Log-gamma and Log-normal distributions, are commonly-used heavy-tailed distributions. The Student's t distribution with finite degrees of freedom (DOF) has a higher peak around the centre and fatter tails than a normal distribution. [Liesenfeld and Jung \(1997\)](#) proposed a stochastic volatility model based on the Student's t distribution and employed a simulated maximum likelihood approach to estimate

the stochastic volatility specifications. Pivac (2006) proposed a procedure of Value-at-Risk forecast under the assumption that financial asset returns follow the Student's t distribution when applying the GARCH model. A risk estimation model based on the Student's t distribution with a fixed DOF was suggested by Zumbach (2007). However, fixing the tail thickness to a given value for all assets makes little sense and leads to a significant limitation. Furthermore, the classic Student's t distribution is symmetric, which is not proper in the case where there is a significant distributional asymmetry. In order to capture skewness, Theodossiou (1998) developed a skewed version of the Generalized t distribution and applied it to financial data. Fernández and Steel (1998) presented a general method that transforms any continuous unimodal and symmetric distribution into a skewed one for a fat-tailed distribution by changing the scale at each side of the mode. The skewed distribution f_s is given by

$$f_s(y|\gamma) = \frac{2\gamma}{\gamma^2 + 1} \left\{ f\left(\frac{y}{\gamma}\right) I_{[0,\infty)}(y) + f(y\gamma) I_{(-\infty,0]}(y) \right\},$$

where $\gamma > 0$ is a skewness parameter, $f(\cdot)$ a univariate density symmetric about zero and $I_c(\cdot)$ the indicator function. Note that $\gamma = 1$ gives the symmetric distribution as $f_s(y|\gamma = 1) = f(y)$, and a value of $\gamma > 1$ (< 1) indicates a right (left) skewness. A continuous random variable X with location and scale parameter $\mu \in \mathbb{R}$ and $\sigma > 0$ can be represented as $X = \mu + \sigma y$ and its density function is then given by

$$f_s(x|\gamma) = \frac{2\gamma}{(\gamma^2 + 1)\sigma} \left\{ f\left(\frac{x - \mu}{\sigma\gamma}\right) I_{[\mu,\infty)}(x) + f\left(\frac{(x - \mu)\gamma}{\sigma}\right) I_{(-\infty,\mu]}(x) \right\}.$$

Thus, $f_s(s|\gamma)$ is the skewed version of the location-scale density $f(\cdot)$ preserving the mode μ .

A logarithmic version of the skew-normal and skew- t distributions have been introduced by Walls (2005) for modelling film returns. These models were further studied by Pitt (2010) in the economic accomplishments of individual members in a performing rights organization.

Stable distributions have been widely used since the 1960s. Mandelbrot (1963) and Fama (1965) suggested the class of stable Paretian distributions as an extension to the Gaussian hypothesis. The application of stable Paretian distributions to financial data can be found in Curto et al. (2003). However, more empirical scrutiny produced evidence that the observed financial return distribution is not consistent with those predicted by the stable Paretian distribution (Rachev et al., 2005). Hence, alternative classes of distributions were suggested. Mittnik and Rachev (1993) considered the Weibull distribution, the Geometric Stable distribution and other ones for returns of financial assets. Rachev et al. (2005) introduced the Tempered Stable distribution which is

achieved by modifying only the tails of stable distributions so that they remain thicker than the Gaussian tails but do not lead to an infinite volatility; see also [Kim et al. \(2010\)](#). In risk management, however, [Champagnat et al. \(2013\)](#) pointed out that the stable hypothesis seems too strong, and suggested power law-based models.

[Eberlein and Keller \(1995\)](#) assumed the hyperbolic distribution for returns. [Dutta et al. \(2002\)](#) applied the g-and-h distribution and Generalized Beta Distributions of the Second Kind (GB2) to model the skewness and kurtosis of short rates. [Manning et al. \(2005\)](#) applied a three parameter generalized gamma distribution to handle cross-section heavy-tailed data. [Sun et al. \(2008\)](#) introduced copulas to model the dependencies over time, and heavy-tailed distributions, such as GB2, to model the marginal distributions with application to Wisconsin nursing homes data. Copulas have also been applied to risk management by [Rosenberg and Schuermann \(2006\)](#). There is also a large literature on Lévy processes in finance; see [Kou \(2014\)](#) and [Tankov \(2004\)](#) and the references therein. [Maiboroda and Markovich \(2004\)](#) employed Hill's estimator ([Hill et al., 1975](#)) to estimate the tail based on a Pareto distribution with application to Web data. [Holan and McElroy \(2010\)](#) developed a consistent, flexible estimator of the tail exponent based on a Fourier series estimator and employed the logarithm of Parzen's density-quantile function to separately estimate the left and right tail exponent. [Bruffaerts et al. \(2014\)](#) applied a simple rank-preserving transformation on the original data so that the transformed observations can be adjusted by a so-called Tukey g-and-h distribution, and further created a boxplot for variables with a continuous and unimodal distribution that might be skewed and/or heavy-tailed. [Rubio and Steel \(2014\)](#) applied a Bayesian method to model heavy-tailed and skewed distributions based on a five-parameter distribution obtained by using a two-piece probability density function. [Ehlers \(2015\)](#) represented the commonly-used parametric models by using scale mixtures which enable efficient Bayesian estimation via Markov chain Monte Carlo methods.

The advantages of the parametric approach lay in fast computing and easy interpretation. However, they may suffer in performance owing to model misspecification and hence are not always reliable. Nonparametric method makes fewer assumptions on the process of an underlying asset and tend to be more robust against specification errors.

2.4.3 Nonparametric models for heavy-tailed distributions

Many nonparametric methods for heavy-tailed distributions are also proposed using the idea of kernel estimation. [Markovitch and Krieger \(2000\)](#) considered a Parzen-Rosenblatt kernel estimate and a histogram with variable bin width called polygram to estimate the long-tailed density functions

with world wide web traffic data. Takada et al. (2001) showed that, for univariate heavy-tailed density estimation, the logspline estimator of Kooperberg and Stone (1991, 1992) and the adaptive kernel density estimator of Silverman (1986) outperform the fixed bandwidth kernel estimator and the Hermite series estimator. They further pointed out that the logspline estimator exhibits good convergence independent of the tail behaviour of the target density. The adaptive kernel density estimator of Silverman (1986) is defined as

$$\hat{f}_{ak} = \frac{1}{n} \sum_{i=1}^n \frac{1}{h\lambda_i} \phi\left(\frac{X - x_i}{h\lambda_i}\right),$$

where λ_i is a local bandwidth factor which makes the bandwidth $h\lambda_i$ smaller around the center and larger in the tails.

Properties of some nonparatmetric Value-at-Risk estimators have been investigated by Chen and Tang (2005), in which an estimation procedure of standard errors was proposed based on kernel estimation of the spectral density of a derived series. Boundary kernel was designed by Markovich (2005) to improve upon estimation on distribution tails. With applications to insurance, Buch-Larsen et al. (2005) proposed an estimator obtained by transforming the data with a modification of the Champernowne cumulative density function and applied the classical kernel density estimator to the transformed data. This method involves determining the threshold level between large and small losses. The Champernowne distribution is a generalization of the logistic distribution that was introduced by Champernowne (1953) which developed the distribution to describe the logarithm of income. Denote by $F_{\text{Cha}}(x; \alpha, M, c)$ the modified Champernowne distribution functions with parameters (α, M, c) and $x \geq 0$, it is a heavy-tailed distribution converging to a Pareto distribution as x tends to infinity. Buch-Larsen et al. (2005) suggested a natural way to estimate the parameter M as the empirical median by recognizing $F_{\text{Cha}}(M) = 0.5$, and estimate (α, c) by maximizing the log likelihood function. They also recommended the classical kernel density estimator (Silverman, 1986; Wand and Jones, 1994) on the transformed data. Buch-Larsen et al. (2005) denoted this resulting transformation kernel density estimator based on the Champernowne distribution by KMCE and compared it with the estimators of Bolancé et al. (2003) and Clements et al. (2003). They recommended the KMCE estimator for the heavy-tailed distributions. Buch-Kromann (2009) further proposed a conditional maximum likelihood method to estimate the parameters of the modified Champernowne distribution to improve the performance of the Champernowne transformed kernel density estimator in the tail. They showed that the KMCE obtained this way outperforms in general the benchmark estimators in terms of tail performance. A Monte Carlo simulation study has been carried out to show that the new KMCE estimator provides a density

estimate on the full density superior to the g-and-h distribution and different from the generalized pareto distribution (GPD) estimator, which only provides a density estimate above the threshold.

Markovich (2006) also introduced a preliminary transformation of the data and applied the kernel estimator to estimate a heavy-tailed distribution with application to world wide web traffic data. A double transformation kernel density estimator for estimating a skewed and heavy-tailed distribution was proposed by Bolancé et al. (2008). Saulo et al. (2013) introduced a new nonparametric kernel method for estimating asymmetric densities based on generalized skew-Birnbaum-Saunders distributions with application to air pollutant concentration data. With applications to income, Charpentier and Flachaire (2014) proposed a kernel density estimator with data logarithmic transformations.

However, all above-mentioned estimators involve choosing bandwidth, and some of them need to choose a threshold. For instance, the GPD estimator and the estimator of Buch-Kromann (2009). The threshold is important for the estimator and is hard to pick. Some efforts have been made to avoid kernel-based methods. A joint parametric estimator, where the tail is approximated by a general Pareto-like family, and nonparametric, where the main body is fitted by a finite linear combination of some basic functions, was proposed by Markovitch and Krieger (2002) for heavy-tailed distributions. This estimator chooses a specific tail quantile as boundary between the tail and the body. Venturini et al. (2008) proposed an efficient general Bayesian approach for the estimation of heavy-tailed distributions based on a mixture of Gamma distributions with applications to health services. Unfortunately, this estimator only handles positive values. It also needs to choose an appropriate number of mixture components m . A small value of m may create a severe restriction to the model when data has a large mean. On the other hand, too large a value of m may cause numerical problems. Anderson-Bergman (2014) introduced a new, more flexible shape constraint, or “inverse convex”, for survival analysis and other types of heavy tailed data. This estimator is very attractive for survival analysis for censored data, but it is not smooth.

2.5 Nonparametric Mixtures

Later in the thesis, we will relate shape-restricted density estimation to nonparametric mixture estimation. We thus give a general description of this kind estimation. For a probability measure G on $\Omega \in \mathbb{R}$,

$$f(x; G) = \int_{\Omega} f(x; \theta) dG(\theta),$$

is the density of a univariate nonparametric mixture of $\{f(x; \theta) : x \in \mathcal{X}, \theta \in \Omega\}$ with mixing distribution G . Given a random sample x_1, \dots, x_n from density f , the log-likelihood of G is given by

$$l(G) = \sum_{i=1}^n \log \{f(x_i; G)\}.$$

Any probability measure \hat{G} on Ω that maximizes $l(G)$ among all probability measures on Ω is called a nonparametric maximum likelihood estimator of G (NPMLE). The NPMLE \hat{G} is known to be discrete with some support set. Further, the number of points contained in this support set is no more than that of distinct values in the sample; see [Laird \(1978\)](#) and [Lindsay et al. \(1983\)](#). For $\theta \in \Omega$, let δ_θ be the unit measure that assigns mass 1 to the point θ . Then for a discrete G , it can be written as $G(\theta) = \sum_{j=1}^m \pi_j \delta_{\theta_j}$, where $\theta_j \in \Omega$ and $\pi_j > 0$ for $j = 1, \dots, m$, $\sum_{j=1}^m \pi_j = 1$. The nonparametric maximum likelihood estimation of a mixing distribution is thus the following optimization problem:

$$\text{maximize } l(G), \quad \text{subject to } G \in \mathcal{G},$$

where \mathcal{G} is the class of all probability measures defined on the measurable space (Θ, B) , and B contains all singletons of θ . An important property of the log-likelihood $l(G)$ is that it is a concave function of G on the convex set \mathcal{G} . This makes global maximization of $l(G)$ possible.

A key feature in the theory of the NPMLE is the gradient function which is a special directional derivative of the log-likelihood function. For each probability measure H on Ω , the directional derivative of $l(G)$ in direction of H is defined as

$$d(H; G) = \lim_{\varepsilon \rightarrow 0} \frac{l((1 - \varepsilon)G + \varepsilon H) - l(G)}{\varepsilon} = \sum_{i=1}^n \frac{f(x_i; H)}{f(x_i; G)} - n.$$

For $\theta \in \Omega$, the directional derivative $d(\theta, G)$ of $l(G)$ in direction of θ is defined as $d(\theta; G) = d(\delta_\theta; G)$. When considered as a function of θ , $d(\theta; G)$ is called the gradient function. Apparently,

$$d(\theta; G) = \sum_{i=1}^n \frac{f(x_i; \theta)}{f(x_i; G)} - n.$$

Let $l(\hat{G}) > -\infty$, then \hat{G} is a NPMLE of G if and only if $d(\theta; \hat{G}) \leq 0$ for all $\theta \in \Omega$, owing to the general equivalent theorem. Furthermore, it satisfies:

$$\sup_{\theta} d(\theta; G) \geq l(\hat{G}) - l(G).$$

We refer the reader to [Lindsay \(1995\)](#), [Lesperance and Kalbfleisch \(1992\)](#) and [Wang \(2007\)](#) for more theoretical results.

There are many algorithms for computing the nonparametric maximum likelihood estimate. [Laird \(1978\)](#) proposed an expectation-maximization algorithm which seeks to find the maximum likelihood estimate of the marginal likelihood by iteration. This algorithm uses a large number of initial support points. [Fedorov \(1972\)](#), [Wu \(1982\)](#) and [Groeneboom et al. \(2008\)](#) developed the vertex-direction-method which has the drawbacks of slow convergence and the appearance of clusters sometimes. Other methods including the vertex exchange method ([Böhning, 1985](#)) and the intra-simplex direction method ([Lesperance and Kalbfleisch, 1992](#)) may also converge very slowly. The speed of convergence is important in an iterative computational method. [Wang \(2007\)](#) proposed the constrained Newton method (CNM) which is demonstrated to be very fast and stable for computing the NPMLE for a mixture. At each iteration, [Wang \(2007\)](#) uses the gradient function to find new support points, updates all mixing proportions via a quadratically convergent method and discards the support points with mass 0. It also uses the back-tracking line search strategy guarded by the Armijo rule to ensure the monotonic increase of log-likelihood. The CNM is much faster than the other NPMLE methods, especially when the number of support points is large. The convergence of this algorithm is also theoretically established. New algorithms will be developed in this thesis for shape-restricted nonparametric density estimation, all based on the central idea of the CNM.

Chapter 3

A Fast Algorithm for Log-concave Density Estimation

3.1 Introduction

The study of nonparametric density estimation under qualitative assumptions such as monotonicity or convexity on certain subset of its domain has received considerable attention recently. Such assumptions are usually plausible or sometimes direct consequences of the problem itself under investigation (Wang et al., 2005). Further, by imposing these constraints, it can obtain fully automatic nonparametric estimators without any tuning parameter which can be difficult to choose.

The nonparametric maximum likelihood estimate (NPMLE) for a nonincreasing density function f on $[0, -\infty)$ is known as the Grenander (1956) estimator which is given explicitly by the left derivative of the least concave majorant of the empirical cumulative distribution function. For the monotone density estimator, its asymptotic distribution theory has been established by Rao (1969) and its applications can be found in Barlow (1972) and Robertson et al. (1988). The construction of monotone density estimator and its properties can be straightforwardly extended to estimate a unimodal density with true mode known. However, one may encounter another problem in this situation that the density estimator turns out to be inconsistent at the mode, the alleged spiking phenomenon (Woodroffe and Sun, 1993; Meyer and Woodroffe, 2004). Moreover, the extension can not be carried out if the mode is unknown, in which case the likelihood has no boundary and thus the NPMLE does not exist (Birgé, 1997).

In this chapter, we study an attractive and natural alternative shape constraint on the density

f , the log-concavity. A log-concave density function f on $\mathcal{X} \subset \mathbb{R}$ is defined as

$$f(x) = e^{\varphi(x)}, \quad (3.1)$$

where $\varphi : \mathcal{X} \rightarrow (-\infty, \infty)$ is a concave function that satisfies $\int e^{\varphi(x)} dx = 1$. Note that the class of log-concave densities is a subclass of unimodal densities. They are often referred to as “strongly unimodal” densities, in the sense that the convolution of a log-concave density with any unimodal density is still unimodal (Ibragimov, 1956). The allure of log-concave assumption is that the NPMLE of a log-concave density exists, is consistent and can be computed without any tuning parameter (Walther, 2002, 2009; Pal et al., 2007). By imposing log-concavity, one can also avoid aforementioned spiking problem (Dümbgen and Rufibach, 2009). Moreover, log-concave densities can adaptively capture skewness, subexponential tails and nondecreasing hazard rates (maybe exponential tails) (Karlin, 1968; Barlow and Proschan, 1975). They form a fairly rich and flexible family which entails most of the commonly-used parametric densities, such as normal, uniform, Gamma(r, γ) for $r \geq 1$, Beta(a, b) for $a, b \geq 1$ and logistic density (Bagnoli and Bergstrom, 2005; Balabdaoui et al., 2009). Hence assuming log-concavity offers a flexible nonparametric alternative to purely parametric models. These desirable properties of log-concave densities make them practically very useful in many fields, such as economics (An, 1996, 1998; Bagnoli and Bergstrom, 2005), finance (Chen, 2014), sampling (Gilks and Wild, 1992), reliability theory (Barlow and Proschan, 1975) and nonparametric Bayesian analysis (Brooks et al., 1998).

In order to investigate the NPMLE $\hat{\varphi}$ in the univariate setting, let us consider a random sample of size n independently drawn from a univariate log-concave density. Since we only need to deal with order statistics, for notational simplicity, let $x_1 \leq x_2 \leq \dots \leq x_n$ be the order statistics. Then the NPMLE $\hat{\varphi}(x)$ exists, is unique and is a piecewise linear continuous function on $[x_1, x_n]$ with knots only at some of x_1, \dots, x_n , and $\hat{\varphi}(x) = -\infty$ on $\mathbb{R} \setminus [x_1, x_n]$, i.e., no positive probability mass will be given to the outside of $[x_1, x_n]$ (Walther, 2002; Pal et al., 2007; Rufibach, 2007).

There are several algorithms proposed in the literature for finding $\hat{\varphi}$. Some algorithms are developed based on maximizing, with respect to a piecewise linear, concave φ ,

$$\begin{aligned} \sum_{i=1}^n \varphi(x_i) - n \int e^{\varphi(x)} dx &= n\varphi_1 + \sum_{i=2}^n (n-i+1)(x_i - x_{i-1})h_i \\ &\quad - ne^{\varphi_1} \sum_{i=2}^n \frac{e^{\sum_{k=2}^i (x_k - x_{k-1})h_k} - e^{\sum_{k=2}^{i-1} (x_k - x_{k-1})h_k}}{h_i}, \end{aligned} \quad (3.2)$$

i.e., under the constraint that $\mathbf{h} = (h_1, h_2, \dots, h_n)^T$ belongs to the cone

$$\{\mathbf{h} \in \mathbf{R}^n : h_n \leq \dots \leq h_2\},$$

where $h_1 = \varphi_1 = \varphi(x_1)$ and $h_i = \{\varphi(x_i) - \varphi(x_{i-1})\} / (x_i - x_{i-1})$, $i = 2, \dots, n$, be the slope between x_{i-1} and x_i . This is because maximizing function (3.2) is equivalent to likelihood maximization over φ subject to $\int e^{\varphi(x)} dx = 1$ (Silverman, 1982). Note that equation (3.2) is a nonlinear/nonquadratic approximation (with a Lagrange multiplier term) to the log-likelihood function. It is known that the optimization problem for function (3.2) can be solved by the log-barrier algorithm (Terlaky and Vial, 1998) which is proposed for the general optimization problem under linear constraints or the iterative convex minorant algorithm (ICMA) (Groeneboom and Wellner, 1992; Jongbloed, 1998; Walther, 2002; Rufibach, 2007) which is developed for convex optimization problem over a convex cone.

To speed up the ICMA, Rufibach (2007) complements the line search by a Hermite interpolation in the Newton procedure and compares the modified ICMA to the log-barrier algorithm. Since in each iteration the ICMA uses all observed points as knots for the pool-adjacent-violators algorithm (PAVA) (Ayer et al., 1955; Robertson et al., 1988) to update the changes of slope, it slows down rapidly as the sample size increases. Furthermore, the ICMA works by approximating the target function with a second order Taylor expansion, in which the off-diagonal partial derivatives are ignored. This may lead to many more iterations needed to find the solution.

The active set algorithm (ASA) (Fletcher, 1987) was introduced to compute the log-concave NPMLE by Dümbgen et al. (2007) based on the idea of the vertex direction and support reduction algorithm (Groeneboom et al., 2008). The ASA improves upon the speed of updating the changes of slope by considering only a much smaller subset of the active set of parameters during optimization, rather than all n parameters, except when choosing a new parameter to add to the active set. It finds and adds one new knot in each iteration and requires finitely many iterations to find all necessary knots. This algorithm is effective for solving small to medium-scale problems, but its speed of convergence is affected by the number of knots, which increases with the sample size.

Anderson-Bergman (2014) provided a new method for finding the log-concave NPMLE. Named as log concave NPMLE algorithm (LC), it combines the ideas of Fletcher (1987) and Dümbgen et al. (2007). Its main advantage over the ASA is for handling interval-censored data. Our studies show that for exact observations, the LC algorithm does not significantly outperform the ASA and may become slower than the latter for a large sample size. Unlike the ASA, the LC algorithm directly

maximizes the log likelihood function

$$l(\varphi) = \sum_{i=1}^n \varphi(x_i) - n \log \left\{ \int e^{\varphi(x)} dx \right\}. \quad (3.3)$$

Both the ASA and the LC algorithm use ICMA for maximization over the active set given by the linear constraints of concavity, where standard Newton's method is difficult to apply as it would not respect the boundaries. The main difference between them is that the LC algorithm includes both the univariate and multivariate steps in optimizing the parameters, while the ASA only has the multivariate step. The LC algorithm may be affected more by the number of knots. The main problem in computing the NPMLE of a log-concave density is how to efficiently locate all relevant knots and update the changes of slope at the knots.

In this chapter, we propose a fast algorithm for finding the NPMLE of a log-concave density. It is methodologically different from the existing algorithms for this task, as it maximizes directly log-likelihood function (3.3), without resorting to function (3.2) or ICMA. It is an extension of the constrained Newton method (CNM) (Wang, 2007) which was proposed for computing the NPMLE of a mixing distribution. Our new algorithm is iterative and consists of three steps in each iteration. Firstly, it locates multiple knots, but not too many, by applying the gradient function which is defined by a special directional derivative of the log likelihood function. Adding multiple knots helps reduce the number of iterations. Secondly, a quadratically convergent method is used to update the changes of slope, which requires only non-negativity constraints and is thus numerically very stable. Finally, it discards the knots that do not change the slope. It is shown that the sequence of iterates produced by our algorithm is guaranteed to converge to the NPMLE. Numerical studies also show that it outperforms all existing algorithms in terms of computing time.

Some useful functions and properties of the NPMLE of a log-concave density are introduced in Section 3.2. Section 3.3 first describes how to update the changes of slope for a given set of knots and then presents the new algorithm. The convergence of the algorithm is established in Section 3.4. Numerical studies and real-world applications are given in Section 3.5. Section 3.6 gives some concluding remarks.

3.2 Characterization of the Nonparametric Maximum Likelihood Estimate

3.2.1 Parametrizations and log-likelihood function

We propose a general log-concave function based on the maximum likelihood method from the observations in this section. Since [Walther \(2002\)](#) and [Rufibach \(2007\)](#) showed that the NPMLE of φ must be piecewise linear and could only have knots at some observation points, we let $\mathcal{X} = [x_1, x_n]$, $\mathcal{P} = \{x_1, \dots, x_n\}$ and \mathcal{K} the family of concave functions $\varphi: \mathcal{X} \rightarrow \mathbb{R}$ which is linear on each interval $[x_k, x_{k+1}]$, $1 \leq k < n$, and $\varphi := -\infty$ on $\mathbb{R} \setminus \mathcal{X}$. Hence, the piecewise linearity enables us to write $\varphi \in \mathcal{K}$ as

$$\varphi(x) = \alpha x - \sum_{j=1}^m \pi_j (x - \theta_j)_+ - C, \quad (3.4)$$

where $x \in \mathcal{X}$, $\alpha \in \mathbb{R}$, $\pi_j \geq 0$, $\theta_j \in \mathcal{P}$ for $j = 1, \dots, m$, and the normalizing constant C is given by

$$C = \log \left\{ \int_{\mathcal{X}} e^{\alpha x - \sum_{j=1}^m \pi_j (x - \theta_j)_+} dx \right\}.$$

A subscript “+” means that

$$a_+ = \begin{cases} a & \text{if } a > 0; \\ 0 & \text{if } a \leq 0, \end{cases}$$

which has precedence over other operations, such as exponentiation. Denoting $\boldsymbol{\pi} = (\pi_1, \dots, \pi_m)^T$ the vector of positive masses, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)^T$ the vector of knots. Note that the slope can only change at a knot.

For $\varphi \in \mathcal{K}$, density (3.1) is written as

$$f(x; \varphi) = \frac{e^{\alpha x - \sum_{j=1}^m \pi_j (x - \theta_j)_+}}{\int_{\mathcal{X}} e^{\alpha x - \sum_{j=1}^m \pi_j (x - \theta_j)_+} dx}. \quad (3.5)$$

The log-likelihood function of φ is thus given by

$$l(\varphi) = \sum_{i=1}^n \left\{ \alpha x_i - \sum_{j=1}^m \pi_j (x_i - \theta_j)_+ \right\} - n \log \left\{ \int_{\mathcal{X}} e^{\alpha x - \sum_{j=1}^m \pi_j (x - \theta_j)_+} dx \right\}. \quad (3.6)$$

Denoting $\boldsymbol{\beta} = (\alpha, \boldsymbol{\pi}^T)^T$, then any $\varphi \in \mathcal{K}$ is completely defined by its knot set $\boldsymbol{\theta}$ and its coefficient vector $\boldsymbol{\beta}$. One may also write density (3.5) as $f(x; \boldsymbol{\beta}, \boldsymbol{\theta})$ and log-likelihood (3.6) as $l(\boldsymbol{\beta}, \boldsymbol{\theta})$. We aim to maximize $l(\varphi)$ over the space \mathcal{K} , or equivalently, $l(\boldsymbol{\beta}, \boldsymbol{\theta})$, which includes the finding of the dimension m .

3.2.2 Characterization

In the context of fitting a nonparametric mixture model, Wang (2007) uses a gradient function to find new support points which is a special directional derivative of the log-likelihood function. In this chapter, we consider a similar gradient function for finding new knots. Consider a $\varphi \in \mathcal{K}$ and an arbitrary $\theta \in \mathcal{P}$ and let $e_\theta(x) = -(x - \theta)_+$ be a basis function. Define the gradient function by

$$\begin{aligned} d(\theta; \varphi) &\equiv \left. \frac{\partial l(\varphi + \varepsilon e_\theta)}{\partial \varepsilon} \right|_{\varepsilon=0^+} = \lim_{\varepsilon \rightarrow 0^+} \frac{l(\varphi + \varepsilon e_\theta) - l(\varphi)}{\varepsilon} \\ &= - \sum_{i=1}^n (x_i - \theta)_+ + n \int_{x \in \mathcal{X}} (x - \theta)_+ e^{\varphi(x)} dx \\ &= \sum_{i=1}^n e_\theta(x_i) - n \mathbf{E}_f(e_\theta(X)), \end{aligned} \quad (3.7)$$

where $f = f(x; \varphi)$ and \mathbf{E}_f is the expectation with respect to f . For completeness of notation, letting $e_{(0)}(x) = x$, we also define

$$\begin{aligned} d_0(\alpha; \varphi) &\equiv \left. \frac{\partial l(\varphi + \varepsilon e_{(0)})}{\partial \varepsilon} \right|_{\varepsilon=0^+} = \sum_{i=1}^n x_i - n \int_{x \in \mathcal{X}} x e^{\varphi(x)} dx \\ &= \sum_{i=1}^n e_{(0)}(x_i) - n \mathbf{E}_f(e_{(0)}(X)). \end{aligned} \quad (3.8)$$

Moreover, for an arbitrary $v \in \mathcal{K}$, denote by α_v the first element of its coefficient vector. Similarly, let us define

$$\begin{aligned} d(v; \varphi) &\equiv \left. \frac{\partial l(\varphi + \varepsilon v)}{\partial \varepsilon} \right|_{\varepsilon=0^+} \\ &= d_0(\alpha; \varphi) \alpha_v + \int_{\mathcal{P}} d(\theta; \varphi) dv(\theta). \end{aligned}$$

Further, we have

$$\begin{aligned} d(v - \varphi; \varphi) &\equiv \left. \frac{\partial l(\varphi + \varepsilon(v - \varphi))}{\partial \varepsilon} \right|_{\varepsilon=0^+} \\ &= d(v; \varphi) - d(\varphi; \varphi). \end{aligned}$$

3.2.3 Theoretical properties

Some properties of the equivalent log-likelihood function (3.2) and the uniform consistency of the estimator have been described by Dümbgen and Rufibach (2009). Anderson-Bergman (2014)

3.2. Characterization of the Nonparametric Maximum Likelihood Estimate

provided some properties of the log-likelihood function (3.3). In our parametrization, we give some further properties of the log-likelihood and gradient functions, which will be needed later for establishing the convergence of our algorithm. Denote $\mathcal{K}(l_0) = \{\varphi \in \mathcal{K} : l(\varphi) \geq l_0\}$ for any given $l_0 > -\infty$.

Proposition 3.2.1. *$l(\varphi)$ is concave on \mathcal{K} and any non-empty level set $\mathcal{K}(l_0)$ is convex.*

Proof. The proof is similar to that given by Dümbgen and Rufibach (2009). For $\varphi, v \in \mathcal{K}$ and $\epsilon \in (0, 1)$,

$$\left. \frac{\partial l(\varphi + \epsilon(v - \varphi))}{\partial \epsilon} \right|_{\epsilon=0^+} = \sum_{i=1}^n (v(x_i) - \varphi(x_i)) - n \int_{x \in \mathcal{X}} (v(x) - \varphi(x)) e^{\varphi(x)} dx,$$

$$\begin{aligned} \left. \frac{\partial^2 l(\varphi + \epsilon(v - \varphi))}{\partial^2 \epsilon} \right|_{\epsilon=0^+} &= -n \left(\int_{x \in \mathcal{X}} (v(x) - \varphi(x))^2 e^{\varphi(x)} dx - \left(\int_{x \in \mathcal{X}} (v(x) - \varphi(x)) e^{\varphi(x)} dx \right)^2 \right) \\ &= -n \left(\mathbb{E}_f (v(X) - \varphi(X))^2 - \mathbb{E}_f^2 (v(X) - \varphi(X)) \right), \end{aligned}$$

where $f = f(x; \varphi)$. Note that according to the Cauchy-Schwarz inequality, the latter expression is nonpositive, with equality holding only when $v = \varphi$ almost everywhere on \mathcal{X} . Furthermore, if $\varphi, v \in \mathcal{K}(l_0)$, then

$$l(\varphi + \epsilon(v - \varphi)) \geq \min \{l(\varphi), l(v)\} \geq l_0,$$

which means $\varphi + \epsilon(v - \varphi) \in \mathcal{K}(l_0)$, that is, $\mathcal{K}(l_0)$ is convex. \square

Proposition 3.2.2. *$f(x; \varphi)$ is bounded below and above for all $x \in \mathcal{X}$ and all $\varphi \in \mathcal{K}(l_0)$.*

Proof. Since $f(x)$ is a log-concave function, then $\forall x_i, x_j \in \mathcal{X}$ we have

$$f((1 - \lambda)x_i + \lambda x_j) \geq f(x_i)^{1-\lambda} f(x_j)^\lambda, \quad \lambda \in [0, 1].$$

Denote by x_m the mode of f and by f_{\min} and f_{\max} , respectively, the minimum and maximum values of $f(x)$ on \mathcal{X} . Then

$$\begin{aligned} 1 &= \int_{x_1}^{x_n} f(x) dx \\ &= \int_{x_1}^{x_m} f(x) dx + \int_{x_m}^{x_n} f(x) dx \\ &\geq \frac{r_1(f_{\max} - f_{\min})}{\log f_{\max} - \log f_{\min}}, \end{aligned} \tag{3.9}$$

3.2. Characterization of the Nonparametric Maximum Likelihood Estimate

where $r_1 = x_n - x_1$. Since the right-hand side of inequality (3.9) increases with f_{\min} , using $f_{\min} f_{\max}^{n-1} \geq l_0$ gives

$$1 \geq \frac{r_1(f_{\max} - l_0/f_{\max}^{n-1})}{\log f_{\max} - \log(l_0/f_{\max}^{n-1})}.$$

With $f_{\max} \geq \sqrt[n]{l_0}$, consider if $f_{\max} \geq n^{-1}\sqrt[n]{l_0}$, the above inequality becomes

$$n \log f_{\max} - \log(l_0) - r_1(f_{\max} - 1) \geq 0. \quad (3.10)$$

The right-hand side of inequality (3.10) is unimodal and has a unique maximum at n/r_1 . If the maximum is negative, then inequality (3.10) does not hold and hence $f_{\max} \leq n^{-1}\sqrt[n]{l_0}$. If the maximum is nonnegative, then solving inequality (3.10) gives

$$f_{\max} \leq w \equiv -\frac{nW_{-1}\left(-\frac{r_1}{n}e^{-\frac{r_1}{n}\sqrt[n]{l_0}}\right)}{r_1},$$

where $W_{-1}(z)$ is the lower branch of a Lambert W -function which is the inverse of the function $z = We^W$. Therefore,

$$f_{\max} \leq U_f \equiv \max\left\{n^{-1}\sqrt[n]{l_0}, w\right\}.$$

This also implies that f_{\min} has a lower bound, say, L_f . □

Lemma 3.2.1. For all $\varphi \in \mathcal{K}(l_0)$, $\sum_{j=1}^m \pi_j$ and $|\alpha|$ are bounded above, where $(\alpha, \pi_1, \dots, \pi_m)^T$ is the coefficient vector of φ .

Proof. Given $\varphi \in \mathcal{K}(l_0)$, from density (3.5), we have

$$\frac{e^{\alpha x_1}}{\int_{x_1}^{x_n} e^{\alpha x} dx} \leq f(x_1) \leq \frac{e^{\alpha x_1}}{\int_{x_1}^{x_2} e^{\alpha x} dx}.$$

This implies that

$$\frac{\alpha}{e^{\alpha r_2} - 1} \geq L_f \quad \text{and} \quad \frac{\alpha}{e^{\alpha r_1} - 1} \leq U_f,$$

where $r_2 = x_2 - x_1$. Solving the above inequalities gives

$$\frac{-W\left(-U_f r_1 e^{-U_f r_1}\right)}{r_1} - U_f \leq \alpha \leq \frac{-W_{-1}\left(-L_f r_2 e^{-L_f r_2}\right)}{r_2} - L_f, \quad (3.11)$$

3.2. Characterization of the Nonparametric Maximum Likelihood Estimate

where $W(z)$ is the principal branch of a Lambert W -function. Analogously, we can obtain that

$$f(x_n) \leq \frac{e^{(\alpha - \sum_{j=1}^m \pi_j)x_n}}{\int_{x_{n-1}}^{x_n} e^{(\alpha - \sum_{j=1}^m \pi_j)x} dx},$$

which gives

$$\frac{(\alpha - \sum_{j=1}^m \pi_j)e^{(\alpha - \sum_{j=1}^m \pi_j)r_3}}{e^{(\alpha - \sum_{j=1}^m \pi_j)r_3} - 1} \geq L_f,$$

where $r_3 = x_n - x_{n-1}$. Using the lower and upper bounds of α given in (3.11), we have

$$\sum_{j=1}^m \pi_j \leq \frac{-W_{-1}(-L_f r_2 e^{-L_f r_2})}{r_2} - \frac{W_{-1}(-L_f r_3 e^{-L_f r_3})}{r_3} - 2L_f.$$

□

Lemma 3.2.2. $\forall \varphi, v \in \mathcal{K}(l_0)$, it holds that

$$\begin{aligned} l(v) - l(\varphi) &\leq d(v; \varphi) - d(\varphi; \varphi) \\ &= (\alpha_v - \alpha_\varphi) d_0(\alpha; \varphi) + \int_{\theta} d(\theta; \varphi) d(v - \varphi)(\theta). \end{aligned}$$

Proof. This follows easily from the concavity of the log-likelihood function and the fact that

$$\frac{\partial^2 l(\varphi)}{\partial^2 \alpha} = -n \text{Var}_f(X) \leq 0,$$

where $f = f(x; \varphi)$ and Var_f is the variance with respect to f .

□

Theorem 3.2.1. $\hat{\varphi}$ maximizes $l(\varphi)$ if and only if

$$d_0(\alpha; \hat{\varphi}) = 0 \quad \text{and} \quad d(\theta; \hat{\varphi}) = \begin{cases} \leq 0, & \forall \theta \in \mathcal{P}, \\ = 0, & \forall \theta \in \text{supp}(\hat{\varphi}), \end{cases}$$

where $\text{supp}(\hat{\varphi})$ is the support set of $\hat{\varphi}$.

Proof. Sufficiency follows from Lemma 3.2.2.

Necessity can be established as follows. Denote by $\hat{\alpha}$ the first element of coefficient vector of $\hat{\varphi}$. For the first condition, it holds owing to the fact that $\hat{\alpha}$ is an interior point. For the second condition, $\forall \theta \in \mathcal{P}$, $\exists \varepsilon > 0$ such that $\hat{\varphi} + \varepsilon \theta \in \mathcal{K}$, it must hold that $l(\hat{\varphi} + \varepsilon \theta) - l(\hat{\varphi}) \leq 0$, since $\hat{\varphi}$ maximizes $l(\varphi)$. According to function (3.7), it apparently gives $d(\theta; \hat{\varphi}) \leq 0$. Furthermore, $\hat{\varphi}$

maximizes l , and hence

$$\int_{\theta} d(\theta; \hat{\varphi}) d\hat{\varphi}(\theta) = \hat{\alpha} d_0(\alpha; \hat{\varphi}) + \int_{\theta} d(\theta; \hat{\varphi}) d\hat{\varphi}(\theta) = d(\hat{\varphi}; \hat{\varphi}) = \lim_{\varepsilon \rightarrow 0} \frac{l(\hat{\varphi} + \varepsilon \hat{\varphi}) - l(\hat{\varphi})}{\varepsilon} = 0.$$

This means that $d(\theta; \hat{\varphi}) = 0$ at every support point of $\hat{\varphi}$. □

3.3 Computation

Let us discuss some computational aspects of a log-concave density function estimation and present a new algorithm for computing the NPMLE. Some existing algorithms are also reviewed in Section 3.3.1.

3.3.1 Main idea

As for the computational aspects, the main difficulty is how to efficiently handle the optimization problem of a log-concave function. The log-barrier algorithm (Terlaky and Vial, 1998) can be directly employed to solve this kind of problem. A logarithm penalty function and a trade-off parameter $\mu > 0$ are introduced to obtain a barrier function by freeing the constraints. For each fixed μ , the barrier function can be solved by damped Newton method. By decreasing the value of μ in a controlled way, a sequence of vector solutions can be collected, known as the central path. The optimal solution of the original problem can be found through the central path as μ is driven down to 0. This algorithm works for the general optimization problem under linear constraints, without taking into account much of the features of the problem.

Rufibach (2007) successfully employs the ICMA for this problem. The core idea is to optimize the quadratic approximation to the function under a Newton procedure by using the pool-adjacent-violators algorithm (PAVA) (Ayer et al., 1955; Robertson et al., 1988). It also supplies the line search by Hermite interpolation to improve the convergence speed of the algorithm. However, this algorithm uses all the observations as potential knots for updating the changes of slope via the PAVA. Its performance deteriorates quickly as the sample size increases.

In order to improve the speed of computation, Dümbgen et al. (2007) introduces the ASA (Fletcher, 1987) for computing the log-concave MLE. It follows a basic procedure to replace the current candidate solution with a “conditional” optimal one. In each iteration, it examines its directional derivative function, adds one new knot and updates the changes of slope by optimizing function (3.2). Repeat these steps until no new knot can be found. This algorithm is effective for solving small to medium-scale problems, but it may converge slowly when many knots are needed.

The idea of the LC algorithm for exact observations is similar to that of the ASA. The difference is that the former is more direct and includes both the univariate and multivariate steps in optimization procedure. It is more efficient for small-sized problems and takes fewer iterations to find the solution compared with the ASA. However, it requires a longer time in each iteration for a large sample size and can be slow in the situation where many knots are needed.

The key idea of our algorithm is as follows. In each iteration, it locates multiple knots, but not too many, by applying the gradient function. Knots found and added in this way may reduce the iterations taken when many knots are needed. Then a quadratically convergent method is used to update the change of slope by easing the linear constraints with only non-negativity constraints which is numerically more stable. At last, it discards the knots where there are no changes of slope. Numerical studies show that it outperforms the existing algorithms.

3.3.2 Updating masses

Let us first consider how to update β when θ is held fixed. Since the second-order Taylor series expansion of the log-likelihood function in the neighbourhood of β is applied, we give some derivatives of the log-likelihood function.

The first derivatives of the log-likelihood are

$$\begin{aligned}\frac{\partial l(x; \beta, \theta)}{\partial \alpha} &= \sum_{i=1}^n x_i - nE_f(X), \\ \frac{\partial l(x; \beta, \theta)}{\partial \pi_j} &= - \sum_{i=1}^n (x_i - \theta_j)_+ + nE_f((X - \theta_j)_+).\end{aligned}$$

where $f = f(x; \beta, \theta)$. The second derivatives are

$$\begin{aligned}\frac{\partial^2 l(x; \beta, \theta)}{\partial^2 \alpha} &= n(E_f^2(X) - E_f(X^2)), \\ \frac{\partial^2 l(x; \beta, \theta)}{\partial \alpha \partial \pi_j} &= n \left(E_f(X(X - \theta_j)_+) - E_f(X)E_f(X - \theta_j)_+ \right), \\ \frac{\partial^2 l(x; \beta, \theta)}{\partial \pi_i \partial \pi_j} &= n \left(E_f(X - \theta_i)_+ E_f(X - \theta_j)_+ - E_f((X - \theta_i)_+(X - \theta_j)_+) \right).\end{aligned}$$

Let β^* be an updating vector of β , $\eta = \beta^* - \beta$ and

$$\begin{aligned}s_i(\beta, \theta) &\equiv \frac{\partial \{\log f(x_i; \beta, \theta)\}}{\partial \beta} \\ &= (x_i - E_f(X), -(x_i - \theta_1)_+ + E_f\{(X - \theta_1)_+\}, \dots, -(x_i - \theta_m)_+ + E_f\{(X - \theta_m)_+\})^T.\end{aligned}$$

3.3. Computation

Let $\mathbf{S}^T \equiv \mathbf{S}(\boldsymbol{\beta}, \boldsymbol{\theta})^T \equiv (s_1(\boldsymbol{\beta}, \boldsymbol{\theta}), \dots, s_n(\boldsymbol{\beta}, \boldsymbol{\theta}))$ and $\mathbf{H} \equiv \mathbf{H}(\boldsymbol{\beta}, \boldsymbol{\theta})$, the Hessian matrix. Then

$$\mathbf{S}^T \mathbf{1} = \frac{\partial l(x; \boldsymbol{\beta}, \boldsymbol{\theta})}{\partial \boldsymbol{\beta}} \quad \text{and} \quad \mathbf{H} = \frac{\partial^2 l(x; \boldsymbol{\beta}, \boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T},$$

where $\mathbf{1} = (1, \dots, 1)^T$. Note that denoting by $\mathbf{e} = (e_{(0)}(x), e_{\theta_1}(x), \dots, e_{\theta_m}(x))^T$, $-\mathbf{H}$ can be expressed as

$$n \begin{pmatrix} \text{Var}_f(X) & \text{Cov}_f(X, -(X - \theta_1)_+) & \cdots & \text{Cov}_f(X, -(X - \theta_m)_+) \\ \text{Cov}_f(X, -(X - \theta_1)_+) & \text{Var}_f(-(X - \theta_1)_+) & \cdots & \text{Cov}_f(-(X - \theta_1)_+, -(X - \theta_m)_+) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}_f(X, -(X - \theta_m)_+) & \text{Cov}_f(-(X - \theta_1)_+, -(X - \theta_m)_+) & \cdots & \text{Var}_f(-(X - \theta_m)_+) \end{pmatrix} \\ = n \text{Var}(\mathbf{e})$$

where Cov_f is the covariance with respect to $f = f(x; \boldsymbol{\beta}, \boldsymbol{\theta})$. Since $-\mathbf{H}$ is a symmetric, positive semi-definite matrix, let $-\mathbf{H} = \mathbf{R}^T \mathbf{R}$, where \mathbf{R} is some square matrix. Expanding $l(\boldsymbol{\beta}^*, \boldsymbol{\theta})$ in the Taylor series about $\boldsymbol{\beta}$ to the second order gives

$$\begin{aligned} l(\boldsymbol{\beta}, \boldsymbol{\theta}) - l(\boldsymbol{\beta}^*, \boldsymbol{\theta}) &\approx -\mathbf{1}^T \mathbf{S} \boldsymbol{\eta} - \frac{1}{2} \boldsymbol{\eta}^T \mathbf{H} \boldsymbol{\eta} \\ &= \frac{1}{2} \|\mathbf{R} \boldsymbol{\beta}^* - \mathbf{R} \boldsymbol{\beta} - \mathbf{R}^{-T} \mathbf{S}^T \mathbf{1}\|^2 - \frac{1}{2} \mathbf{1}^T \mathbf{S} \mathbf{R}^{-1} \mathbf{R}^{-T} \mathbf{S}^T \mathbf{1}, \end{aligned}$$

where $\|\cdot\|$ denotes the L_2 -norm. Maximizing $l(\boldsymbol{\beta}^*, \boldsymbol{\theta})$ in the neighbourhood of $\boldsymbol{\beta}$ can be replaced approximately with the following linear regression problem:

$$\min_{\boldsymbol{\beta}^*} \|\mathbf{R} \boldsymbol{\beta}^* - \mathbf{R} \boldsymbol{\beta} - \mathbf{R}^{-T} \mathbf{S}^T \mathbf{1}\|^2, \quad \text{subject to } \boldsymbol{\pi}^* \geq 0. \quad (3.12)$$

This is precisely the type of problem for which the NNLS algorithm was developed; see [Lawson and Hanson \(1974\)](#). In our computation, the first element α^* in $\boldsymbol{\beta}^*$ is not restricted, and this can be solved by the function `pnnls()` in the R package `lse1` ([Wang et al., 2015](#)). Note that the log-likelihood function is approximated by a quadratic function, which is optimized by solving problem (3.12).

Numerically, \mathbf{H} may become singular, or negative semidefinite. Therefore, in our implementation its spectral decomposition

$$\mathbf{H} = -\mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^T$$

is considered, where Λ is diagonal and contains only the positive eigenvalues of $-\mathbf{H}$, and \mathbf{U} has the corresponding eigenvectors as its columns. Note that zero eigenvalues are excluded from Λ and thus \mathbf{U} may not be a square matrix. We hence use $\mathbf{R} = \Lambda^{\frac{1}{2}}\mathbf{U}^T$, and the Moore-Penrose generalized inverse $\mathbf{U}\Lambda^{-\frac{1}{2}}$ for the inverse \mathbf{R}^{-1} . An eigenvalue is treated as zero, if it is less than the largest eigenvalue times a small threshold value, say, 10^{-15} .

Denoting by $\beta + \eta$ the constrained solution to problem (3.12), we complement the line search by Armijo's rule to ensure a monotonic increase of the log-likelihood after each iteration. The inequality

$$l(\beta + \sigma^k \eta, \theta) \geq l(\beta, \theta) + \rho \sigma^k \nabla l(\beta, \theta)^T \eta, \quad 0 < \rho < \frac{1}{2}, \quad (3.13)$$

is examined to find the smallest satisfying integer k . Then the vector $\beta + \sigma^k \eta$ becomes to be the new β in the next iteration. Wang (2010) suggests using $\rho = \frac{1}{3}$ and $\sigma = \frac{1}{2}$.

3.3.3 Algorithm

In this section, we present how to compute the NPMLE of a log-concave function and give details of our algorithm. It is an extension of the CNM algorithm (Wang, 2007) for computing the NPMLE of a mixing distribution. One critical component of our new algorithm is how to expand and reduce the knot set θ . The gradient function $d(\theta; \varphi)$ is introduced here for finding the new knots.

As for expanding the knot set, the most important is to choose new knots. Let us consider an estimate φ with m knots and divide $\mathcal{X} = [x_1, x_n]$ into $m + 1$ disjoint intervals by using the m knots as breakpoints. In each interval, find the θ' that has the largest gradient value and add it to the set of knots θ , if it is not already in the set. After adding the new knots, we update the mass vector by solving problem (3.12), and then reduce the knot set by removing the elements of θ with zero proportion. Our algorithm for log-concave density estimation is described as follows.

Algorithm 1 (CNMLCD). Set $s = 0$. From an initial estimate φ_0 with a finite number of knots and $l(\varphi_0) > -\infty$, repeat the following steps.

1. Compute $d(\theta; \varphi_s)$ for all $\theta \in \mathcal{P}$ and, as described above, find new knots $\theta'_{s1}, \dots, \theta'_{sp}$.
2. Set $\theta_s^+ = (\theta_s^T, \theta'_{s1}, \dots, \theta'_{sp})^T$ and $\beta_s^+ = (\beta_s^T, 0, \dots, 0)^T$. Compute β_{s+1}^- by solving problem (3.12) and conducting line search (3.13).
3. Discard all knots with zero changes of slope, which gives β_{s+1} and θ_{s+1} of φ_{s+1} . Stop if $l(\varphi_{s+1}) - l(\varphi_s) \leq \text{tolerance}$. Set $s = s + 1$.

3.4 Convergence

The consistency of the nonparametric maximum likelihood estimator of a log-concave density function has been established by [Pal et al. \(2007\)](#) and [Dümbgen and Rufibach \(2009\)](#). In the later paper, the efficiency of the estimator has also been investigated. However, Algorithm 1 is developed based on the idea of CNM, which is different from that of the ASA. Hence the establishment of its convergence below has some similarity to the theoretical work given in [Wang \(2007\)](#). A particular difficulty here is that the total mass $\sum_{j=1}^m \pi_j$ is not restricted to 1, unlike the mixture distribution case.

In each iteration, when updating β and θ of φ to β^* and θ^* of φ^* , suppose $\theta'_1, \dots, \theta'_p$ are the new knots found in step 1, and denote $\theta^+ = (\theta^T, \theta'_1, \dots, \theta'_p)^T$ and $\beta^+ = (\beta^T, 0, \dots, 0)^T$. Let β^{*-} be the updated vector for the changes of slopes for θ^+ after step 2. Discarding the knots from θ^+ with zero masses in β^{*-} in step 3 gives θ^* , with mass vector β^* . In addition, let $\eta = \beta^{*-} - \beta^+$, $S^+ = S(\beta^+, \theta^+)$ and $R^+ = R(\beta^+, \theta^+)$.

Lemma 3.4.1. $\eta^T (R^+)^T R^+ \eta$ is bounded above for all $\varphi \in \mathcal{K}(l_0)$ and all directions η which satisfies that $\varphi' \in \mathcal{K}(l_0)$, where φ' has $(\eta + \beta^+, \theta^+)$.

Proof. For any $\varphi \in \mathcal{K}(l_0)$, denote

$$\begin{aligned} e &= (x, -(x - \theta_1^+)_+, \dots, -(x - \theta_m^+)_+)^T, \\ \nu &= E_{f^+}(e) = (E_{f^+}(X), -E_{f^+}\{(X - \theta_1^+)_+\}, \dots, -E_{f^+}\{(X - \theta_m^+)_+\})^T, \end{aligned}$$

where $f^+ = f(x; \beta^+, \theta^+)$. Note that $(R^+)^T R^+ = n\{-\nu\nu^T + E_{f^+}(ee^T)\}$. Hence we have

$$\begin{aligned} \eta^T (R^+)^T R^+ \eta &= n \left[-E_{f^+}^2(e^T \eta) + E_{f^+} \{(e^T \eta)^2\} \right] \\ &= n \text{Var}_{f^+} [\varphi(X; \beta^{*-}, \theta^+) - \varphi(X; \beta^+, \theta^+)] \\ &= n \text{Var}_{f^+} \left[\log \left\{ \frac{f(X; \beta^{*-}, \theta^+)}{f(X; \beta^+, \theta^+)} \right\} \right] \\ &\leq n \left[\log \left\{ \frac{\sup_{\psi \in \mathcal{K}(l_0), x \in \mathcal{X}} f(x; \psi) \sup_{x \in \mathcal{X}} f(x; \beta^+, \theta^+)}{\inf_{\psi \in \mathcal{K}(l_0), x \in \mathcal{X}} f(x; \psi) \inf_{x \in \mathcal{X}} f(x; \beta^+, \theta^+)} \right\} \right]^2. \end{aligned} \quad (3.14)$$

It is thus bounded above, which completes the proof. \square

Lemma 3.4.2. The Armijo search that is used always succeeds with in a finite number of steps independent of s .

Proof. To minimize $\|\mathbf{R}^+\boldsymbol{\eta} - (\mathbf{R}^+)^{-T}(\mathbf{S}^+)^T\mathbf{1}\|^2$ under the constraint, let $\boldsymbol{\eta} = t\mathbf{d}$, where \mathbf{d} the unit vector in the same direction as $\boldsymbol{\eta}$. In this direction, the constrained minimum point must be located at or before the unconstrained minimum point which leads to $t \leq \mathbf{1}^T \mathbf{S}^+ \mathbf{d} / (\mathbf{d}^T (\mathbf{R}^+)^T \mathbf{R}^+ \mathbf{d})$. Substituting \mathbf{d} by $\boldsymbol{\eta}/t$ gives

$$\boldsymbol{\eta}^T (\mathbf{R}^+)^T \mathbf{R}^+ \boldsymbol{\eta} \leq \mathbf{1}^T \mathbf{S}^+ \boldsymbol{\eta}.$$

For any $0 < \rho < \frac{1}{2}$, there is a $\lambda > 0$ such that if when $\|\mathbf{R}^+\boldsymbol{\eta}\| \leq \lambda$, we have

$$\begin{aligned} l(\boldsymbol{\beta}^+ + \boldsymbol{\eta}, \boldsymbol{\theta}^+) - l(\boldsymbol{\beta}, \boldsymbol{\theta}^+) &= l(\boldsymbol{\beta}^+ + \boldsymbol{\eta}, \boldsymbol{\theta}^+) - l(\boldsymbol{\beta}^+, \boldsymbol{\theta}^+) \\ &= \mathbf{1}^T \mathbf{S}^+ \boldsymbol{\eta} - \frac{1}{2} \boldsymbol{\eta}^T (\mathbf{R}^+)^T \mathbf{R}^+ \boldsymbol{\eta} + o(\|\mathbf{R}^+\boldsymbol{\eta}\|^2) \\ &\geq \mathbf{1}^T \mathbf{S}^+ \boldsymbol{\eta} - \frac{1}{2} \boldsymbol{\eta}^T (\mathbf{R}^+)^T \mathbf{R}^+ \boldsymbol{\eta} - \left(\frac{1}{2} - \rho\right) \boldsymbol{\eta}^T (\mathbf{R}^+)^T \mathbf{R}^+ \boldsymbol{\eta} \\ &\geq \rho \mathbf{1}^T \mathbf{S}^+ \boldsymbol{\eta}. \end{aligned}$$

The Armijo rule is satisfied.

If $\|\mathbf{R}^+\boldsymbol{\eta}\| > \lambda$, then $\|\sigma^k \mathbf{R}^+\boldsymbol{\eta}\| \leq \lambda$ can be satisfied by $\bar{k} = \max\left(\left\lceil \log_{\sigma}\left(\frac{\lambda}{\sqrt{U}}\right) \right\rceil, 0\right)$, since $\|\mathbf{R}^+\boldsymbol{\eta}\| \leq \sqrt{U}$, where U is the upper bound on the right-hand side of inequality (3.14). Hence \bar{k} is an upper bound on the number of steps for Armijo's rule to be satisfied in all situations. \square

Theorem 3.4.1. Suppose that φ_s is a sequence created by Algorithm 1 and $\hat{\varphi}$ is the NPMLE. Then $l(\varphi_s) \rightarrow l(\hat{\varphi})$ monotonically as $s \rightarrow \infty$.

Proof. In Algorithm 1, the sequence $\{l(\varphi_s)\}$ is a bounded and increasing sequence. According to the monotone convergence theorem, the least upper bound of the set $\{l(\varphi_s) : s \in \mathbb{N}\}$ is the limit of $\{l(\varphi_s)\}$. From the proof in Lemma 3.4.2, with the non-negative definiteness of $(\mathbf{R}_s^+)^T \mathbf{R}_s^+$, we have

$$\begin{aligned} l(\varphi_{s+1}) - l(\varphi_s) &\geq \rho \sigma^{\bar{k}} \mathbf{1}^T \mathbf{S}_s^+ \boldsymbol{\eta}_s \\ &\geq \rho \sigma^{\bar{k}} \left(\mathbf{1}^T \mathbf{S}_s^+ \boldsymbol{\eta}_s - \frac{1}{2} \boldsymbol{\eta}_s^T (\mathbf{R}_s^+)^T \mathbf{R}_s^+ \boldsymbol{\eta}_s \right). \end{aligned} \quad (3.15)$$

Denote by $\boldsymbol{\eta}_{sj}$ the direction from $\boldsymbol{\beta}_s^+$ to \mathbf{e}_j , a vector whose component is only changed ΔJ at the j th component, where $|\Delta J| = 1$. For any $\epsilon \in [0, 1]$, owing to the optimality of $\boldsymbol{\eta}_s$, it holds that

$$\|\mathbf{R}_s^+ \boldsymbol{\eta}_s - (\mathbf{R}_s^+)^{-T} (\mathbf{S}_s^+)^T \mathbf{1}\|^2 \leq \|\epsilon \mathbf{R}_s^+ \boldsymbol{\eta}_{sj} - (\mathbf{R}_s^+)^{-T} (\mathbf{S}_s^+)^T \mathbf{1}\|^2.$$

3.4. Convergence

Based on inequality (3.15) and Lemma 3.4.1, expanding both sides gives

$$\begin{aligned} l(\varphi_{s+1}) - l(\varphi_s) &\geq \rho\sigma^{\bar{k}}\epsilon(\mathbf{1}^T \mathbf{S}_s^+ \boldsymbol{\eta}_{sj} - \frac{\epsilon}{2} \boldsymbol{\eta}_{sj}^T (\mathbf{R}_s^+)^T \mathbf{R}_s^+ \boldsymbol{\eta}_{sj}) \\ &\geq \rho\sigma^{\bar{k}}\epsilon(\mathbf{1}^T \mathbf{S}_s^+ \boldsymbol{\eta}_{sj} - \frac{\epsilon}{2} U). \end{aligned} \quad (3.16)$$

Let $g(\varphi_s) = \max \left\{ |d_\alpha(\alpha; \varphi_s)|, |d(\theta_{\varphi_s}; \varphi_s)|, \sup_{\theta} d(\theta; \varphi_s) \right\}$, where θ_{φ_s} is any knot in φ_s . Now assume that $g(\varphi_s)$ does not approach 0 as $s \rightarrow \infty$. Consider the following three situations.

(1) $\lim_{s \rightarrow +\infty} |d_\alpha(\alpha; \varphi_s)| \geq \tau > 0$.

Note that $\mathbf{1}^T \mathbf{S}_s^+ \boldsymbol{\eta}_{s1} = \Delta J d_\alpha(\alpha; \varphi_s)$, then if $d_\alpha(\alpha; \varphi_s) \geq \tau$, let $\Delta J = 1$; or if $d_\alpha(\alpha; \varphi_s) \leq -\tau$, let $\Delta J = -1$, then inequality (3.16) becomes

$$l(\varphi_{s+1}) - l(\varphi_s) \geq \rho\sigma^{\bar{k}}\epsilon(\tau - \frac{\epsilon}{2} U) \quad (3.17)$$

(2) $\lim_{s \rightarrow +\infty} |d(\theta_{\varphi_s}; \varphi_s)| \geq \tau > 0$.

Analogously, we can obtain inequality (3.17)

(3) $\lim_{s \rightarrow +\infty} \sup_{\theta} d(\theta; \varphi_s) \geq \tau > 0$.

Let $j (> 1)$ in inequality (3.16) correspond to the θ that maximizes $d(\theta; \varphi_s)$, which is in θ_s^+ owing to step 1 of Algorithm 1. Let $\Delta J = 1$, using the fact that

$$\mathbf{1}^T \mathbf{S}_s^+ \boldsymbol{\eta}_{sj} = d(\theta; \varphi_s),$$

inequality (3.17) can be also obtained.

Since inequality (3.17) holds in all of the three situations above, without loss of generality choose $\epsilon = \tau/(\tau + U)$, then it holds that

$$l(\varphi_{s+1}) - l(\varphi_s) \geq \frac{\rho\sigma^{\bar{k}}\tau^2(2\tau + U)}{2(\tau + U)^2}. \quad (3.18)$$

The right-hand side of inequality (3.18) is a positive value that is independent of s , showing that the sequence $\{l(\varphi_s)\}$ is divergent, which contradicts the fact of its convergence. Hence $g(\varphi_s)$ approaches 0 as $s \rightarrow \infty$. From Lemma (3.2.2), we know that

$$l(\hat{\varphi}) - l(\varphi_s) \leq g(\varphi_s) \left(|\hat{\alpha} - \alpha_s| + \sum_{j=1}^{\hat{m}} \hat{\pi}_j + \sum_{j=1}^m \pi_{sj} \right). \quad (3.19)$$

Therefore, Lemma 3.2.1 implies that $l(\varphi_{s+1}) \rightarrow l(\hat{\varphi})$ as $s \rightarrow \infty$, which completes the proof. \square

3.5 Numerical Studies

3.5.1 Setup

To compare with the proposed CNMLCD algorithm, four other algorithms are included: the log-barrier algorithm of [Terlaky and Vial \(1998\)](#) and [Rufibach \(2007\)](#), the modified ICMA of [Rufibach \(2007\)](#) which is obtained by function `icmaLogCon()` in the R package `logcondens` ([Dümbgen and Rufibach, 2011](#)), the active set algorithm (ASA) of [Dümbgen et al. \(2007\)](#) which is also available in the R package `logcondens` by function `logConDens()`, and the LC algorithm of [Anderson-Bergman \(2014\)](#) which is available in the R package `logconPH`, by function `logconcave()` with `covariates` left blank.

According to our experience, `logConDens()` may return estimates of low precision owing to its termination criterion. In order to give a reasonably fair comparison, we terminate the internal function `activeSetLogCon()` function only if $l_s - l_{s-1} \leq \epsilon_1$, where ϵ_1 varies with sample sizes, l_s being the log-likelihood in step s . The precision tolerance in the function `MLE()` is reset to $\epsilon_2 = 10^{-15}$. The other arguments used in the package `logcondens` take their default values. The arguments for Log-barrier and ICMA are given in [Table 3.1](#), whose meanings are described in [Rufibach \(2007\)](#). The arguments in `logconcave()` also take their default values.

Tab. 3.1: Settings for the algorithms.

Algorithm	Parameters				
Log-barrier	τ	θ	μ	T1	T2
	0.9	0.1	0.1	8000	20
ICMA				T1	T2
				8000	20
ASA	ϵ_2				
	10^{-15}				

All computations were carried out in R (version 3.1.1) ([R Core Team, 2015](#)) on a computer with a 3.4 GHz Intel Core i5-3570 central processing unit. Each algorithm was started with the initial density being uniform on $[x_1, x_n]$. Our implementation of the CNMLCD algorithm is available in the R package `cnmlcd` ([Liu and Wang, 2015](#)), by function `cnmlcd()`.

3.5.2 Simulation studies

The simulation studies were carried out to investigate the performance of the CNMLCD algorithm in different scenarios. To be able to compare the performance of the algorithms, we repeated the computation for 10 random samples drawn in each case, and each case has a sample size $n = 100, 1000, 10000$ or 100000 , and each sample is drawn from one of three distributions: Normal(0, 1), Gamma(2, 1) and Beta(5, 2). The value of ϵ_1 is set to $10^{-8}, 10^{-7}, 10^{-6}, 10^{-5}$ corresponding to each sample size. In each case, we always first executed the CNMLCD algorithm to achieve a maximum log-likelihood value as large as numerically possible, which is denoted by \hat{l} . Then we executed the ASA, and the maximum log-likelihood value it achieved is used as the critical value to terminate the other methods when executed on the same data set, unless the maximum number of iterations, 8000, is reached first. However, the LC algorithm is terminated by its own stopping criterion. Their numbers of iterations and CPU times were recorded.

In Table 3.2, we report for each algorithm the mean number of knots (m), the mean number of iterations taken (s), the mean (t) and standard deviation ($sd(t)$) of CPU times, and the mean log-likelihood difference from the NPMLE ($\hat{l} - l_s$). Figure 3.1 also shows the mean CPU times taken by the algorithms for Gamma(2, 1) with different sample size. The log-barrier algorithm failed to achieve the critical log-likelihood value, even when the sample size is small ($n = 100$). The ICMA also performed poorly when $n = 100$, and it took far longer time than the LC, CNMLCD and ASA. For a larger sample size (≥ 1000), we had to exclude the Log-barrier and ICMA for their poor performance. As the sample size increases, the number of knots needed increases and the speed of the LC algorithm slows down much more quickly than the ASA and CNMLCD, even though it takes fewer iterations than ASA, and we had to also exclude the LC algorithm for $n \geq 10,000$. As the sample size increases and so does the knots needed, the ASA algorithm required increasingly more iterations and longer times than the CNMLCD algorithm, which is consistently the fastest in all cases studied.

In Figure 3.2, histograms and log-concave density estimates for a typical sample of size 1000 are plotted in the left panels, with knots designated by solid points, and the corresponding gradient curves are shown in the right panels. It can be seen that the convergence conditions specified in Theorem 3.2.1 are well satisfied.

3.5.3 Real-world data

To further demonstrate our algorithm, we also apply it to analyse real-world data sets.

Reliability data

Tab. 3.2: Performance of algorithms in simulation studies

n	Law	Algorithm	m	s	t	sd(t)	$\hat{l} - l_s$
100	Gamma	Log-barrier	100.0	8000.0	230.910	215.189	8.85×10^0
		ICMA	5.0	7615.0	16.810	4.010	9.89×10^{-3}
		LC	4.9	6.1	0.012	0.011	2.33×10^{-12}
		ASA	4.9	7.9	0.019	0.009	3.41×10^{-14}
		CNMLCD	4.9	8.6	0.006	0.008	1.14×10^{-14}
	Beta	LC	6.0	6.5	0.017	0.008	1.79×10^{-12}
		ASA	6.0	9.3	0.022	0.013	4.68×10^{-13}
		CNMLCD	6.0	9.5	0.007	0.009	2.76×10^{-13}
	Normal	LC	6.2	7.6	0.023	0.007	5.25×10^{-12}
ASA		6.2	10.3	0.022	0.009	5.83×10^{-14}	
CNMLCD		6.2	9.5	0.007	0.010	2.42×10^{-14}	
1,000	Gamma	LC	9.0	11.5	0.355	0.150	2.09×10^{-11}
		ASA	9.0	17.7	0.072	0.012	2.09×10^{-12}
		CNMLCD	9.0	11.0	0.013	0.009	1.73×10^{-12}
	Beta	LC	10.3	14.6	0.494	0.122	7.57×10^{-5}
		ASA	10.4	21.0	0.090	0.026	8.91×10^{-12}
		CNMLCD	10.4	13.2	0.019	0.012	6.38×10^{-12}
	Normal	LC	10.7	14.9	0.544	0.184	2.57×10^{-12}
		ASA	10.7	21.0	0.086	0.010	1.61×10^{-12}
		CNMLCD	10.7	11.7	0.018	0.009	6.14×10^{-13}
10,000	Gamma	ASA	15.0	36.0	0.695	0.058	1.58×10^{-6}
		CNMLCD	15.1	13.0	0.152	0.025	1.10×10^{-7}
	Beta	ASA	18.0	42.0	0.823	0.084	1.47×10^{-6}
		CNMLCD	17.9	17.0	0.222	0.050	3.13×10^{-7}
	Normal	ASA	17.3	40.2	0.752	0.100	8.24×10^{-6}
		CNMLCD	17.3	14.2	0.177	0.025	1.53×10^{-6}
100,000	Gamma	ASA	25.9	56.0	8.488	1.020	7.46×10^{-3}
		CNMLCD	26.8	11.9	2.241	0.338	2.12×10^{-3}
	Beta	ASA	30.1	69.3	10.919	0.859	1.98×10^{-3}
		CNMLCD	30.2	18.3	3.773	0.677	7.00×10^{-4}
	Normal	ASA	30.0	62.8	9.383	1.018	2.49×10^{-4}
		CNMLCD	30.3	15.7	3.128	0.348	9.77×10^{-5}

The first is the reliability data from [Dümbgen and Rufibach \(2011\)](#), which has 786 observations. The reliability data was collected as part of a consulting project at the Institute for Mathematical Statistics and Actuarial Science at the University of Bern ([Dümbgen and Rufibach, 2009](#)). In the project, a company asked the Monte Carlo experiments to predict the reliability of a certain device they provided. The reliability depends in a certain deterministic way on five different and independent random input parameters. For each input parameter, a sample is available and the goal

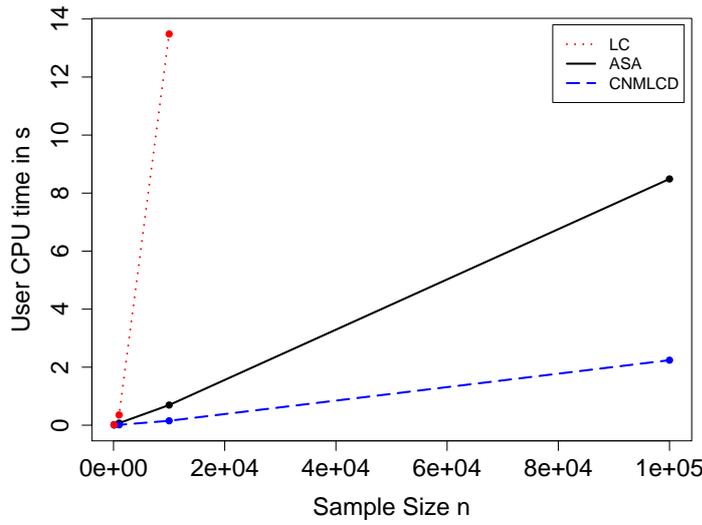


Fig. 3.1: CPU times for Gamma(2, 1) with different sample sizes.

is to find a suitable distribution. The data here only contains the first of these input parameters. [Dümbgen and Rufibach \(2009\)](#) also applied a smoothing estimator of a log-concave function to this data. The piecewise linear log-density function for this data is given by

$$\begin{aligned} \hat{\varphi}(x) = & 0.046x - 0.032(x - 1483.36)_+ - 0.006(x - 1601.19)_+ - 0.001(x - 1630.10)_+ - \\ & 0.006(x - 1685.12)_+ - 0.014(x - 1721.52)_+ - 0.007(x - 1777.78)_+ - \\ & 0.011(x - 1785.12)_+ - 0.021(x - 1804.76)_+ - 84373.01. \end{aligned}$$

Figure 3.3 provides plots of the density estimates, their logarithms and the gradient curves after the zeroth, second and final iteration, respectively. The solid points are the knots found by CNMLCD. Histograms and log-concave density estimates found by CNMLCD are given in the top panels, and the corresponding log density plots and gradient curves are given in the middle row and bottom panels, respectively. The gradient curve after the final iteration shows that CNMLCD has found the NPMLE of the log-concave function, according to Theorem 3.2.1.

Financial data

The other two data sets are the 252 daily log returns $\log(P_{i+1}/P_i)$ (from 03/01/2011 to 03/01/2012), where P_i is the close price of day i of S&P 500 index, and the 4786 daily log volatilities $\log(V_i)$ (from 03/01/1995 to 03/01/2014) where V_i is the VIX index close levels of day i , which were retrieved from Yahoo Finance. It is very common in the literature that the stock price is assumed to follow a random walk in continuous time. The renowned Black-Scholes model assumes that the

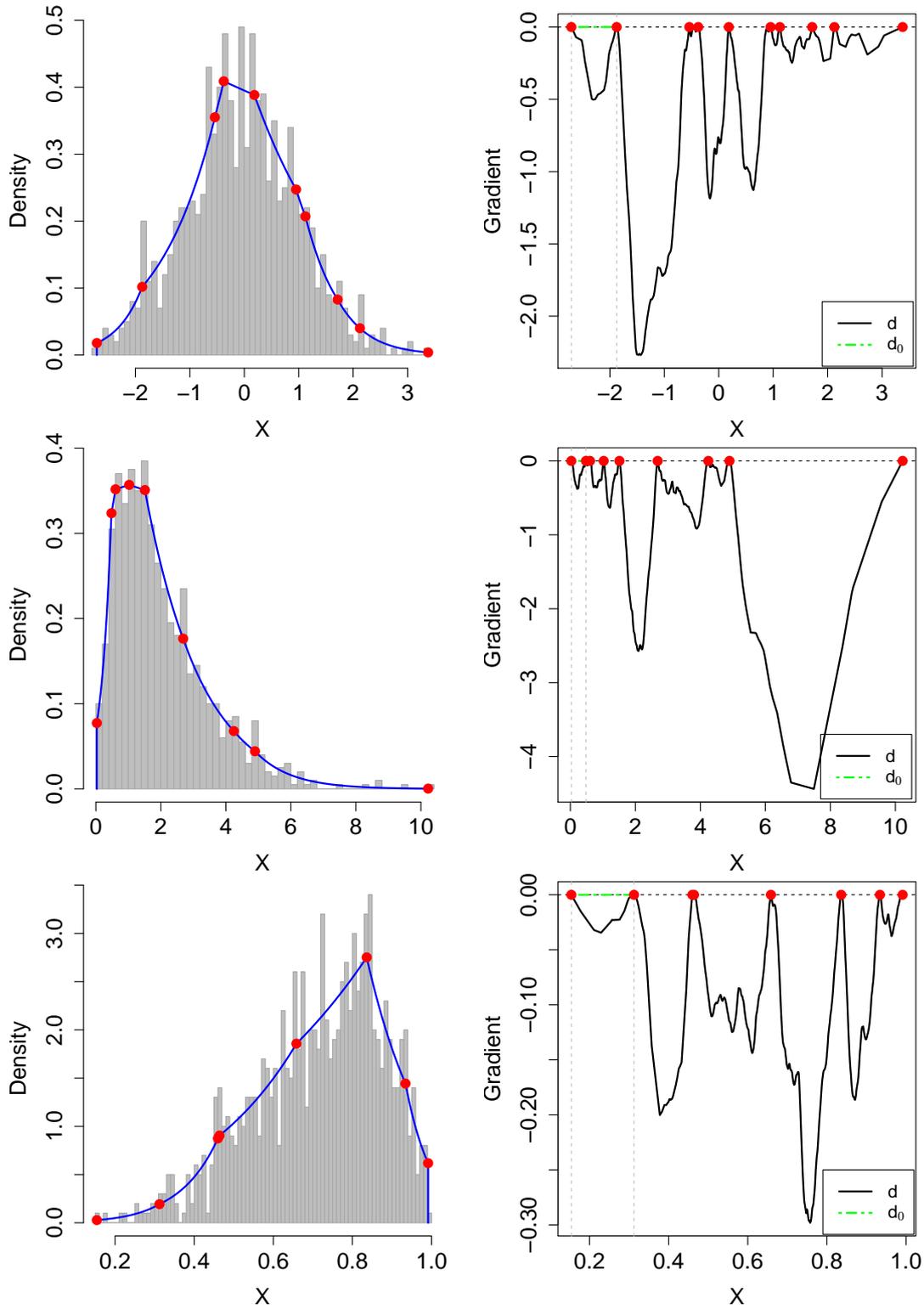


Fig. 3.2: In the left panels, histograms and log-concave density estimates from typical samples simulated from, respectively, Normal(0, 1) (top panels), Gamma(2, 1) (middle panels) and Beta(5, 2) (bottom panels). Right panels show the corresponding gradient curves. Knots are shown by solid points.

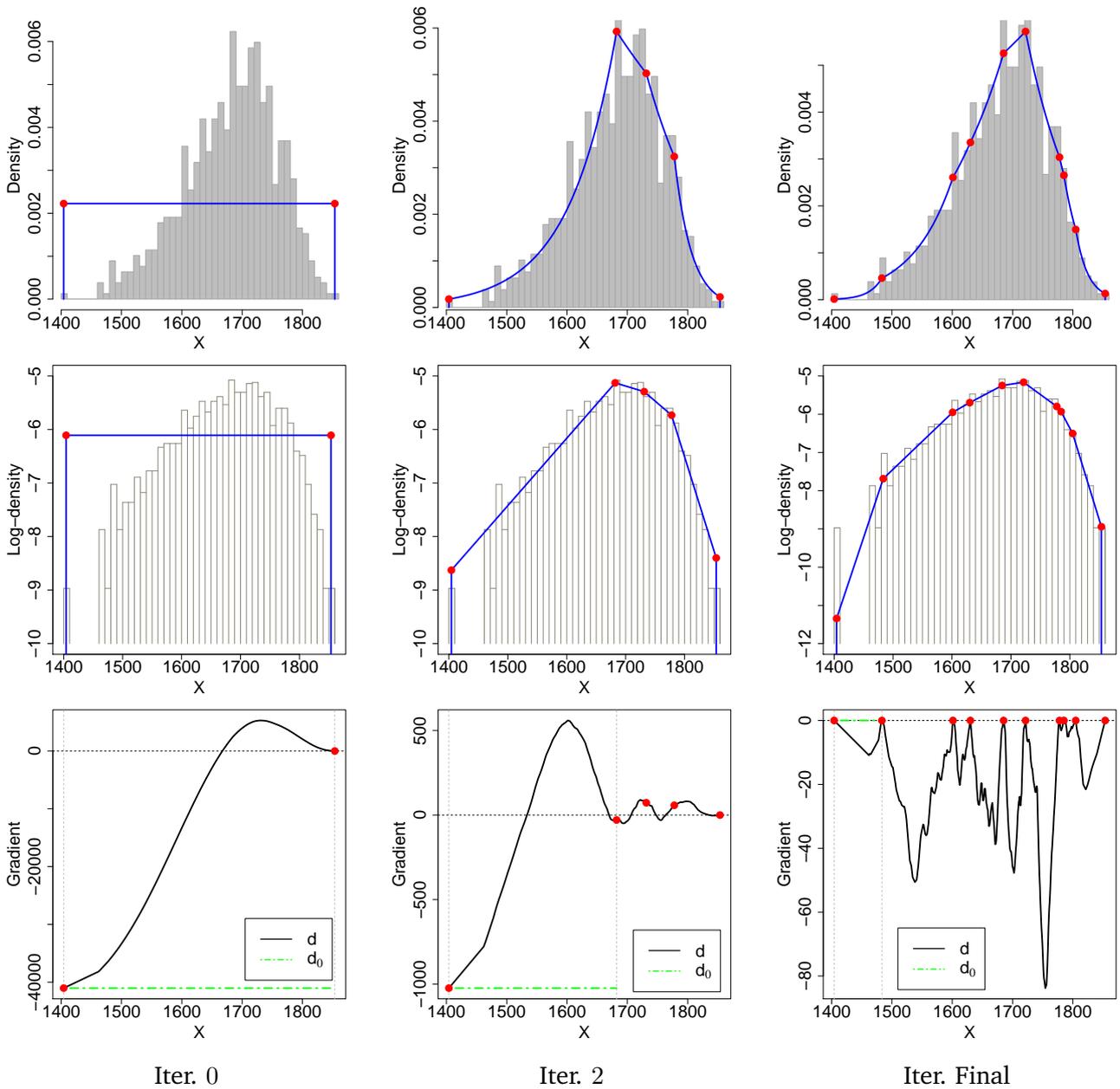


Fig. 3.3: Histogram and density estimates (top panels), log density plots (middle row panels) with gradient curves (bottom panels) for the reliability data correspond to the zeroth, second and final iteration found by CNMLCD, with knots indicated by solid points.

underlying asset price process follows a geometric Brownian motion and therefore, the risk-neutral distribution is a log-normal distribution. That is, the log return follows a normal distribution. [Chen \(2014\)](#) proposed semiparametric time series models with log-concave innovations. Moreover, the S&P 500 index is considered one of the best indicators of the state of the market and economy ([Markowitz and Usmen, 1996](#)). The VIX calculation measures 30-day expected volatility of the S&P 500 Index and is quoted as an annualized standard deviation ([Chicago Board Options Ex-](#)

change, 2009). More and more derivatives are explicitly sensitive to future (both implied and instantaneous) volatility levels. By assuming that the volatility of the underlying price is a stochastic process, it becomes possible to model derivatives more accurately. Cizeau et al. (1997) also pointed out that the volatility distribution can be very well described by a log-normal function.

For the daily log-returns of S&P 500, the piecewise linear log-density function is given by

$$\hat{\varphi}(x) = 88.176x - 28.682(x + 0.001)_+ - 169.006(x - 0.003)_+ - 9.574.$$

The density, the log-density and the gradient curve after the zeroth, second and final iterations are shown in Figure 3.4. One can see from the final iteration that the convergence conditions are also achieved.

For the daily log volatilities data, the piecewise linear log-density function is given by

$$\begin{aligned} \hat{\varphi}(x) = & 49.29x - 31.59(x - 2.33)_+ - 9.57(x - 2.41)_+ - 5.83(x - 2.44)_+ - 1.44(x - 2.50)_+ - \\ & 0.13(x - 3.03)_+ - 1.67(x - 3.07)_+ - 3.01(x - 3.19)_+ - 3.93(x - 4.25)_+ - 175.75. \end{aligned}$$

The density, the log density and the gradient curve after the zeroth, third and final iterations are shown in Figure 3.5. The convergence conditions are satisfied by the gradient curve after the final iteration.

Comparisons

We have demonstrated that CNMLCD can well find the NPMLE of a log-concave density function. When considering the efficiency, we also compare the different algorithms for finding the NPMLE of a log-concave density function. Only the CNMLCD and ASA were compared for these real-world data sets. As shown in Table 3.3, ASA and CNMLCD terminated after a similar number of iterations of the first two data sets, but the CPU time of CNMLCD is about one-fourth that of ASA. For the third data set that has a large sample size, the ASA algorithm took 6.4 times the time needed by CNMLCD.

3.6 Summary

A new algorithm CNMLCD has been proposed for computing the NPMLE of a log-concave density in this chapter. Being an extension of the CNM method that was proposed for computing the NPMLE of a mixing distribution, the new algorithm maximizes the log-likelihood function by modifying the set of knots using the gradient function and updating the changes of slope through a quadratically

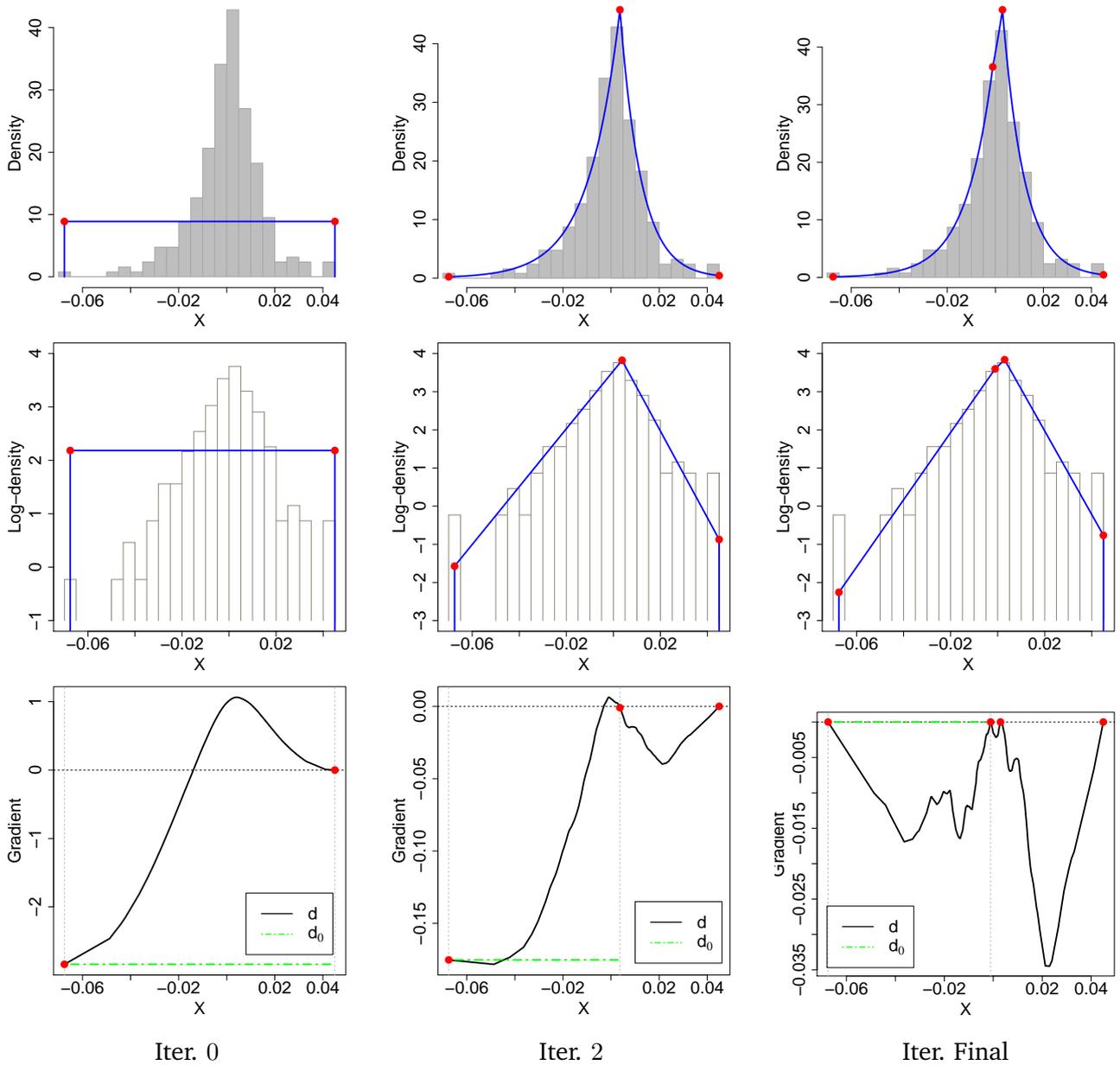


Fig. 3.4: Histogram and density plots (top panels), log density plots (middle row panels) with gradient curves (bottom panels) for the daily log returns of S&P 500 index correspond to the zeroth, second and final iteration found by CNMLCD, with knots indicated by solid points.

convergent method. The characterization of the NPMLE has been studied and the convergence of this new algorithm is also theoretically established. It is numerically compared with four algorithms available in the literature, Log-barrier, ICMA, ASA and LC algorithms, and clearly outperforms the other four in terms of computing time. Its outperformance is more significant for large-sized samples.

For future work, there can be several directions. It seems fairly straightforward to extend it to

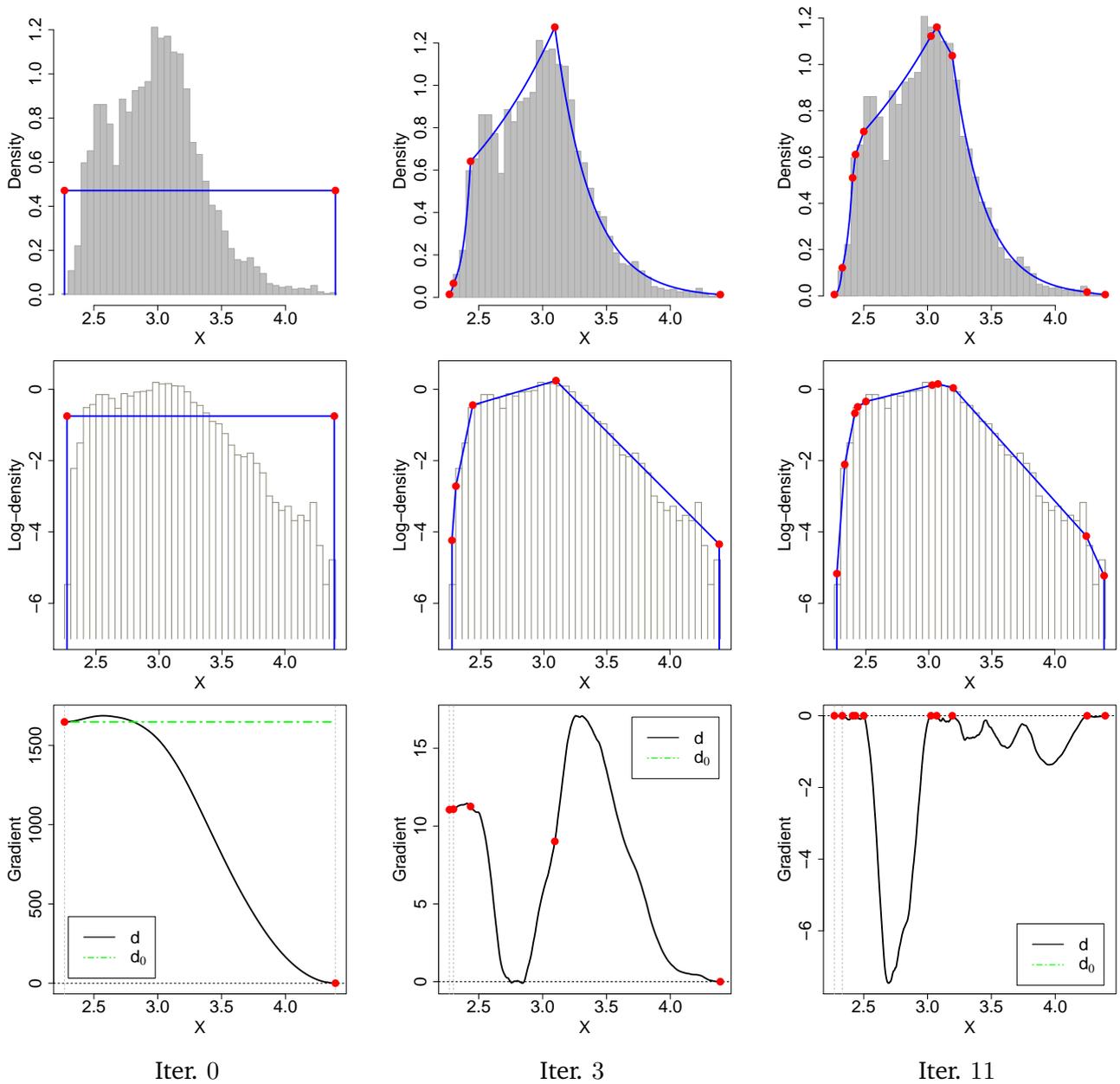


Fig. 3.5: Histogram and density plots (top panels), log density plots (middle row panels) with gradient curves (bottom panels) for the daily log volatilities correspond to the zeroth, third and final iteration found by CNMLCD, with knots indicated by solid points.

estimation under other shape constraints, such as convexity (An, 1998). It may also be extensible to maximum likelihood estimation of multidimensional log-concave density (Cule et al., 2010). Considering the fast computation, a new test of the log-concavity may be built based on the CNMLCD. Furthermore, how to add smoothness to log-concave density estimation in some controlled manner is also a very interesting topic. In order to achieve a smooth version estimator, one can generalize the NPMLE approach by applying a higher-order smoothness of an estimator. We undertake

Tab. 3.3: Results of the real-world data sets.

n	Data set	Algorithm	m	s	t	$\hat{l} - l_s$
252	Log-return	ASA	4	9	0.021	2.274×10^{-13}
		CNMLCD	4	10	0.006	0
786	Reliability	ASA	10	17	0.061	5.457×10^{-12}
		CNMLCD	10	12	0.014	2.728×10^{-12}
4786	Log-volatility	ASA	10	25	0.154	9.800×10^{-11}
		CNMLCD	10	10	0.024	6.685×10^{-11}

detailed studies of some of these extensions in the subsequent chapters.

Chapter 4

Nonparametric Smooth Estimation of a Log-concave Density function

4.1 Introduction

Density estimation under shape restrictions has drawn an increasing research attention in the past decade or so. Such methods offer a great advantage that the density estimator does not require a tuning parameter, the value of which is often difficult to properly determine, as in, say, kernel density estimation. Shape-restricted density estimation can be traced back to [Grenander \(1956\)](#), who studied maximum likelihood estimation of a non-increasing density with support $[0, \infty)$. The Grenander estimator can easily be extended to estimating a unimodal density when the mode is known, where one just needs to restrict the density to be non-decreasing (or non-increasing) to the left (or right) of the mode. If the mode is unknown, then its estimation is more tricky and the mode can be estimated consistently, through a penalty ([Woodroffe and Sun, 1993](#)) or a shape restriction ([Meyer and Woodroffe, 2004](#)).

Estimation of a log-concave density f , i.e., $\log(f)$ being concave, has been studied by [Pal et al. \(2007\)](#), [Rufibach \(2007\)](#), [Dümbgen et al. \(2007\)](#), [Dümbgen and Rufibach \(2009\)](#) and [Cule et al. \(2010\)](#). It is known that its nonparametric maximum likelihood estimator (NPMLE) \hat{f} has support on the convex hull of the observations and $\log(\hat{f})$ is piecewise linear. Discontinuities in the first order derivatives may be deemed unsatisfactory if a smoothness assumption is warranted. In the broadest sense, smoothing is the very essence of statistics. To smooth is to sand away the rough edges from a set of data, to offer substantially improved estimation performance, particularly for

small-sized samples, where the convex hull of the data is likely to be rather small. A smooth estimate also has a more attractive visual appearance. For these reasons, throughout this chapter we investigate smoothed versions of the log-concave density estimator. Nonparametric smoothing techniques provide researchers with flexible tools for analysing data. Unfortunately, there is not much literature available on the subject of the smooth log-concave density estimation so far. In the following, we first give a review of smooth unimodal density estimators in Section 4.1.1 and then a review of smooth log-concave density estimation in Section 4.1.2. In Section 4.1.3, an overview of our new methods is given to smooth log-concave density estimation.

4.1.1 Smooth estimation of a unimodal density function

As well known, the kernel-based method is one of the simplest nonparametric approaches and is of wide applicability, particularly in the univariate case. Kernel-based methods for smooth estimation of a unimodal density function can be found in Fougères (1997), Cheng et al. (1999), Eggermont and LaRiccia (2000) and Hall and Huang (2002). Dümbgen and Rufibach (2009) proposed a smooth log-concave density estimator by convolving their nonparametric maximum likelihood density estimator with a Gaussian density, which preserves the log-concavity shape constraint. This estimator was further studied in d -dimensional situation by Chen and Samworth (2013) in which a new hypothesis test of log-concavity of multivariate distribution has been developed. Chen and Samworth (2013) also pointed out that when the true density is log-concave, the smoothed log-concave estimator outperforms by a considerable margin the kernel density estimator with the optimal Integrated Squared Error (ISE) bandwidth. Some kernel-based methods also involve data sharpening techniques that can improve upon the performance of the standard kernel density estimator; we refer the interested reader to Braun and Hall (2001), Hall and Kang (2005) and Wolters (2009). Of course, the kernel-based estimation approach is flexible and easy to interpret. However, choosing the appropriate tuning parameter, such as the smoothing or bandwidth parameter, is far from trivial. If the bandwidth is too small, it tends to under-smooth the sparse regions; if it is too large, it tends to over-smooth the dense regions. Hence, when the data are not uniformly distributed throughout the range of interest, it is impossible to choose a common bandwidth that is large enough to avoid the introduction of spurious features in the tails of the density, and that is small enough to show important features in the central portion (main body). A remedy has been proposed for this deficiency in the fixed bandwidth kernel estimators by Wand et al. (1991) in which the modified procedure is similar to kernel density estimation with an adaptive or variable bandwidth. Other approaches, such as the spline-based and likelihood-based methods, become the

alternative choice. [Bickel and Fan \(1996\)](#) obtained a smooth unimodal density estimate by solving an isotone cubic spline regression problem. A smooth version based on the spline method can also be found in [Meyer \(2012\)](#). [Turnbull and Ghosh \(2014\)](#) introduced a mixture of Beta densities to achieve a smooth unimodal density estimator.

4.1.2 Smooth estimation of a log-concave density function

In this section, a brief summary of the smooth methods that are available in the literature so far to log-concave density estimation is given. A density function f is said to be log-concave, if $\log f$ is concave on its domain. Consider a random sample of size n independently drawn from a univariate log-concave density f . It is convenient to define some standard notation, except where otherwise stated. Denote by $x_1 \leq x_2 \leq \dots \leq x_n$ the order statistics of the sample, by F_n the corresponding empirical distribution function and by $\hat{\sigma}^2$ an unbiased estimator of the variance of the distribution. We know that the NPMLE \hat{f} exists, is unique and is a piecewise linear, continuous function on $[x_1, x_n]$; see [Walther \(2002\)](#), [Pal et al. \(2007\)](#) or [Rufibach \(2007\)](#). Since log-concavity is preserved under convolution, the smooth estimate \hat{f}^* remains log-concave if the applied kernel has this property; see [Karlin, 1968](#)). By realizing this fact, [Dümbgen and Rufibach \(2009\)](#) introduced a smooth log-concave density estimator, defined as

$$\hat{f}^*(x) = \int \phi_{\hat{\gamma}}(x - y) d\hat{F}(y),$$

where \hat{F} is the distribution function corresponding to the NPMLE \hat{f} and $\phi_{\hat{\gamma}}$ denotes the normal density of $\mathcal{N}(0, \hat{\gamma}^2)$. The bandwidth $\hat{\gamma}$, as shown in [Dümbgen and Rufibach \(2009\)](#), is chosen based on the fact that the variance of \hat{f}^* should coincide with $\hat{\sigma}^2$, hence,

$$\hat{\gamma}^2 = \hat{\sigma}^2 - \text{Var}(\hat{F}).$$

Note that the discontinuities of \hat{f} at x_1 and x_n are smoothed out by $\hat{f}^*(x)$. However, \hat{f}^* highly depends on and is rather close to the \hat{f} ; see [Chen and Samworth \(2013\)](#). This fact may allude the limitation for the tail distribution of \hat{f}^* . [Rufibach \(2012\)](#) also employed this version to develop a new smooth estimator of the ROC curve based on the log-concavity assumption of the constituent distributions.

4.1.3 Overview

Letting $\varphi = \log f$, the log-likelihood function is written as

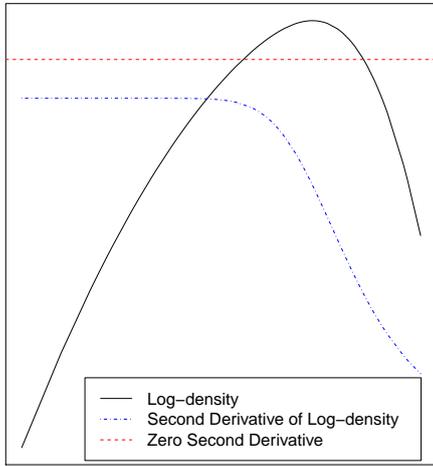
$$l(\varphi) = \sum_{i=1}^n \varphi(x_i).$$

Smoothness conditions are usually imposed on the density function and its derivatives. Furthermore, obtaining a smooth density estimator of f is equivalent to achieving a smooth estimator $\hat{\varphi}$ of φ , which is concave and at least first continuously differentiable in the domain with the constraint $\int e^{\varphi(x)} dx = 1$. We shall specialize to the problem of imposing concavity and continuous derivative on the log-density function φ .

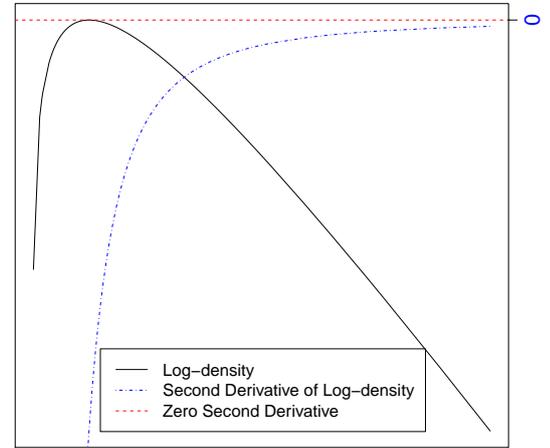
From the continuity aspect, the key property of spline functions is that they are continuous at the knots, and if all the knots are distinct, their derivatives are also continuous. The smoothness can be offered by imposing a degree on the spline. If the construction of spline with continuous derivatives (at least first order) is required, then the degree of the spline must be at least equal to 2. In our approach, we introduce a likelihood-based method by employing a B-spline with degree 2 for estimating φ , i.e., the logarithm of the density function is approximated by a piecewise quadratic function which is constructed by patching pieces of algebraic functions.

From the shape aspect, mathematically the first derivative of a function can tell us if the function is increasing or decreasing, and the second derivative of a function gives information about the shape of the first derivative and information about the concavity or convexity of the original function. In this sense, the concavity restriction can be fulfilled by imposing second derivative properties. We speak of a concave curve if the second derivative is negative. Furthermore, satisfying the concavity restriction without sacrificing the flexibility of the function form is a challenging task as shown in [Ryan and Wales \(2000\)](#) and references therein. To meet the global concavity restriction will cause other problems. Hence imposing concavity restriction can only be done locally. The conditions for local concavity are more stringent than is often realized. From the view point of the algorithmic feasibility, we consider to impose four kinds of restrictions on the behaviour of the second derivative of the log-density φ . The first is that the second derivative of φ is decreasing, the second that it is increasing, the third that it is first increasing and then decreasing, and the last that it is first decreasing and then increasing. In this way, we consider four types of smooth concave functions as shown in [Figure 4.1](#). Note that the first two types are special cases of the last two types. Since we will use a piecewise quadratic function to estimate φ , the second derivative of an estimate φ only changes at a knot. Of course, this is more restrictive than just imposing concavity.

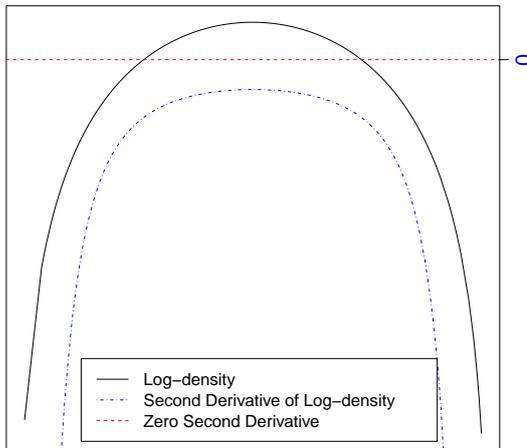
Nevertheless, as examined in Section 4.2.5, commonly-used parametric log-concave distributions all fall in this family of four types, which suggests wide-ranging applicability of the family.



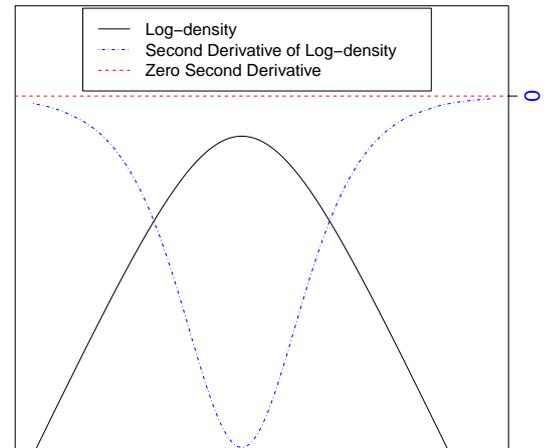
(a) Second derivative decreasing



(b) Second derivative increasing



(c) Second derivative increasing and then decreasing



(d) Second derivative decreasing and then increasing

Fig. 4.1: Four types of log-concave density functions depending on different behaviours of the second derivative.

This new approach requires likelihood maximization that involves infinite-dimension parameters and hence an infinite number of constraints. We also propose a new algorithm to solve these resulting optimization problems, by solving quadratic programming problems repeatedly and replacing the infinite number of constraints with a finite number of constraints that are chosen

dynamically. Our new method will be applied to both simulated and real-world datasets.

4.2 Maximum Likelihood Estimation under Smoothness Assumption

We know that the NPMLE of a log-concave density has a piecewise linear form for log-density φ . To achieve smoothness, we thus consider using a piecewise quadratic function for φ . Since it is impossible to obtain a piecewise quadratic estimator by direct likelihood maximization only under the log-concavity constraint, we have to impose extra constraints. This motivates us to constrain the change of the second derivative of φ , while preserving log-concavity.

Notice that the piecewise linear estimator (3.4) has its φ with a decreasing first derivative, i.e., φ' decreases with x . Therefore, straightaway by replacing linear functions with quadratic ones we obtain a piecewise quadratic estimator that has a decreasing φ'' . Further, reversing the direction of monotonicity gives an estimator that has an increasing φ'' . These two estimators are automatically log-concave and can be found directly by likelihood maximization.

To extend this idea, we take one further step, by allowing φ'' to change the direction of monotonicity once. This means that φ'' can be first decreasing and then increasing, or that φ'' first increasing and then decreasing. This expands the family of our smooth log-density estimators and seemingly meets most practical needs. We thus have four piecewise quadratic estimators, with their details given below.

4.2.1 φ'' decreasing

We firstly consider the situation when φ'' decreases as x increases. Let $\mathcal{I} = [x_1, x_n]$. Hence

$$\varphi(x) = \alpha_1 x - \alpha_2 x^2 - \sum_{j=1}^m \pi_j (x - \theta_j)_+^2 - C_1, \quad (4.1)$$

where $\alpha_1 \in \mathbb{R}$, $\alpha_2, \pi_j \geq 0$, $\theta_j \in \mathcal{I}$ for $j = 1, \dots, m$, and

$$C_1 = \log \left\{ \int_{\mathcal{X}} e^{\alpha_1 x - \alpha_2 x^2 - \sum_{j=1}^m \pi_j (x - \theta_j)_+^2} dx \right\}.$$

Denoting $\boldsymbol{\pi} = (\pi_1, \dots, \pi_m)^T$ the vector of positive masses, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)^T$ the vector of knots. Note that the second derivative can only change value at a knot. We use \mathcal{K}_D for the family of log-concave densities defined by function (4.1) and name the maximum likelihood estimator as Piecewise Quadratic Estimator 1 (PQ1).

4.2. Smoothness Assumption

For $\varphi \in \mathcal{K}_D$, the density function is given by

$$f(x; \varphi) = e^{\varphi(x)}, \quad (4.2)$$

which must be log-concave. If there is no knot, f is simply a normal density function. In general, it is a “piecewise truncated normal density function”. The log-likelihood function of φ is thus given by

$$l(\varphi) = \sum_{i=1}^n \left(\alpha_1 x_i - \alpha_2 x_i^2 - \sum_{j=1}^m \pi_j (x_i - \theta_j)_+^2 \right) - nC_1. \quad (4.3)$$

The nonparametric maximum likelihood estimation of the concave function φ , under the above concave shape restriction, is thus the following optimization problem:

$$\begin{aligned} & \text{maximize } l(\varphi) \\ & \text{subject to } \varphi \in \mathcal{K}_D. \end{aligned} \quad (4.4)$$

To be consistent in notations with later chapters, let us denote $\boldsymbol{\beta} = (\alpha_1, \alpha_2, \boldsymbol{\pi}^T)^T$ and $\boldsymbol{\vartheta} = \boldsymbol{\theta}$. Then any $\varphi \in \mathcal{K}_D$ is completely defined by its $\boldsymbol{\vartheta}$ and $\boldsymbol{\beta}$, with m implicitly assumed known. It is easy to show that \mathcal{K}_D is a convex set and $l(\varphi)$ is a concave function on \mathcal{K}_D . By maximizing the log-likelihood function (4.3) over all functions in \mathcal{K}_D , we obtain the smooth nonparametric maximum likelihood estimate (SNPMLE) $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\vartheta}})$ of $(\boldsymbol{\beta}, \boldsymbol{\vartheta})$, i.e., the PQ1 estimator.

Our approach involves locating new knots. A gradient function is introduced to achieve this purpose. Consider a $\varphi \in \mathcal{K}_D$, the gradient function is defined below in aid of the basis function $e_{1,\xi}(x) = -(x - \xi)_+^2$ for $\xi \in \mathcal{I}$:

$$\begin{aligned} d_1(\xi; \varphi) &\equiv \left. \frac{\partial l(\varphi + \varepsilon e_{1,\xi})}{\partial \varepsilon} \right|_{\varepsilon=0^+} \\ &= - \sum_{i=1}^n (x_i - \xi)_+^2 + n \int_{\mathcal{X}} (x - \xi)_+^2 e^{\varphi(x)} dx \\ &= \sum_{i=1}^n e_{1,\xi}(x_i) - nE_{\varphi}(e_{1,\xi}(X)), \end{aligned} \quad (4.5)$$

where E_{φ} is the expectation with respect to $f(\varphi) = e^{\varphi}$. Note that d_1 is a piecewise quadratic

function of ξ . Letting $e_{0,1}(x) = x$, $e_{0,2}(x) = -x^2$, we also define

$$\begin{aligned} d_{0,j}(\varphi) &\equiv \left. \frac{\partial l(\varphi + \varepsilon e_{0,j})}{\partial \varepsilon} \right|_{\varepsilon=0^+} \\ &= \sum_{i=1}^n e_{0,j}(x_i) - nE_{\varphi}(e_{0,j}(X)), \quad \text{for } j = 1, 2. \end{aligned} \quad (4.6)$$

Notice that the above gradients are all of the same form.

4.2.2 φ'' increasing

Let us move onto the second situation when φ'' increases with x . It is a straightforward extension of the PQ1 estimator in Section 4.2.1. To reverse the direction of the monotonicity of φ'' , we can simply use

$$\varphi(x) = \alpha_1 x - \alpha_2 x^2 - \sum_{j=1}^p \omega_j (\tau_j - x)_+^2 - C_2, \quad (4.7)$$

where $\alpha_1 \in \mathbb{R}$, $\alpha_2, \omega_j \geq 0$, $\tau_j \in \mathcal{I}$ for $j = 1, \dots, p$, and

$$C_2 = \log \left\{ \int_{\mathcal{X}} e^{\alpha_1 x - \alpha_2 x^2 - \sum_{j=1}^p \omega_j (\tau_j - x)_+^2} dx \right\}.$$

Denote by $\boldsymbol{\omega} = (\omega_1, \dots, \omega_p)^T$ the vector of positive masses and by $\boldsymbol{\tau} = (\tau_1, \dots, \tau_p)^T$ the vector of knots. Similarly, we use \mathcal{K}_I for the family of log-concave densities defined by function (4.7) and name the maximum likelihood estimator as Piecewise Quadratic Estimator 2 (PQ2). It is obvious that if $\varphi(x) \in \mathcal{K}_I$, then $\varphi(-x) \in \mathcal{K}_D$.

Given $\varphi \in \mathcal{K}_I$, the density and log-likelihood functions can be derived in the same way as in functions (4.2) and (4.3), respectively. Further, consider the basis function $e_{2,\xi}(x) = -(\xi - x)_+^2$ for $\xi \in \mathcal{I}$, the gradient function is defined as

$$\begin{aligned} d_2(\xi; \varphi) &\equiv \left. \frac{\partial l(\varphi + \varepsilon e_{2,\xi})}{\partial \varepsilon} \right|_{\varepsilon=0^+} \\ &= \sum_{i=1}^n e_{2,\xi}(x_i) - nE_{\varphi}(e_{2,\xi}(X)), \end{aligned} \quad (4.8)$$

which is of the same form as $d_1(\xi; \varphi)$ in function (4.5). Further, $d_{0,j}(\varphi)$, $j = 1, 2$, also has the same form as function (4.6). Of course, all expectations here are with respect to the density defined by the φ given in function (4.7).

Denote $\boldsymbol{\beta} = (\alpha_1, \alpha_2, \boldsymbol{\omega}^T)$ and $\boldsymbol{\vartheta} = \boldsymbol{\tau}$. Then by maximizing the log-likelihood function over all

concave functions in \mathcal{K}_I , we obtain the SNPML estimator $(\widehat{\beta}, \widehat{\vartheta})$ of (β, ϑ) , i.e., the PQ2 estimator. Both PQ1 and PQ2 estimators are based on that the second derivative of the log-density changes monotonically. Taking a step further from the assumption of a monotone second derivative, we consider two cases. One is that the second derivative of the log-density first increases and after reaching its maximum, decreases. The other does the opposite, with the second derivative first decreasing and then increasing. Apparently, both new estimators have PQ1 and PQ2 as their special cases.

4.2.3 φ'' increasing and then decreasing

Now let us consider the situation when φ'' first increases and, after reaching its maximum, decreases, as x increases. Combining functions (4.1) and (4.7), we define

$$\varphi(x) = \alpha_1 x - \alpha_2 x^2 - \sum_{j=1}^p \omega_j (\tau_j - x)_+^2 - \sum_{j=1}^m \pi_j (x - \theta_j)_+^2 - C_3, \quad (4.9)$$

where $\vartheta = (\boldsymbol{\tau}^T, \boldsymbol{\theta}^T)^T$ denotes the knot vector of φ with its corresponding coefficient vector $\beta = (\alpha_1, \alpha_2, \boldsymbol{\omega}^T, \boldsymbol{\pi}^T)^T$. Here, $\alpha_2, \omega_j, \pi_j \geq 0$ for all j , $\alpha_1 \in \mathbb{R}$. In particular, $\boldsymbol{\tau} = (\tau_1, \dots, \tau_p)^T$, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)^T$, $\boldsymbol{\omega} = (\omega_1, \dots, \omega_p)^T$, $\boldsymbol{\pi} = (\pi_1, \dots, \pi_m)^T$, and

$$C_3 = \log \left\{ \int_{\mathcal{X}} e^{\alpha_1 x - \alpha_2 x^2 - \sum_{j=1}^p \omega_j (\tau_j - x)_+^2 - \sum_{j=1}^m \pi_j (x - \theta_j)_+^2} dx \right\}.$$

In addition, we let

$$x_1 \leq \tau_1 < \dots < \tau_p \leq \theta_1 < \dots < \theta_m \leq x_n, \quad (4.10)$$

which ensures that φ'' first increases and then decreases with x .

Note that each τ_j ($1 \leq j \leq p$) indicates a knot at which the second derivative of φ changes to a greater value, while each θ_j ($1 \leq j \leq m$) a knot where φ has the value of its second derivative changed to a smaller one. Between τ_p and θ_1 , φ has the largest second derivative $-2\alpha_2$. We use \mathcal{K}_{ID} for the family defined by function (4.9) and name the maximum likelihood estimator as Piecewise Quadratic Estimator 3 (PQ3).

The density and log-likelihood functions can also be written in terms of $\varphi \in \mathcal{K}_{ID}$. Further, two gradient functions of τ and θ are needed for respectively, the part with an increasing second derivative and the part with a decreasing second derivative. In the aid of basis functions $e_{1,\xi}(x)$ and $e_{2,\xi}(x)$, $d_1(\xi; \varphi)$ and $d_2(\xi; \varphi)$ are the same defined as in functions (4.5) and (4.8), respectively. Note that here d_1 is defined on $[\tau_p, x_n]$ and d_2 on $[x_1, \theta_1]$. Further, $d_{0,j}(\varphi)$, $j = 1, 2$, still has the

same form as function (4.6). The SNPML ($\widehat{\beta}, \widehat{\vartheta}$) is also obtained by maximizing the corresponding log-likelihood function.

4.2.4 φ'' decreasing and then increasing

The fourth situation is when φ'' first decreases and, after reaching its minimum, increases as x increases. This means

$$\varphi(x) = \alpha_1 x - \alpha_2 x^2 + \sum_{j=1}^p \omega_j (\tau_j - x)_+^2 + \sum_{j=1}^m \pi_j (x - \theta_j)_+^2 - C_4, \quad (4.11)$$

where the parameters are restricted the same way as for function (4.9), except that the normalizing constant C_4 here is given by

$$C_4 = \log \left\{ \int_{\mathcal{X}} e^{\alpha_1 x - \alpha_2 x^2 + \sum_{j=1}^p \omega_j (\tau_j - x)_+^2 + \sum_{j=1}^m \pi_j (x - \theta_j)_+^2} dx \right\}.$$

Also, denote by $\tau, \theta, \omega, \pi, \vartheta$ and β the same way as in Section 4.2.3. The ordering of the knots also remains the same as 4.10. Note that here each τ_j ($1 \leq j \leq p$) indicates a knot where φ'' is decreasing and each θ_j ($1 \leq j \leq m$) a knot where φ'' is increasing. Between τ_p and θ_1 , $\varphi'' = -2\alpha_2$, which is the smallest.

Unlike PQ3, we need additional constraints to ensure log-concavity, i.e., $\varphi'' \leq 0$ for all $x \in \mathcal{X}$. Denote by L and U the lower and upper boundaries of $\mathcal{X} \subset \mathbb{R}$, respectively, which can be finite or infinite. Then we have

$$\varphi''(L) \equiv \lim_{x \rightarrow L} \frac{\partial^2 \varphi(x)}{\partial x^2} = 2 \left(-\alpha_2 + \sum_{j=1}^p \omega_j \right), \quad (4.12)$$

$$\varphi''(U) \equiv \lim_{x \rightarrow U} \frac{\partial^2 \varphi(x)}{\partial x^2} = 2 \left(-\alpha_2 + \sum_{j=1}^m \pi_j \right). \quad (4.13)$$

We need to impose the following two conditional constraints:

$$\varphi''(L) \leq 0 \quad \text{and} \quad \varphi''(U) \leq 0. \quad (4.14)$$

Similarly, \mathcal{K}_{DI} is used to denote the family defined above and the resulting maximum likelihood estimator is named Piecewise Quadratic Estimator 4 (PQ4).

The density and log-likelihood functions can be written in terms of $\varphi \in \mathcal{K}_{DI}$. Two gradient functions are also defined in the aid of basis functions $e_{3,\xi}(x) = (x - \xi)_+^2$ and $e_{4,\xi}(x) = (\xi - x)_+^2$:

$$d_j(\xi; \varphi) \equiv \sum_{i=1}^n e_{j,\xi}(x_i) - nE_{\varphi}(e_{j,\xi}(X)), \text{ for } j = \{3, 4\}. \quad (4.15)$$

Note that d_3 is defined on $[\tau_p, x_n]$ and d_4 on $[x_1, \theta_1]$. The gradient functions for α_1 and α_2 are algebraically the same as function (4.6). For all four estimators, notationally, φ is exchangeable by its coefficient vector β and knot vector ϑ , namely $l(\varphi) \equiv l(\beta, \vartheta)$.

4.2.5 Log-concavity properties of commonly-used distributions

As pointed out in Section 4.1.3, our estimators are more restrictive than the one only subject to log-concavity, but it seems that all commonly-used parametric distributions that are log-concave, e.g., those discussed in Bagnoli and Bergstrom (2005), have their φ'' following one of the above four trends. This means that the proposed estimators cover a wide range of log-concave distributions and these distributions can thus be consistently estimated. In addition, the smoothness of these estimators helps reduce estimation variance, as compared with the piecewise linear estimator. In terms of φ'' , we have examined parametric log-concave distributions as many as we can and found that they all conform with the proposed families above. The location parameter is ignored, as it does not affect the shape of the distribution.

Uniform distribution

The uniform distribution, defined on the interval $[a, b]$, has density $f(x) = 1/(b - a)$. The second derivative of the log-density is zero everywhere on $[a, b]$. Therefore it falls into any of the four types.

Exponential distribution

The exponential distribution has support on $[0, \infty)$ and density function $f(x) = \lambda e^{-\lambda x}$. The second derivative of the log-density is zero, so it belongs to any of the four types.

Normal distribution

The normal distribution has density $f(x) = e^{-x^2/(2\sigma^2)}/(\sqrt{2\pi}\sigma)$. The first derivative of $\log\{f(x)\}$ is $-x/\sigma^2$ and the second derivative is $-1/\sigma^2$, indicating that it is of any of the four types.

Logistic distribution

The logistic distribution has density $f(x) = e^{-x/s}/\{s(1 + e^{-x/s})^2\}$. Calculation shows the second derivative of the log-density is $-2f(x)/s$, which decreases on $(-\infty, 0]$ and increases on $[0, \infty)$. It is covered by the PQ4 estimator.

Extreme-Value distribution

The distribution often referred to as the extreme value distribution is the limiting distribution of the minimum of a large number of unbounded identically distributed random variables. This is sometimes known as the Gumbel distribution, or as a Type I extreme value distribution. It has density function $f(x) = e^{x/\lambda} \exp(-e^{x/\lambda})/\lambda$. The second derivative of the log-density is $-e^{x/\lambda}/\lambda^2$ which decreases in its domain. This case belongs to the PQ1 estimator.

Weibull distribution

The Weibull distribution with parameter $k > 0$ has density function $f(x) = k(x/\lambda)^{k-1} e^{-(x/\lambda)^k}/\lambda$, which is defined for $x \in (0, \infty)$ and hence is log-concave for $k > 1$. The second derivative of the log-density is $(1 - k)x^{-2}(1 + kx^k/\lambda^k)$. If $1 < k \leq 2$, the second derivative increases in the domain which is the case of the PQ2 estimator. If $k > 2$, the second derivative first increases and then decreases, and this is the situation of estimator PQ3.

Gamma distribution

The Gamma distribution has density function

$$f(x) = \frac{x^{k-1} e^{-x/\theta}}{\theta^k \Gamma(k)},$$

which is defined for $x \in (0, \infty)$, where $\theta > 0$, $k > 0$ and $\Gamma(\cdot)$ is the gamma function. Calculation shows that the second derivative of the log-density is $(1 - k)/x^2$, which increases in its domain when $k > 1$ and is zero when $k = 1$. Estimator PQ2 can be applied in this situation.

Beta distribution

The beta distribution has density function

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

which is defined for $x \in (0, 1)$, where $a > 0$, $b > 0$ and $B(\cdot)$ is the beta function. It is log-concave when $a \geq 1$ and $b \geq 1$. The second derivative of the log-density is $(1 - a)/x^2 + (1 - b)/(1 - x)^2$ which first increases and then decreases in its domain with $a \geq 1$ and $b \geq 1$. Estimator PQ3 can readily handle this case.

4.2.6 Theoretical properties

The properties for these four estimators are similar and are hence presented together, with differences given where ever needed. Since the value of the log-likelihood function can not be increased by taking any arbitrarily small valid step away from the maximum likelihood estimate $\hat{\varphi}$ as restricted by \mathcal{K} , where \mathcal{K} can be \mathcal{K}_D , \mathcal{K}_I , \mathcal{K}_{ID} or \mathcal{K}_{DI} , $\hat{\varphi}$ must satisfy the following first-order conditions:

- (i) $d_{0,1}(\hat{\varphi}) = 0$.
- (ii) $d_{0,2}(\hat{\varphi}) = 0$, if $\hat{\alpha}_2 > 0$;
 $d_{0,2}(\hat{\varphi}) \leq 0$, if $\hat{\alpha}_2 = 0$.
- (iii) $d_j(\xi; \hat{\varphi}) \leq 0$, for $\xi \in \hat{\mathcal{S}}_j$.
- (iv) $d_j(\xi; \hat{\varphi}) = 0$, for $\xi \in \text{supp}_j(\hat{\varphi})$.

In condition (iii), by $\hat{\mathcal{S}}_j$ we mean the potential support space for the j -th gradient function. For example, for PQ1 there is only d_1 and $\hat{\mathcal{S}}_1$ is just \mathcal{I} , and for PQ2 there is only d_2 and $\hat{\mathcal{S}}_2 = \mathcal{I}$. For PQ3, $\hat{\mathcal{S}}_1 = [\hat{\tau}_p, x_n]$ and $\hat{\mathcal{S}}_2 = [x_1, \hat{\theta}_1]$, and for PQ4, $\hat{\mathcal{S}}_3 = [\hat{\tau}_p, x_n]$ and $\hat{\mathcal{S}}_4 = [x_1, \hat{\theta}_1]$. In condition (iv), by $\text{supp}_j(\hat{\varphi})$ we mean the j -th support set of $\hat{\varphi}$, i.e., the support points for the basis functions of the j -th form.

4.3 Computation

The computation of the proposed estimators requires maximizing the log-likelihood function. The challenge in solving this problem is that it has no closed-form solution, and an iterative method must be used. In this section, we focus on the computational problems and present new algorithms for computing the SNPMLE estimators of a log-concave density function. The algorithms for these four estimators are based on a similar idea, and hence we discuss them together except where differences exist and will be pointed out. The overall structure of the algorithm is similar to that of CNMLCD in the sense that it also has two alternating steps: the first is to expand and reduce the knot vector ϑ and the second is to update the coefficient vector β . There are a couple of tricky issues when dealing with the PQ3 and PQ4 estimators. One is that the PQ3 or PQ4 estimator involves an overlapping interval where two directional gradient functions need to be considered. The other is that estimator PQ4 requires two extra linear constraints.

4.3.1 Derivatives of the log-likelihood function

Given φ , the first partial derivative vector and Hessian matrix of the log-likelihood are given in the following. In all situations, the first partial derivatives with respect to α_1 and α_2 are of the same forms as follows:

$$\frac{\partial l(\varphi)}{\partial \alpha_j} = d_{0,j}(\varphi), \quad j = 1, 2,$$

where φ is the concave function in the corresponding situation. Further,

$$\frac{\partial l(\varphi)}{\partial \gamma_{i,j}} = d_i(\xi_{i,j}; \varphi),$$

where $\xi_{i,j}$ is the j -th knot in i -th support set of φ and $\gamma_{i,j}$ the corresponding mass. For PQ1, there is only one support set and thus $i = 1$, and for PQ2 there is also one support set and $i = 2$. For PQ3, $i = 1, 2$, and for PQ4, $i = 3, 4$. Further, denote by

$$\mathbf{e} = (e_{0,1}(x), e_{0,2}(x), e_{i,\xi_{i,j}}(x))^T,$$

then the Hessian matrix can be expressed as $-n\text{Var}_\varphi(\mathbf{e})$, where Var_φ is the variance operator with respect to density e^φ .

4.3.2 Updating coefficients

We now present the method to updating the coefficient vector from β to β^* when ϑ is held fixed. Let the first partial derivative vector and Hessian matrix of the log-likelihood be, respectively,

$$\mathbf{g} \equiv \mathbf{g}(\beta, \vartheta) = \frac{\partial l(\beta, \vartheta)}{\partial \beta},$$

$$\mathbf{H} \equiv \mathbf{H}(\beta, \vartheta) = \frac{\partial^2 l(\beta, \vartheta)}{\partial \beta \partial \beta^T}.$$

Let $\mathbf{H} = -\mathbf{R}^T \mathbf{R}$, where \mathbf{R} is some square matrix which is obtained in the same way as described in Section 3.3. Expanding $l(\beta^*, \vartheta)$ in the Taylor series about β to the second order gives

$$\begin{aligned} l(\beta, \vartheta) - l(\beta^*, \vartheta) &\approx -\mathbf{g}^T \boldsymbol{\eta} + \frac{1}{2} \boldsymbol{\eta}^T \mathbf{R}^T \mathbf{R} \boldsymbol{\eta} \\ &= \frac{1}{2} \|\mathbf{R}\beta^* - \mathbf{R}\beta - \mathbf{R}^{-T} \mathbf{g}\|^2 - \frac{1}{2} \mathbf{g}^T \mathbf{R}^{-1} \mathbf{R}^{-T} \mathbf{g}, \end{aligned}$$

4.3. Computation

where $\boldsymbol{\eta} = \boldsymbol{\beta}^* - \boldsymbol{\beta}$ and $\|\cdot\|$ denotes the L_2 -norm. Maximizing $l(\boldsymbol{\beta}^*, \boldsymbol{\theta})$ in the neighbourhood of $\boldsymbol{\beta}$ can be replaced approximately with the following linear regression problem:

$$\min_{\boldsymbol{\beta}^*} \left\| \mathbf{R}\boldsymbol{\beta}^* - \mathbf{R}\boldsymbol{\beta} - \mathbf{R}^{-T}\mathbf{g} \right\|^2, \quad (4.16)$$

where $\boldsymbol{\beta}^*$ is constrained by the conditions that correspond to each of the four estimators, as discussed in Section 4.2.

Specifically, for estimator PQ1, PQ2 or PQ3, only the constraint $\beta_{-1}^* \geq 0$ needs to be satisfied, where $\beta_{-1}^* \geq 0$ is $\boldsymbol{\beta}^*$ without its first element α_1^* , i.e., α_1^* is left unconstrained. Problem (4.16) can then be solved by the `pnnls()` function in the R package `lsei` (Wang et al., 2015). This R function solves a linear squares problem subject to nonnegativity, but it allows for some parameters unconstrained. To ensure a monotone increase of the log-likelihood, a step-halving line search can subsequently be conducted.

For estimator PQ4, two extra linearity constraints are required according to (4.14), i.e.,

$$-\alpha_2^* + \sum_{j=1}^p \omega_j^* \leq 0, \quad (4.17)$$

$$-\alpha_2^* + \sum_{j=1}^m \pi_j^* \leq 0. \quad (4.18)$$

If both constraints are inactive for the current iterate φ , i.e., $-\alpha_2 + \sum_{j=1}^p \omega_j < 0$ and $-\alpha_2 + \sum_{j=1}^m \pi_j < 0$, then the computation proceeds the same way as for PQ1, PQ2 or PQ3, only under the constraint $\beta_{-1}^* \geq 0$. If, however, one or both constraints are active for φ , we need to eliminate a coefficient using each active constraint, e.g., replacing ω_p^* with $\alpha_2^* - \sum_{j=1}^{p-1} \omega_j^*$, or replacing π_m^* with $\alpha_2^* - \sum_{j=1}^{m-1} \pi_j^*$. Note that the gradient vector and Hessian matrix are only with respect to the coefficient vector without the eliminated coefficients.

With either formulation, the `pnnls()` function can then be used to find a solution, say, $\boldsymbol{\beta}^*$, which may or may not satisfy constraints (4.17) and (4.18). If it does not, one can backtrack the solution so that both constraints are satisfied, typically one becoming active. In more detail, this backtracking step works as follows:

1. Set $k_l = 1$ and $k_u = 1$.
2. If $\varphi^{*''}(L) > 0$, reset

$$k_l = \frac{\varphi''(L)}{\varphi''(L) - \varphi^{*''}(L)},$$

and if $\varphi^{*''}(U) > 0$, reset

$$k_u = \frac{\varphi''(U)}{\varphi''(U) - \varphi^{*''}(U)}.$$

3. Reset $\beta^* = (1 - k)\beta + k\beta^*$, where $k = \min(k_l, k_u)$.

After this, the usual step-halving backtracking line search is conducted to ensure that the log-likelihood increases monotonically.

If one is interested in finding the MLE of β with ϑ held fixed, then the above process can be iterated indefinitely until the final estimate of β is found. It can be easily established that the solution must be the MLE in this case, which has a concave log-likelihood function on a convex set. Note that under the nonnegativity constraint, some coefficients may turn out to be exactly zero.

4.3.3 Expanding and reducing knot sets

We now turn to the second step: expanding and reducing knot sets. Choosing the optimal number and positions of knots is a complex task. In our approach, new knots are found in aid of the gradient function. The general idea for the new algorithm is that in each iteration, it expands the knot set ϑ by including all local maxima of the gradient function. Such local maxima are located by the Newton-bisection method used in Wang (2007), that requires the first and second derivatives of the gradient function.

Consider the PQ1 and PQ2 estimators, and take estimator PQ1 as an example. The gradient function $d_1(\xi; \varphi)$ (or $d_2(\xi; \varphi)$ for PQ2) is a piecewise quadratic function and it is simple to find all of its local maxima on an interval $[k_l, k_u]$, where k_l is the lower boundary for the knot set, and k_u is the upper boundary.

For estimators PQ3 and PQ4, the situation gets slightly trickier, as each estimator has two gradient functions involved and thus two knot sets, τ and θ . The dealings with the two estimators are similar. Take estimator PQ4 as an example, which has gradient functions $d_4(\xi, \varphi)$ and $d_3(\xi; \varphi)$. The domain \mathcal{X} can be divided into three intervals: (i) the increasing part on $[k_l, \tau_p)$, (ii) the middle part on $[\tau_p, \theta_1]$ and (iii) the decreasing part on $(\theta_1, k_u]$.

In the first step of the new algorithm, we expand the two knot sets that correspond to the second-derivative increasing and decreasing parts of a log-concave function by finding and adding new knots in each of the two corresponding intervals. For the increasing part on $[k_l, \tau_p)$, we use the gradient function $d_4(\xi; \varphi)$ and find all of its local maxima, and add them to the increasing-part knot set τ . For the decreasing part on $(\theta_1, k_u]$, the gradient function $d_3(\xi; \varphi)$ is applied and all of its local maxima are found and added to the decreasing-part knot set θ .

For the middle part on $[\tau_p, \theta_1]$, we employ both gradient functions $d_4(\xi; \varphi)$ and $d_3(\xi; \varphi)$ if $\tau_p \neq \theta_1$.

For each gradient function, the point corresponding to the greatest value of each gradient function is recorded, say, τ' for d_4 and θ' for d_3 . We do not add any new knot if both gradient values $d_4(\tau')$ and $d_3(\theta')$ are nonpositive. Otherwise, there are two situations that may be encountered, either $\tau' < \theta'$ or $\tau' > \theta'$, as shown in Figure 4.2. If $\tau' < \theta'$, then we add the two knots to τ and θ , respectively. Of course, a maximum may occur at the boundary. If $\tau' > \theta'$, then we only add τ' to the increasing-part knot set τ if $d_4(\tau', \varphi) > d_3(\theta', \varphi)$, or else only θ' to the decreasing-part knot set θ .

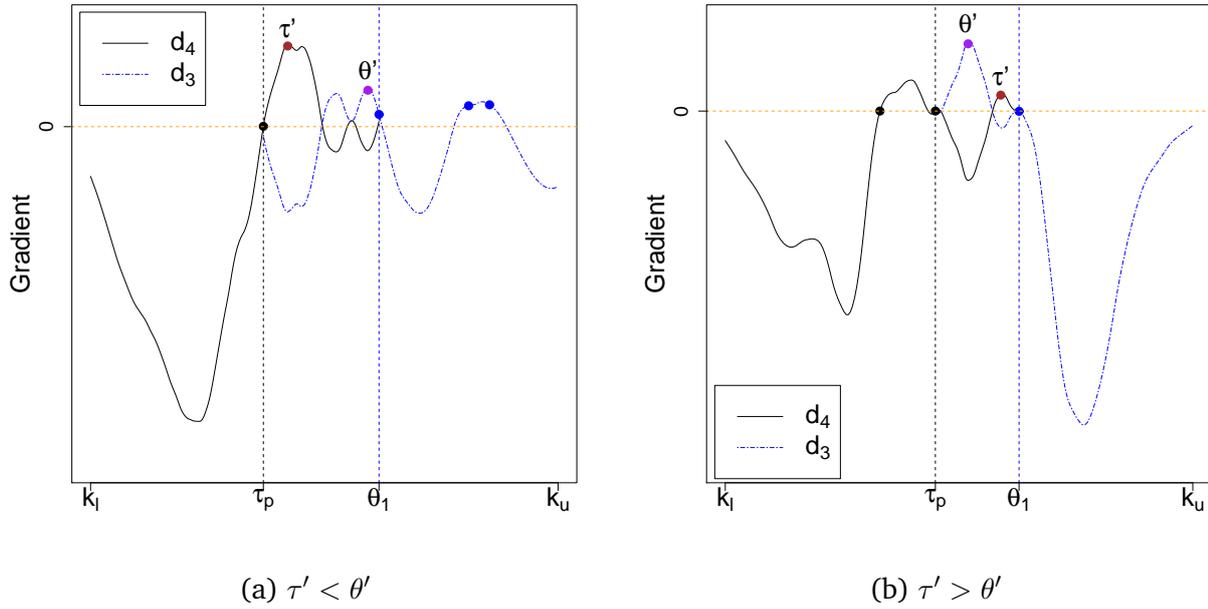


Fig. 4.2: Finding knots for the middle part.

After adding new knots, we update the coefficient vector β by solving problem (4.16), then reduce the knot sets of ϑ by removing the elements that have zero masses. Furthermore, when no information is known about the knot boundaries k_l and k_u , we set $k_l = x_1 + 2s/\sqrt{n}$, $k_u = x_n - 2s/\sqrt{n}$, where s and n are the standard deviation and size of the sample, respectively. The boundaries chosen this way, in our experience, can offer moderate smoothness outside $[x_1, x_n]$ and we found no problems arising from this treatment.

4.3.4 Algorithm

The algorithms for the above four shape-restricted smooth estimators are based on one general idea and are summarized as follows.

Algorithm 2 (CNMLCDS). Set $s = 0$. From an initial estimate φ_0 with a finite number of knots and $l(\varphi_0) > -\infty$, repeat the following steps.

1. Compute all local maxima of the gradient functions, which gives new knots, $\theta'_{s1}, \dots, \theta'_{sq}$ for PQ1, $\tau'_{s1}, \dots, \tau'_{st}$ for PQ2, or $\tau'_{s1}, \dots, \tau'_{st}$ and $\theta'_{s1}, \dots, \theta'_{sq}$ for PQ3 and PQ4 .
2. Set $\boldsymbol{\theta}'_s = (\theta'_{s1}, \dots, \theta'_{sq})^T$ and/or $\boldsymbol{\tau}'_s = (\tau'_{s1}, \dots, \tau'_{st})^T$, and $\boldsymbol{\vartheta}_s^+ = (\boldsymbol{\tau}_s^T, \boldsymbol{\tau}_s'^T, \boldsymbol{\theta}_s^T, \boldsymbol{\theta}_s'^T)^T$ and $\boldsymbol{\beta}_s^+ = (\alpha_1, \alpha_2, \boldsymbol{\omega}_s^T, \mathbf{0}^T, \boldsymbol{\pi}_s^T, \mathbf{0}^T)^T$.
3. Compute $\boldsymbol{\beta}_{s+1}^-$ by solving problem (4.16) ($\boldsymbol{\beta}_{s+1}^- = \boldsymbol{\beta}^*$), execute the backtracking step described in Section 4.3.3, and conduct a step-halving line search.
4. Discard all knots with zero masses in $\boldsymbol{\beta}_{s+1}^-$, which gives $\boldsymbol{\beta}_{s+1}$ and $\boldsymbol{\vartheta}_{s+1}$ of φ_{s+1} . Stop if $l(\varphi_{s+1}) - l(\varphi_s) \leq \text{Tolerance}$. Set $s = s + 1$.

The tolerance for the stopping criterion in step 4 is set to 10^{-7} in our numerical studies.

4.4 Assessing Log-concavity

One can never be certain that a particular sample has been drawn from a log-concave distribution. It is thus desirable to be able to test the log-concavity assumption. In this section, we turn to the construction of statistical tests for log-concavity. An (1996) proposed tests for increasing hazard rates and the *new-is-better-than-used* property. Walther (2002) presented a methodology for detecting mixing in the log-concave model which can be extended to test log-concavity. Bootstrapping allows estimation of the sampling distribution of almost any statistic using random sampling methods. It also has been used for testing log-concavity, based on various test statistics (Cule et al., 2010; Hazelton, 2011; Chen and Samworth, 2013).

4.4.1 Bootstrapping

In order to examine

$$H_0 : \text{The true density is log-concave,}$$

we make use of the log-concave density estimator and apply the bootstrap test as follows:

- (a) Compute the log-concave maximum likelihood density estimate \hat{f} from the piecewise linear estimator (PL) studied in Chapter 3 from the given sample, and denote by \hat{F} the corresponding distribution function.

- (b) Compute the test statistic $t_0 = T(x_1, \dots, x_n; \hat{F})$.
- (c) For $s = 1, \dots, B$, draw an independent bootstrap sample $x_{s1}^*, \dots, x_{sn}^*$ from the null sampling distribution \hat{F} . From each bootstrap sample, compute first the piecewise linear, log-concave density estimate \hat{F}_s^* and then the test statistic $t_s^* = T(x_{s1}^*, \dots, x_{sn}^*; \hat{F}_s^*)$.
- (d) The p -value is given by $p = (B + 1)^{-1} \sum_{s=1}^B \{t_s^* \geq t_0\}$.

4.4.2 Test statistics

There are many statistics that can be used here for T . For example, it can be the Kolmogorov-Smirnov statistic, which is given by

$$\text{KS}(x_1, \dots, x_n; \hat{F}) = \sup_{x \in \mathcal{X}} |F_n(x) - \hat{F}(x)|, \quad (4.19)$$

where $F_n(x)$ the empirical distribution function given by the sample x_1, \dots, x_n . To be a bit more precise, $F_n(x)$ is defined as

$$F_n(x) = \frac{1}{n} [\text{Number of observations} \geq x]. \quad (4.20)$$

Similarly, we can further use the mean, standard deviation, skewness or kurtosis as the test statistic to assess how the PL estimator fits to a particular sample. Taking the mean as an example, the statistic is given by

$$\text{mean}(x_1, \dots, x_n; \hat{F}) = |\text{mean}(x_1, \dots, x_n) - \text{mean}(\hat{F})|.$$

We will apply the bootstrap test with these statistics to real-world data in Section 4.6, and denote their resulting p -values by p_{ks} , p_{mean} , p_{sd} , p_{skew} and p_{kurt} , respectively.

4.5 Simulation Studies

4.5.1 Between piecewise quadratic estimators

Setup

We proposed four smooth estimators for a log-concave density function in Section 4.2. As mentioned in Section 4.2.1, the second derivative of estimator PQ1 behaves oppositely to that of estimator PQ2, that is, for any $\varphi(x) \in \mathcal{K}_I$, we have $\varphi(-x) \in \mathcal{K}_D$. Since PQ1 and PQ2 are only special

cases of PQ3 and PQ4, in this section we examine how well PQ3 or PQ4 can perform in the situation where data conforms with PQ1. The conclusions drawn certainly also hold if data conforms with PQ2.

Two distributions are used below, the truncated normal (TN) and the skew normal (SN), as listed in Table 4.1. Both distributions are log-concave. The log-density of the truncated normal has a constant second derivative and that of the skew normal has an decreasing second derivative, as shown in Figure 4.3.

Tab. 4.1: Two distributions of Type 1.

Distribution	Notation	Density Function	Range	Parameters
Truncated Normal	$TN(\mu, \sigma, a, b)$	$\frac{\exp\{-0.5(x - \mu)^2/\sigma^2\}}{\int_a^b \exp\{-0.5(x - \mu)^2/\sigma^2\}} dx$	$x \in [a, b]$	$\mu, a, b \in \mathbb{R}, \sigma \geq 0$
Skew Normal	$SN(\mu, \sigma, \alpha)$	$\frac{\exp\{-0.5(x - \mu)^2/\sigma^2\}}{\pi\sigma} \int_{-\infty}^{\alpha(x-\mu)/\sigma} e^{-t^2/2} dt$	$x \in \mathbb{R}$	$\mu, \alpha \in \mathbb{R}, \sigma \geq 0$

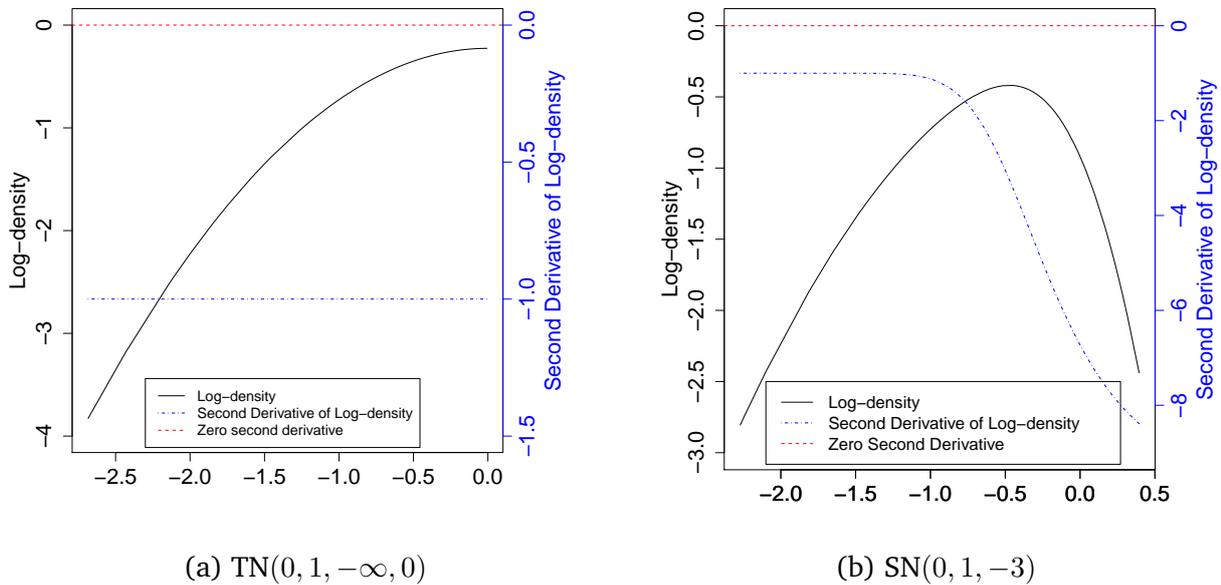


Fig. 4.3: Log-density and its second derivative of a truncated normal (left) and a skew normal (right).

All computations were carried out in R (version 3.1.1) (R Core Team, 2015) on a computer with a 3.4 GHz Intel Core i5-3570 central processing unit. Our implementation of the CNMLCDS algorithm is available in the R package `cnmlcd` (Liu and Wang, 2015), by function `cnmlcds()`.

Performance measures

To evaluate the performance of a density estimator, the distance between the true density and an estimated one can be used. Here, two distance functions, the integrated squared error (ISE) and Hellinger distance (HD), are used, which are given by

$$\begin{aligned} \text{ISE}(f, \hat{f}) &= \int_{\mathcal{X}} \{f(x) - \hat{f}(x)\}^2 dx, \\ \text{HD}(f, \hat{f}) &= \int_{\mathcal{X}} \{f(x)^{\frac{1}{2}} - \hat{f}(x)^{\frac{1}{2}}\}^2 dx, \end{aligned}$$

where \hat{f} is an estimate of the true density f . We repeated the computation for 100 random samples drawn from each distribution with a sample size $n = 1000$, and calculated the mean integrated squared error (MISE) and mean Hellinger distance (MHD).

Results

The estimated MISE and MHD along with their standard deviations are reported in Table 4.2. It shows that the PQ1 estimator performs best in both scenarios in terms of both MISE and MHD (highlighted in boldface). This is not surprising since data conforms well with PQ1 and PQ1 is the simplest model of the three. It is obvious that if more information is known about the underlying distribution, more restrictions can be correctly imposed on a density estimator, which would make the estimation more accurate. PQ3 and PQ4 are more general and flexible than PQ1, and still performed generally well; see the boxplots in Figure 4.4. For this reason, we will only use PQ3 and PQ4 in the remaining numerical studies. Note also that there is little difference in the performance of the PQ3 and PQ4 estimators in both scenarios.

Tab. 4.2: Simulation results for two distributions in terms of the MISE ($\times 10^{-3}$) and MHD ($\times 10^{-3}$), with standard deviations given in parentheses.

Estimator	Distribution			
	TN(0, 1, $-\infty$, 0)		SN(0, 1, -3)	
	MISE	MHD	MISE	MHD
PQ1	2.63 (1.47)	1.51 (0.82)	0.94 (0.74)	1.13 (0.69)
PQ3	3.04 (1.61)	1.93 (1.00)	1.07 (0.78)	1.37 (0.75)
PQ4	3.07 (1.77)	1.71 (0.94)	1.10 (0.76)	1.30 (0.68)

4.5.2 Against other estimators

Setup

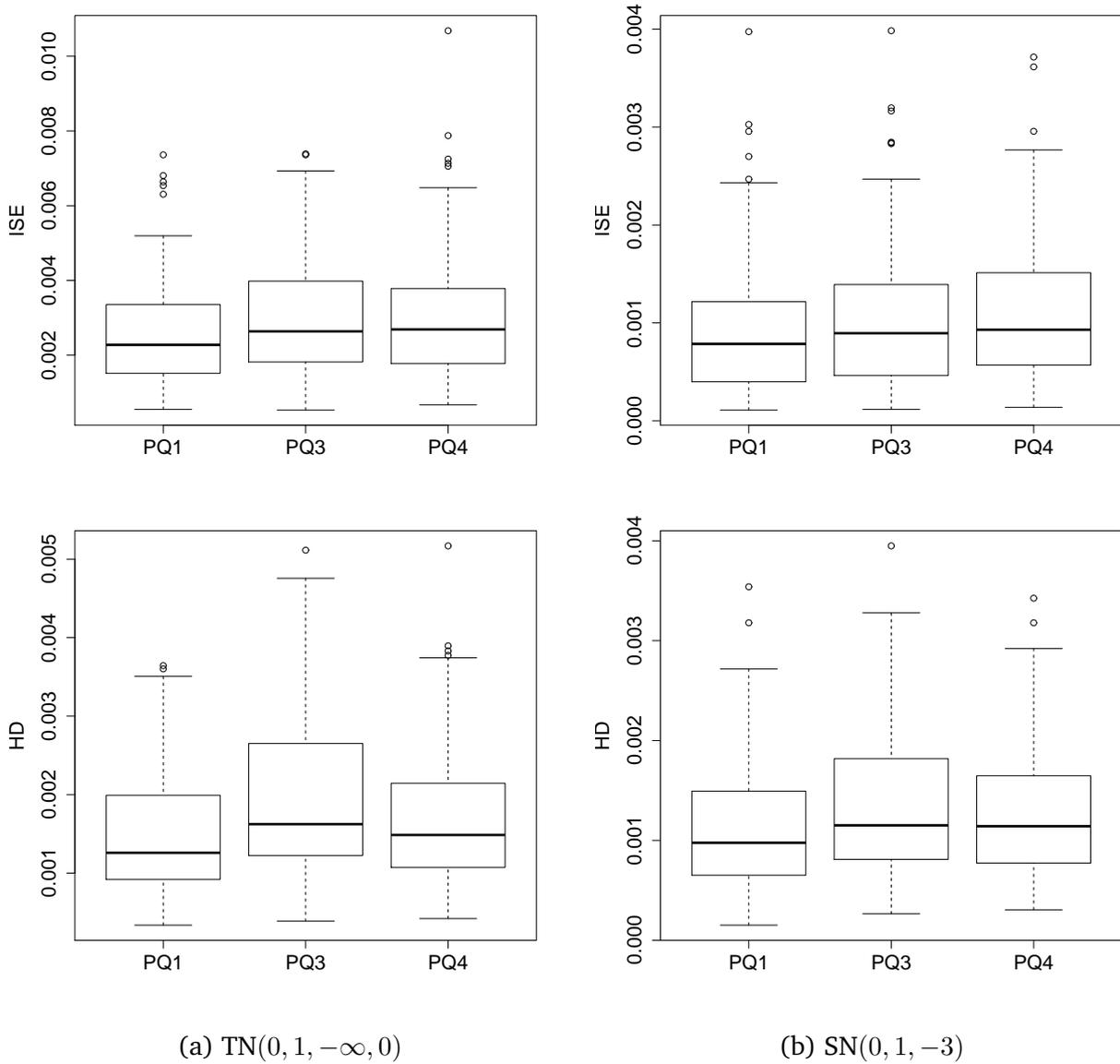


Fig. 4.4: Boxplots for the ISE (top panel) and HD (bottom panel) of different estimators for a truncated normal and a skew normal.

In this section, another simulation study is carried out to compare the performance of several non-parametric density estimators. Besides PQ3 and PQ4, we also combine PQ3 and PQ4 as estimator PQC which simply takes the better estimate of PQ3 and PQ4, i.e., the one with a larger likelihood value. This combined estimator is more flexible and convenient for an end user. Clearly, PQC is the maximum likelihood estimator among all in $\mathcal{K}_{ID} \cup \mathcal{K}_{DI}$, which contains \mathcal{K}_D and \mathcal{K}_I . This is the family of log-concave densities, with the second derivative of the log-density function changing monotonicity at most once on its domain \mathcal{X} .

Two other existing estimators are included: the piecewise linear estimator (PL) as computed

by the algorithm proposed in Chapter 3, and the smooth log-concave density estimator (LCDS) of Dümbgen and Rufibach (2009). Furthermore, in our numerical study, we shall not consider estimators which are for unimodal density estimation; see, e.g., Meyer (2012). This is because estimators under the log-concavity restriction would naturally outperform those not under the restriction, in the situation where underlying densities are log-concave.

The PL and PQC estimators are computed by functions `cnmlcd()` and `cnmlclds()`, respectively, in the R package `cnmlcd` (Liu and Wang, 2015) with default settings. To obtain a smooth log-concave density estimator, Dümbgen and Rufibach (2009) consider the convolution between the NPMLE and a Gaussian density. An implementation of their algorithm is available in the R package `logcondens` (Dümbgen and Rufibach, 2011), by function `logConDens()`. It first estimates the NPMLE by using the active set algorithm (ASA) of Dümbgen et al. (2007), and then conducts the convolution. The argument `smoothed` in `logConDens()` is set to `TRUE`. The other arguments of the relevant functions take their default values.

In order to compare the performance of the estimators, four distributions are considered here, the exponential (EP), gamma (GM), beta (BT) and logistic (LG) distributions, whose densities are given in Table 4.3. Note that the exponential and logistic distributions are always log-concave, the gamma distribution is log-concave for $k \geq 1$ and the beta distribution is log-concave for $\alpha \geq 1$ and $\beta \geq 1$. Figure 4.5 shows the log-density and its second derivative in each of the four cases to be studied.

Tab. 4.3: Four distributions of different types.

Distribution	Notation	Density Function	Range	Parameters
Exponential	$EP(\lambda)$	$\lambda e^{-\lambda x}$	$x \geq 0$	$\lambda > 0$
Gamma	$GM(k, \theta)$	$x^{k-1} e^{-x/\theta} / \{\theta^k \Gamma(k)\}$	$x > 0$	$\theta, k > 0$
Beta	$BT(\alpha, \beta)$	$x^{\alpha-1} (1-x)^{\beta-1} / B(\alpha, \beta)$	$x \in (0, 1)$	$\alpha, \beta > 0$
Logistic	$LG(\mu, s)$	$e^{-(x-\mu)/s} / \{s(1 + e^{-(x-\mu)/s})^2\}$	$x > 0$	$\mu \in \mathbb{R}, s > 0$

Results

For each of the cases shown in Figure 4.5, 100 random samples were generated, each of size 1000, and their estimation results are reported in Table 4.4. Each entry in the table is an empirical MISE or MHD value, with its standard error given in parentheses. The smallest of the expected losses of these estimators for each given density is highlighted in boldface. Boxplots for the ISE and HD of different estimators are also shown in Figures 4.6 and 4.7. The estimated densities and log-

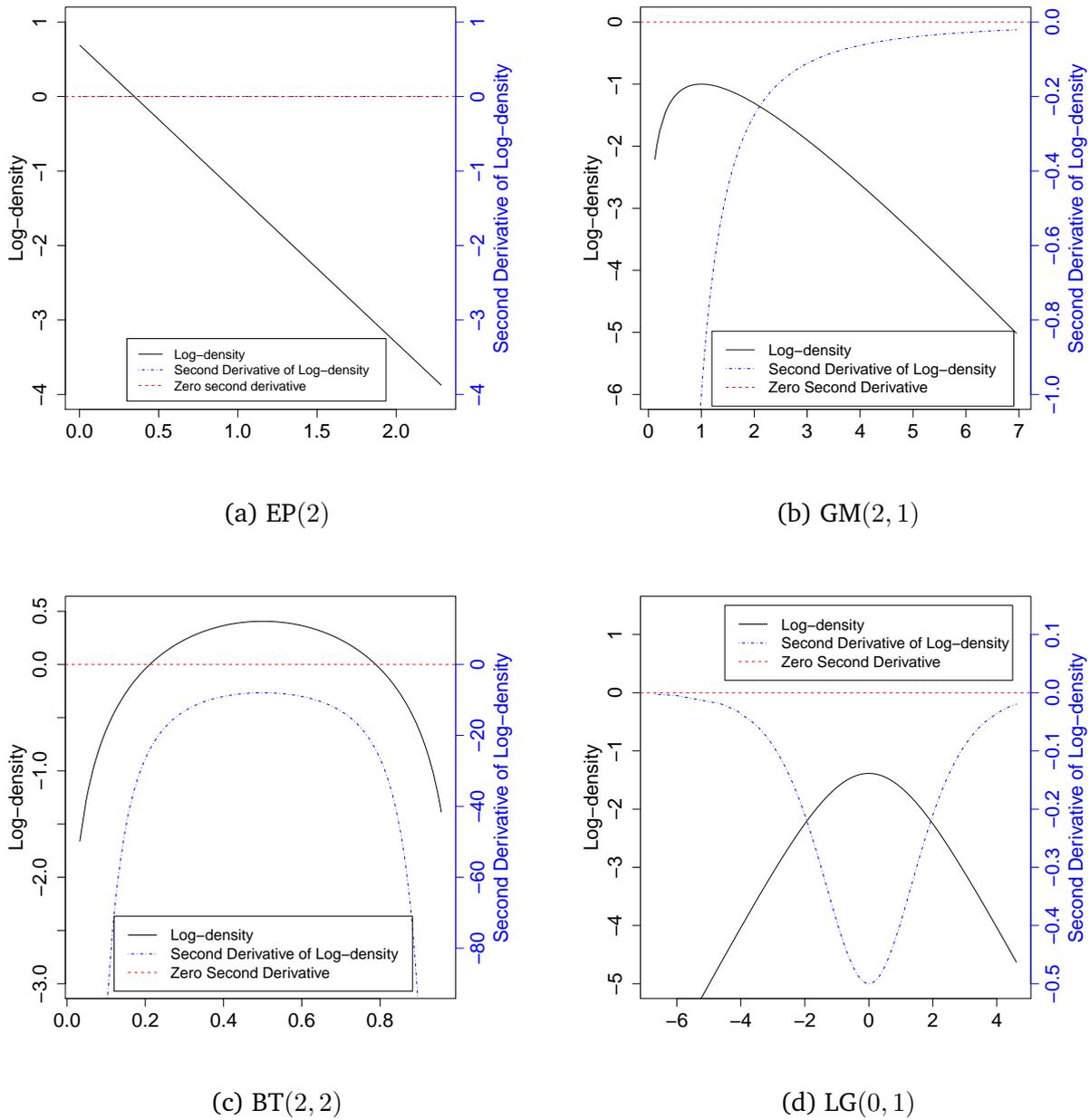


Fig. 4.5: Log-density and the second derivatives of exponential, gamma, beta and logistic distributions.

densities are displayed in Figures 4.8 and 4.9, each for a typical sample of size 1000, along with the true distribution (Truth).

These results show that the smooth estimators generally outperform the piecewise linear estimator. Of the four smooth estimators, the three new ones outperform the LCDS, except in one case, BT(2, 2) in terms of MHD, where the LCDS is marginally better than PQ3 (and PQ3). It is worth pointing out that the shape restriction imposed by estimator PQ3 is not appropriate for LG(0, 1),

4.6. Real-world Data

Tab. 4.4: Simulation results for four scenarios in terms of the MISE ($\times 10^{-3}$) and MHD ($\times 10^{-3}$), with standard errors given in parentheses.

<i>Estimator</i>	<i>Distribution</i>			
	EP(2)	GM(2, 1)	BT(2, 2)	LG(0, 1)
MISE				
PL	4.16 (0.28)	1.13 (0.06)	5.58 (0.25)	0.73 (0.04)
LCDS	69.66 (2.24)	1.06 (0.06)	4.83 (0.23)	0.60 (0.04)
PQ3	4.01 (0.26)	0.91 (0.05)	4.22 (0.21)	1.08 (0.05)
PQ4	4.52 (0.27)	0.91 (0.05)	20.62 (0.43)	0.53 (0.04)
PQC	4.06 (0.26)	0.91 (0.05)	4.22 (0.21)	0.53 (0.04)
MHD				
PL	2.62 (0.13)	3.60 (0.15)	3.69 (0.15)	3.44 (0.15)
LCDS	13.94 (0.43)	2.86 (0.10)	1.63 (0.06)	2.85 (0.13)
PQ3	1.16 (0.07)	1.75 (0.09)	1.74 (0.08)	4.47 (0.06)
PQ4	1.10 (0.07)	1.57 (0.08)	6.79 (0.11)	1.44 (0.07)
PQC	1.15 (0.07)	1.75 (0.09)	1.74 (0.08)	1.44 (0.07)

hence its performance is understandably inferior to the others. This is also the case when estimator PQ4 is applied to BT(2, 2). This is exactly the reason that we propose the combined estimator PQC, which chooses the better of the two and performs well in all cases in terms of both MISE and MHD. Note that LCDS performs worst in the case of EP(2) which is extremely right skewed. This can perhaps be attributed to the chosen bandwidth. A fixed bandwidth may not be large enough for the tail, or small enough for the central part. Figures 4.8 and 4.9 further show that, except in the case of BT(2, 2), LCDS always fails to estimate the tail well, whereas estimator PQC performs very well.

4.6 Real-world Data

Real-world datasets are used to further study the estimators. The first real dataset comprises the reliability data from a consulting case for predicting the reliability of a certain device. The other three are the six-month, one-year and ten-year daily log returns, $\log(P_{i+1}/P_i)$, where P_i is the closing price of day i of S&P 500 index, retrieved from Yahoo Finance.

Setup



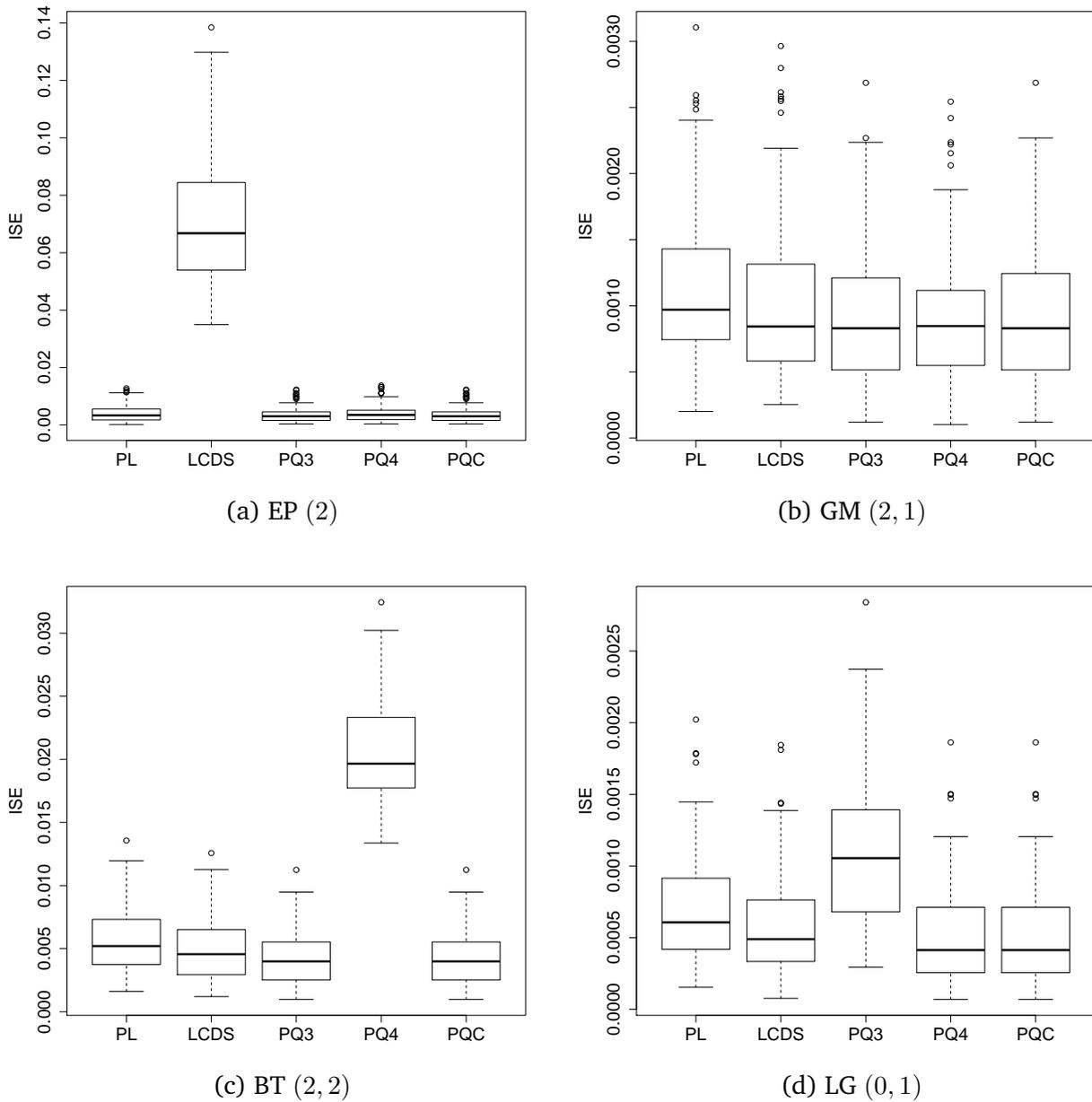


Fig. 4.6: Boxplots for the ISE of different estimators for exponential, gamma, beta and logistic distributions.

In this study, we compare the performance of three estimators: PL, LCDS and PQC. The QQ and density plots are used here for inspecting how closely an estimator fits to a dataset and for identifying the differences among the three estimators, because visual inspection is usually an excellent place to start with. Despite this, we will also look at more quantitative approaches to assess their performance. However, in the empirical studies for real-world data, we do not know the true underlying density function, and thus the loss functions given in Section 4.5.1 can not be exactly utilized. As a substitute, we modify the loss functions by replacing the true density f with the

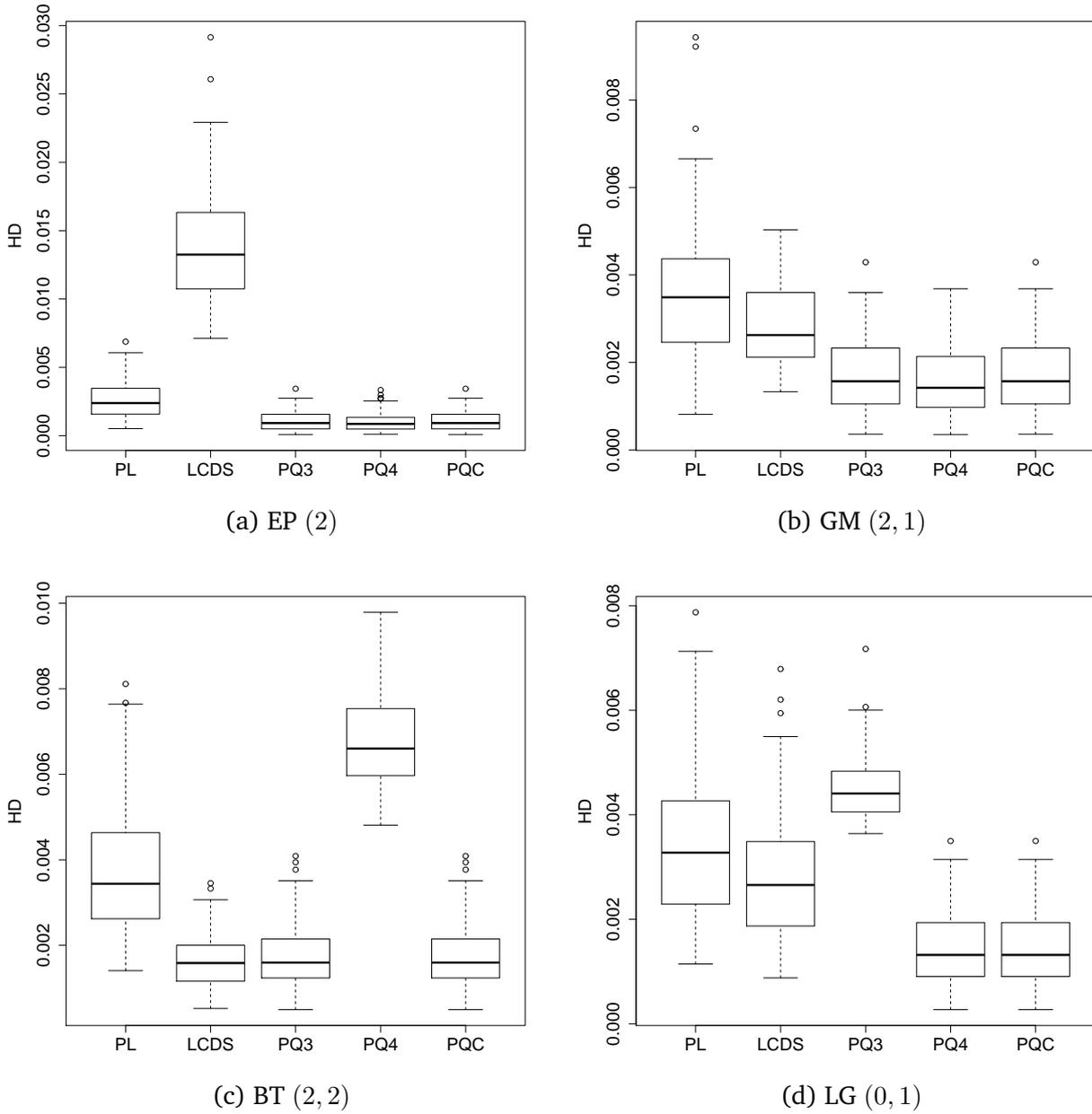


Fig. 4.7: Boxplots for the HD of different estimators for exponential, gamma, beta and logistic distributions.

empirical probability mass function \hat{f}_n , for a test set of size n . Two loss functions, the ISE and Kullback-Leibler (KL) divergence (Novak, 2011), are thus given by, respectively,

$$\text{ISE}(\hat{f}_n, \hat{f}) = \int_{\mathcal{X}} \{\hat{f}(x)\}^2 dx - \frac{2}{n} \sum_{i=1}^n \hat{f}(x_i),$$

$$\text{KL}(\hat{f}_n, \hat{f}) = -\frac{1}{n} \sum_{i=1}^n \log\{\hat{f}(x_i)\},$$

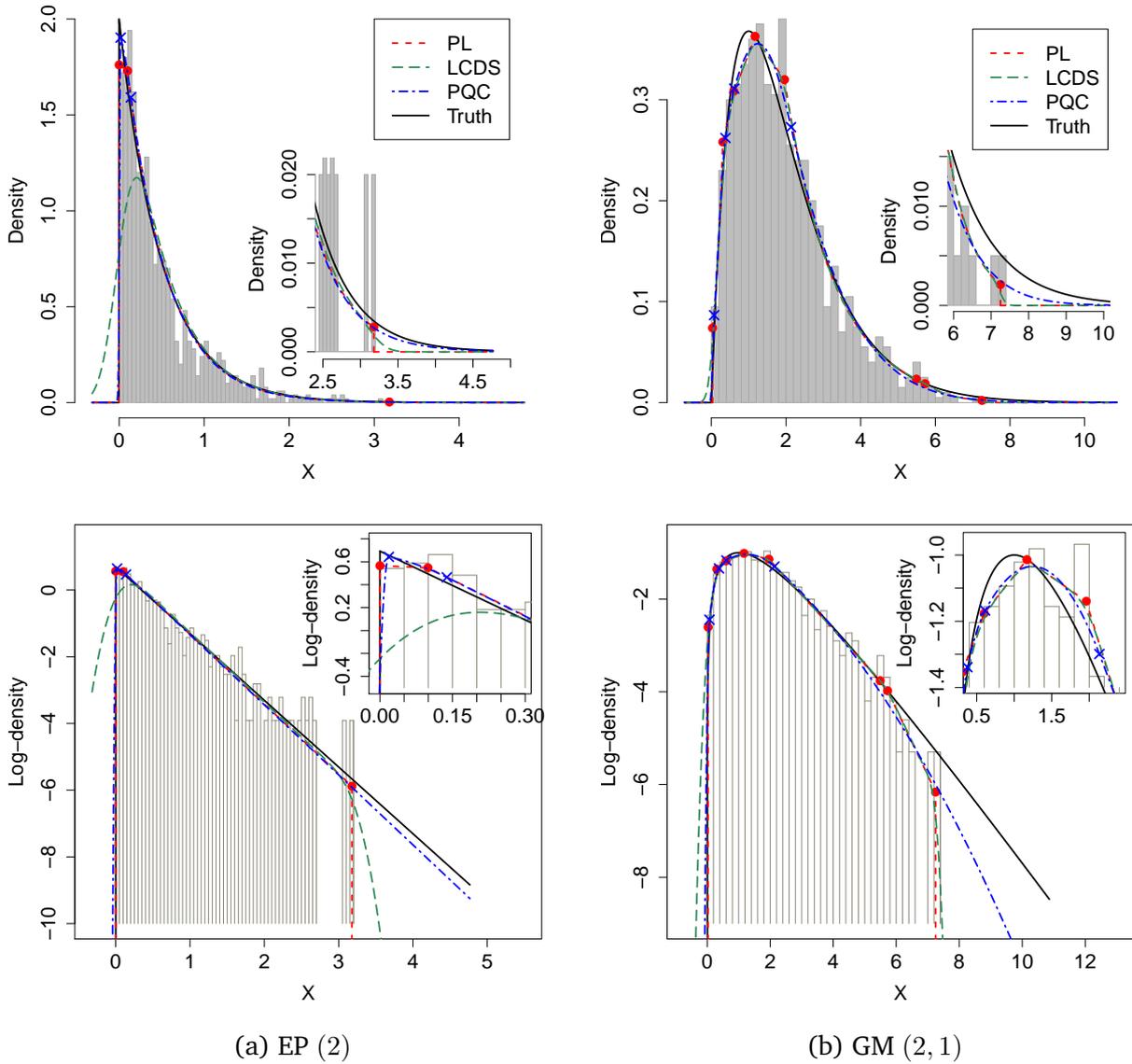


Fig. 4.8: Histograms and log-concave density estimates for exponential and gamma distributions. The solid and cross points represent the knots of the PL and PQC estimates, respectively.

where \hat{f}_n denotes the empirical mass function from a test set of size n and \hat{f} is a density estimator based on a training set. Additive constants are excluded from the above functions.

Furthermore, the Kolmogorov-Smirnov and Anderson-Darling statistics (Anderson and Darling, 1954) are also introduced for evaluating the performance. These two statistic-based criteria are

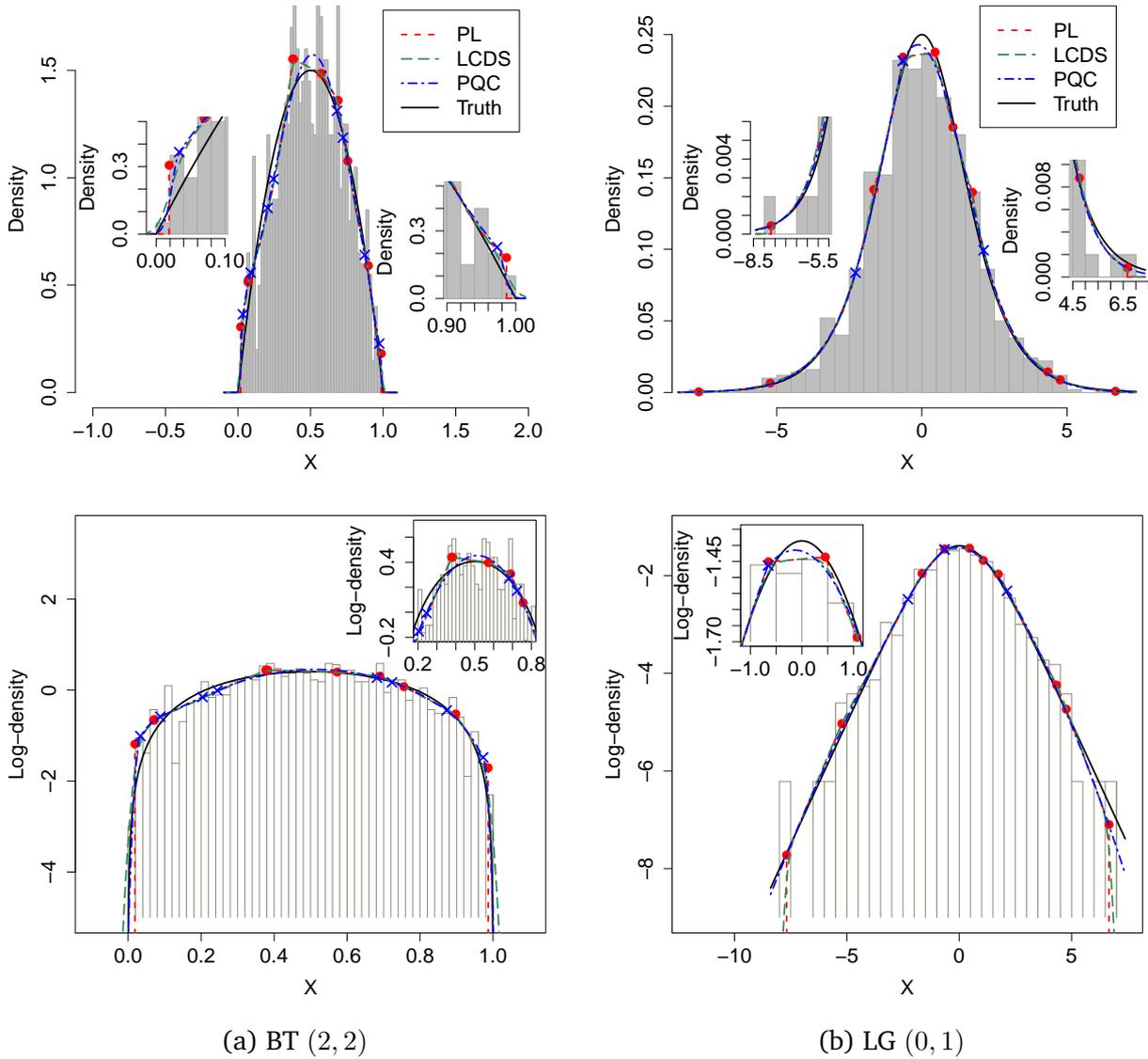


Fig. 4.9: Histograms and log-concave density estimates for beta and logistic distributions. The solid and cross points represent the knots of the PL and PQC estimates, respectively.

given by, respectively,

$$KS(\hat{F}_n, \hat{F}) = \sup_{\mathcal{X}} |\hat{F}_n(x) - \hat{F}(x)|,$$

$$AD(\hat{F}_n, \hat{F}) = n \int_{-\infty}^{+\infty} \frac{\hat{F}_n(x) - \hat{F}(x)}{\hat{F}(x)(1 - \hat{F}(x))} d\hat{F}(x),$$

where \hat{F}_n denotes the empirical distribution function from a test set of size n and \hat{F} is the CDF of an estimate from a training set.

4.6.1 Reliability data

The reliability data consists of 786 observations, as also used by [Dümbgen and Rufibach \(2011\)](#). It was collected as part of a consulting project at the Institute for Mathematical Statistics and Actuarial Science at the University of Bern. Some descriptive statistics of this dataset are given in [Table 4.5](#), where SD stands for the standard deviation. The mean is obtained by R function `mean()`, standard deviation by `sd()`, and skewness and kurtosis by, respectively, `skewness()` and `kurtosis()` in R package `HyperbolicDist`. It can be seen that the reliability data is left skewed and has a similar kurtosis to that of the standard normal distribution.

Tab. 4.5: Descriptive statistics of reliability data.

<i>n</i>	<i>Mean</i>	<i>SD</i>	<i>Skewness</i>	<i>Kurtosis</i>
786	1681.07	76.28	-0.50	2.95

[Table 4.6](#) lists several p -values, whose meanings have been described in [Section 4.4](#). Each p -value is based on 9999 bootstrap samples. The value of p_{ks} 0.62 indicates that it seems safe to assume that the underlying density is log-concave. Other p -values further show that the PL estimator fits to this data well.

Tab. 4.6: p -values from bootstrap test for reliability data.

<i>Variable</i>	p_{ks}	p_{mean}	p_{sd}	p_{skew}	p_{kurt}
reliability	0.618	0.990	0.781	0.683	0.332

Results

[Figure 4.10](#) shows the log-concave density estimate and the gradient function at the zeroth, fifth and final iteration, respectively, as computed by CNMLCDS. Note that due to the constraint $\alpha_2 \geq 0$, the negative value of $d_{0,2}$ at convergence implies that $\widehat{\alpha}_2$ is zero. Furthermore, the piecewise quadratic log-density function for reliability data is given by

$$\widehat{\varphi}(x) = 0.015x - 0.00013(1510.68 - x)_+^2 - 0.00010(1513.73 - x)_+^2 - 0.000035(x - 1561.37)_+^2 - 0.000081(x - 1678.38)_+^2 - 0.00025(x - 1775.89)_+^2 - 0.00079(x - 1835.63)_+^2 - 29.37.$$

The QQ and density plots for the PL, LCDS and PQC estimates are displayed in [Figures 4.11](#) and [4.12](#), respectively. From the QQ plots, one can see that all three estimators fit well to the data and exhibit little difference among them. However, from the density plots, differences can be seen at

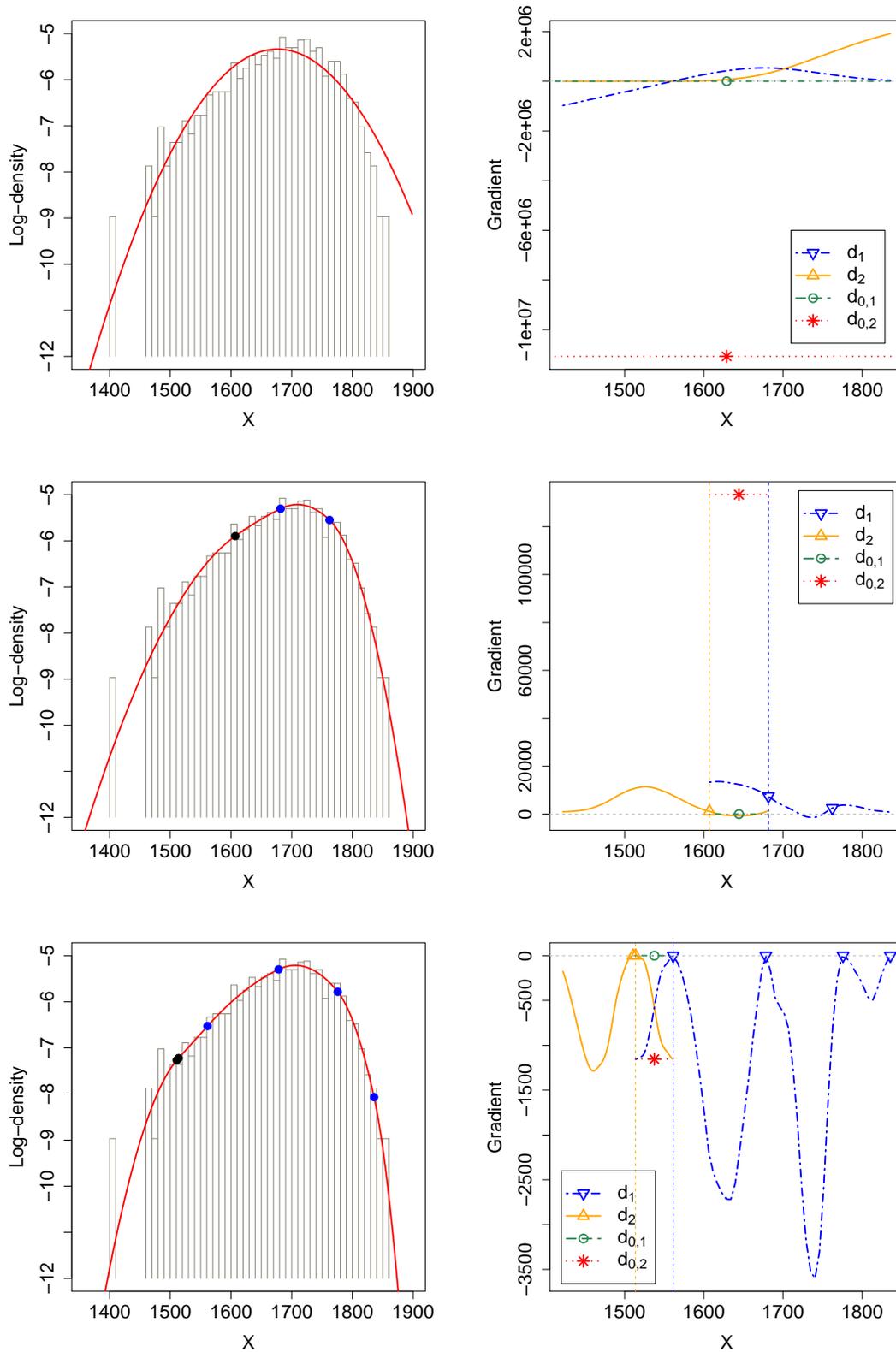


Fig. 4.10: Log-density plots (left panels) with gradient curves (right panels) for the reliability data, which correspond to the zeroth (top panels), the fifth (middle panels) and the final (bottom panel) iteration of CNMLCDS, with knots indicated by solid points.

the peak and tails. The PQC estimate has a slightly lower peak and heavier tails than those of the PL and LCDS estimates.

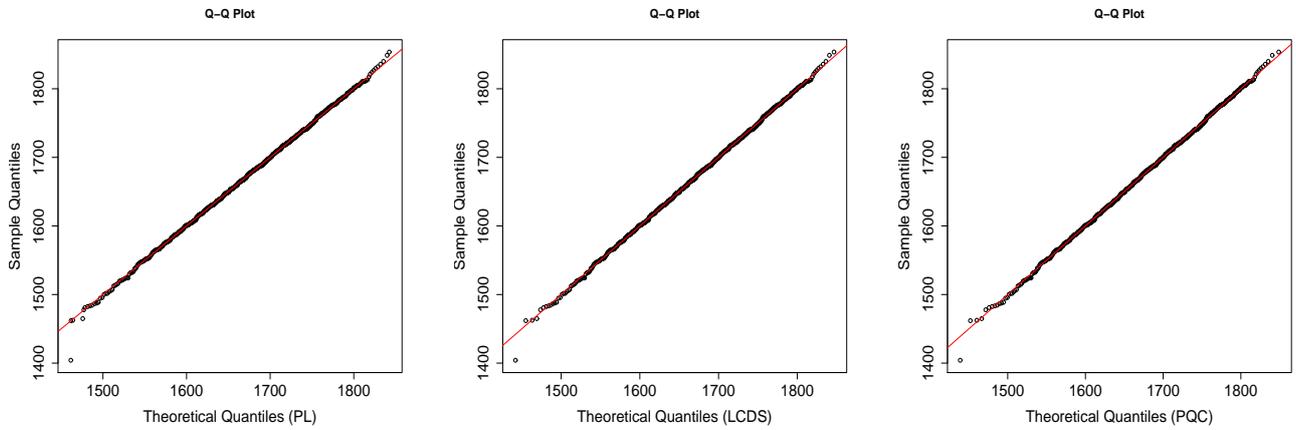


Fig. 4.11: Q-Q plots for the reliability data from different estimates: the PL (right panel), LCDS (middle panel) and PQC (left panel).

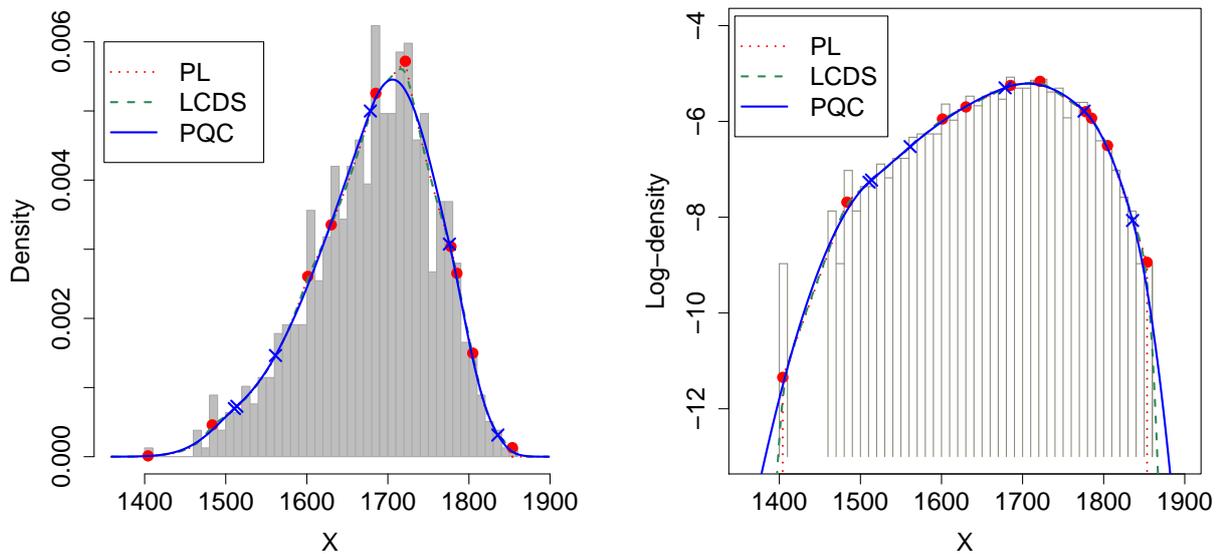


Fig. 4.12: Histograms and log-concave density estimates for the reliability data. The solid and cross points represent the knots of the PL and PQC estimates, respectively.

In order to further investigate the performance of different estimators for this dataset, we run 2-fold cross-validation, and conduct the paired t -test over 100 replications. The results are summarised in Table 4.7. The p -values less than 5% are highlighted in boldface. Note that the PL estimates may give infinite values of KL and AD. The reason is that the PL estimate gives zero mass

outside the range of the training data. Generally speaking, the smooth estimators are better than the piecewise linear estimator, especially in terms of KL and AD. Furthermore, the PQC outperforms the LCDS by all criteria. In this case, the LCDS chooses a very small bandwidth to avoid putting too much mass on the tails (Dümbgen et al., 2011). However, too small a bandwidth leads to poor estimation around the tails.

Tab. 4.7: Paired t -test based on cross-validation results for the reliability data, where MD stands for the mean difference.

(I) Est.	(J) Est.	MD (I-J)	Std. Error	95% Confidence Interval		p -value
				Lower Bound	Upper Bound	
ISE ($\times 10^{-4}$)						
PQC	PL	-11.469	1.137	-13.725	-9.213	$< 2.2 \times 10^{-16}$
	LCDS	-5.792	0.811	-7.402	-4.182	1.6×10^{-10}
LCDS	PL	-5.677	0.447	-6.563	-4.790	$< 2.2 \times 10^{-16}$
KL						
PQC	PL	$-\infty$	-	-	-	0
	LCDS	-0.033	0.003	-0.040	-0.026	9.7×10^{-16}
LCDS	PL	$-\infty$	-	-	-	0
KS ($\times 10^{-4}$)						
PQC	PL	-10.560	2.294	-15.118	-6.012	1.2×10^{-5}
	LCDS	-6.511	1.878	-10.237	-2.784	7.8×10^{-4}
LCDS	PL	-4.054	1.199	-6.433	-1.675	1.0×10^{-3}
AD						
PQC	PL	$-\infty$	-	-	-	0
	LCDS	-0.052	0.016	-0.084	-0.019	2.0×10^{-3}
LCDS	PL	$-\infty$	-	-	-	0

4.6.2 Log-return data

Three financial datasets are analysed here and they are the daily log-returns of S&P 500 index from 09 October 2014 to 09 April 2015, 02 January 2014 to 31 December 2014 and 02 January 2004 to 31 December 2014, respectively, i.e., for six months, one year and ten years, respectively. In finance, a return is the profit on an investment. There are many types of returns, such as net return, gross return and rate of return. From the theoretical and algorithmic aspect, the logarithmic return is the most commonly-used in the analysis of financial time series data. Furthermore, the S&P 500 index

is considered one of the best indicators of the state of American market and economy (Markowitz and Usmen, 1996). Table 4.8 provides some descriptive statistics of the datasets. They indicate that the six-month data does not demonstrate skewness or heavy-tailedness given the smaller size of the sample, but the other two are clearly left skewed and heavy-tailed, as compared with the normal distribution. It also suggests that the longer period the data spans, the higher the kurtosis, or the heavier the tails.

Tab. 4.8: Descriptive statistics of log returns.

<i>Data</i>	<i>n</i>	<i>Mean</i> ($\times 10^{-3}$)	<i>SD</i> ($\times 10^{-3}$)	<i>Skewness</i>	<i>Kurtosis</i>
Six-month	124	0.654	8.367	0.075	2.976
One-year	252	0.428	7.170	-0.424	4.269
Ten-year	2769	0.222	12.464	-0.338	14.568

The p -values for these three datasets are given in Table 4.9, as computed in the same way as for the reliability data in Section 4.6.1. It can be seen that at the 5% significance level, it is reasonable to assume log-concavity for the six-month ($p_{ks} = 0.685$) and one-year ($p_{ks} = 0.083$) datasets. However, we detected a significant departure from log-concavity for the ten-year dataset. The small value of p_{ks} (< 0.001) implies that a log-concave density estimator is not suitable for this ten-year log-return data. Although the log-concave density estimator PL is well fitted for its mean and skewness ($p_{mean} = 0.992$ and $p_{skew} = 0.712$, respectively), it can not capture the large standard deviation and the high kurtosis ($p_{sd} < 0.001$ and $p_{kurt} < 0.001$, respectively).

Tab. 4.9: p -values from the bootstrap test for log-return data.

<i>Data</i>	p_{ks}	p_{mean}	p_{sd}	p_{skew}	p_{kurt}
Six-month	0.685	0.995	0.662	0.815	0.467
One-year	0.083	0.990	0.512	0.547	0.139
Ten-year	< 0.001	0.992	< 0.001	0.712	< 0.001

Results

The gradient function and log-density plots at the convergence of the CNMLCDS algorithm for the three datasets are shown in Figure 4.13. The values of $d_{0,1}$, $d_{0,2}$ and the maximum of the gradient functions are all zero, which indicates that CNMLCDS has successfully found the SNPMLE for each

data. Furthermore, for the six-month data, the log-density is given by

$$\hat{\varphi}(x) = -70.46x - 196711x^2 + 192847(-0.00040 - x)_+^2 + 192883(x + 0.00002)_+^2 + 4.10.$$

For the one-year and ten-year data, they are given by, respectively,

$$\begin{aligned} \hat{\varphi}(x) &= 71.75x - 280780x^2 + 280567(-0.00018 - x)_+^2 + 265277(x - 0.00014)_+^2 + \\ &14951(x - 0.0077)_+^2 + 4.34, \end{aligned}$$

$$\begin{aligned} \hat{\varphi}(x) &= 300.96x - 123986x^2 + 123986(0.00073 - x)_+^2 + 18569(x - 0.00092)_+^2 + \\ &105417(x - 0.0019)_+^2 + 3.93. \end{aligned}$$

Figure 4.14 shows the QQ plots of the estimates for all datasets. Note that the ten-year data is so heavy-tailed that the log-concavity shape restriction is simply not appropriate. For the six-month and one-year datasets, more details about the differences can be seen in the density and log-density plots shown in Figure 4.15. The major differences among these estimates are at the centre and tails.

We also ran 2-fold cross-validation, and conducted the paired t -test over 100 replications to obtain the results in Table 4.10 for the six-month data and in Table 4.11 for the one-year data. The p -values less than 5% are highlighted in boldface. Note that for both datasets, the infinite KL and AD values of the PL estimates are due to the same reason as discussed in Section 4.6.1.

From Table 4.10, for the six-month data set we can see that the PQC is similar to the LCDS, only slightly worse in one out of four measures. For the one-year data set, as summarized in Table 4.11, the LCDS performs worst in terms of ISE and KS. It is also significantly worse than the PQC in terms of AD. The major reason may still be due to the bandwidth selected. In order to smooth out the discontinuities near the boundaries of the data (x_1 and x_n), the LCDS needs to choose a bandwidth for the Gaussian density function. The choice of the bandwidth is guided by that the variance of the LCDS estimate coincides with the unbiased estimate of the variance of the sample. If the sample shows no heavy tails, the bandwidth chosen will be small enough to capture the important feature in the main body and avoid putting too much mass at the tails. If the sample shows heavy tails and a high peak, a relatively large bandwidth will be chosen to give sufficient mass for the tails, at the sacrifice of estimation at the main body. In such a case, it is very difficult to choose one bandwidth that is suitable for both the center and the tails of a distribution. In fact, the log-returns of most financial assets exhibit outlying observations, which is a characteristic of heavy tails. Accurate estimation of the tails of a distribution is crucial for analysing financial time series data and for risk management.

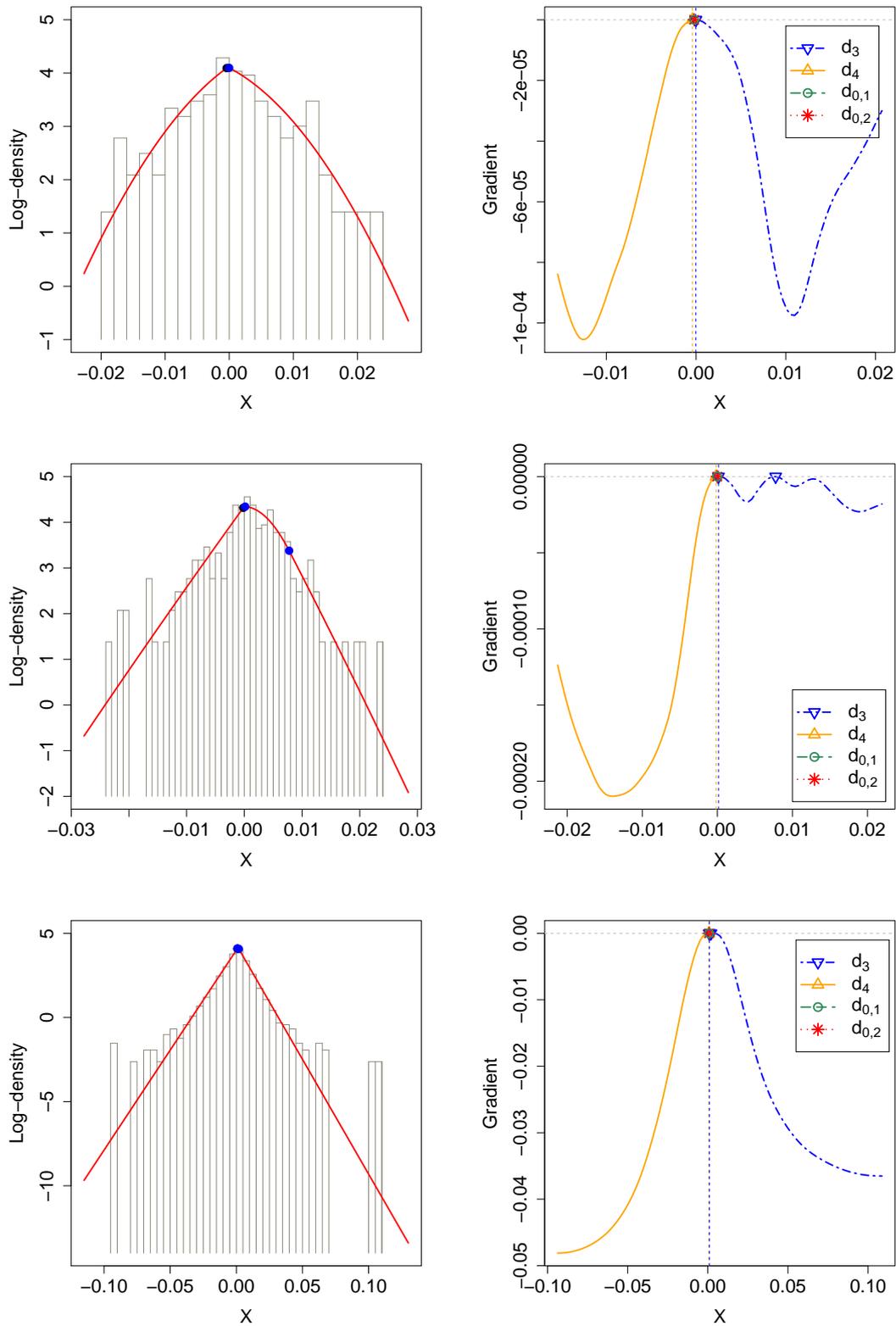


Fig. 4.13: Log-density plots (left panels) with gradient curves (right panels) for the log returns of S&P 500 index: six-month (upper panels), one-year (middle panels) and ten-year (bottom panels) computed by CNMLCDS, with knots indicated by solid points.

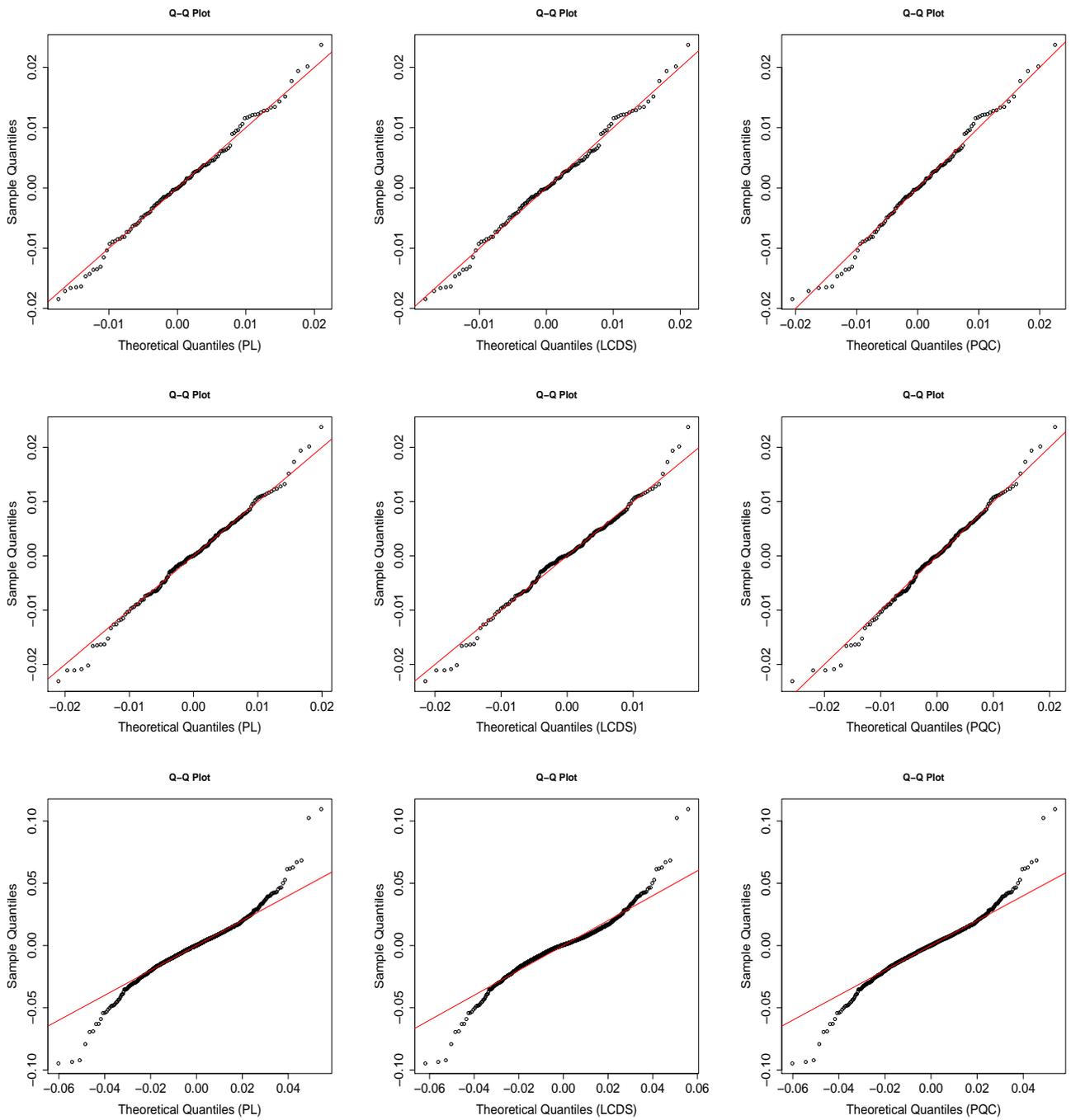


Fig. 4.14: Q-Q plots for the daily log-returns of S&P 500 index: six-month (top panels), one-year (middle row panels and ten-year (bottom panels), from different estimates: the PL (left panels), LCDS (middle column panels) and PQC (right panels).

4.6.3 Timings

In this section, we simply give the speed of the CNMLCDS algorithm for computing the smooth estimator PQC. Table 4.12 gives the CPU times taken by the CNMLCDS for three real-world datasets.

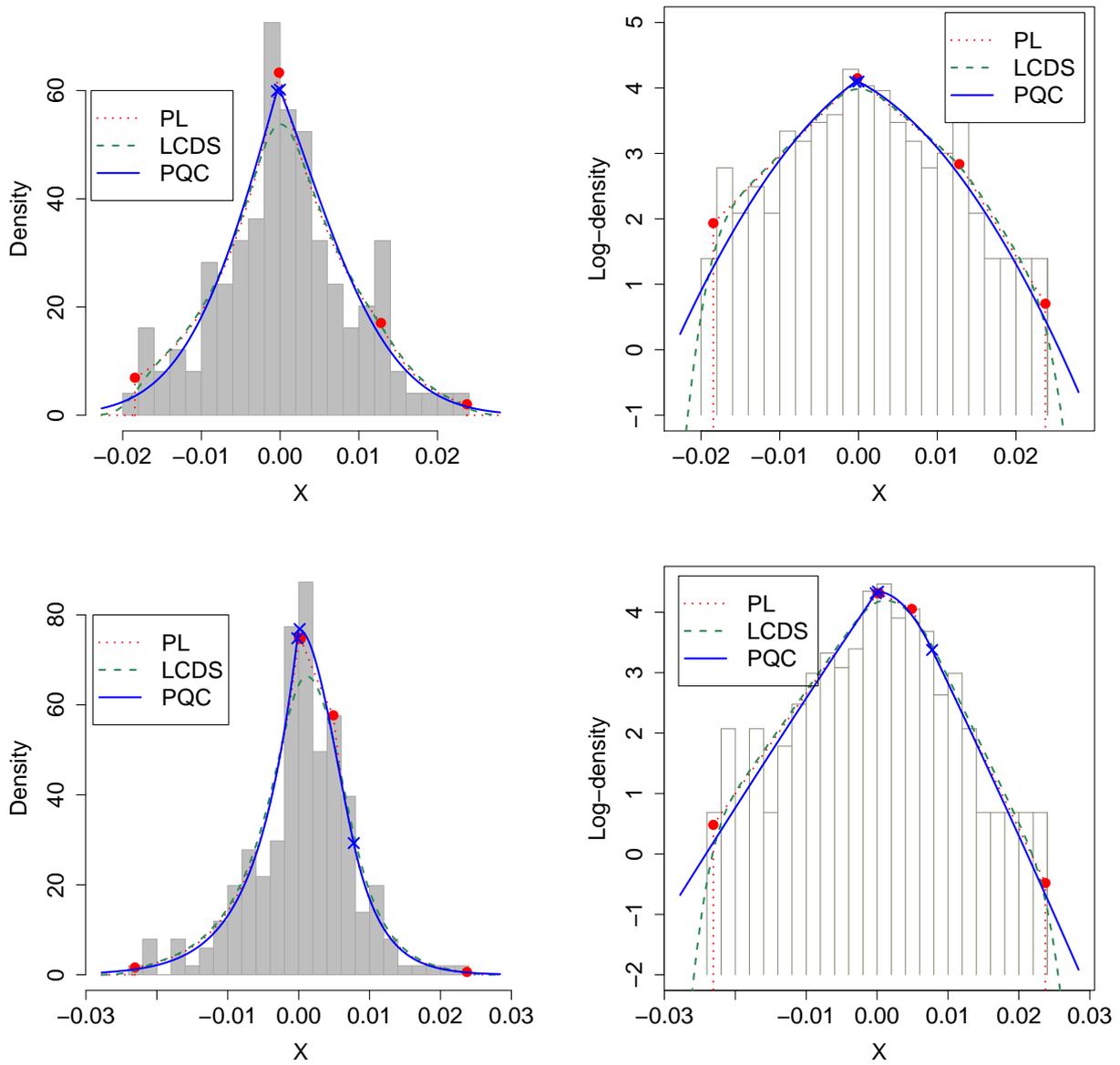


Fig. 4.15: Histograms and log-concave density estimates for the daily log returns of S&P 500 index: six-month (upper panels) and one-year (lower panels). The solid and cross points represent the knots of the PL and PQC estimates, respectively.

Compared with those given in Section 3.5.3, the CNMLCDS algorithm for computing estimator PQC is understandably slower than that of the CNMLCD algorithm for computing the NPMLE of a log-concave density function. However, the time required by the CNMLCDS is still not too much.

Note that the LCDS estimate is the convolution between the NPMLE of a log-concave density function and a Gaussian density, whose standard deviation is obtained in a special way. Once the NPMLE is found, the LCDS estimate is defined at the same time. From this point of view, the time required by ASA or our CNMLCD for computing the NPMLE is much less than that required by

4.7. Summary

Tab. 4.10: Paired t -test based on cross-validation results for the estimators for the six-month log-return data set, where MD stands for mean difference.

(I) Est.	(J) Est.	MD (I-J)	Std. Error	95% Confidence Interval		p -value
				Lower Bound	Upper Bound	
ISE						
PQC	PL	0.012	0.081	-0.149	0.174	0.88
	LCDS	0.277	0.128	0.023	0.532	0.03
LCDS	PL	-0.265	0.082	-0.429	-0.101	2.0×10^{-3}
KL $\times 10^{-3}$						
PQC	PL	$-\infty$	-	-	-	0
	LCDS	0.402	3.526	-6.593	7.398	0.91
LCDS	PL	$-\infty$	-	-	-	0
KS ($\times 10^{-3}$)						
PQC	PL	1.133	0.708	-0.272	2.537	0.11
	LCDS	1.684	1.286	-0.867	4.235	0.193
LCDS	PL	-0.551	0.871	-2.280	1.177	0.53
AD						
PQC	PL	$-\infty$	-	-	-	0
	LCDS	0.094	0.051	-0.007	0.195	0.07
LCDS	PL	$-\infty$	-	-	-	0

CNMLCDS for the PQC. However, it takes a much longer time for an LCDS estimate than for a PQC estimate to evaluate the density value at an arbitrary point. The reason is that the convolution required by the LCDS estimate for density evaluation is very costly, while one computes a density value straight away, given a PQC estimate.

4.7 Summary

In this chapter, we studied the smooth nonparametric estimation of a log-concave density function. Several new smooth estimators have been proposed and the algorithms for their computation are presented, which extend the CNM method. We further introduced a new log-concavity test based on the KS statistic by applying the piecewise linear estimator, i.e., the nonparametric maximum likelihood estimator.

In the first part of simulation studies, several piecewise quadratic estimators have been analysed

4.7. Summary

Tab. 4.11: Paired t -test based on cross-validation results for the estimators for the one-year log-return data set, where MD stands for mean difference.

(I) Est.	(J) Est.	MD (I-J)	Std. Error	95% Confidence Interval		p -value
				Lower Bound	Upper Bound	
ISE						
PQC	PL	0.061	0.044	-0.028	0.149	0.18
	LCDS	-0.456	0.087	-0.629	-0.283	9.6×10^{-7}
LCDS	PL	0.517	0.075	0.367	0.666	5.9×10^{-10}
KL						
PQC	PL	$-\infty$	-	-	-	0
	LCDS	-0.005	0.014	-0.033	0.226	0.70
LCDS	PL	$-\infty$	-	-	-	0
KS ($\times 10^{-3}$)						
PQC	PL	-0.285	0.573	-1.421	0.852	0.62
	LCDS	-4.187	1.653	-7.467	-0.907	0.01
LCDS	PL	3.903	1.217	1.488	6.318	2.0×10^{-3}
AD						
PQC	PL	$-\infty$	-	-	-	0
	LCDS	-0.127	0.050	-0.226	-0.027	0.01
LCDS	PL	$-\infty$	-	-	-	0

Tab. 4.12: Running times (s) of CNMLCDS for real-world data.

Data	Reliability	Six-month	One-year
Time	0.64	0.31	0.54

and a combined estimator has been proposed for its flexibility and convenience from an end user's point of view. In the second part, three estimators, PL, LCDS and PQC, have been compared for density estimation. Simulation results indicate that in general, the smooth estimators have better performance than the piecewise linear one, and the PQC estimator outperforms the LCDS. While the LCDS can fit well to some data, such as BT(2, 2) and GM(2, 1), its performance deteriorates when data is extremely skewed or has even slightly heavy tails, such as EP(2) and LG(0, 1). Our study further confirm that wrong shape restrictions lead to bad results, and correct restrictions give more accurate ones.

For the real-world datasets, four assessment criteria have been employed to evaluate the per-

formance of different density estimators. Generally speaking, the PL estimator performs worst with little prediction power at the tails. Both the LCDS and PQC can adaptively capture the skewness, but the LCDS is vulnerable to high kurtosis, which is likely attributable to the bandwidth chosen. One common bandwidth chosen by the LCDS for all the data may not be appropriate everywhere. By contrast, the PQC estimator has no tuning parameter.

Furthermore, we also realized that log-concavity is not a good assumption for the ten-year log-returns dataset which has a very high kurtosis, owing to extremely heavy tails; see also [Meyer \(2012\)](#). Chapter 6 will look into this problem.

Chapter 5

An Application of Log-concave Density Estimation: ROC Curve Estimation

5.1 Introduction

The receiver operating characteristic (ROC) curve is a graphical plot that illustrates the performance of a binary classifier system as its discrimination threshold varies (Collinson, 1998). The curve is about how the sensitivity (the true positive rate) varies against one minus specificity (the false positive rate) at various threshold settings. The ROC curve was firstly developed for detecting enemy objects in battlefields by electrical and radar engineers. ROC analysis since then has been used in psychology, medicine, finance, biometrics and other areas for many decades. It is also increasingly used in machine learning and data mining research; see Gonçalves et al. (2014) and the references therein. Nowadays, ROC analysis has been widely employed in medical setting, especially in diagnostic medicine, where the ROC curves are the most common tools to represent the diagnostic accuracy of a diagnostic test with continuous outcome.

Let X and Y be two independent random variables, denoting the test results for, respectively, a non-diseased subject and a diseased one. Denote by F_0 and F_1 the distribution functions of X and Y , respectively. Then, the ROC curve of the test is a plot of $1 - F_1(t)$ versus $1 - F_0(t)$, for $-\infty < t < \infty$, and can be written as

$$R(u) = 1 - F_1(F_0^{-1}(1 - u)), \quad (5.1)$$

where F_0^{-1} is the inverse function of F_0 and $u \in (0, 1)$ is the false-positive rate corresponding to a cut-off point for positivity.

A fundamental property of an ROC curve of a test is that it must be invariant to any monotone increasing transformation of test results. Hence, any sensible estimation methods should have this property. There have been many parametric, semi-parametric, and nonparametric methods proposed for estimating an ROC curve. The parametric methods assume that F_0 and F_1 belong to parametric distribution families; see [Zweig and Campbell \(1993\)](#) for a discussion on these methods. However, pure parametric methods sometimes do not possess the invariance property. In order to meet this property, semiparametric methods are an alternative choice. The most prominent semiparametric procedure is to assume a parametric form for the ROC curve, but avoids making any additional parametric assumptions about the distribution of test results ([Cai and Moskowitz, 2004](#)). This type of semiparametric methods has the parametric form as

$$R(u) = G(a + bH^{-1}(u))$$

where G and H are some known cumulative distribution functions. The most common choice is the binormal form, that is, $G = H = \Phi$, where Φ is the cumulative distribution (CDF) of a standard normal distribution; see [Metz et al. \(1998\)](#) and [Zhou and Lin \(2008\)](#). Under the binormal models, researchers have made great efforts to estimate the parameters a and b ; see [Metz et al. \(1998\)](#), [Alonzo and Pepe \(2002\)](#), [Pepe and Cai \(2004\)](#) and [Cai and Moskowitz \(2004\)](#). [Zhou and Lin \(2008\)](#) discussed some problems in this kind of approaches and proposed a new profile maximum likelihood method to estimate the two parameters a and b . However, the binormal ROC model sometimes produces inappropriate fits that cross the chance line with degenerate datasets ([Dorfman et al., 1997](#); [Hughes and Bhattacharya, 2013](#)). Furthermore, semiparametric methods may be sensitive to the assumptions and can only provide a limited range of distributional forms.

As for nonparametric methods, the simplest and most commonly used one is the empirical estimator, which is based on plugging empirical estimates into function (5.1). Let x_1, \dots, x_{n_0} be the test responses of the non-diseased subjects and y_1, \dots, y_{n_1} be the test responses of the diseased subjects. Roughly speaking, the empirical functions of $F_0(\cdot)$ and $F_1(\cdot)$ can be defined as, respectively,

$$\bar{F}_0(t) = \#\{x_i \leq t\}/n_0 \quad \text{and} \quad \bar{F}_1(t) = \#\{y_i \leq t\}/n_1. \quad (5.2)$$

Then the empirical estimate of the ROC curve is given by

$$\bar{R}(u) = 1 - \bar{F}_1(\bar{F}_0^{-1}(1 - u)), \quad (5.3)$$

where \bar{F}_0^{-1} is the inverse function of \bar{F}_0 by $\bar{F}_0^{-1}(u) = \inf\{t : \bar{F}_0(t) \geq u\}$. Hsieh et al. (1996) provided strong consistency and strong approximation properties of the empirical ROC curve estimator $\bar{R}(u)$. However, one major weakness of the empirical estimator is that it is not smooth, and thus its interpretation becomes more complex (Jokiel-Rokita and Pulit, 2013). In applications, smoothness of the underlying ROC curve may be a sensible assumption. Smoothing methods provide a powerful methodology for gaining insight into data. The kernel smoothing method was introduced by Zou et al. (1997) to overcome the lack of smoothness of the empirical estimator $\bar{R}(u)$. In Zou et al. (1997), both densities F_0 and F_1 are estimated by kernel estimators. Zou's estimator was later improved by Lloyd (1998) by changing the way of choosing bandwidths. Lloyd and Yong (1999) further showed that the kernel estimator of Lloyd (1998) has better performance than the empirical ROC curve in terms of mean square error. However, Zhou and Harezlak (2002) pointed out that choosing bandwidth in Lloyd's estimators is difficult and may require complicated computation. It may not be practical for a routine used in medical studies. Hence, we will not further consider Lloyd's estimators in this chapter.

In the spirit of Lloyd (1998), Rufibach (2012) proposed an alternative ROC curve estimator based on log-concave density estimates initially introduced by Walther (2002) and Rufibach (2007). Rufibach (2012) also showed that the ROC curve estimator of Rufibach (2012), as a process in u , is asymptotically equivalent to the empirical one $\bar{R}(u)$ if the log-concavity assumption holds. In addition, one main advantage of shape constraint estimation in general is that such estimates are fully automatic without choosing any tuning parameter. In this chapter, we focus on estimating the ROC curve based on the assumption that the constituent distributions are both log-concave.

5.2 ROC Curve Estimation Based on Log-concave Density Estimates

Rufibach (2012) modelled the constituent distributions F_0 and F_1 nonparametrically by applying the nonparametric maximum likelihood estimator (NPMLE) of a log-concave density, that is, the piecewise linear estimator (PL) proposed in Chapter 3. The estimates are then plugged into function (5.1) to obtain an estimate of $R(u)$. They also applied the smooth log-concave density estimator (LCDS) of Dümbgen and Rufibach (2009) to the ROC curve estimation in the same way. Rufibach (2012) pointed out that log-concave ROC curve estimates can indeed be biased, but the

bias is in general constrained to regions of $(0, 1)$ that are small, and is typically smaller than that of the binormal model. [Rufibach \(2012\)](#) also gave an example to illustrate the PL-based and LCDS-based estimates and compared them to the empirical estimate $\bar{R}(u)$ and the binormal estimate. They observed that, in that example, both the PL-based and LCDS-based estimates are better than the binormal model, given that the empirical estimate follows the true curve closely. They even provided some simulation studies to show that the log-concave ROC curve estimate is somewhat robust against misspecification, i.e., either F_0 and / or F_1 are not log-concave.

In the spirit of [Rufibach \(2012\)](#), we introduce a new ROC curve estimator based on the PQC estimator which is proposed in Chapter 4. It is constructed in the similar way to that by [Rufibach \(2012\)](#), in the sense that we first compute log-concave distribution function estimates \widehat{F}_0 and \widehat{F}_1 based on samples x_1, \dots, x_{n_0} and y_1, \dots, y_{n_1} , respectively, by applying the PQC estimator. The estimates are then plugged into function (5.1) to obtain

$$\widehat{R}(u) = 1 - \widehat{F}_1(\widehat{F}_0^{-1}(1 - u))$$

for $u \in (0, 1)$. In the following, we present examples for estimating an ROC curve with both simulated and real data and illustrate the use of the PL, LCDS and PQC estimators in this situation.

5.3 Simulation Studies

Setup

In the simulation studies, three ROC curve estimators based on the PL, LCDS and PQC are compared. The LCDS-based estimator is obtained by function `logConROC()`, with its default settings, in the R package `logcondens` ([Rufibach, 2012](#)).

We have performed a simulation study for the scenarios provided in Table 5.1. Denoting by $N(\mu, \sigma^2)$ the normal distribution with mean μ and variance σ^2 , the first scenario, which has been analysed by [Lloyd and Yong \(1999\)](#), serves as a benchmark for comparing the estimators to the binormal model. The second scenario, which has been used by [Rufibach \(2012\)](#), evaluates the methods for symmetric but non-normal distributions, where $LG(\mu, s)$ denotes the logistic distribution with location μ and scale s .

All computations were carried out in R (version 3.1.1) ([R Core Team, 2015](#)) on a computer with a 3.4 GHz Intel Core i5-3570 central processing unit.

Performance measures

Tab. 5.1: Two scenarios for ROC curve estimation.

Scenario	F_0	n_0	F_1	n_1	Used by
1	N (0, 1)	100	N (1, 1)	100	Lloyd and Yong (1999)
2	LG (0, 1)	100	LG (2, 1)	100	Rufibach (2012)

We evaluate the performance of an estimator $\widehat{R}(\cdot)$ for the ROC curve using the average square errors (ASE) which is defined as

$$\text{ASE}(\widehat{R}) = \int_0^1 \left\{ \widehat{R}(u) - R(u) \right\}^2 du, \quad (5.4)$$

where $R(\cdot)$ is the true ROC curve. The integral can be easily evaluated by a grid of points, $\{u_i : i = 1, \dots, n_{\text{grid}}\}$. This criterion has been used by, e.g., Zhou and Lin (2008) and Rufibach (2012). Following the same settings as in their papers, in our simulation studies we choose the u_i 's to be equidistant on $[0, 1]$, $n_{\text{grid}} = 100$, and carry out 500 simulations for each scenario.

Results

Figure 5.1 shows the density estimates from typical samples in each scenario in the left panels, with the corresponding ROC curves in the right panels. In Scenario 1, difference clearly exists among the three density estimates, but it becomes less obvious in the corresponding ROC curves. In Scenario 2, difference clearly exists not only among the three density estimates, but also in the resulting ROC curves. It shows that in general, the PQC-based estimate is closer to the true curve.

In each scenario, we performed the paired t -test on the ASE's of two estimators. The results are summarised in Table 5.2. The p -values less than 5% are highlighted in boldface. From the results given, one can see that in general, the ROC curve estimators based on the smooth estimators are significantly better than that based on the piecewise linear estimator. While in Scenario 1 there is no significant difference between the PQC and LCDS-based estimators, in Scenario 2 the PQC-based estimator is much better.

5.4 An Example

In this section, we further illustrate our new ROC curve estimator on the pancreases cancer data of Wieand et al. (1989) which is from Mayo Clinic and used for the biomarker study for detecting pancreatic cancer. This data set has been studied for ROC curve estimation in Zhou and Harezlak (2002), Zhou and Lin (2008) and Rufibach (2012). It was created by taking measurements in sera from 51 "control" patients with pancreatitis and 90 "cases" with pancreatic cancer using two

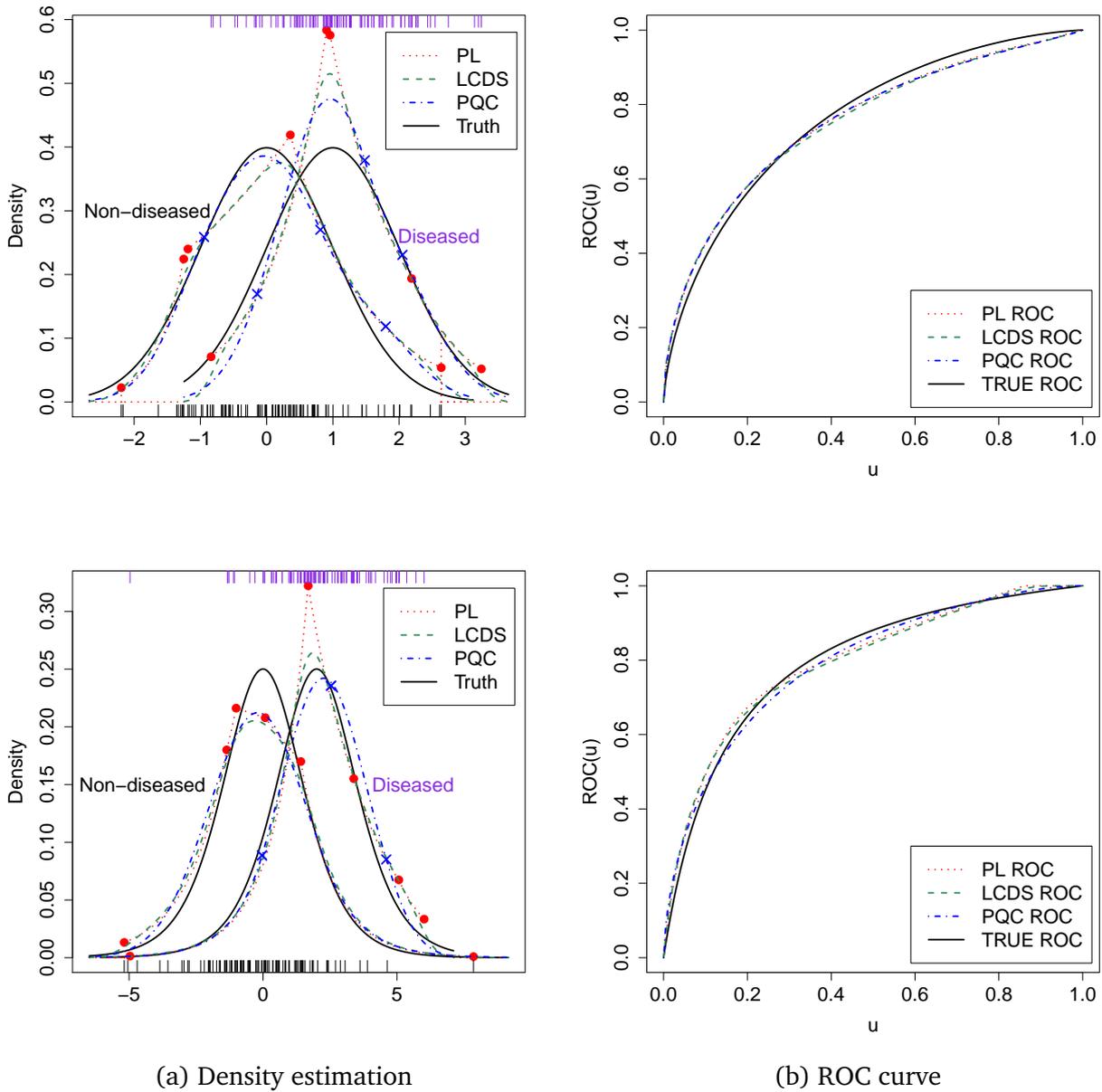


Fig. 5.1: (a) The log-concave density estimates for the simulated data: Scenario 1 (upper panels) and Scenarios 2 (lower panels). The solid and cross points represent the knots for the PL estimate and PQC estimate of a log-concave density function, respectively. (b) The corresponding induced ROC curves.

biomarkers: the antigenic determinant, designated as CA 125 (Bast Jr et al., 1983), and carbohydrate antigen, designated as CA 19-9 (Del Villano et al., 1983). CA 19-9 has higher sensitivity (Wieand et al., 1989; Zhou and Lin, 2008) and was used by Rufibach (2012) for comparing the log-concave ROC curve estimates with binormal and empirical estimates. Rufibach (2012) further pointed out that it is plausible to assume a log-concave density for the logarithm of CA 19-9 mea-

5.4. An Example

Tab. 5.2: Paired t -test results based on ASE values in simulated studies, where MD stands for mean difference.

(I) Est.	(J) Est.	MD (I-J)	Std. Error	95% Confidence Interval		p -value
				Lower Bound	Upper Bound	
N(0, 1) vs. N(1, 1) ($\times 10^{-5}$)						
PQC	PL	-7.671	2.053	-11.706	3.637	2.1×10^{-4}
	LCDS	-1.618	1.819	-5.193	1.957	0.37
LCDS	PL	-6.053	0.975	-7.969	-4.138	1.1×10^{-9}
LG(0, 1) vs. LG(2, 1) ($\times 10^{-4}$)						
PQC	PL	-5.704	0.529	-6.743	-4.666	$< 2.2 \times 10^{-16}$
	LCDS	-1.617	0.484	-2.568	-0.666	9.0×10^{-4}
LCDS	PL	-4.087	0.182	-4.444	-3.730	$< 2.2 \times 10^{-16}$

measurements since the tests by Hazelton (2011) yield large p -values ($p = 0.84$ for the control group and $p = 0.83$ for the case group) when assessing the null hypothesis of log-concavity. Here, we used both logarithms of CA 19-9 (log-CA 19-9) and CA 125 (log-CA 125) to compare the three non-parametric estimators: the PL, LCDS and PQC-based estimators. The datasets are available in the R package logcondens (Dümbgen and Rufibach, 2011; Rufibach, 2012).

Pancreases Data

Some descriptive statistics of the data are given in Table 5.3. For log-CA 19-9 measurements, it shows that the distribution of controls is right skewed, but that of cases is likely not skewed. For log-CA 125 measurements, both controls and cases seem to have right-skewed distributions. Furthermore, for the same group (controls or cases), the kurtosis of log-CA 125 is higher than that of log-CA 19-9.

Tab. 5.3: Descriptive statistics of pancreases data.

Variable	n	Mean	SD	Skewness	Kurtosis
log-CA 19-9					
controls	51	2.472	0.865	0.651	2.594
cases	90	5.415	2.342	-0.038	1.994
log-CA 125					
controls	51	2.666	0.782	1.337	4.387
cases	90	3.261	0.989	1.144	5.103

Log-concavity assumption

Table 5.4 gives the p -values based on the bootstrap test described in Section 4.4 by using 9999 bootstrap samples. For log-CA 19-9 measurements, the p_{ks} for controls is 0.887 which means that it is safe to assume that the underlying density of the controls is log-concave. For the cases, the same assumption can also be made since $p_{ks} = 0.417$. Other p -values indicate that the PL estimator is suitable for both groups. For log-CA 125 measurements, the p -values also support the assumption of log-concavity for each group under the 5% significance level.

Tab. 5.4: p -values from bootstrap test for pancreases data.

Variable	p_{ks}	p_{mean}	p_{sd}	p_{skew}	p_{kurt}
log-CA 19-9					
controls	0.887	0.988	0.743	0.764	0.632
cases	0.417	0.979	0.777	0.972	0.966
log-CA 125					
controls	0.130	0.967	0.377	0.273	0.374
cases	0.303	0.960	0.522	0.129	0.066

Fitted ROC curve

Figure 5.2 shows the log density plots and gradient curves at the convergence for the log-CA 19-9 measurements provided by CNMLCDS proposed in Chapter 4. The gradient curve for each group shows that the SNPMLE has been correctly found and the negative value of $d_{0,2}$ implies $\alpha_2 = 0$ after the last iteration. The plots for log-CA 125 measurements are given in Figure 5.3. The negative value of $d_{0,2}$ in 5.3(a) also means $\alpha_2 = 0$ after the final iteration. More preciously, the piecewise quadratic log-density functions for control and case groups of log-CA 19-9, are given by, respectively,

$$\begin{aligned} \hat{\varphi}(x) &= -0.67x - 4.24(1.89 - x)_+^2 - 0.10(x - 3.06)_+^2 - 1.53(x - 4.01)_+^2 + 0.71, \\ \hat{\varphi}(x) &= 0.07x - 1.03(2.13 - x)_+^2 - 0.05(x - 5.40)_+^2 - 0.94(x - 8.83)_+^2 - 2.38. \end{aligned}$$

Those for control and case groups of log-CA 125 are also given by, respectively,

$$\begin{aligned} \hat{\varphi}(x) &= -1.27x - 1.01(2.32 - x)_+^2 - 4.36(2.36 - x)_+^2 - 0.77(x - 4.60)_+^2 + 2.73, \\ \hat{\varphi}(x) &= 215.02x - 55.84x^2 + 55.84(1.88 - x)_+^2 + 53.38(x - 1.89)_+^2 + 1.92(x - 2.53)_+^2 + \\ & 0.46(x - 3.78)_+^2 + 0.09(x - 4.78)_+^2 - 208.91. \end{aligned}$$

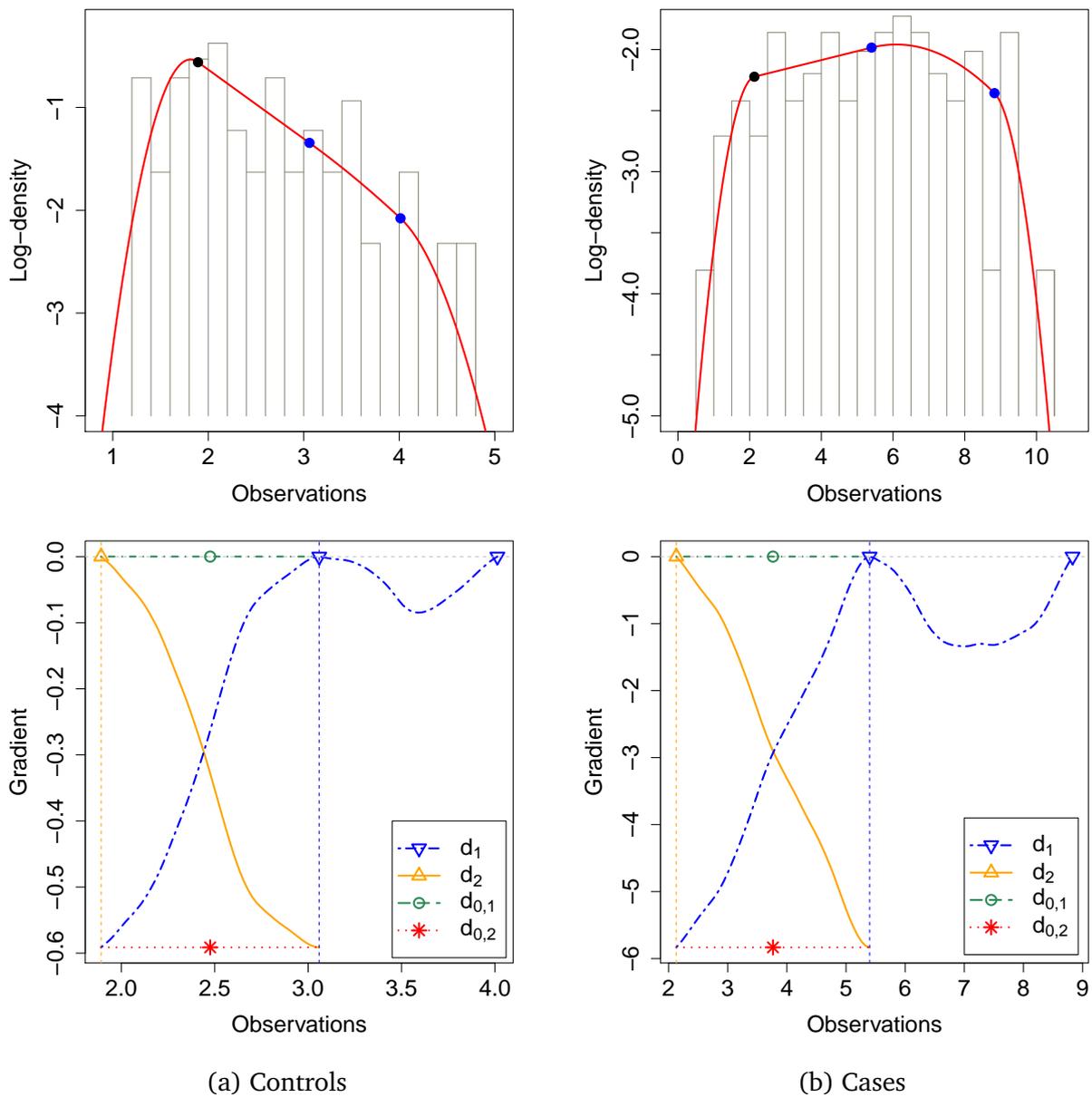


Fig. 5.2: Log-density plots (upper panels) with gradient curves (lower panels) for the pancreas data (log-CA 19-9): controls (left panels) and cases (right panels), found by CNMLCDS, with knots indicated by solid points.

In order to compare the different log-concave density estimators, QQ plots are shown in Figures 5.4 and 5.5, for log-CA 19-9 and log-CA 125, respectively. We can see that all three estimators fit well to each group of log-CA 19-9, with very little difference. However, for log-CA 125, differences can be found among these three estimators, especially for the controls. Generally speaking, each log-concave density estimator performs better for log-CA 19-9 than for log-CA 125. This conclusion is consistent with the bootstrap test results in Table 5.4 which show relatively larger p -values for log-CA 19-9 compared to those for log-CA 125.

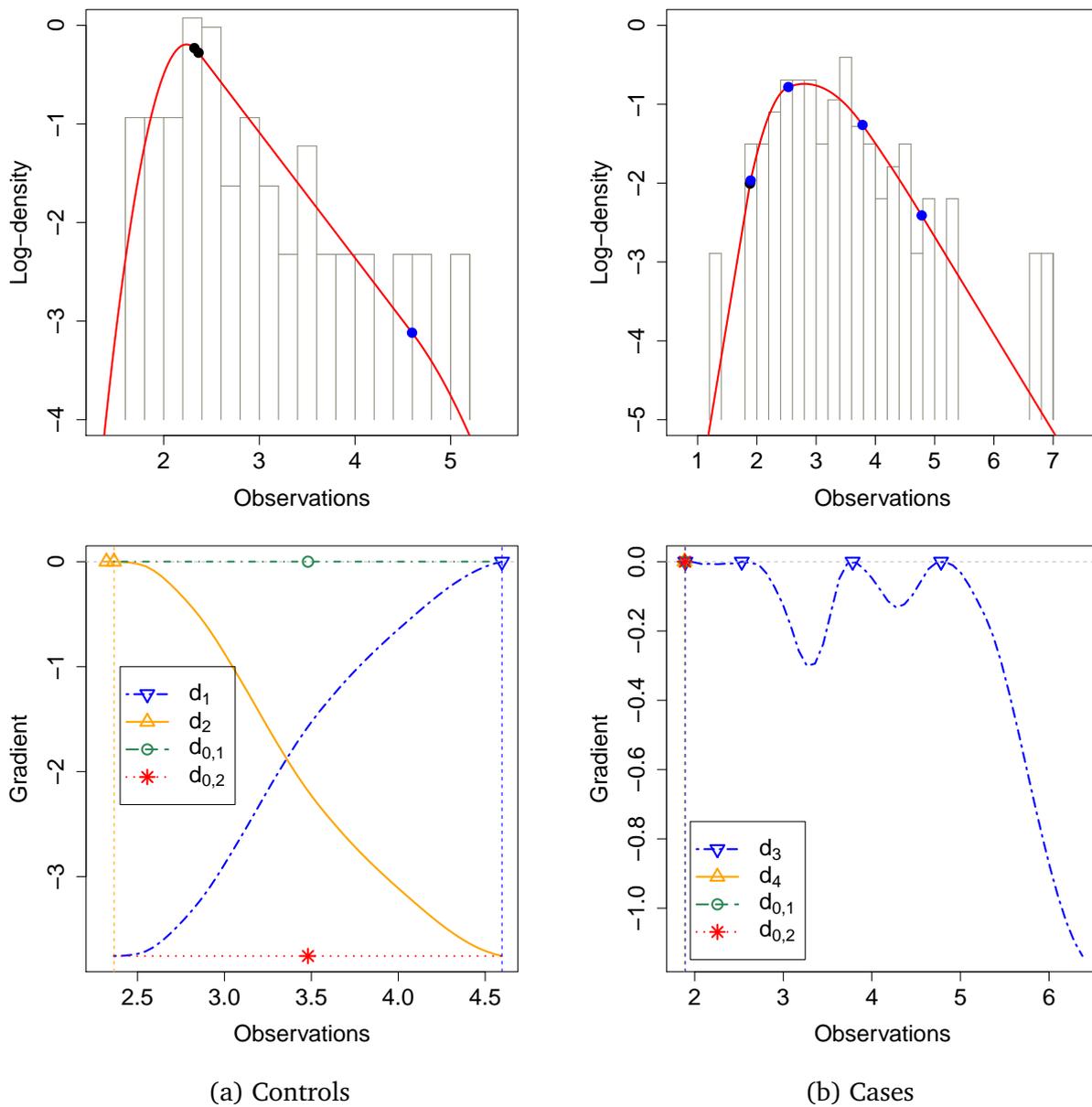


Fig. 5.3: Log-density plots (upper panels) with gradient curves (lower panels) for the pancreas data (log-CA 125): controls (left panels) and case (right panels), found by CNMLCDS, with knots indicated by solid points.

Finer differences can be found in the density plots and the induced ROC curves shown in Figure 5.6. From the top-left panel of Figure 5.6, for each group of log-CA 19-9, the differences among the three estimates are mainly around the peak and the boundaries of the domain of the observations. Note that all three estimates are able to capture the obvious skewness of the data. Furthermore, it is clear that the smooth estimates can not only alleviate the sharp changes in the density as shown by the piecewise linear estimate but also smooth out at the data boundaries.

5.4. An Example

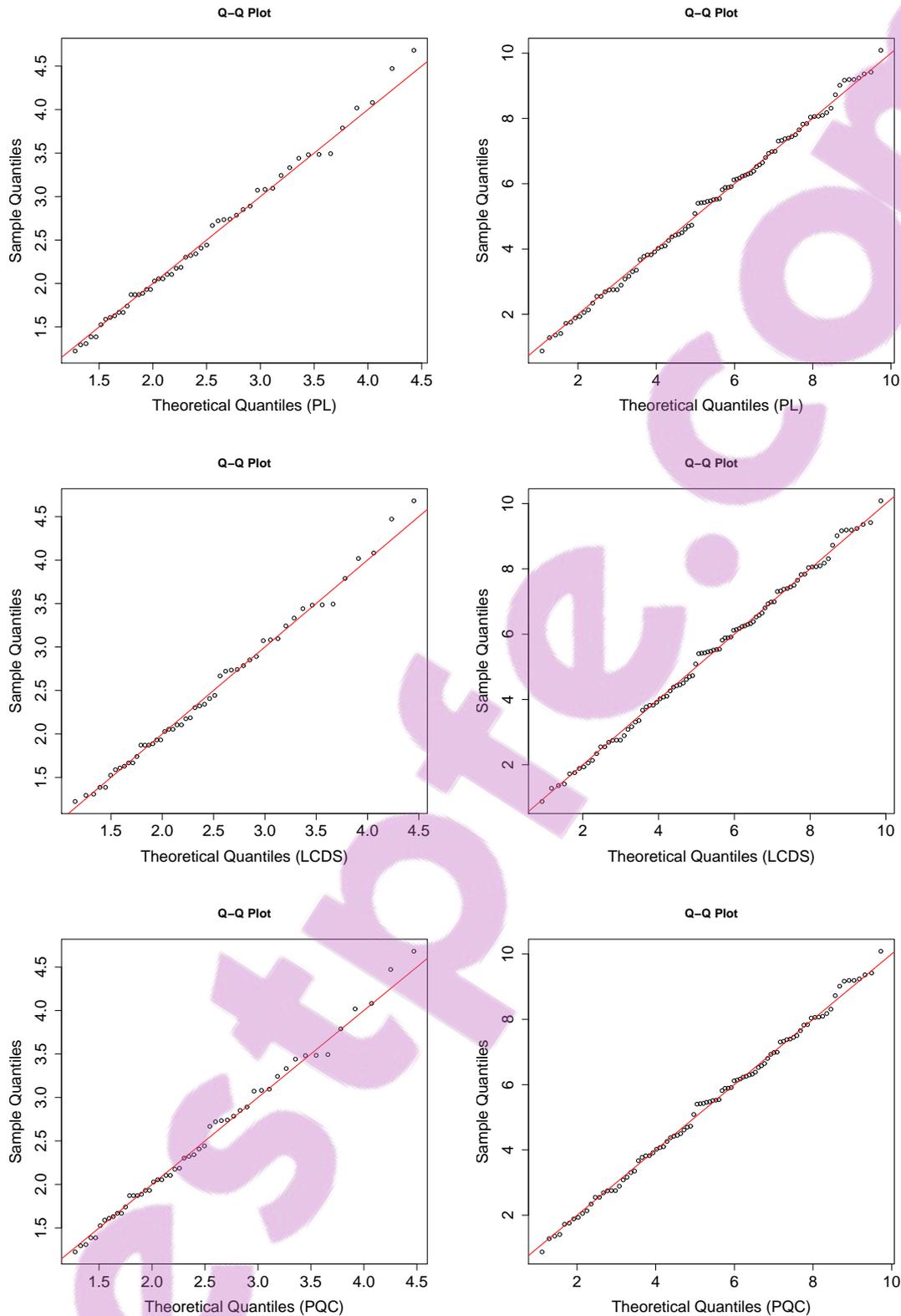


Fig. 5.4: Q-Q plots for the log-CA 19-9 measurements: controls (left panels) and cases (right panels), from different estimates: PL (top panels), LCDS (middle panels) and PQC (bottom panels).

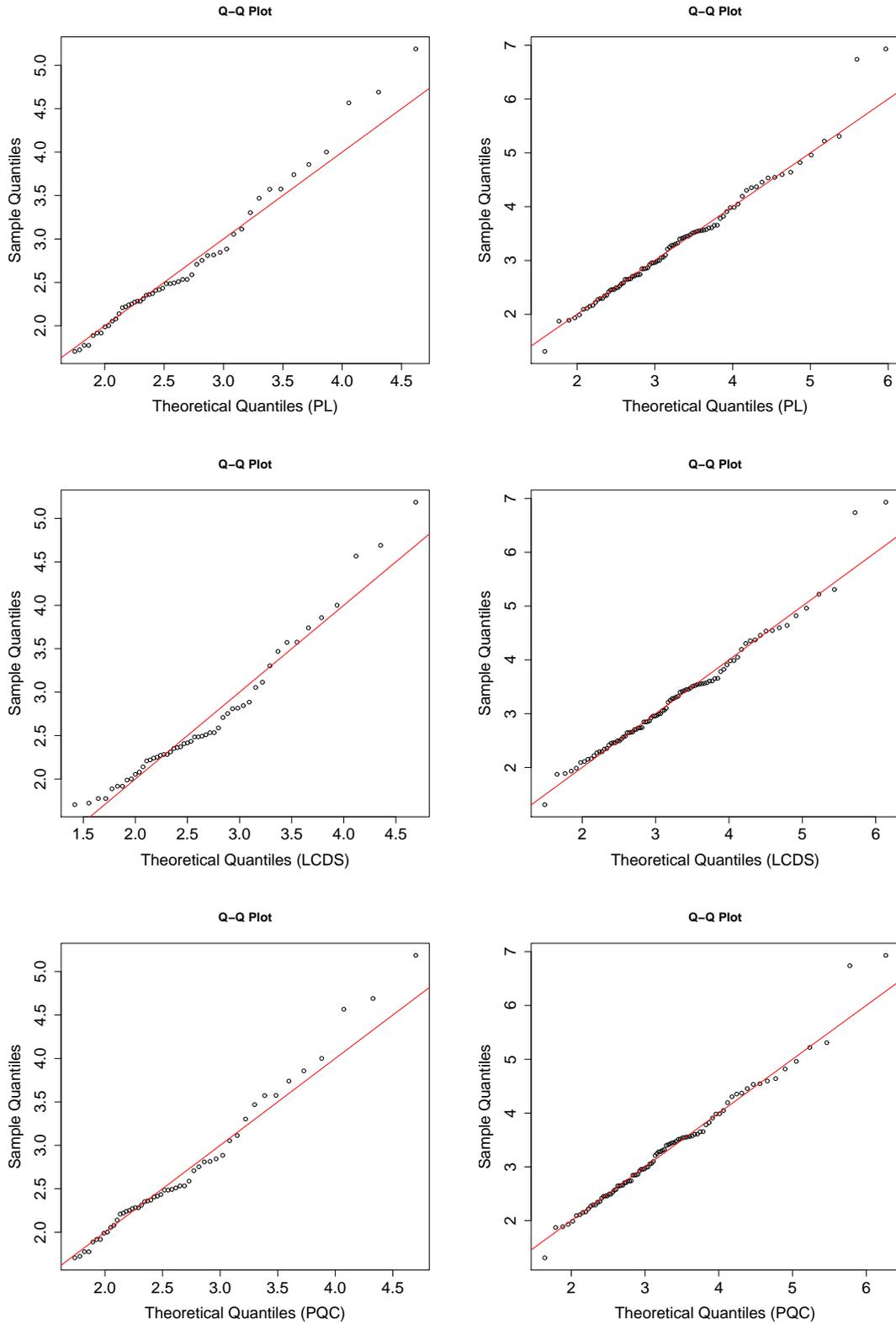


Fig. 5.5: Q-Q plots for the log-CA 125 measurements: controls (left panels) and cases (right panels), from different estimates: PL (top panels), LCDS (middle panels) and PQC (bottom panels).

However the differences between the density estimates become less obvious between the induced ROC curves, shown in the top-right panel of Figure 5.6.

From the bottom-left panel of Figure 5.6, for each group of log-CA 125, the differences among these three estimates become more pronounced. The large differences in density estimation also lead to the large differences in the induced ROC curves on the bottom-right panel of Figure 5.6. The induced ROC curves from the PQC and PL-based estimators are almost similar everywhere and both close to the empirical curve. However, the estimated curve from the LCDS-based estimator has a lower sensitivity for $u > 0.4$ than the other two and the difference becomes larger as u increases.

Performance measures

For real-world data studies, apart from visual inspection we also try to seek some appropriate quantified criteria to evaluate the performance among these estimators for ROC curve estimation. However, we have the same problem as discussed in Section 4.6 that the true function is unknown, here the ROC curve function. Hence, we can not directly utilize the measure given in (5.4). Instead, we replace the true function with the empirical ROC curve function, i.e.,

$$\text{ASE}(\widehat{R}, \widehat{R}_e) = \int_0^1 \left\{ \widehat{R}(u) - \widehat{R}_e(u) \right\}^2 du, \quad (5.5)$$

where \widehat{R} and \widehat{R}_e are the estimated and empirical ROC curves, respectively. Treating the ROC curve as a cumulative distribution function, the KS statistic can also be used to measure the error, as follows,

$$\text{KS}(\widehat{R}, \widehat{R}_e) = \sup_{u \in (0,1)} |\widehat{R}(u) - \widehat{R}_e(u)|. \quad (5.6)$$

Results

We estimate both the ROC curve \widehat{R} and the empirical ROC curve \widehat{R}_e using the whole data in criteria (5.5) and (5.6). The results obtained this way have been summarised in Table 5.5, with each smallest loss value highlighted in boldface. We can see that the three estimators are similar, only with the PQC slightly better in terms of ASE for log-CA 125 measurements. These results may be good enough to indicate which estimator is the best for estimation for this dataset, but they may lack the power to tell which estimator is the best for prediction.

In order to further compare the different estimators in terms of prediction, let us estimate the ROC curve \widehat{R} using training data and the empirical ROC curve \widehat{R}_e based on test data in the criterion

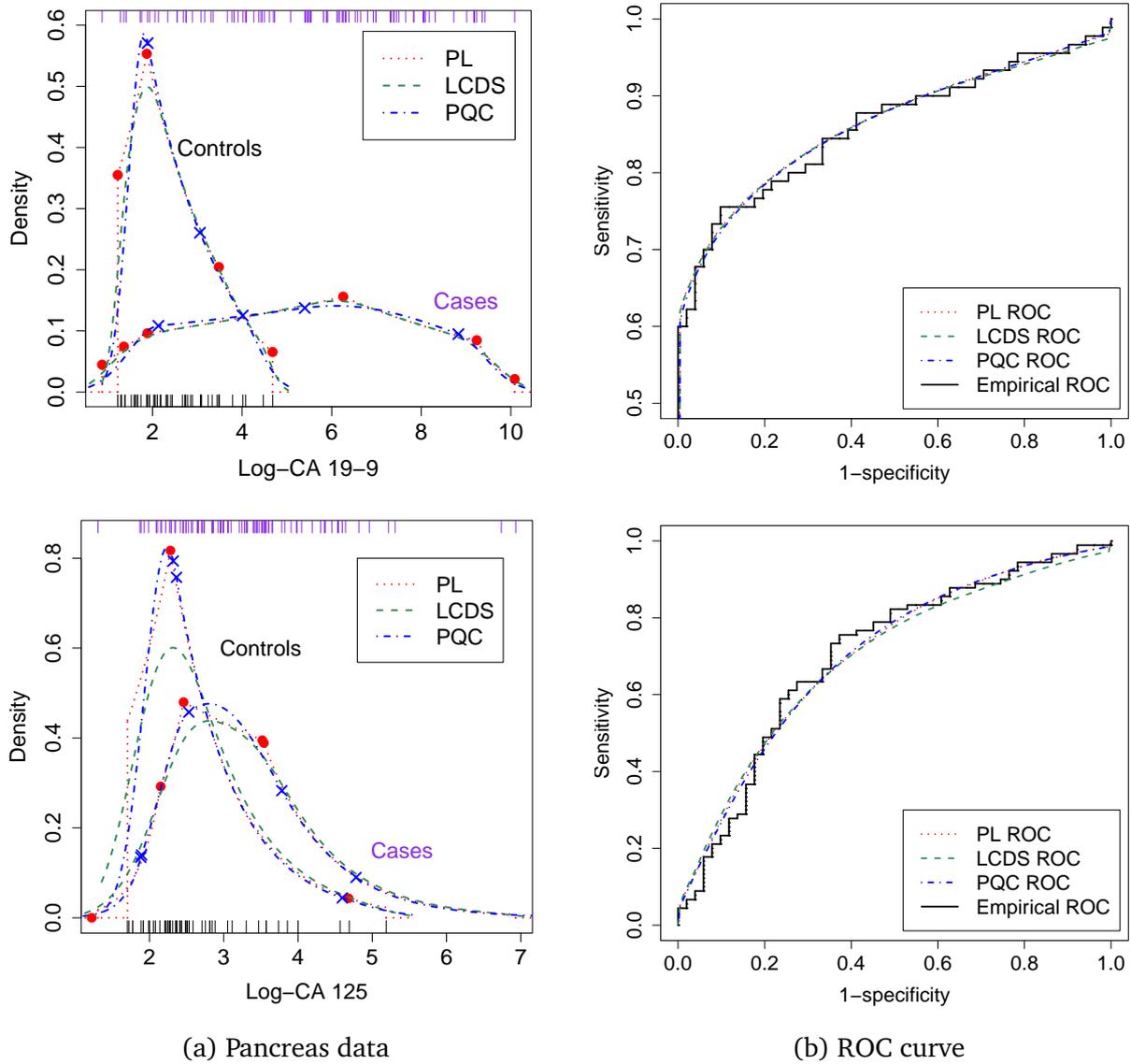


Fig. 5.6: (a) The log-concave density estimates for the log-CA 19 – 9 (upper panels) and log-CA 125 (lower panels). The solid and cross points represent the knots for the PL estimate and PQC estimate of a log-concave density function, respectively. (b) The corresponding induced ROC curves from the different estimates.

(5.5). Furthermore, the ROC curve involves two groups, the controls and cases, and hence we can also use the misclassification rate (MR) to compare different estimators. The error rate metrics for

Tab. 5.5: ASE ($\times 10^{-3}$) and KS results for the estimators for the pancreases data.

Estimator	log-CA 19-9		log-CA 125	
	ASE	KS	ASE	KS
PL	3.706	0.600	1.008	0.080
LCDS	3.709	0.600	1.274	0.084
PQC	3.707	0.600	0.891	0.083

a two-class classification problem can be summarized by a confusion matrix, given by

$$\begin{array}{c}
 \text{Predicted Controls} \quad \text{Predicted Cases} \\
 \text{Controls} \left(\begin{array}{cc} a & b \\ c & d \end{array} \right) \\
 \text{Cases}
 \end{array} \quad (5.7)$$

where each entry is the number of observations that meet the specified condition. In particular, a and d are those for correct predictions, and b and c for incorrect ones. The MR is then given by $\frac{b+c}{a+b+c+d}$. Each prediction is made based on comparing the two measurements $n_0 \hat{f}_0(x)$ and $n_1 \hat{f}_1(x)$, where $\hat{f}_0(\cdot)$ and $\hat{f}_1(\cdot)$ are the density estimates for controls and cases of sizes n_0 and n_1 , respectively. For a new observation x , if $n_0 \hat{f}_0(x) > n_1 \hat{f}_1(x)$, it is predicted as a control, or else as a case, if otherwise.

For each biomarker (log-CA 19-9 or log-CA 125), we ran a 2-fold cross-validation on controls and cases to compute the ASE and MR for each method. A paired t -test was then executed over 100 replications between each pair of methods. The results are given in Table 5.6, in terms of ASE's, and in Table 5.7, in terms of MR's. The p -values less than 5% are highlighted in boldface.

Overall, the PL-based estimator performs worst. From Table 5.6, we can see that the PQC-based estimator outperforms the LCDS-based estimator for log-CA 19-9 measurements. There is no significant difference between them for log-CA 125 measurements. From Table 5.7, no significant difference has been detected between the PQC- and LCDS-based estimators for log-CA 19-9 measurements, but the PQC-based one is much better for log-CA 125 measurements.

5.5 Summary

In this chapter, ROC curve estimation has been studied. Parametric, semiparametric and nonparametric methods in this area have been reviewed. We focused on the nonparametric methods and when the constituent distributions are log-concave. The new smooth log-concave density estimator

5.5. Summary

Tab. 5.6: Paired *t*-test results for ASE values for the pancreases data, where MD stands for mean difference.

(I) Est.	(J) Est.	MD (I-J)	Std. Error	95% Confidence Interval		<i>p</i> -value
				Lower Bound	Upper Bound	
log-CA 19-9 ($\times 10^{-4}$)						
PQC	PL	-15.641	1.484	-18.586	-12.697	$< 2.2 \times 10^{-16}$
	LCDS	-12.388	1.730	-15.821	-8.956	1.4×10^{-10}
LCDS	PL	-3.253	0.854	-4.947	-1.559	2.4×10^{-4}
log-CA 125 ($\times 10^{-4}$)						
PQC	PL	-3.115	0.916	-4.932	-1.298	9.7×10^{-4}
	LCDS	0.908	2.236	-3.529	5.345	0.69
LCDS	PL	-4.023	1.873	-7.739	-0.307	0.03

Tab. 5.7: Paired *t*-test results for MR (%) values for the pancreases data, where MD stands for mean difference.

(I) Est.	(J) Est.	MD (I-J)	Std. Error	95% Confidence Interval		<i>p</i> -value
				Lower Bound	Upper Bound	
log-CA 19- 9 ($\times 10^{-5}$)						
PQC	PL	-3.620	0.377	-4.367	-2.872	7.9×10^{-16}
	LCDS	0.352	0.192	-0.297	0.734	0.07
LCDS	PL	-3.972	0.372	-4.710	-3.234	$< 2.2 \times 10^{-16}$
log-CA 125						
PQC	PL	-3.880	0.351	-4.576	-3.185	$< 2.2 \times 10^{-16}$
	LCDS	-1.028	0.267	-1.558	-0.498	2.1×10^{-4}
LCDS	PL	-2.852	0.318	-3.482	-2.222	1.8×10^{-14}

PQC proposed in Chapter 4 has been applied to estimate the densities of both case and control data. Three ROC curve estimators, PL-, LCDS-and PQC-based, have been compared. The performance of these methods are assessed via simulation and on a medical dataset.

Simulation results indicate that in general, the ROC curve estimators based on the smooth estimators are better than that based on the piecewise linear one, and further the PQC-based estimator is slightly better than the LCDS-based one. For the pancreases data, tests for log-concavity have been carried out for both the case and control data. Three assessment criteria have been employed to evaluate the performance of the different estimators. The results are consistent in simulation and real-world data studies.

5.5. Summary

For future work, since concavity is known as a characteristic of proper ROC curves ([Dorfman et al., 1997](#)), it would be interesting to apply shape-restricted regression methods to ROC curve estimation; see [Gonçalves et al. \(2014\)](#) and the references therein.

Chapter 6

Nonparametric Estimation for Heavy-tailed Distributions under Shape Restrictions

6.1 Empirical Motivation

Financial decisions are commonly made based on the expected returns and risks of alternative investment opportunities. How to accurately model the process of return formation and evaluate risk in financial markets thus becomes very important for investment. A vast amount of research and theory has been developed to understand the behaviour of returns.

Our main motivation for using nonparametric density estimation under shape restrictions in finance comes from fitting an asset return distribution. Such estimators can hopefully allow us to handle data with high peak, skewness and heavy-tailedness. Log-return, the natural logarithm of the simple gross return of an asset, is popularly studied for its tractable statistical properties. Consider the daily log-returns of S&P 500 index (SPX) from March 23, 1995 to March 23, 2015, as shown in Figure 6.1. The SPX is an American stock market index based on the market capitalizations of 500 large companies having common stock listed on the NYSE or NASDAQ. The components and their weightings are determined by S&P Dow Jones indices. As pointed out by [Markowitz and Usmen \(1996\)](#), SPX is the most commonly followed stock market index and is viewed as an important indicator of the state of the market and economy. A histogram of daily log-returns along with the fitted normal density function (with mean 2.87×10^{-4} and standard deviation 1.22×10^{-2}) is shown in Figure 6.1. The minimum and maximum of the daily log-returns are about -9.47×10^{-2} and 1.10×10^{-1} , respectively. Note that for a standard normal random vari-

able Z , $Pr(Z < \frac{-9.47 \times 10^{-2} - 2.87 \times 10^{-4}}{1.22 \times 10^{-2}}) \approx 3.46 \times 10^{-15}$ just too small for a sample of size 5035 daily data. The kurtosis of the data is 11.02. Compared with a normal distribution that has kurtosis 3, the feature of a high peak and two heavy tails of the data, i.e. being leptokurtic, is quite clear. This is not only true for the SPX but also for almost all financial asset returns, e.g. the world wide stock indices, individual stocks, foreign exchange rates and interest rates; see [Mittnik et al. \(1998\)](#), [Curto et al. \(2003\)](#) and [Meerschaert and Scheffler \(2003\)](#). The skewness of the data is -0.24 , suggesting that the left tail is slightly heavier than the right tail. The histogram also suggests the unimodality of the underlying density function. Actually, unimodality is widely cited as a reasonable assumption in finance and econometrics and is a standard feature of the commonly-used distributions for fitting an asset return, such as the t -, log-normal, log-gamma and stable distributions.

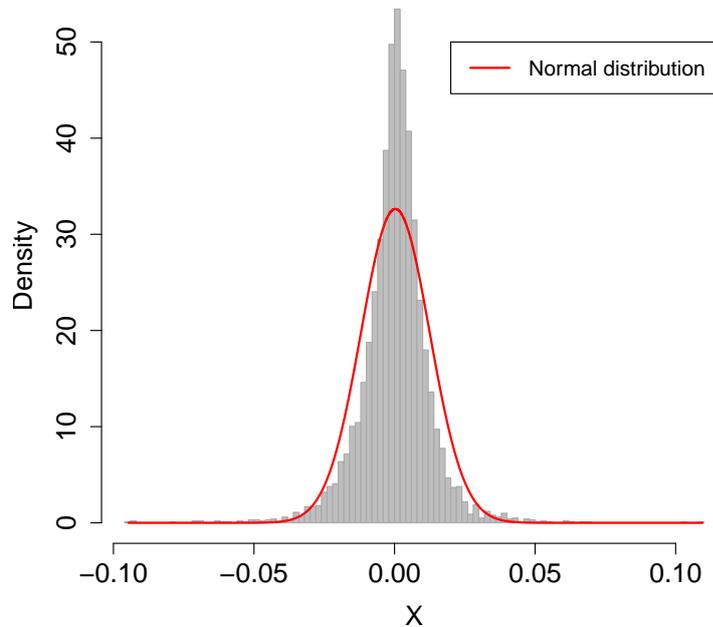


Fig. 6.1: The histogram of SPX daily log-returns (from March 23, 1995 to March 23, 2015) and the fitted normal density.

6.2 Introduction

Over the last few decades, a plenty of research has been conducted to deal with heavy-tailed distributions based on parametric models, including the t -, stable, Pareto and Weibull distributions. These parametric models have the advantages of easy interpretation and fast fitting, but they also have the drawback of potential misspecification.

6.2.1 Nonparametric estimation of heavy-tailed distributions

Nonparametric estimation of heavy-tailed distributions has been investigated by many researchers. Two nonparametric estimators, the Parzen-Rosenblatt kernel density estimator and the histogram with variable bin width, have been considered by [Markovitch and Krieger \(2000\)](#) for analysing World Wide Web traffic. [Takada et al. \(2001\)](#) performed a comparative study of nonparametric density estimators and concluded that the adaptive bandwidth kernel estimator of [Silverman \(1986\)](#) and the logspline (LS) approach of [Kooperberg and Stone \(1991, 1992\)](#) are superior for fitting heavy-tailed densities to the fixed bandwidth kernel estimator and the Hermite series estimator. The adaptive kernel method improves upon the classical kernel density estimator by introducing a local bandwidth factor which makes the bandwidth to be smaller around the center and larger in the tails. The LS method models a log-density function as a cubic spline having knots at selected order statistics of the sample, with the sample minimum and maximum selected as the first and last knots, respectively. A cubic spline is also restricted to be linear to the left of the first knot and to the right of the last knot, that is, its density estimate is exponential to the left of the first knot and to the right of the last knot. The parameters are estimated by maximum likelihood. Further, the knots between the first and last knots can be selected in a knot addition and deletion procedure, and the final model is determined by the BIC criterion ([Koo et al., 1999](#)). This estimator may perform well when both tails are heavy, but poorly when only one or none is heavy. A mixture of gamma distributions was proposed by [Venturini et al. \(2008\)](#) for the estimation of heavy-tailed distributions. However, it involves choosing the number of mixture components and only deals with positive variables.

[Markovitch and Krieger \(2002\)](#) proposed a joint parametric-nonparametric estimation approach for estimating a heavy-tailed distribution. The tail of the distribution is estimated by a general Pareto-like family and its body by a nonparametric method using a finite linear combination of trigonometric functions. It has both the advantage of a parametric model for the tail and that of a nonparametric method for the main body. However, it has an inconvenience to choose the boundary between the tail and body; see also [Barron et al. \(1992\)](#). A semiparametric model for heavy-tailed distributions was proposed by [Buch-Larsen et al. \(2005\)](#). They transformed the data using a modified Champernowne distribution and then applied the classical kernel density estimator to the transformed data. This semiparametric model was further studied by [Buch-Kromann \(2009\)](#), that can only handles positive values. The idea of transforming the original data has also been employed by [Markovich \(2006\)](#) and [Charpentier and Flachaire \(2014\)](#).

6.2.2 Nonparametric estimation of unimodal distributions

As mentioned in Section 6.1, the univariate financial data almost always has a unimodal leptokurtic density. For modelling a density with mode known, Fougères (1997) proposed a unimodal estimator based on a unimodal rearrangement of the kernel estimator. Data sharpening was employed for constructing a unimodal estimator by Braun and Hall (2001), who provided no theoretical support or clear guidance for choosing the distance function. It was further studied by Hall and Kang (2005), who provided both the theoretical and numerical properties based on the L_1 distance function. Dümbgen and Rufibach (2009) applied convolution to their nonparametric maximum likelihood density estimate with a Gaussian density to produce a smooth log-concave density estimate. However, log-concavity can not capture the heavy tail feature (Meyer, 2012). All models mentioned above are kernel-based, and we know that finding the optimal bandwidth h from data is highly challenging in practice. For too small an h , the estimate is too noisy, exhibiting highly various and extraneous wiggles. For too large an h , the estimate is over-smoothing and may miss key features and wash out fine details.

To preserve unimodality, spline estimation is an alternative approach. The estimators of Fougères (1997) and Dümbgen and Rufibach (2009) have been further studied by Meyer (2012), who proposed a smooth unimodal regression spline estimator (UMRS) for unimodal density estimation with a known mode. Meyer (2012) firstly proposed quadratic splines for the monotonic case, and with a known mode, further estimated the decreasing right-hand side of the mode and the increasing left-hand side of the mode. When the mode is unknown, Meyer suggested a plug-in estimator based on polynomial kernel density estimation for estimating the mode. The number and positions of knots were chosen by data-driven compromises and the parameters were obtained by minimizing the integrated square error. Turnbull and Ghosh (2014) applied a mixture of Beta densities to estimate a unimodal density function and employed the AIC, the BIC and the Condition Number (CN) criterion to choose the number of weights. The mixing weights were computed via quadratic programming subject to linear inequality constraints. However, if the number of weights is not large enough, the mixture of Beta densities can not well fit to heavy-tailed distributions. A new and more flexible shape constraint, named as “inverse convex”, was introduced by Anderson-Bergman (2014) for survival analysis and other types of heavy-tailed data; unfortunately, this estimator is not smooth.

6.2.3 Overview

In general, the tail of a probability density function can not be estimated by pure nonparametric methods without imposing further assumptions on the tail shape. We thus shall restrict ourselves to nonparametric estimation of a density function under unimodality and heavy-tailed shape assumptions. The most difficult for modelling the whole unimodal and heavy-tailed distribution is how to accurately capture the tails without losing the important information in the main body of the density. In the same spirit as by [Markovitch and Krieger \(2002\)](#), we separately consider the tail and main body of the density function, but avoid choosing a parametric model, the boundary between tail and body, or any tuning parameter. Our approach leads to a full automatic nonparametric estimator, in particular a nonparametric distribution, whose tails are fitted under log-convexity restriction and whose main body is fitted under the log-concavity restriction. The estimator automatically splices together, in a smooth and seamless way, the parts that have log-concavity or log-convexity shapes. We speak of a concave curve if its second derivative is negative, and a convex curve if its second derivative is positive. Consider a Student's t -distribution as shown in [Figure 6.2](#). The second derivative of its log-density function is positive at the tails and negative in the main part. More precisely, the second derivative of its log-density function first increases, changing values from zero to positive, and then decreases, from positive to negative, as x increases at the left-hand side of the mode. At the right-hand side of the mode, the second derivative first increases, from negative to positive, and then decreases, from positive to zero, as x increases. We would like to model such a change of shape using a nonparametric approach by applying a piecewise quadratic function, as detailed next.

6.3 Maximum Likelihood Estimation for Unimodal Heavy-tailed Distributions

6.3.1 The log-convex-concave-convex density estimator

In order to obtain a piecewise quadratic estimator by likelihood maximization under the heavy-tailedness and unimodality restrictions, we impose extra constraints on the change of second derivative of a log-density function φ . We consider a φ that its second derivative has four changing parts as x increases. That is, the second derivative first increases from zero and then decreases from positive to negative at the left-hand side of the mode. At the right-hand side of the mode, it first increases from negative to positive and then decreases to zero. The piecewise quadratics

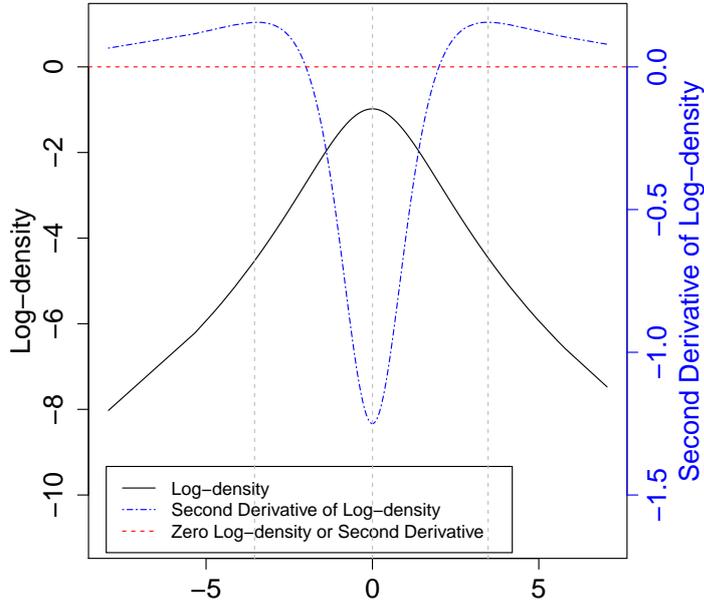


Fig. 6.2: Log-density function and the second derivative of a student- t distribution.

enables us to write φ as

$$\begin{aligned} \varphi(x) = & \alpha_1 x - \alpha_2 x^2 - \sum_{j=1}^q \rho_j (v_j - x)_+^2 + \sum_{j=1}^p \omega_j (\tau_j - x)_+^2 \\ & + \sum_{j=1}^m \pi_j (x - \theta_j)_+^2 - \sum_{j=1}^k \lambda_j (x - \mu_j)_+^2 - C, \end{aligned} \quad (6.1)$$

where C is the normalizing constant satisfying $\int_{\mathcal{X}} e^{\varphi(x)} dx = 1$.

Define $\varphi^{(0)}(x) \equiv \alpha_1 x - \alpha_2 x^2$, $\varphi^{(1)}(x) \equiv -\sum_{j=1}^q \rho_j (v_j - x)_+^2$, $\varphi^{(2)}(x) \equiv \sum_{j=1}^p \omega_j (\tau_j - x)_+^2$, $\varphi^{(3)}(x) \equiv \sum_{j=1}^m \pi_j (x - \theta_j)_+^2$ and $\varphi^{(4)}(x) \equiv -\sum_{j=1}^k \lambda_j (x - \mu_j)_+^2$, then we can write

$$\varphi(x) = \sum_{j=0}^4 \varphi^{(j)}(x) - C.$$

Note that if $\varphi^{(1)}$ and $\varphi^{(4)}$ are missing, φ becomes PQ4 family proposed in Chapter 4.

Always, we let $x_1 \leq v_1 < \dots < v_q < \tau_1 < \dots < \tau_p \leq \theta_1 < \dots < \theta_m < \mu_1 < \dots < \mu_k \leq x_n$. Note that each v_j ($1 \leq j \leq q$) or θ_j ($1 \leq j \leq m$) indicates a knot at which the second derivative of φ changes to a greater value, while each τ_j ($1 \leq j \leq p$) or μ_j ($1 \leq j \leq k$) a knot where φ has the value of its second derivative changed to a smaller one.

Denote by $\vartheta = (\mathbf{v}^T, \boldsymbol{\tau}^T, \boldsymbol{\theta}^T, \boldsymbol{\mu}^T)^T$ a knot vector of φ with its corresponding coefficient vector $\boldsymbol{\beta} = (\alpha_1, \alpha_2, \boldsymbol{\rho}^T, \boldsymbol{\omega}^T, \boldsymbol{\pi}^T, \boldsymbol{\lambda}^T)^T$, all restricted to be positive, except α_1 , a free parameter. In particular, $\mathbf{v} = (v_1, \dots, v_q)^T$, $\boldsymbol{\tau} = (\tau_1, \dots, \tau_p)^T$, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)^T$, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_k)^T$, $\boldsymbol{\rho} = (\rho_1, \dots, \rho_q)^T$, $\boldsymbol{\omega} = (\omega_1, \dots, \omega_p)^T$, $\boldsymbol{\pi} = (\pi_1, \dots, \pi_m)^T$, $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_k)^T$.

Like PQ4 in Chapter 4, we also need additional constraints to meet the basic unimodal density function properties. Denote by L and U the lower and upper boundaries of $\mathcal{X} \subset \mathbb{R}$, respectively, which can be finite or infinite. Letting

$$\varphi''(L) \equiv \lim_{x \rightarrow L} \frac{\partial^2 \varphi(x)}{\partial x^2} = 2 \left(-\alpha_2 - \sum_{j=1}^q \rho_j + \sum_{j=1}^p \omega_j \right), \quad (6.2)$$

$$\varphi''(U) \equiv \lim_{x \rightarrow U} \frac{\partial^2 \varphi(x)}{\partial x^2} = 2 \left(-\alpha_2 + \sum_{j=1}^m \pi_j - \sum_{j=1}^k \lambda_j \right). \quad (6.3)$$

Furthermore, denote $I_l = -\alpha_2 + \sum_{j=1}^p \omega_j$ and $I_r = -\alpha_2 + \sum_{j=1}^m \pi_j$. Note that $I_l > 0$ (or $I_r > 0$) indicates that the left (or right) side of φ has a convex part, respectively. We need to impose the following different additional constraints in different situations:

(I) Either

(a) $I_l \leq 0$, i.e., the left side of φ is always concave;

or

(b) $I_l > 0$, i.e., the left side of φ has a convex part, and then

(i) if $L = -\infty$, then $\varphi''(L) = 0$ and $\varphi'(L) = \alpha_1 + 2 \sum_{j=1}^q \rho_j v_j - 2 \sum_{j=1}^p \omega_j \tau_j > 0$;

(ii) if $L > -\infty$, then $\varphi''(L) \geq 0$ and $\varphi'(L) = \alpha_1 + 2 \sum_{j=1}^q \rho_j v_j - 2 \sum_{j=1}^p \omega_j \tau_j + L \varphi''(L) \geq 0$.

(II) Either

(a) $I_r \leq 0$, i.e., the right side of φ is always concave;

or

(b) $I_r > 0$, i.e., the right side of φ has a convex part, and then

(i) if $U = +\infty$, then $\varphi''(U) = 0$ and $\varphi'(U) = \alpha_1 - 2 \sum_{j=1}^m \pi_j \theta_j + 2 \sum_{j=1}^k \lambda_j \mu_j < 0$;

(ii) if $U < +\infty$, then $\varphi''(U) \geq 0$ and $\varphi'(U) = \alpha_1 - 2 \sum_{j=1}^m \pi_j \theta_j + 2 \sum_{j=1}^k \lambda_j \mu_j + U \varphi''(U) \leq 0$.

Let \mathcal{K} be the family of φ defined by function (6.1) with above constraints (I) and (II) satisfied and name the maximum likelihood estimator as the log-convex-concave-convex (LCCC) density estimator.

6.3.2 Maximum likelihood estimation and its characterization

Given $\varphi \in \mathcal{K}$, the density function is given by $f(x; \varphi) = e^{\varphi(x)}$ and the log-likelihood function by $l(\varphi) = \sum_{i=1}^n \varphi(x_i)$. The nonparametric maximum likelihood estimation of φ restricted to \mathcal{K} is to solve the following optimization problem:

$$\text{maximize } l(\varphi), \quad \text{subject to } \varphi \in \mathcal{K}. \quad (6.4)$$

By maximizing $l(\varphi)$ over all functions in \mathcal{K} , we obtain the nonparametric maximum likelihood estimate $(\widehat{\beta}, \widehat{\vartheta})$ of (β, ϑ) , i.e., the LCCC estimator. Notationally, φ is exchangeable with its coefficient vector β and knot vector ϑ , namely $l(\varphi) \equiv l(\beta, \vartheta)$.

Gradient functions are important for computing the nonparametric maximum likelihood estimate here. Four gradient functions are needed, one for each of the four shape-restricted parts. In the aid of the basis functions $e_{1,\xi}(x) = -(\xi - x)_+^2$, $e_{2,\xi}(x) = (\xi - x)_+^2$, $e_{3,\xi}(x) = (x - \xi)_+^2$ and $e_{4,\xi}(x) = -(x - \xi)_+^2$, they are defined as

$$\begin{aligned} d_j(\xi; \varphi) &\equiv \left. \frac{\partial l(\varphi + \varepsilon e_{j,\xi})}{\partial \varepsilon} \right|_{\varepsilon=0^+} \\ &= \sum_{i=1}^n e_{j,\xi}(x_i) - nE_\varphi(e_{j,\xi}(X)), \end{aligned} \quad (6.5)$$

where $j = \{1, 2, 3, 4\}$ and E_φ is the expectation with respect to $f(\varphi) = e^\varphi$. Note that d_1 is defined on $[x_1, \tau_1]$, d_2 on $[v_q, \theta_1]$, d_3 on $[\tau_p, \mu_1]$ and d_4 on $[\theta_m, x_n]$. Letting $e_{0,1}(x) = x$ and $e_{0,2}(x) = -x^2$, we have

$$d_{0,j}(\varphi) \equiv \frac{\partial l}{\partial \alpha_j} = \sum_{i=1}^n e_{0,j}(x_i) - nE_\varphi(e_{0,j}(X)), \quad \text{for } j = 1, 2.$$

For computing the LCCC, the log-likelihood function $l(\varphi)$ potentially may have multiple local maxima. A local maximum $\widehat{\varphi}$ must satisfy the following first-derivative conditions:

- (a) $d_{0,1}(\widehat{\varphi}) = 0$.
- (b) $d_{0,2}(\widehat{\varphi}) = 0$, if $\widehat{\alpha}_2 > 0$;
 $d_{0,2}(\widehat{\varphi}) \leq 0$, if $\widehat{\alpha}_2 = 0$.

(c) $d_j(\xi; \widehat{\varphi}) \leq 0$, for $\xi \in \widehat{\mathcal{S}}_j$, $j = 1, 2, 3, 4$.

(d) $d_j(\xi; \widehat{\varphi}) = 0$, for $\xi \in \text{supp}_j(\widehat{\varphi})$, $j = 1, 2, 3, 4$.

where \mathcal{S}_j means the potential support space for the j -th gradient function and $\text{supp}_j(\widehat{\varphi})$ the j -th support set of $\widehat{\varphi}$, i.e., the support points for the basis functions of the j -th part $\widehat{\varphi}^{(j)}$.

6.3.3 Relations to parametric distributions

We remarked in Section 4.2.5 that the uniform, exponential, normal, logistic, extreme-value, Weibull (for $k \geq 1$), gamma (for $k \geq 1$) and beta (for $a \geq 1$ and $b \geq 1$) distributions are all log-concave. Here we also examine some commonly-used parametric distributions and will see that they all belong to the new family of the LCCC estimator. Note that all the distributions that fall in the family of PQ4 in Chapter 4 are special cases of the LCCC family, i.e., when there exists no log-convex tail. We do not examine those situations again and emphasize on the distributions which have a log-convex tail. The location parameter is ignored for the same reason as mentioned in Section 4.2.5.

Weibull distribution

The Weibull distribution with parameter $k > 0$ has density $f(x) = k(x/\lambda)^{k-1}e^{-(x/\lambda)^k}/\lambda$, which is defined for $x \in (0, \infty)$. The second derivative of its log-density is $(1 - k)x^{-2}(1 + kx^k/\lambda^k)$ which is positive and decreasing when $k < 1$ (for the negative third derivative). That is, the log-density has a convex tail with a decreasing second derivative when $k < 1$.

Gamma distribution

The Gamma distribution has density function

$$f(x) = \frac{x^{k-1}e^{-x/\theta}}{\theta^k\Gamma(k)},$$

which is defined for $x \in (0, \infty)$, where $\theta > 0$, $k > 0$ and $\Gamma(\cdot)$ is the gamma function. The second derivative of log-density is $(1 - k)/x^2$, which is positive and decreasing when $k < 1$.

Pareto distribution

The Pareto distribution is defined over the non-negative real number and has density function $f(x) = \beta x^{-\beta-1}\lambda^\beta$, where $\beta > 0$ and $x \geq \lambda$. The second derivative of log-density is $(1 + \beta)/x^2 > 0$, which yields a convex tail, and decreases in the domain.

Log-normal distribution

The log-normal density function is $f(x) = e^{-(\log x)^2/(2\sigma^2)}/(x\sqrt{2\pi}\sigma)$. A bit of calculation shows that the second derivative of the log-density is $(\sigma^2 - 1 + \log x)/(x\sigma)^2$, which is negative on $(0, e^{1-\sigma^2})$ and positive on $(e^{1-\sigma^2}, \infty)$. Furthermore, it increases on $(0, e^{1.5-\sigma^2})$ and decreases on $(e^{1.5-\sigma^2}, \infty)$.

Student's t -distribution

The Student's t -distribution is defined on the entire real line with density

$$f(x) = \frac{\left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}}{\sqrt{\nu}B(0.5, \nu/2)},$$

where $B(a, b)$ is the beta function and ν is referred to as the number of degrees of freedom. The second derivative is $-(\nu+1)(\nu-x^2)/(\nu+x^2)^2$, which is negative on the central interval $[-\sqrt{\nu}, \sqrt{\nu}]$ and positive on $(-\infty, \sqrt{\nu}]$ and $[\sqrt{\nu}, \infty)$. Furthermore, the second derivative increases on $(-\infty, -\sqrt{3\nu}]$ and $[0, \sqrt{3\nu}]$ and decreases on $[-\sqrt{3\nu}, 0]$ and $[\sqrt{3\nu}, \infty)$. Note that the Cauchy distribution is the Student's t -distribution with 1 degree of freedom.

6.4 Computation

We now discuss the computational aspect for solving problem (6.4). Since there is no closed-form solution, an iterative method must be used. A new algorithm is presented below for computing the LCCC estimate. Its basic idea is similar to that of the CNM (Wang, 2007) in the sense that it also involves two alternating steps: the first is to expand and reduce the knot vector ϑ and the second to update the coefficient vector β . Here special considerations need to be given to the different constraints in different situations, as specified in Section 6.3.1.

6.4.1 Derivatives of the log-likelihood function

Given φ , the first partial derivatives of the log-likelihood are given as follows:

$$\frac{\partial l(\varphi)}{\partial \alpha_j} = d_{0,j}(\varphi), \quad j = 1, 2,$$

and for the i -th part $\varphi^{(i)}$, $i = 1, 2, 3, 4$,

$$\frac{\partial l(\varphi)}{\partial \gamma_j} = d_i(\xi_j; \varphi),$$

where ξ_j is the j -th knot of $\varphi^{(i)}$ and γ_j the corresponding mass.

Furthermore, denote by

$$\mathbf{e} = (e_{0,1}(x), e_{0,2}(x), e_{i,\xi_{i,j}}(x))^T, \quad i = 1, 2, 3, 4,$$

where $\xi_{i,j}$ is the j -th knot of the i -th part $\varphi^{(i)}$. Then the Hessian matrix is defined by $-n\text{Var}_\varphi(\mathbf{e})$, where Var_φ is the variance operator with respect to density e^φ .

6.4.2 Updating coefficients

Let us first consider updating the coefficient vector from β to β^* with ϑ fixed by employing the second-order Taylor series expansion of the log-likelihood function in the neighbourhood of β . Let the first partial derivative vector and the Hessian matrix of log-likelihood be, respectively,

$$\mathbf{g} \equiv \mathbf{g}(\beta, \vartheta) = \frac{\partial l(\beta, \vartheta)}{\partial \beta},$$

$$\mathbf{H} \equiv \mathbf{H}(\beta, \vartheta) = \frac{\partial^2 l(\beta, \vartheta)}{\partial \beta \partial \beta^T},$$

then the second-order Taylor series about β is given by

$$l(\beta, \theta) - l(\beta^*, \theta) \approx -\mathbf{g}^T \boldsymbol{\eta} - \frac{1}{2} \boldsymbol{\eta}^T \mathbf{H} \boldsymbol{\eta},$$

where $\boldsymbol{\eta} = \beta^* - \beta$ indicates a direction away from β . Let $\mathbf{H} = -\mathbf{R}^T \mathbf{R}$, where \mathbf{R} is some square matrix which is obtained in the same way as described in Section 3.3. Maximizing $l(\beta^*, \theta)$ in the neighbourhood of β with θ fixed is equivalent to the following linear regression problem:

$$\min_{\beta^*} \left\| \mathbf{R}\beta^* - \mathbf{R}\beta - \mathbf{R}^T \mathbf{g} \right\|^2, \quad (6.6)$$

where β^* is constrained to satisfy the conditions specified in Section 6.3.1 and $\| \cdot \|$ denotes the L_2 -norm. Several constraints are required as discussed in Section 6.3.1, hence several steps are needed as follows, where $I_l = -\alpha_2 + \sum_{j=1}^p \omega_j$ and $I_r = -\alpha_2 + \sum_{j=1}^m \pi_j$. Further, the infinite boundaries L and U are considered here, the procedure for finite ones is similar.

First, if both $I_l < 0$ and $I_r < 0$, then directly employ the `pnnls()` function to solve the optimization problem (6.6) only under $\beta_{-1}^* \geq 0$, where β_{-1}^* is β^* without its first element α_1^* . Otherwise, we need to eliminate a coefficient by replacing ω_p^* with $\alpha_2^* + \sum_{j=1}^q \rho_j^* - \sum_{j=1}^{p-1} \omega_j^*$, if $I_l \geq 0$; or replacing π_m^* with $\alpha_2^* - \sum_{j=1}^{m-1} \pi_j^* + \sum_{j=1}^k \lambda_j^*$, if $I_r \geq 0$. Then the `pnnls()` function can be used

to find a solution, say, β^* . Let $I_l^* = -\alpha_2^* + \sum_{j=1}^p \omega_j^*$, $\varphi^{*''}(L) = 2(-\alpha_2^* + \sum_{j=1}^p \omega_j^* - \sum_{j=1}^q \rho_j^*)$, $I_r^* = -\alpha_2^* + \sum_{j=1}^m \pi_j^*$ and $\varphi^{*''}(U) = 2(-\alpha_2^* + \sum_{j=1}^m \pi_j^* - \sum_{j=1}^k \lambda_j^*)$.

Second, in order to preserve the unimodality of φ^* , the algorithm then needs to check the following constraints:

- (a) Either $I_l^* \leq 0$ or $I_l^* > 0$ and $\varphi^{*''}(L) = 0$.
- (b) Either $I_r^* \leq 0$ or $I_r^* > 0$ and $\varphi^{*''}(U) = 0$.

If any is violated, a backtracking step is conducted so that both these two constraints (a) and (b) are satisfied. The backtracking procedure is similar as in Section 4.3.2.

In the third step, in order to ensure that φ^* is a log-density function, the algorithm further checks whether $\varphi^{*'}(L) = \alpha_1^* + 2 \sum_{j=1}^q \rho_j^* v_j^* - 2 \sum_{j=1}^p \omega_j^* \tau_j^* > 0$ and $\varphi^{*'}(U) = \alpha_1^* - 2 \sum_{j=1}^m \pi_j^* \theta_j^* + 2 \sum_{j=1}^k \lambda_j^* \mu_j^* < 0$ are violated. If any is, the backtracking step is executed again to make these two constraints be satisfied.

Finally, the usual step-halving backtracking is conducted to ensure that that log-likelihood increases monotonically.

6.4.3 Expanding and reducing knot sets

We now turn to how to expand and reduce the knot sets. The general idea is similar to that in Section 4.3.3. In each iteration, the knot set \mathcal{V} is expanded by including all local maxima of the gradient functions and after solving problem (6.6) is reduced by discarding the redundant knots with zero masses.

For estimator LCCC, we need to handle four knot sets ν , τ , θ and μ at the same time. It has four gradient functions $d_j(\xi; \varphi)$, $j = \{1, 2, 3, 4\}$. The procedure of enlarging the knot sets \mathcal{V} needs to be carried out on seven intervals, respectively, $[x_1, v_q)$, $[v_q, \tau_1)$, $[\tau_1, \tau_p)$, $[\tau_p, \theta_1)$, $(\theta_1, \theta_m]$, $(\theta_m, \mu_1]$ and $(\mu_1, x_n]$.

We employ different gradient functions on different intervals, as shown in table 6.1. In the first step of the new algorithm, we expand the four knot sets, ν , τ , θ and μ , by finding and adding new knots in each of the four corresponding intervals, the first, third, fifth and seventh interval.

Tab. 6.1: Intervals with corresponding gradient functions.

Interval	$[x_1, v_q)$	$[v_q, \tau_1)$	$[\tau_1, \tau_p)$	$[\tau_p, \theta_1)$	$(\theta_1, \theta_m]$	$(\theta_m, \mu_1]$	$(\mu_1, x_n]$
Gradient functions	d_1	$d_1 \ \& \ d_2$	d_2	$d_2 \ \& \ d_3$	d_3	$d_3 \ \& \ d_4$	d_4

For the overlapped intervals, we take the second interval $[v_q, \tau_1)$ as an example and describe it in more detail. We employ both gradient functions $d_1(\xi; \varphi)$ and $d_2(\xi; \varphi)$ if $v_q \neq \tau_1$. For each gradient function, the point corresponding to the greatest value of each gradient function is recorded, say, v' for d_1 and τ' for d_2 . We do not add any new knot if both gradient values $d_1(v')$ and $d_2(\tau')$ are non-positive. Otherwise, we only add one knot with larger gradient value to the corresponding knot set. In more detail, we add v' to the knot vector \mathbf{v} if $d_1(v') > d_2(\tau')$, or else, τ' to the knot vector $\boldsymbol{\tau}$. The same way to add new knots for the forth interval $[\tau_p, \theta_1)$ and the sixth interval $(\theta_m, \mu_1]$.

After adding new knots, we update the coefficient vector $\boldsymbol{\beta}$ by solving problem (6.6), then reduce the knot sets of $\boldsymbol{\vartheta}$ by removing the elements with zero masses.

6.4.4 Algorithm

The algorithm for computing the LCCC estimator of a unimodal and heavy-tailed distribution is outlined as follows.

Algorithm 3 (CNMLCCC). Set $s = 0$. From an initial estimate φ_0 with a finite number of knots and $l(\varphi_0) > -\infty$, repeat the following steps.

1. Compute all local maxima of the gradient functions as specified above, which gives new knots $v'_{s1}, \dots, v'_{sd}, \tau'_{s1}, \dots, \tau'_{st}, \theta'_{s1}, \dots, \theta'_{sh}$ and $\mu'_{s1}, \dots, \mu'_{sb}$.
2. Set $\mathbf{v}'_s = (v'_{s1}, \dots, v'_{sd})^T$, $\boldsymbol{\tau}'_s = (\tau'_{s1}, \dots, \tau'_{st})^T$, $\boldsymbol{\theta}'_s = (\theta'_{s1}, \dots, \theta'_{sh})^T$, $\boldsymbol{\mu}'_s = (\mu'_{s1}, \dots, \mu'_{sb})^T$, and $\boldsymbol{\vartheta}'_s = (\mathbf{v}'_s, \mathbf{v}'_s, \boldsymbol{\tau}'_s, \boldsymbol{\tau}'_s, \boldsymbol{\theta}'_s, \boldsymbol{\theta}'_s, \boldsymbol{\mu}'_s, \boldsymbol{\mu}'_s)^T$ and $\boldsymbol{\beta}'_s = (\alpha_1, \alpha_2, \boldsymbol{\rho}'_s, \mathbf{0}^T, \boldsymbol{\omega}'_s, \mathbf{0}^T, \boldsymbol{\pi}'_s, \mathbf{0}^T, \boldsymbol{\lambda}'_s, \mathbf{0}^T)^T$.
3. Compute $\boldsymbol{\beta}_{s+1}^-$ by solving problem (6.6), execute the backtracking steps and conduct a step-halving line search.
4. Discard all knots with zero masses in $\boldsymbol{\beta}_{s+1}^-$, which gives $\boldsymbol{\beta}_{s+1}$ and $\boldsymbol{\vartheta}_{s+1}$ of φ_{s+1} . Stop if $l(\varphi_{s+1}) - l(\varphi_s) \leq \text{Tolerance}$. Set $s = s + 1$.

The tolerance for the stopping criterion in step 4 is set to 10^{-6} in our numerical studies.

6.5 Bootstrap Test

In order to examine that a particular data set comes from a special density, let us consider bootstrap testing for identifying the functional form (6.1) implied in the LCCC estimator. The test statistic used can be, e.g., the Anderson-Darling (AD) statistic, as similarly done in Section 4.4. The AD procedure proposed by Anderson and Darling (1954) is a general test to compare the fit of an

observed cumulative distribution function (CDF) to an expected CDF. The AD statistic is defined as

$$A^2 = n \int_{-\infty}^{+\infty} \frac{(F_n(x) - \widehat{F}(x))}{\widehat{F}(x)(1 - \widehat{F}(x))} d\widehat{F}(x), \quad (6.7)$$

where \widehat{F} is an estimated CDF and \widehat{F}_n the empirical CDF given by the sample. Anderson and Darling (1954) showed that given a sample, the statistic can be approximated by

$$A^2 = -n - \sum_{i=1}^n \frac{2i-1}{n} \left\{ \ln \widehat{F}(x_{(i)}) + \ln (1 - \widehat{F}(x_{(n+1-i)})) \right\}, \quad (6.8)$$

where $x_{(1)}, \dots, x_{(n)}$ are the order statistics of the sample. Compared with the Kolmogorov-Smirnov (KS) distance, the AD statistic gives more weight to observations at the tails of the distribution.

Suppose $x_1, \dots, x_n \stackrel{iid}{\sim} P_0$, and we seek a size $\alpha \in (0, 1)$ test of

$H_0 : P_0$ has the same functional form implied in the LCCC estimator.

against

$H_1 : P_0$ does not have the same functional form implied in the LCCC estimator.

Using the AD statistic, the bootstrap test proceeds as follows.

- (a) Compute the LCCC estimate \widehat{f} from the given sample, and denote by \widehat{F} the corresponding CDF.
- (b) Compute the test statistic \widehat{A}^2 using (6.8).
- (c) Use \widehat{F} as the null sampling distribution. For $s = 1, \dots, S$, draw an independent bootstrap samples, $x_{s1}^*, \dots, x_{sn}^*$, from \widehat{F} . For each bootstrap sample, first compute the estimate \widehat{f}_s^* and then the test statistic \widehat{A}_s^2 .
- (d) Reject H_0 , if $p \equiv (S + 1)^{-1} \sum_{s=1}^S \left\{ |\widehat{A}^2| \leq |\widehat{A}_s^2| \right\} < \alpha$.

Here p is an estimated p -value of the above bootstrap test. Let p_{ad} be the p -value obtained by using the AD statistic. Similarly, we can use the KS, mean, standard deviation, skewness and kurtosis statistics to assess how the LCCC estimator fits to a particular data. We denote by p_{ks} , p_{mean} , p_{sd} , p_{skew} and p_{kurt} the corresponding p -values.

6.6 Simulation Studies

A simulation study is carried out to illustrate the performance of several nonparametric density estimators for heavy-tailed distributions: the standard kernel estimator (KER), the logspline estimator of [Koo et al. \(1999\)](#) (LS), the unimodal estimator of [Meyer \(2012\)](#) (UMRS), the unimodal estimator of [Turnbull and Ghosh \(2014\)](#) (UMBP) and the LCCC estimator. Furthermore, we do not consider the joint parametric-nonparametric estimator of [Markovitch and Krieger \(2002\)](#), the Champernowne transformed kernel density estimators of [Buch-Larsen et al. \(2005\)](#) and [Buch-Kromann \(2009\)](#). The reasons are that the estimator of [Markovitch and Krieger \(2002\)](#) needs to choose a boundary between tail and center which is highly influential on the resulting estimator, and both of the transformed kernel estimators only handle positive values. In addition, the estimator of [Buch-Kromann \(2009\)](#) also needs a threshold for where the distribution should start with and this is very hard to decide from observed values.

Setup

The standard kernel estimate is obtained through the R function `density()` with the default bandwidth value, that is, $h = 0.9n^{-0.2} \min(\sigma, \frac{QR}{1.34})$ ([Silverman, 1986](#)), where σ is the standard deviation, QR the interquartile range and n the sample size.

The logspline (LS) estimator of [Koo et al. \(1999\)](#) applies cubic splines to estimate the log-density function, which always has two exponential tails outside the range of the sample. The estimator is available by function `logspline()` in the R package `logspline`. Note that the logspline estimator is not necessarily unimodal.

The key idea of the UMRS estimator ([Meyer, 2012](#)) is to apply spline regression to both sides of the mode. An R implementation for UMRS is available at <http://www.stat.colostate.edu/~meyer/denspline.htm>. The code is only for the case where the mode is known. The main function `unimodalspline()` has three input arguments without default settings: `x` (the vector of univariate data), `mode` and `support`. For `support`, [Meyer \(2012\)](#) suggested to choose it as the range of `x`-values provided by the R function `density()`. For `mode`, [Meyer \(2012\)](#) advised a polynomial kernel density estimation for estimating the mode. In order to use the algorithm of [Meyer \(2012\)](#), the mode is estimated by function `mlv()` in R package `modeest`. The argument `method` in `mlv()` is set to `kernel`, the `boot` to `TRUE` and `R` (bootstrap) to 150.

The UMBP estimator of [Turnbull and Ghosh \(2014\)](#) is a mixture of Beta densities. The mixing weights are estimated by minimizing a criterion based on the Anderson-Darling test statistic, which makes use of quadratic programming techniques subject to linear inequality constraints.

The number of weights is selected based on the Condition Number (CN)(more details will be given in real data study), AIC or BIC. In our study, the UMBP estimates are obtained by the R function `umd()` with default parameter settings, in particular, CN-base, which is available at <http://www.stat.ncsu.edu/information/library/papers/supp.2650/umd.R>. Note that the support in the default setting is given by $[x_1 - s/\sqrt{n}, x_n + s/\sqrt{n}]$, where x_1 and x_n are, respectively, the minimum and maximum of the sample, and s is its standard deviation. The support provided this way is narrower than the range of x -values provided by the R function `density()`.

In order to study the performance of the estimators, two heavy-tailed distributions are considered, the log-normal (LN) and Student's t (T) distributions, the densities of which are listed in Table 6.2. Note that for the log-normal distribution, the kurtosis is $e^{4\sigma^2} + 2e^{3\sigma^2} + 3e^{2\sigma^2} - 3$ and the mode is $e^{\mu - \sigma^2}$. For the Student's t -distribution, the kurtosis is $6/(\nu - 4) + 3$ for $\nu > 4$ and ∞ for $2 < \nu \leq 4$, and the mode is 0. Moreover, the Student's t -distribution is approximately the standard normal distribution as ν approaches infinity. Figure 6.3 shows the log-densities and their second derivatives of LN (0, 1) and T (5), the two cases to be studied.

Tab. 6.2: Two heavy-tailed distributions we simulated from, where Γ is the gamma function.

Distribution	Notation	Density Function	Range	Parameters
Log-normal	$\text{LN}(\mu, \sigma)$	$\frac{\exp\{-0.5(\log x - \mu)^2/\sigma^2\}}{x\sigma\sqrt{2\pi}}$	$x > 0$	$\sigma > 0, \mu \in \mathbb{R}$
Student's t	$\text{T}(\nu)$	$\frac{\Gamma((\nu + 1)/2)}{\sqrt{\nu\pi}\Gamma(\nu/2)} \left(1 + \frac{x^2}{\nu}\right)^{-(1+\nu)/2}$	$x \in \mathbb{R}$	$\nu > 2$

Results

For each distribution shown in Figure 6.3, 100 random samples were generated for, respectively, a sample size of 100 and 1000, and the estimation results are summarized in Table 6.3. Each entry in the table is an empirical MISE or MHD value, with its standard error given in parentheses. The smallest of the expected losses of the estimators is highlighted in boldface for each given density and sample size. Boxplots for the ISE and HD of different estimators are also shown in Figure 6.4 for sample size 100 and Figure 6.5 for sample size 1000. The estimated densities and log-densities are displayed in Figure 6.6, each for a typical sample of size 1000, along with the true distribution (Truth).

These results show that the LCCC estimator that allows for log-convexity at the tails dominated the other ones, except in one case, i.e., T(5) with sample size 100 in terms of MHD, where LS performed slightly better. It is clear that the introduced tail shape constraint is indeed helpful for

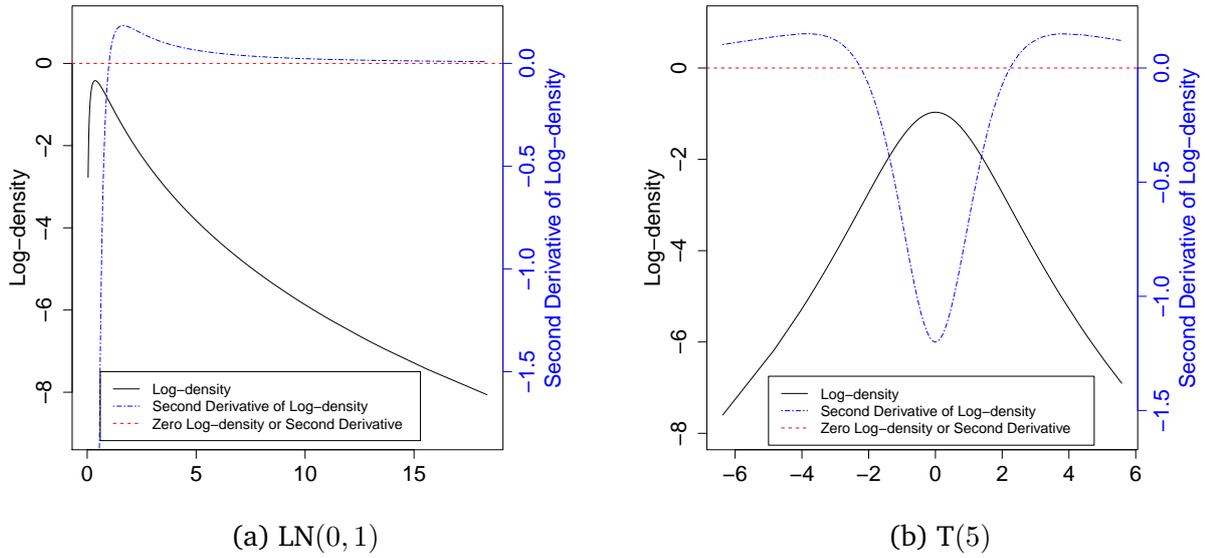


Fig. 6.3: Log-density and its second derivative of a log-normal (left) and a Student's t (right) distributions.

Tab. 6.3: Results of simulation studies for two heavy-tailed distributions, in terms of the MISE ($\times 10^{-2}$) and MHD ($\times 10^{-2}$), with standard errors given in parentheses.

Estimator	Distribution			
	LN (0, 1)		T (5)	
$n = 100$				
	MISE	MHD	MISE	MHD
KER	2.896 (0.118)	5.178 (0.100)	0.559 (0.033)	2.250 (0.071)
LS	2.111 (0.199)	1.976 (0.121)	0.815 (0.074)	1.363 (0.099)
UMRS	1.818 (0.098)	3.218 (0.100)	0.691 (0.037)	2.636 (0.254)
UMBP	4.280 (0.154)	5.976 (0.126)	1.326 (0.147)	4.538 (0.285)
LCCC	1.368 (0.105)	1.512 (0.074)	0.538 (0.033)	1.514 (0.114)
$n = 1000$				
KER	1.131 (0.024)	1.852 (0.015)	0.115 (0.007)	0.539 (0.012)
LS	0.361 (0.021)	0.369 (0.014)	0.121 (0.010)	0.238 (0.011)
UMRS	1.047 (0.071)	1.351 (0.067)	0.129 (0.007)	1.397 (0.141)
UMBP	5.718 (0.356)	7.382 (0.644)	2.738 (0.247)	6.240 (0.628)
LCCC	0.174 (0.010)	0.234 (0.009)	0.096 (0.007)	0.205 (0.009)

estimating a heavy-tail distribution.

It is worth pointing out that the difference between the LCCC and LS estimators in the log normal distribution is more significant than in the Student's t -distribution. This is owing to the flexibility of the LCCC estimator for dealing with heavy tails. The LS estimator always uses the first and last order statistics as the first and last knots respectively and is thus forced to have two exponential tails outside the range of the sample, while the LCCC estimator is fully automatic and can have two, one, or even no exponential tail, without fixing any knot. The log normal only has a right exponential tail and the Student's t -distribution has two heavy tails. Moreover, one can see from Figure 6.6 that the LS estimate is not necessary unimodal.

Of all compared estimators, the UMBP estimator performed worst by all criteria. This is mainly because the UMBP estimator is a mixture of Beta densities. When the number of mixture components is not large, a mixture of Beta densities can not fit well to a heavy-tailed distribution. This problem was also pointed out by [Turnbull and Ghosh \(2014\)](#) themselves. In order to remedy this issue, they suggested that, instead of using CN-based method, one can manually increase the number of mixture components m , until the p -value obtained from the Kolmogorov-Smirnov test is greater than 0.5. The estimate obtained this way tends to outperform the AIC- or CN-based estimate for fitting to the data which has a high peak and heavy tails. We will apply this remedied UMBP estimator to the real-world data.

Between the KER and UMRS estimators, the latter was better in the log normal scenario but became slightly worse in the Student's t scenario. The KER estimator tends to fluctuate a lot at tails and almost always fails to achieve unimodality. The UMRS estimator always captures the high peak and heavy tails within the support but gives zero mass outside the support. This leads to poor estimation at locations far from the observed values. It is interesting to find that except the UMBP estimator, both MISE and MHD of an estimator decrease as the sample size increases. Because the UMBP is not appropriate for fitting to a heavy-tailed distribution if the number of mixture components is not large.

6.7 Financial Data

Financial data is used to further study the above-mentioned estimators. Three financial assets are considered: the Standard and Poor's 500 index (SPX), the exchange rate between the NZ and US dollars (NZD/USD) and the stock price of Microsoft Corporation (MSFT). More information about the three corresponding datasets is given in Table 6.4 which were downloaded from Yahoo! Finance (<http://finance.yahoo.com>) and Pacific Exchange Rate Service (<http://fx.sauder.ubc.ca/data.html>).

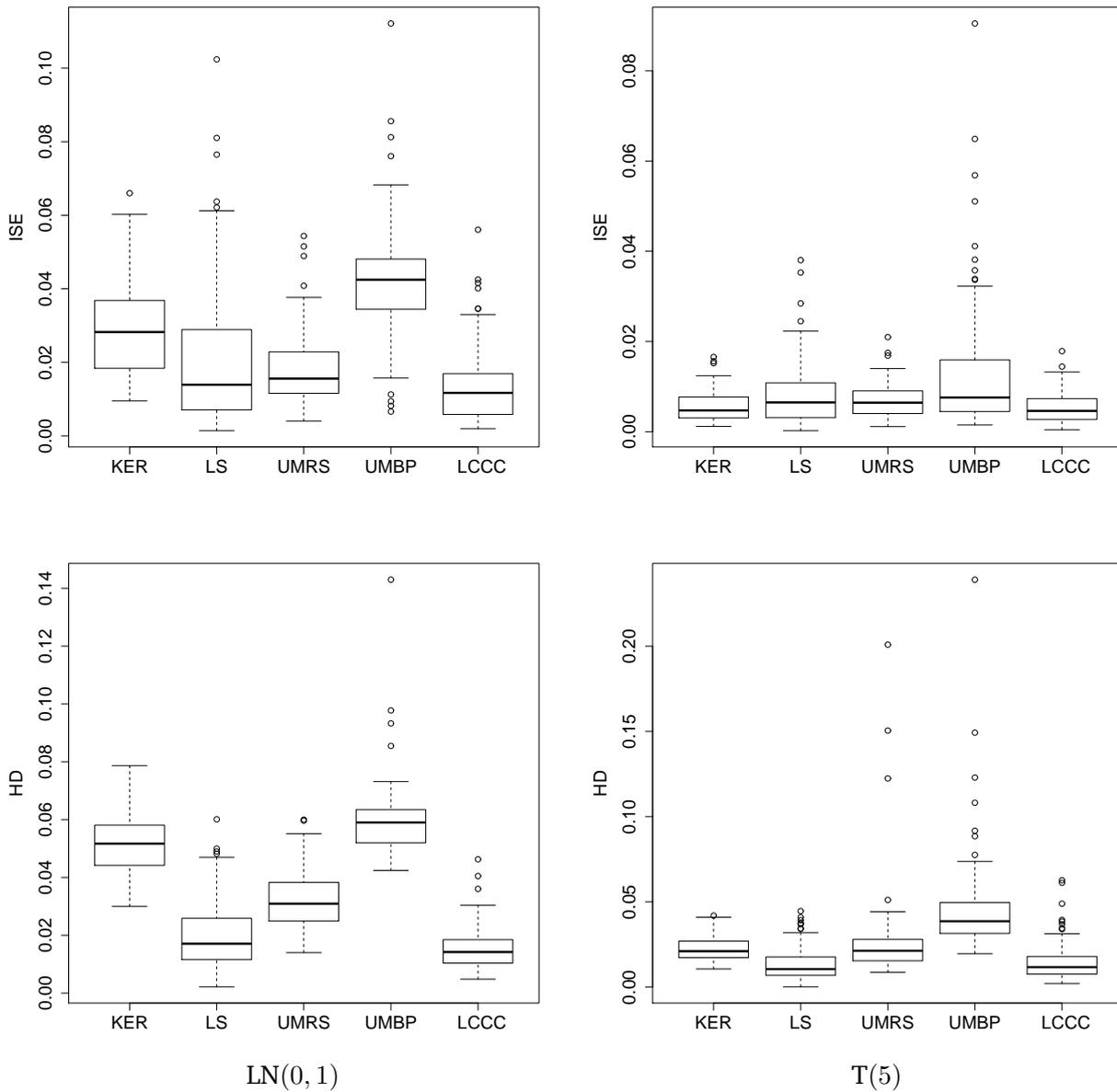


Fig. 6.4: Boxplots for the ISE (upper panels) and HD (lower panels) of different estimators for a log normal and a Student's t with sample size 100.

The log-return is the natural logarithm of the simple gross return of an asset and is most commonly used return in financial study. Note that the log return in day i is defined as

$$x_i = \log \frac{P_i}{P_{i-1}},$$

where P_i is the closing price of an asset in day i . The time series plots for raw data and log-returns are displayed in Figure 6.7.

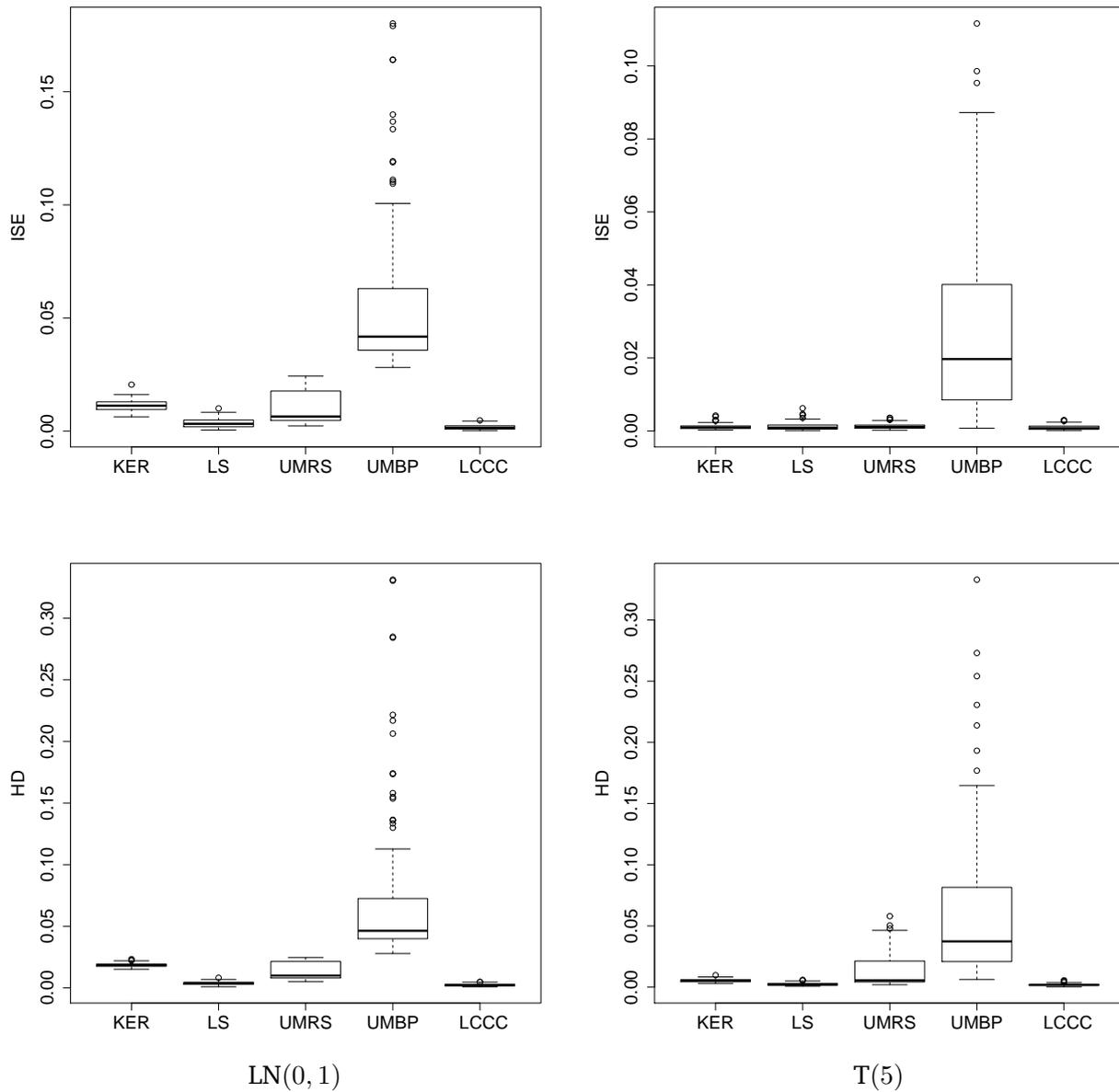


Fig. 6.5: Boxplots for the ISE (upper panels) and HD (lower panels) of different estimators for a log-normal and a Student's t with sample size 1000.

Tab. 6.4: Financial data (closing price).

Name	Observations	Time Period	Frequency	Source
MSFT	2518	28/02/2005 - 27/02/2015	Daily	Yahoo! Finance
SPX	3772	23/03/2000 - 23/03/2015	Daily	Yahoo! Finance
NZD/USD	5012	01/03/1995 - 27/02/2015	Daily	Pacific Exchange Rate Service

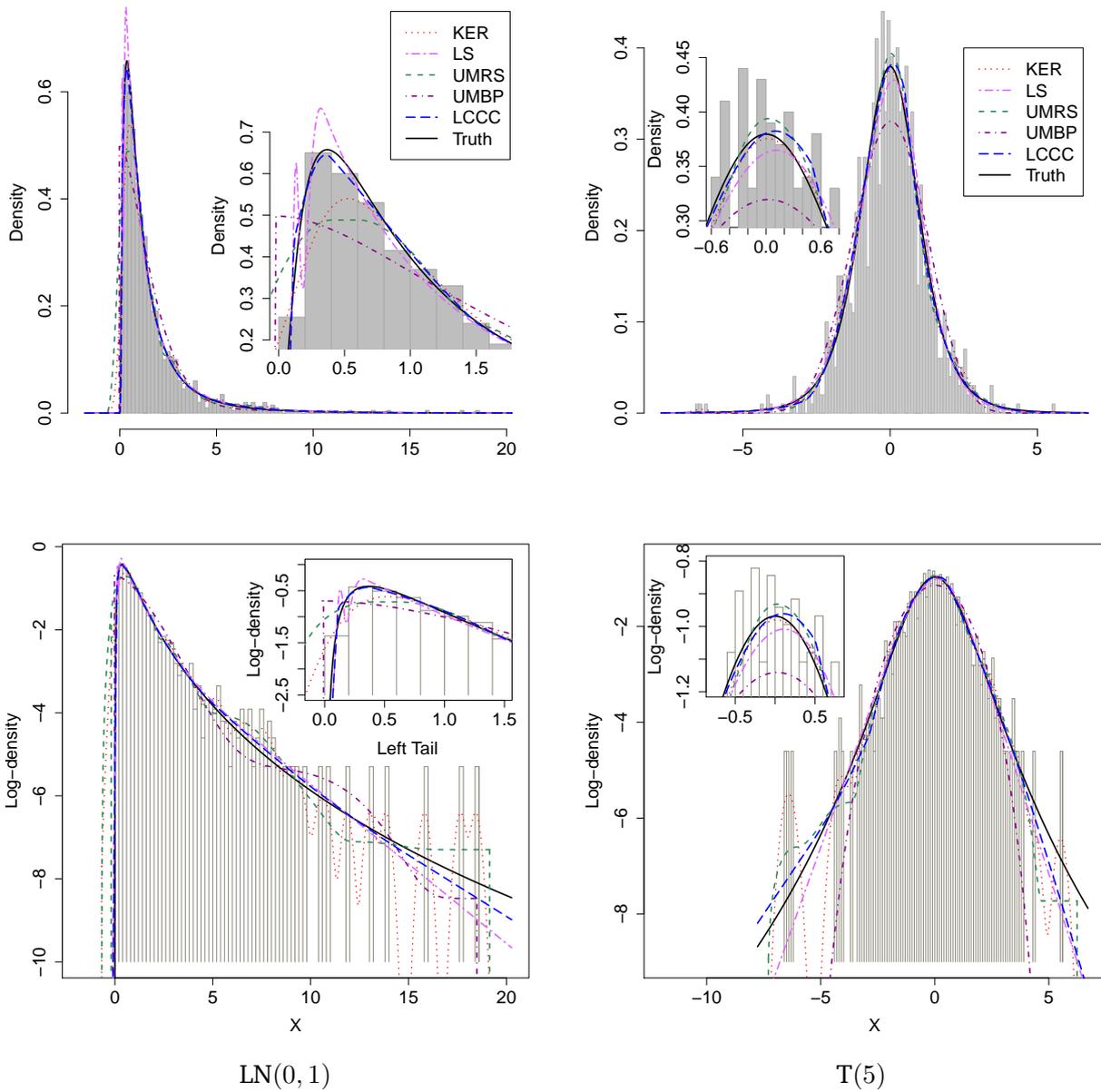


Fig. 6.6: Histograms and varying density estimates for typical log-normal and Student's t samples of size 1000.

6.7.1 Setup

Since the KER estimator almost always has spurious features at the tails, we excluded it from the real-world data study. We should also point out that, for the UMBP estimator, the optimal number of mixture components m is obtained based on the CN method with the default parameter settings in function `umd()`. However, the estimates obtained this way has extremely small p -values given by the KS test for all three datasets. Hence, as suggested by [Turnbull and Ghosh \(2014\)](#), we manually

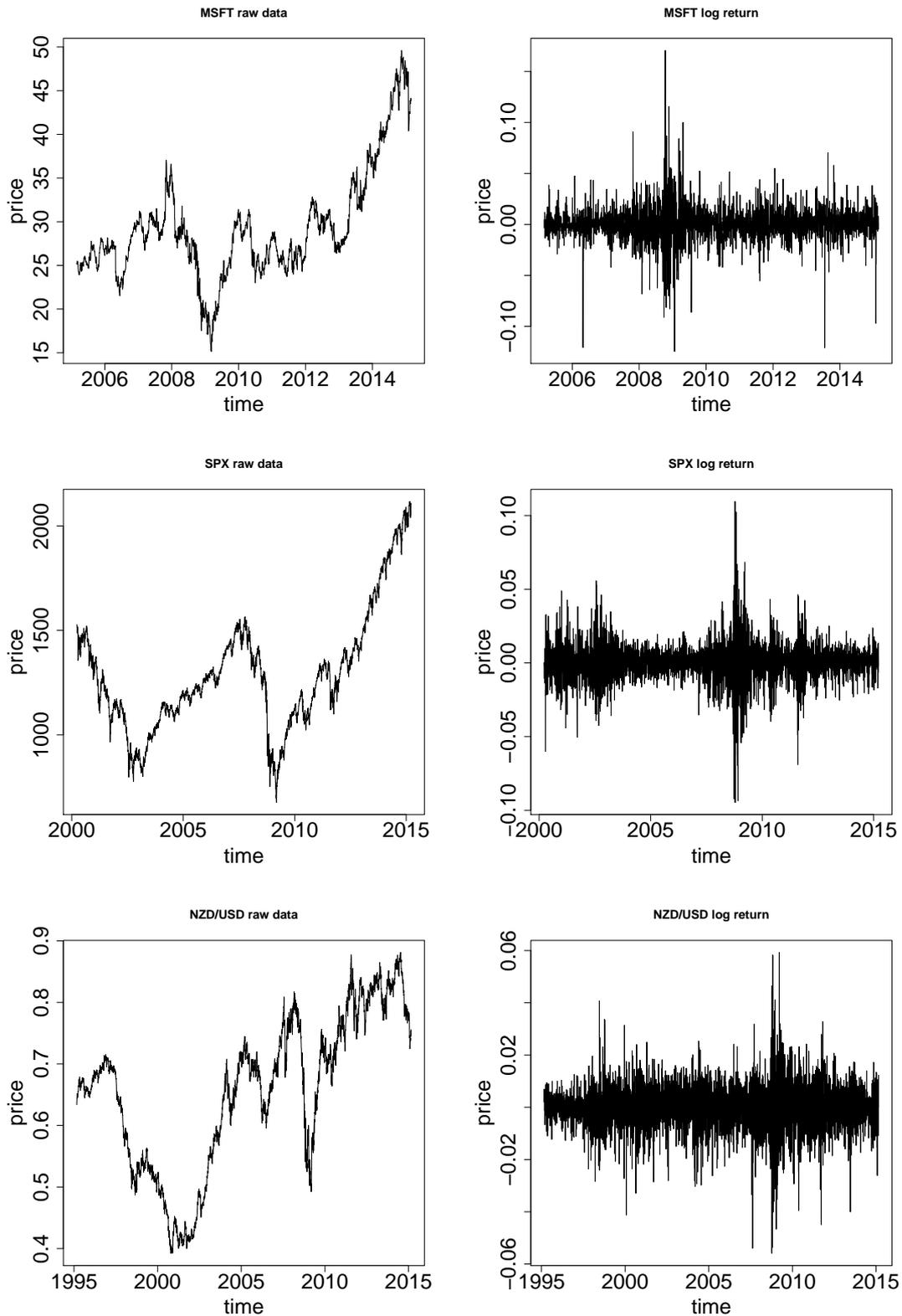


Fig. 6.7: Financial data overview: MSFT (top panels), SPX (middle panels) and NZD/USD (bottom panels).

increased m and found the smallest integer that gave a p -value greater than 0.5. The optimal numbers of mixture components m obtained from both automatic and manual methods are given in Table 6.5, for each of the three datasets. For NZD/USD, however, increasing the value of m can result in numerical errors when executing function `solve.QP()`, and we thus used the smallest m -value that gave a p -value greater than 0.5, without any error occurring.

Tab. 6.5: The optimal number of mixture components m in a UMBP estimate for each financial dataset.

Dataset	CN-based		Manual method	
	m	p -value	m	p -value
MSFT	17	$< 2.2 \times 10^{-16}$	172	0.531
SPX	19	$< 2.2 \times 10^{-16}$	290	0.524
NZD/USD	22	$< 2.2 \times 10^{-16}$	128	0.677

Furthermore, we included a parametric estimator for estimating a skewed Student's t distribution (ST), which was proposed by Fernández and Steel (1998). The density function is given by

$$f_s(x|\gamma) = \frac{2\gamma\Gamma(\frac{\nu+1}{2})}{\sigma\Gamma(\frac{\nu}{2})(\gamma^2+1)\sqrt{\pi\nu}} \left[1 + \frac{1}{\nu} \left(\frac{x-\mu}{\sigma} \right)^2 \left\{ \frac{1}{\gamma^2} I_{[\mu,\infty)}(x) + \gamma^2 I_{[-\infty,\mu)}(x) \right\} \right]^{-\frac{1+\nu}{2}},$$

where μ is the location parameter, σ the scale parameter, ν the shape parameter and γ the skewness parameter. Note that $\gamma = 1$ gives a symmetric distribution and values of $\gamma > 1$ (< 1) indicate right (left) skewness. Ehlers (2015) pointed out that this version for a skewed t -distribution has many advantages: easy and fast calculation of the moments and separate effect of the skewness and tail parameters. Some functions about the distribution are provided in the R package `fGarch`, and the MLE of the parameters can be obtained by function `sstdFit()`.

Besides density and QQ plots, we also employ the four criteria that were used in the real-world data study in Chapter 4 to evaluate the performance of the estimators. Recall the four loss

functions,

$$\begin{aligned} \text{ISE}(\hat{f}_n, \hat{f}) &= \int_{\mathcal{X}} \{\hat{f}(x)\}^2 dx - \frac{2}{n} \sum_{i=1}^n \hat{f}_n(x_i), \\ \text{KL}(\hat{f}_n, \hat{f}) &= -\frac{1}{n} \sum_{i=1}^n \log\{\hat{f}(x_i)\}, \\ \text{KS}(\hat{F}_n, \hat{F}) &= \sup_{\mathcal{X}} |\hat{F}_n - \hat{F}|, \\ \text{AD}(\hat{F}_n, \hat{F}) &= -n - \sum_{i=1}^n \frac{2i-1}{n} \{\ln \hat{F}(x_i) + \ln(1 - \hat{F}(x_{n+1-i}))\}, \end{aligned}$$

where \hat{f}_n denotes the empirical mass function from a test set of size n , \hat{f} a density estimator based on a training set, and \hat{F}_n and \hat{F} are the corresponding CDFs. Some additive constants are ignored in the above expressions.

6.7.2 Testing for financial data

A numerical summary of log-returns in each dataset is given in Table 6.6. They are all centred close to zero and have heavy-tailed (high kurtosis), as compared with a normal distribution. Of the three, MSFT has the highest kurtosis, the largest standard deviation and the least skewness. NZD/USD has the lowest kurtosis, the smallest standard deviation and the largest skewness.

Tab. 6.6: Descriptive statistics of the daily log-returns of financial datasets.

Variable	n	Mean ($\times 10^{-4}$)	SD ($\times 10^{-2}$)	Skewness	Kurtosis
MSFT	2517	2.207	1.724	-0.012	13.802
SPX	3771	0.850	1.273	-0.190	11.262
NZD/USD	5011	0.351	0.815	-0.311	8.022

We have pointed out that log-concavity is not appropriate for the data which has a very high kurtosis and heavy tails in Chapter 4. It is also easy to show that log-concavity is not proper for these three datasets. The LCCC estimate with log-convexity tail is necessary. In order to find out how well the LCCC estimator fit to these three datasets, we have run the bootstrap tests described in Section 6.5 on these three datasets. Each p -value in Table 6.7 is based on 9999 bootstrap samples. Furthermore, each p_{ad} or p_{ks} indicates that it seems safe to assume that the underlying density of the data has the same functional form as that implied by the LCCC estimator. Other p -values also show that the LCCC estimator is very suitable for these data. Furthermore, for each dataset, p_{ad} is always less than p_{ks} , which further shows that the Anderson-Darling test has much more power for

identifying the functional form of a given density.

Tab. 6.7: Test for the daily log-returns of financial datasets.

Data	p_{ad}	p_{ks}	p_{mean}	p_{sd}	p_{skew}	p_{kurt}
MSFT	0.794	0.987	0.995	0.978	0.884	0.982
SPX	0.758	0.986	0.960	0.960	0.964	0.924
NZD/USD	0.753	0.870	0.986	0.948	0.870	0.876

6.7.3 Density estimation

Full density estimation results

The gradient function plots at the convergence of the CNMLCCC algorithm for the three datasets are shown in Figure 6.8. The values of $d_{0,1}$, $d_{0,2}$ and the maximum of the gradient functions are all zero, from which we can know that the CNMLCCC has successfully found the SNPMLE. More precisely, the piecewise quadratic log-density functions for the MSFT, SPX and NZD/USD data are given by, respectively,

$$\begin{aligned}\widehat{\varphi}(x) = & -286.12x - 22180x^2 - 5688(-0.050 - x)_+^2 + 4971(-0.048 - x)_+^2 + 22897(-0.009 - x)_+^2 + \\ & 16590(x + 0.008)_+^2 + 1745(x - 0.002)_+^2 + 4076(x - 0.011)_+^2 + 625(x - 0.061)_+^2 + \\ & 1592(x - 0.073)_+^2 - 2448(x - 0.079)_+^2 + 2.40,\end{aligned}$$

$$\begin{aligned}\widehat{\varphi}(x) = & 107.02x - 91866x^2 - 19847(-0.041 - x)_+^2 + 19578(-0.040 - x)_+^2 + 353(-0.038 - x)_+^2 + \\ & 91782(-0.00004 - x)_+^2 + 83835(x - 0.00084)_+^2 + 7503(x - 0.0067)_+^2 + \\ & 5103(x - 0.0179)_+^2 - 2518(x - 0.0197)_+^2 - \\ & 1725(x - 0.0302)_+^2 - 332(x - 0.099)_+^2 + 3.98,\end{aligned}$$

$$\begin{aligned}\widehat{\varphi}(x) = & -2.24x - 120981x^2 - 1912(-0.039 - x)_+^2 - 1093(-0.020 - x)_+^2 + 3779(-0.016 - x)_+^2 + \\ & 12901(-0.0046 - x)_+^2 + 107306(-0.00025 - x)_+^2 + 114390(x - 0.00029)_+^2 + \\ & 37812(x - 0.0164)_+^2 - 29645(x - 0.0179)_+^2 - 1576(x - 0.0500)_+^2 + 4.24.\end{aligned}$$

Density and log-density function plots from all compared estimates are displayed in Figure 6.9. One can see that the major differences among these estimates are around the peak and tails. For each dataset, the LCCC estimator can not only capture well the peak, but also fit well to the heavy tails and give positive mass outside the range of the sample in a nice, smooth fashion. The LS estimator can moderately capture the heavy tails and give positive mass outside the range of the

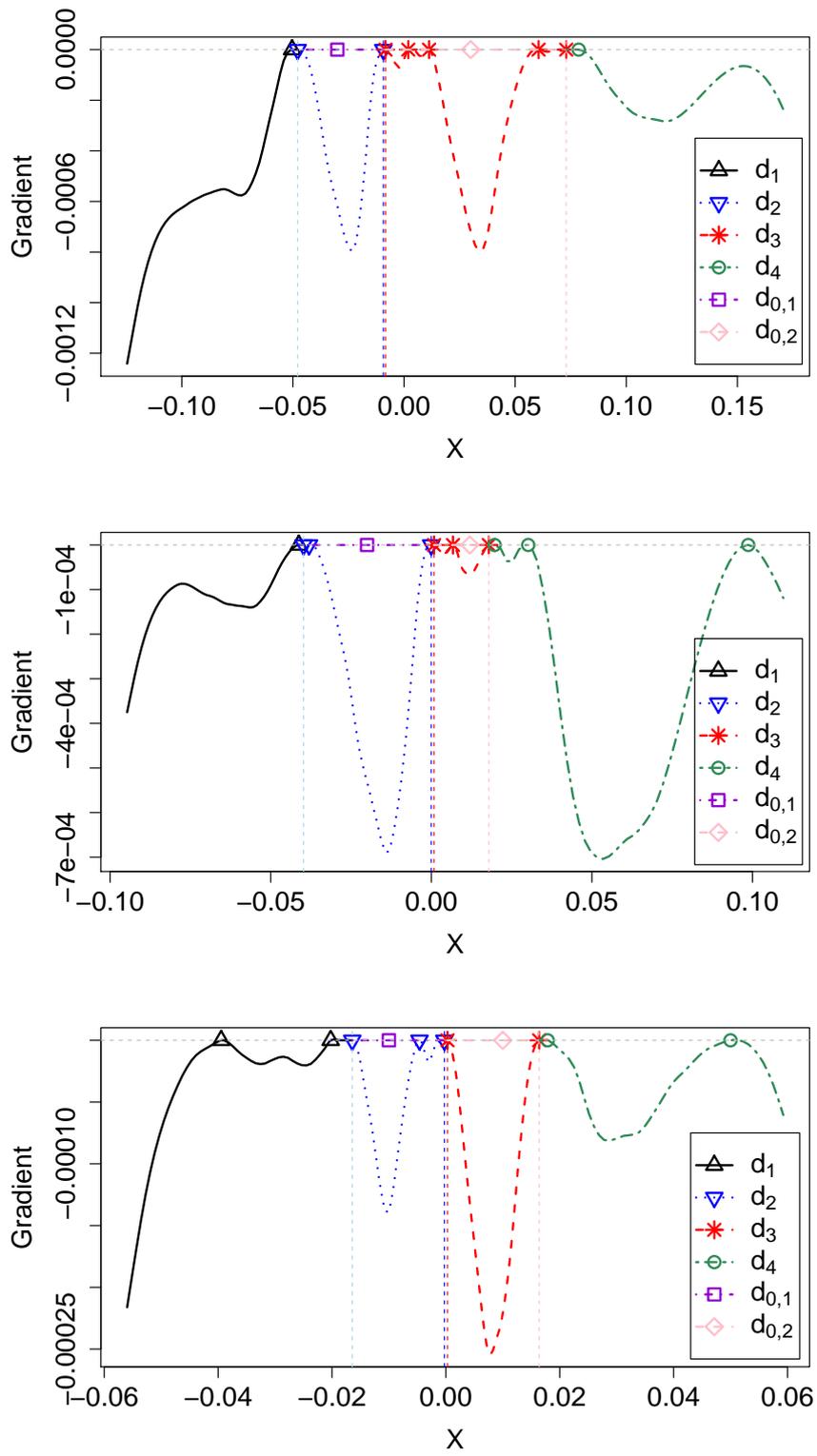


Fig. 6.8: Gradient curves at convergence computed by CNMLCCC for MSFT (top), SPX (middle) and NZD/USD (bottom).

sample, but it can not always capture well the peak, as for the NZD/USD data. The ST estimator behaves similarly to the LCCC estimator at the tails for each dataset, but it may become worse around the peak. Furthermore, both the UMRS and UMBP estimators sharply decrease to zero around the boundaries for every dataset. Note that the UMBP estimator decreases to zero before estimator UMRS. The main reason may be that the UMRS estimator has a wider support that is the range of x -values given by R function `density()` than estimator UMBP, which is support on $[x_1 - s/\sqrt{n}, x_n + s/\sqrt{n}]$. Note that manually increasing the number of mixture components for estimator UMBP significantly improves its performance for fitting high-peaked and heavy-tailed data, but the manual approach is inconvenient and at a higher computation cost.

The estimation results are summarized in Table 6.8. The sample statistics are highlighted in boldface. Note that for the ST estimator, its kurtosis is infinity when its degree of freedom is between 2 and 4, i.e., $2 < \nu \leq 4$, and its skewness is undefined when $\nu \leq 3$. For each dataset, the ST estimator gives apparently an improper estimate of the mean. Overall, the LCCC estimator has the best estimate of the mean and excellent estimates of the standard deviation and skewness. For every dataset, the kurtosis of the LCCC estimate is smaller than that of the ST estimate, and greater than those of the other three nonparametric estimates. In fact, the sample kurtosis for a finite samples can be severely biased, more precisely, downwardly biased; see Bao (2013) and the references therein. The kurtosis, as a measure of tail fatness, could be “substantially underestimated in practice” because usually there are not enough tail observations even in a large sample (Bai and Ng, 2005). From this point of view, the LCCC also provides the best estimate of the kurtosis than other three nonparametric ones. Among the LS, UMRS and UMBP estimators, the LS estimator provides the best estimate of the mean and the UMRS estimator gives good estimate of the skewness.

We also wrote some extra R functions for the UMRS estimator, including a quantile function. Q-Q plots of all the density estimates for each data set are shown in Figure 6.10 for MSFT, in Figure 6.11 for SPX and in Figure 6.12 for NZD/USD. For MSFT data, it can be seen that both the LS and LCCC estimators work well, with small deviations at the tails, whereas the other three estimators are relatively worse. For SPX data, the ST estimator looks worst, and the other four estimators are similar. This clearly suggests the inappropriateness of the used of the parametric estimator ST for this dataset. For NZD/USD data, both the LS and UMRS estimator provide relatively worse fitting, and there is small difference among others.

Cross-validation results

Due to the improper estimation of the mean and skewness by the ST estimator, and the time consuming and troublesome way to choose the optimal number of mixture components in the

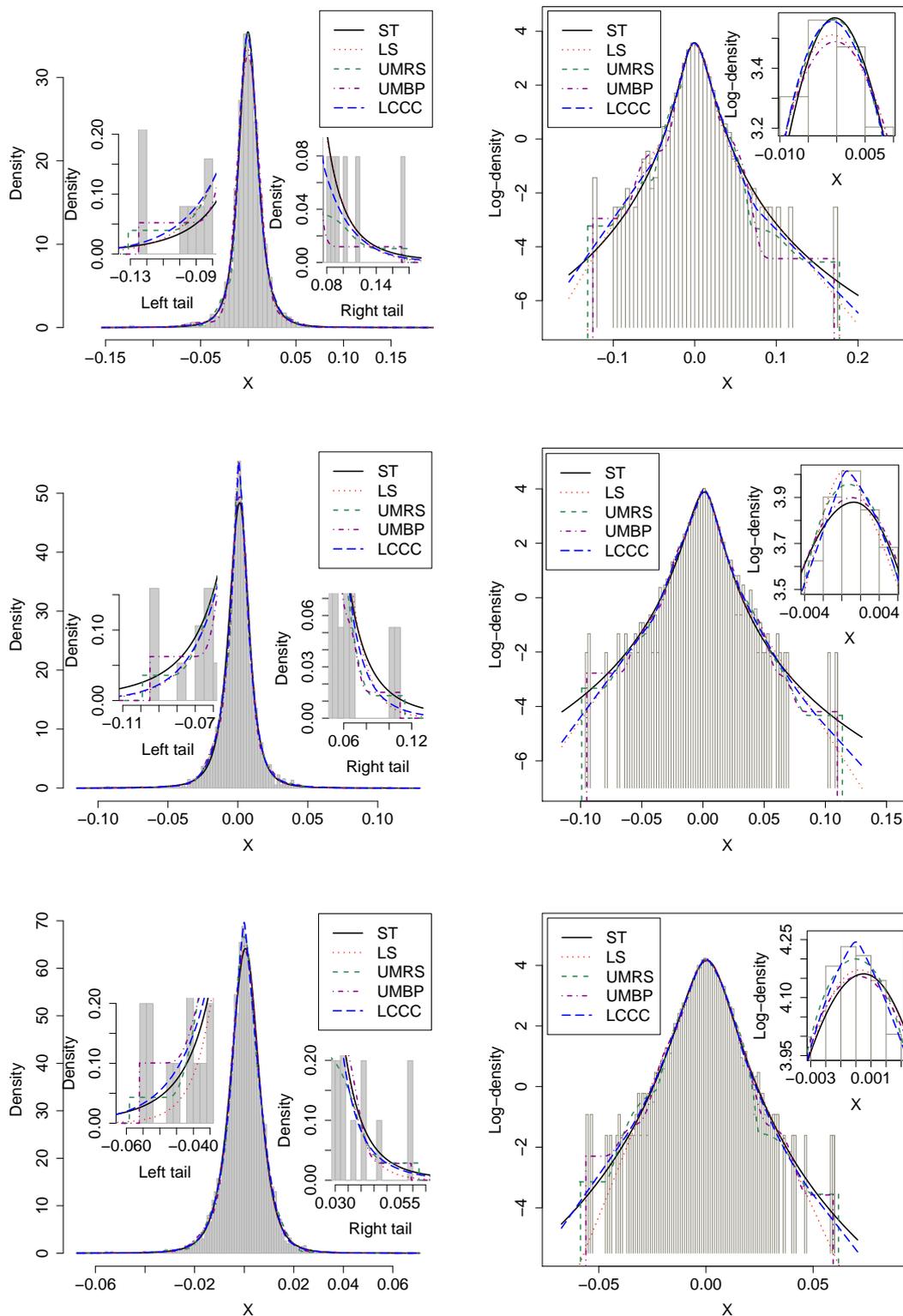


Fig. 6.9: Histograms and different estimates for daily log-returns: MSFT (top panels), SPX (middle panels) and NZD/USD (bottom panels). The left panels are density plots and the the log-density ones.

Tab. 6.8: Model fitting for financial data sets.

<i>Estimator</i>	<i>Mean</i> ($\times 10^{-4}$)	<i>SD</i> ($\times 10^{-2}$)	<i>Skewness</i>	<i>Kurtosis</i>
MSFT				
<i>Sample</i>	2.207	1.724	-0.012	13.802
ST	4.177	1.858	undefined	$+\infty$
LS	2.207	1.718	0.054	13.316
UMRS	2.385	1.697	-0.022	12.926
UMBP	1.600	1.704	-0.186	12.160
LCCC	2.207	1.728	-0.037	15.739
SPX				
<i>Sample</i>	0.850	1.723	-0.190	11.262
ST	-0.338	1.470	undefined	$+\infty$
LS	0.850	1.272	-0.236	11.001
UMRS	0.794	1.265	-0.218	10.195
UMBP	0.590	1.266	-0.309	10.235
LCCC	0.850	1.275	-0.213	12.190
NZD/USD				
<i>Sample</i>	0.351	0.815	-0.311	8.022
ST	0.862	0.827	-0.135	$+\infty$
LS	0.351	0.802	-0.131	6.021
UMRS	0.196	0.803	-0.329	7.389
UMBP	0.249	0.812	-0.366	7.548
LCCC	0.351	0.816	-0.333	8.639

UMBP estimator, we excluded both estimators from the cross-validation studies. For each dataset, we ran 2-fold cross-validation and conducted the paired t -test over 100 replications. The results are summarized in Tables 6.9 – 6.11 for MSFT, SPX and NZD/UAD, respectively. The p -values less than 5% are highlighted in boldface.

Note that the infinite values of KL and AD in the UMRS estimates are due to the zero mass given outside the support provided by a training set. Overall, the UMRS estimator performs worst in terms of KL and AD for all three datasets. The LCCC estimator gives the best performance in all cases, and except in only one case it is not significantly different from the LS estimator when

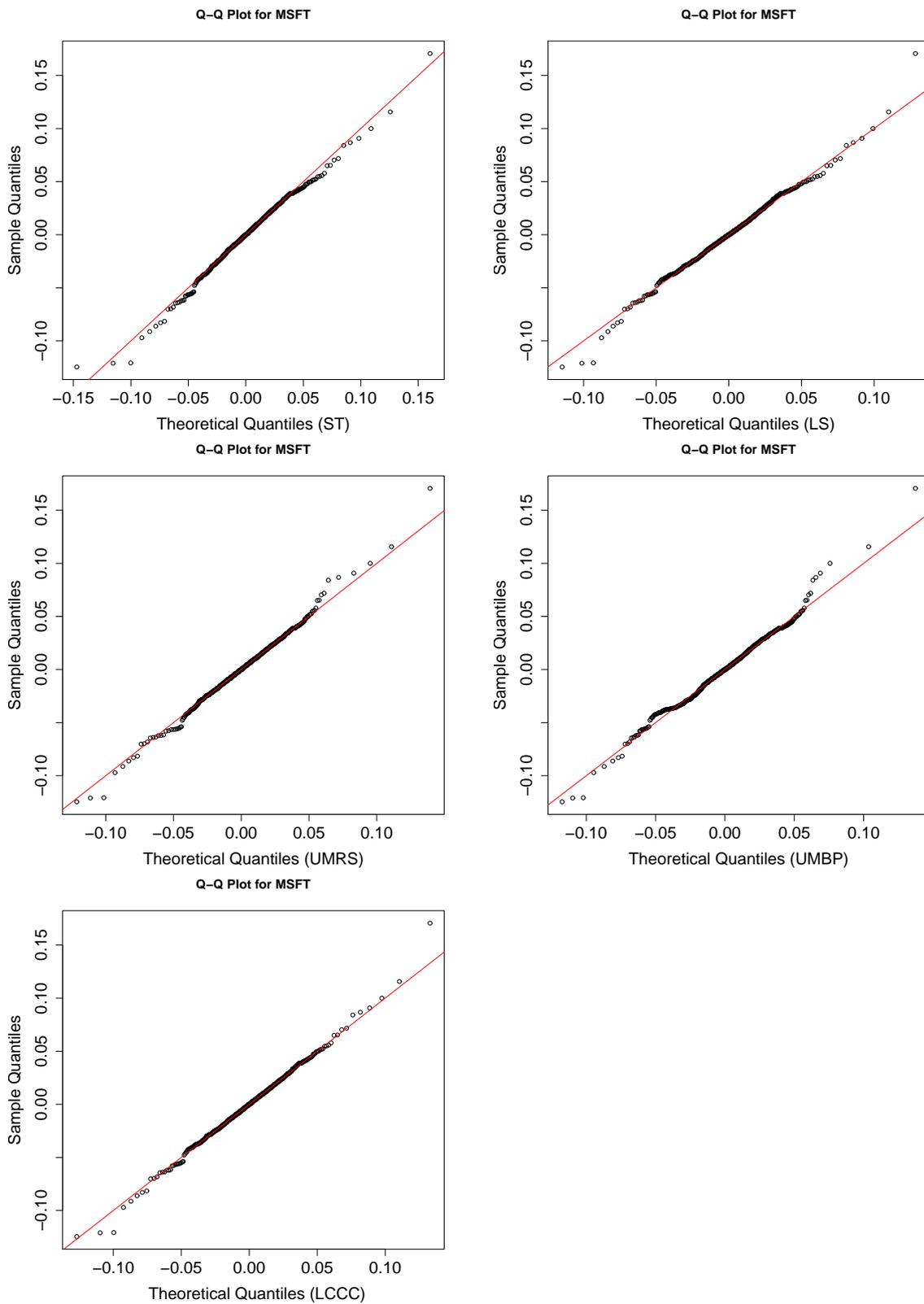


Fig. 6.10: Q-Q plots from different estimates for MSFT data.

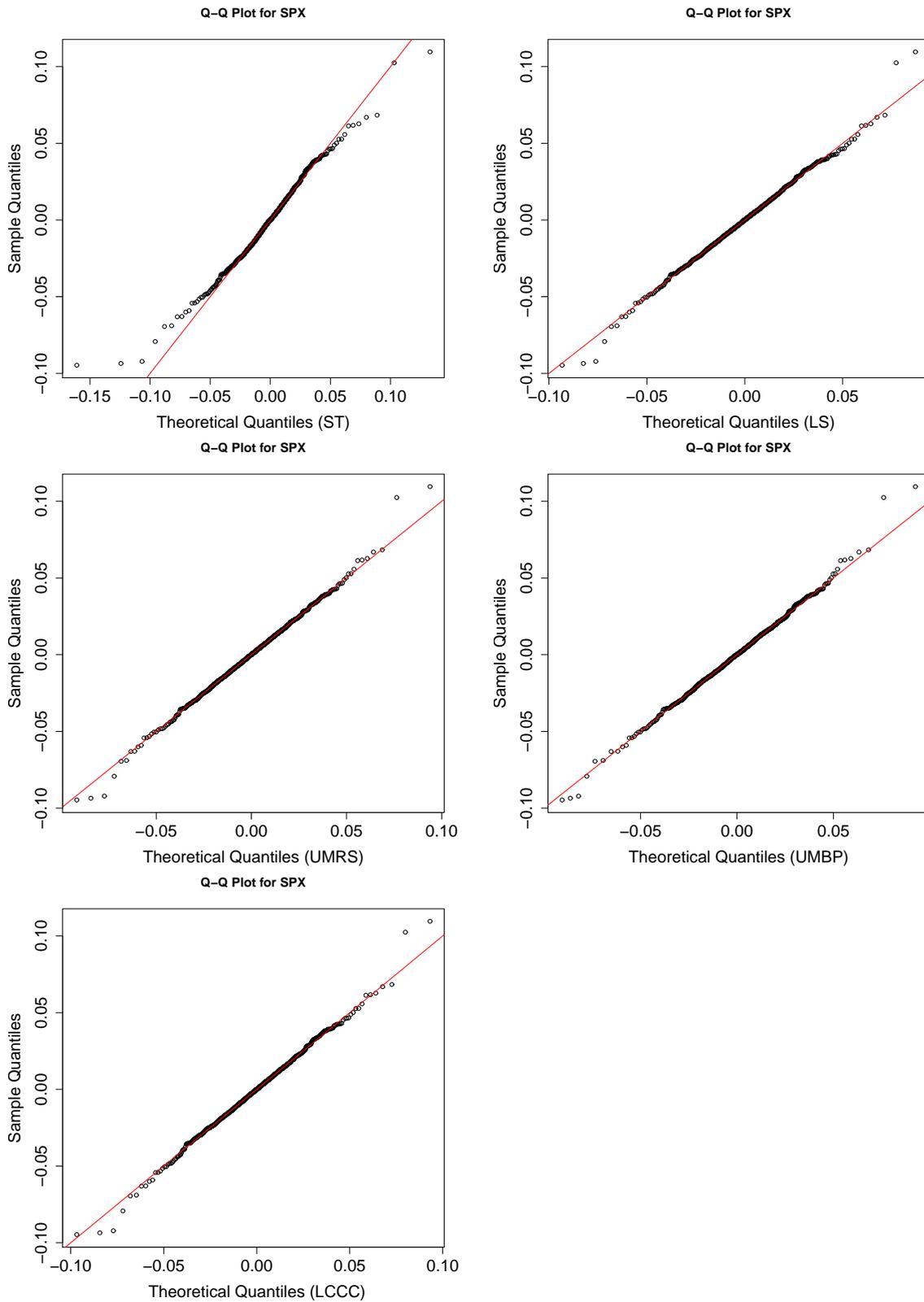


Fig. 6.11: Q-Q plots from different estimates for SPX data.

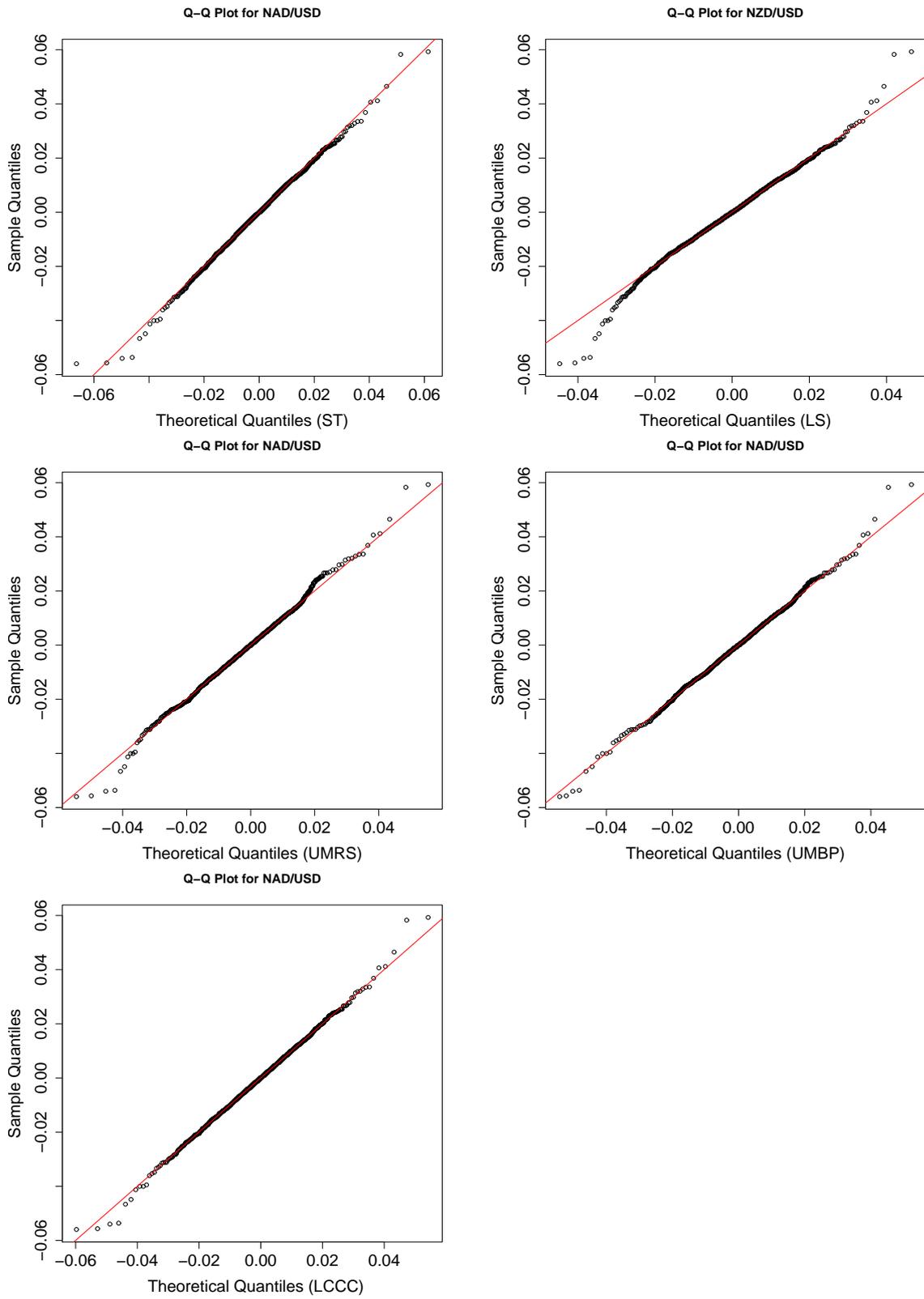


Fig. 6.12: Q-Q plots from different estimates for NZD/USD data.

6.8. Summary

assessed in terms of ISE fro MSFT data. The UMRS and LS estimators perform similarly in terms of ISE and KS for MSFT data. Furthermore, estimator UMRS is better than estimator LS in terms of KS for SPX data, and its outperformance is more significant in terms of both ISE and KS for NZD/USD data. This is likely because MSFT data has little, if any, skewness, but both SPX and NZD/USD are clearly left skewed. From the results in Table 6.8, one can see that the LS estimator provides a worse estimate of the skewness than the UMRS estimator.

Tab. 6.9: Paired t -tests based on cross-validation results for the MSFT dataset, where MD stands for the mean difference. 0 means existing infinity value in the comparison.

(I) Est.	(J) Est.	MD (I-J)	Std. Error	95% Confidence Interval		p-value
				Lower Bound	Upper Bound	
ISE						
LCCC	UMRS	-0.026	0.015	-0.038	-0.014	2.74×10^{-5}
	LS	-0.006	0.031	-0.031	0.019	0.64
UMRS	LS	0.020	0.034	-0.006	0.047	0.14
KL ($\times 10^{-3}$)						
LCCC	UMRS	$-\infty$	-	-	-	0
	LS	-2.520	1.005	-3.318	-1.723	9.46×10^{-9}
UMRS	LS	∞	-	-	-	0
KS ($\times 10^{-3}$)						
LCCC	UMRS	-0.485	0.207	-0.896	-0.074	0.02
	LS	-1.318	0.588	-2.486	-0.151	0.03
UMRS	LS	-0.833	0.572	-1.969	0.302	0.15
AD ($\times 10^{-2}$)						
LCCC	UMRS	$-\infty$	-	-	-	0
	LS	-0.145	0.048	-0.240	-0.051	2.93×10^{-3}
UMRS	LS	∞	-	-	-	0

6.8 Summary

In this chapter, we studied the smooth nonparametric estimation of a unimodal and heavy-tailed distribution under shape restrictions and applied the new method to some financial datasets. A novel idea has been proposed to formulate this new smooth estimator and a algorithm for computing this estimator is proposed based on the CNM method. More precisely, by considering the

6.8. Summary

Tab. 6.10: Paired t -tests based on cross-validation results for the SPX dataset, where MD stands for the mean difference. 0 means existing infinity value in the comparison.

(I) Est.	(J) Est.	MD (I-J)	Std. Error	95% Confidence Interval		p-value
				Lower Bound	Upper Bound	
ISE						
LCCC	UMRS	-0.038	0.017	-0.052	-0.024	3.46×10^{-7}
	LS	-0.052	0.027	-0.073	-0.030	6.78×10^{-6}
UMRS	LS	-0.014	0.028	-0.036	0.008	0.22
KL ($\times 10^{-3}$)						
LCCC	UMRS	$-\infty$	-	-	-	0
	LS	-3.046	0.575	-3.503	-2.590	$< 2.20 \times 10^{-16}$
UMRS	LS	∞	-	-	-	0
KS ($\times 10^{-3}$)						
LCCC	UMRS	-0.670	0.220	-1.107	-0.232	3.04×10^{-3}
	LS	-1.887	0.445	-2.770	-1.003	5.06×10^{-5}
UMRS	LS	-1.217	0.451	-2.112	-0.322	0.01
AD ($\times 10^{-2}$)						
LCCC	UMRS	$-\infty$	-	-	-	0
	LS	-0.283	0.046	-0.375	-0.191	1.83×10^{-8}
UMRS	LS	∞	-	-	-	0

second derivative properties, we imposed different restrictions on the tail and main body of the density function, in particular, log-convexity on the tail and log-concavity on the main body. The new estimator (LCCC) automatically splices together the parts that have log-concavity or log-convexity shapes in a smooth and seamless way. Bootstrap testing for identifying the functional form (6.1) implied in the LCCC estimator is also developed.

In simulation studies, several estimators have been compared, including KER, LS, UMRS, UMBP and LCCC. The results show that the LCCC estimator provides the best estimation of a heavy-tailed distribution. The UMBP estimator performs worst of all due to the number of mixture components chosen. The LS estimator may fail to achieve unimodality and its enforcing of two exponential tails is not always appropriate. The KER, UMRS and UMBP estimators all give zero mass outside the support, which leads to poor estimation of the tails in general.

In the real-world data analysis, three financial datasets have been considered. The bootstrap tests based on the AD and KS statistics have been carried out which indicate that the LCCC estimator

6.8. Summary

Tab. 6.11: Paired t -tests based on cross-validation results for the NZD/USD dataset, where MD stands for the mean difference. 0 means existing infinity value in the comparison.

(I) Est.	(J) Est.	MD (I-J)	Std. Error	95% Confidence Interval		p-value
				Lower Bound	Upper Bound	
ISE						
LCCC	UMRS	-0.060	0.014	-0.072	-0.049	$< 2.20 \times 10^{-16}$
	LS	-0.100	0.032	-0.126	-0.074	9.17×10^{-12}
UMRS	LS	-0.040	0.036	-0.068	-0.011	6.59×10^{-3}
KL ($\times 10^{-3}$)						
LCCC	UMRS	$-\infty$	-	-	-	0
	LS	-2.893	0.508	-3.297	-2.490	$< 2.20 \times 10^{-16}$
UMRS	LS	∞	-	-	-	0
KS ($\times 10^{-3}$)						
LCCC	UMRS	-0.716	0.172	-1.057	-0.375	6.57×10^{-5}
	LS	-2.159	0.391	-2.935	-1.382	2.76×10^{-7}
UMRS	LS	-1.443	0.390	-2.216	-0.670	3.52×10^{-4}
AD ($\times 10^{-2}$)						
LCCC	UMRS	$-\infty$	-	-	-	0
	LS	-0.390	0.056	-0.500	-0.279	2.89×10^{-10}
UMRS	LS	∞	-	-	-	0

can provide a good fit to each data set, when the distribution is unimodal but not log-concave. In this study, one parametric estimator (ST) and four nonparametric ones (LS, UMRS, UMBP, LCCC) are compared. Overall, the LCCC estimator basically dominates the other estimators and shows a great advantage at handling such financial data. The ST estimator does not estimate well the skewness and mean, even though the model involves four parameters. The LS estimator performs well in the case where no clear skewness exists, but it becomes worse when skewness is moderate or large. The performance of the UMBP estimator can be improved by manually increasing the number of mixture components, but it still has difficulty with estimating heavy tails, which is also a problem for the UMRS estimator.

In finance, tail behaviour analysis is vital for risk management. The excellent performance of the LCCC estimator in this regard gives it a great potential of success in this field.

Chapter 7

Heavy Tails and Value at Risk Estimation

7.1 Introduction

Financial asset returns often possess distributions with tails heavier than those of the normal distribution, as shown in the empirical evidence in Chapter 6. In fact, in as early as 1963, [Mandelbrot \(1963\)](#) pointed out that the distribution of asset returns is not well approximated by the Gaussian distribution. Since then many models have been proposed to model heavy-tailed returns of financial assets. The heavy-tailed, highly peaked nature of financial asset returns makes risk management more important in a financial institution. In a traditional way, risk is always measured by volatility. The main problem with volatility, however, is that it treats all uncertainty as risk, regardless of direction. The investors only worry about their losses, but they are not distressed by gains.

Value at risk (VaR) is one of the main indicators for risk management of financial portfolios ([Jorion, 2007](#)). It can also be found in other forms of risk to which the firm is exposed, such as credit, liquidity, and legal risk; see [Bradley and Taqqu \(2003\)](#) and the references therein. A VaR statistic has three components: a time period, a confidence level and a loss amount (or loss percentage). It is defined as the maximum potential loss of financial instrument with a given probability over a certain time period. Statistically, it refers to a quantile which depends crucially on the shape of the return distribution. Usually extreme levels of probability, such as 5% or 1%, are of interest in financial markets. Given a confidence level $\alpha \in (0, 1)$, the VaR of the portfolio at the confidence level α is given by the smallest number l such that the probability that the loss L exceeds l is at most $(1 - \alpha)$. Mathematically, if L is the loss of a portfolio, then $\text{VaR}_\alpha(L)$ is the level

α -quantile, i.e.,

$$\text{VaR}_\alpha(L) = \inf\{l \in \mathbb{R} : P(L > l) \leq 1 - \alpha\} = \inf\{l \in \mathbb{R} : F_L(l) \geq \alpha\},$$

where $F_L(\cdot)$ is the distribution function of loss L . Therefore, VaR is just the quantile of the loss distribution

$$\text{VaR}_\alpha = F_L^{-1}(\alpha). \tag{7.1}$$

If our portfolio has only asset X , the profit-and-loss distribution is X itself. Furthermore, the loss distribution is the distribution of random variable $L = -X$. Since risk management usually studies large losses, we can focus on large losses, at the upper tail of the loss distribution.

7.2 Estimators for Comparison

In this chapter, we restrict our interest to the tail performance of estimators. We include the LS and LCCC in the study below. The UMRS and UMBP estimators, as shown in the full density estimation results in Chapter 6, have difficulty in tail estimation, and hence are not considered.

The study also includes the transformation kernel density estimator based on the Champernowne distributions (KMCE) of Buch-Larsen et al. (2005) which has a good performance on the tails. The KMCE estimator was proposed based on a kernel method by transforming the original data using a modified Champernowne distribution. The Champernowne distribution is a generalization of the logistic distribution and was introduced by Champernowne (1953), who developed the distribution to describe the logarithm of income. In Buch-Larsen et al. (2005), the modified Champernowne density and distribution functions are given by, respectively,

$$f_{\text{Cha}}(x) = \frac{\alpha(x+c)^{\alpha-1} \{(M+c)^\alpha - c^\alpha\}}{\{(x+c)^\alpha + (M+c)^\alpha - 2c^\alpha\}^2},$$

and

$$F_{\text{Cha}}(x) = \frac{(x+c)^\alpha - c^\alpha}{(x+c)^\alpha + (M+c)^\alpha - 2c^\alpha},$$

where $\alpha, M > 0$, and $c, x \geq 0$. The Champernowne distribution is a heavy-tailed distribution converging to a Pareto distribution as $x \rightarrow \infty$.

Lehmann (1983) showed that the empirical median is a robust estimator for the true median

M , especially for heavy-tailed distributions. By recognizing that $F_{\text{Cha}}(M) = 0.5$, Buch-Larsen et al. (2005) estimated the parameter M by the empirical median, and the parameters α and c based on maximum likelihood. Clements et al. (2003) also pointed out that choosing M as the empirical median can significantly simplify the parameter estimation procedure and results in only marginal difference from the estimation by maximum likelihood method.

Buch-Kromann (2009) proposed another method to estimate the parameters of the modified Champernowne distribution, named as the conditional maximum likelihood (CML) method. The new estimator can improve the performance in the tail, but unfortunately it needs to choose a threshold which is crucial for the estimator.

Denote by $(\hat{\alpha}, \hat{M}, \hat{c})$ the estimated parameters as described in Buch-Larsen et al. (2005), and \hat{f}_{Cha} and \hat{F}_{Cha} the resulting modified Champernowne density and distribution functions, using the estimated parameters. Transform the data x_1, \dots, x_n by the distribution function, i.e., $y_i = \hat{F}_{\text{Cha}}(x_i)$. Then apply the classical kernel density estimator (Silverman, 1986; Wand and Jones, 1994) on the transformed data y_1, \dots, y_n ,

$$\hat{f}_t(y) = \frac{1}{nk_y} \sum_{i=1}^n K_h(y - y_i),$$

where $K_h(\cdot) = (1/h)K(\cdot)$ and $K(\cdot)$ is the kernel function. The boundary correction is defined as

$$k_y = \int_{\max(-1, -y/h)}^{\min(1, (1-y)/h)} K(u) du.$$

The estimator of the density for the original data x_1, \dots, x_n is

$$\hat{f}(x) = \frac{\hat{f}_t(\hat{F}_{\text{Cha}}(x))}{|(\hat{F}_{\text{Cha}}^{-1})'(\hat{F}_{\text{Cha}}(x))|} = \frac{1}{nk_{\hat{F}_{\text{Cha}}(x)}} \sum_{i=1}^n K_h\left(\hat{F}_{\text{Cha}}(x) - \hat{F}_{\text{Cha}}(x_i)\right) \hat{f}_{\text{Cha}}(x). \quad (7.2)$$

7.3 Heavy Tails Analysis

Setup

In the simulation study below, we compare the performance of the KMCE, LS and LCCC estimators. For the KMCE estimator, we follow the procedure suggested by Buch-Larsen et al. (2005) to estimating the parameters (α, M, c) . That is, we use the empirical median of the sample to estimate the parameter M and maximize the log-likelihood function, with fixed M , with respect to the other two parameters α and c . Then the standard kernel method, via the R function `density()` with the default bandwidth, is applied to the transformed data. Finally, the density for the original data is

obtained by back transformation, as give in (7.2). For the LS and LCCC estimators, the settings are the same as given in Section 6.6.

The comparison is based on data simulated from two distributions. One is the log normal (LN) with parameters $(\mu, \sigma) = (0, 1)$, and the other the Weibull distribution (WB), with density given by

$$f(x; \lambda, k) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}, \quad x \geq 0,$$

where $k > 0$ is the shape parameter and $\lambda > 0$ the scale parameter. We choose $(k, \lambda) = (0.5, 1)$ here.

As by Buch-Kromann (2009), we use the weighted integrated standard error (WISE) to measure the performance of an estimator. The WISE weights the squared distance between the estimated and true distributions with the value of x^δ . This error measure emphasizes the tail of a distribution, which is very relevant when dealing with economical and financial data. The WISE is given by

$$\text{WISE}_q^\delta(f; \hat{f}) = \int_{x_q}^{\infty} \{f(x) - \hat{f}(x)\}^2 x^\delta dx,$$

where $f(x)$ is the true density and $\hat{f}(x)$ a density estimator. The lower limit x_q is the q -quantile of $f(x)$. The value of δ decides the weight in the error measure: the larger the value of δ , the more weight is put into the tail deviation between the true and the estimated densities. Note that the ISE used in Section 6.6 is the WISE when $\delta = 1$ and $q = 0$. In our study, we choose $(q, \delta) \in \{95\%, 99\%\} \times \{0, 1, 2\}$.

Result

For each distribution, 100 random samples were generated, each of size 100, and the estimation results are reported in Table 7.1. Each entry in the table is an empirical mean WISE (MWISE) with its standard error given in parentheses. The smallest MWISE of these estimators in each study case is highlighted in boldface. Boxplots are shown in Figure 7.1 for LN (0, 1) and Figure 7.2 for WB (0.5, 1). The tail plots (above 95%) are displayed in Figure 7.3, for a typical sample of size 100.

Generally speaking, the KMCE estimator was the worst for tail estimation. It tends over-estimate the heaviness of a tail, and its MWISE relative to those of the other two increases as δ increases. For the Weibull distribution, the LS and LCCC estimators performed similarly and both provide good fits to the heavy tail. For the log normal distribution, however, the former had an inferior performance to the latter.

The R implementation of the LS estimator may sometimes give an error, due to the presence of very similar data values that causes a numerically singular Hessian matrix. If this happens, we

discard the current sample and re-generalize another one to make the results be fairly comparable. By contrast, the LCCC estimator can always be computed without a problem.

Tab. 7.1: Simulation results for two distributions in terms of the MWISE, with standard errors given in parentheses.

Estimator	Density			
	LN(0, 1)		WB(0.5, 1)	
	95%	99%	95%	99%
MWISE $\times 10^{-5}$ ($\delta = 0$)				
KMCE	6.39 (0.73)	1.14 (0.11)	4.06 (0.35)	1.84 (0.09)
LS	6.92 (0.88)	0.55 (0.06)	3.38 (0.52)	0.20 (0.02)
LCCC	4.04 (0.37)	0.37 (0.04)	3.33 (0.37)	0.22 (0.03)
MWISE $\times 10^{-4}$ ($\delta = 1$)				
KMCE	5.53 (0.58)	1.69 (0.17)	10.77 (0.66)	7.68 (0.35)
LS	4.95 (0.59)	0.70 (0.07)	4.41 (0.58)	0.56 (0.06)
LCCC	2.89 (0.25)	0.46 (0.05)	4.27 (0.47)	0.55 (0.09)
MWISE $\times 10^{-3}$ ($\delta = 2$)				
KMCE	5.42 (0.56)	2.90 (0.28)	48.04 (2.23)	43.48 (1.85)
LS	3.86 (0.41)	0.93 (0.08)	6.72 (0.72)	1.78 (0.19)
LCCC	2.31 (0.19)	0.66 (0.07)	6.43 (0.70)	1.75 (0.27)

7.4 VaR estimation

Setup

In this section, we apply the LCCC and LS estimators to VaR estimation, since both are generally good at tail estimation. We estimate the VaR for the three financial datasets used in Chapter 6, at the levels $\alpha = 0.95, 0.99$ as follows:

- (1) Set the loss $L = -X$.
- (2) Estimate the distribution of L by the LS and LCCC estimators.
- (3) Estimate $\text{VaR}_{0.95}$ and $\text{VaR}_{0.99}$ from each model, i.e., respectively, the 95- and 99-percentiles of the loss distribution.

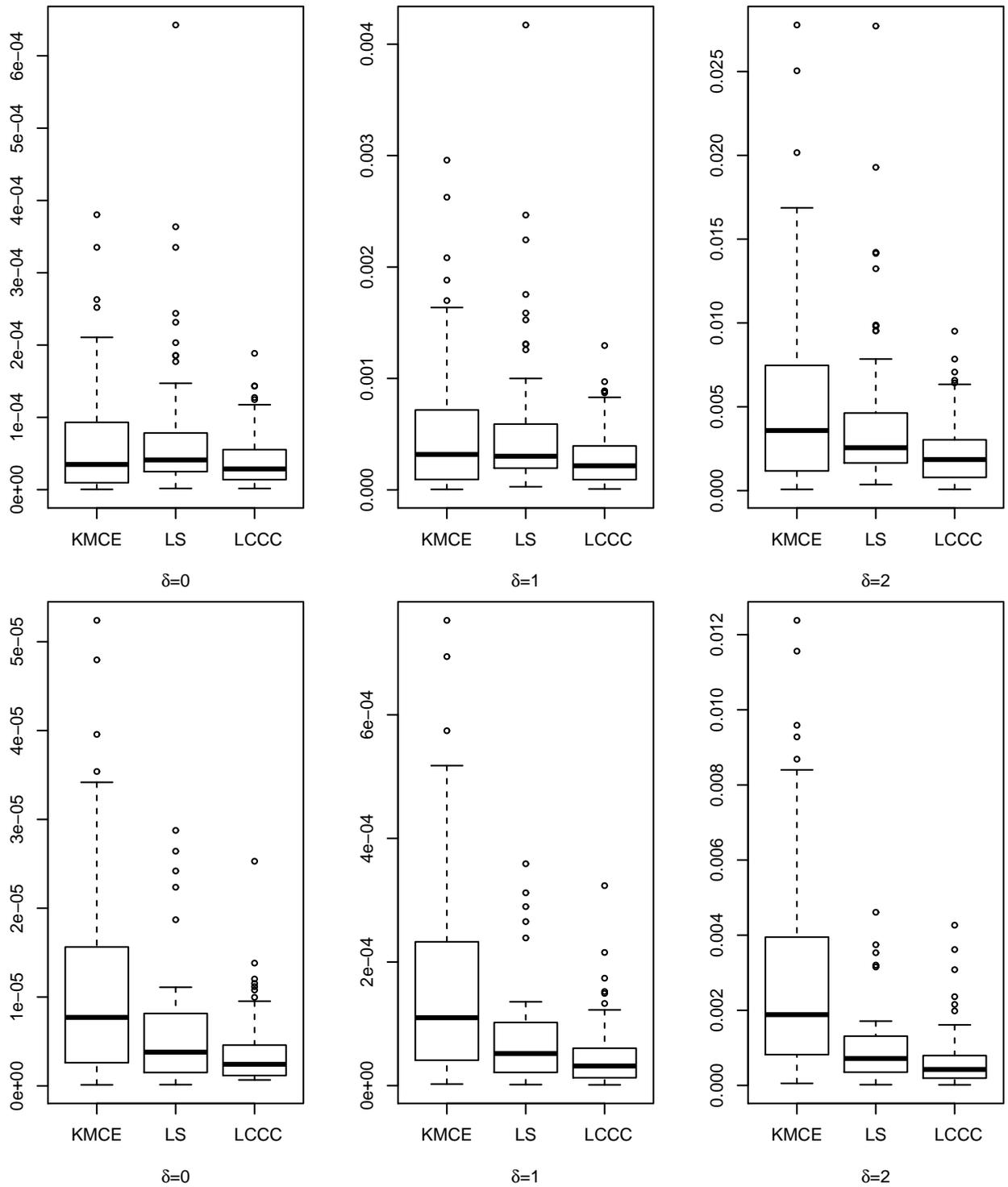


Fig. 7.1: Box plots of the WISE over 100 replications for 95% (upper panels) and 99% (lower panels) tail performances from LN (0, 1).

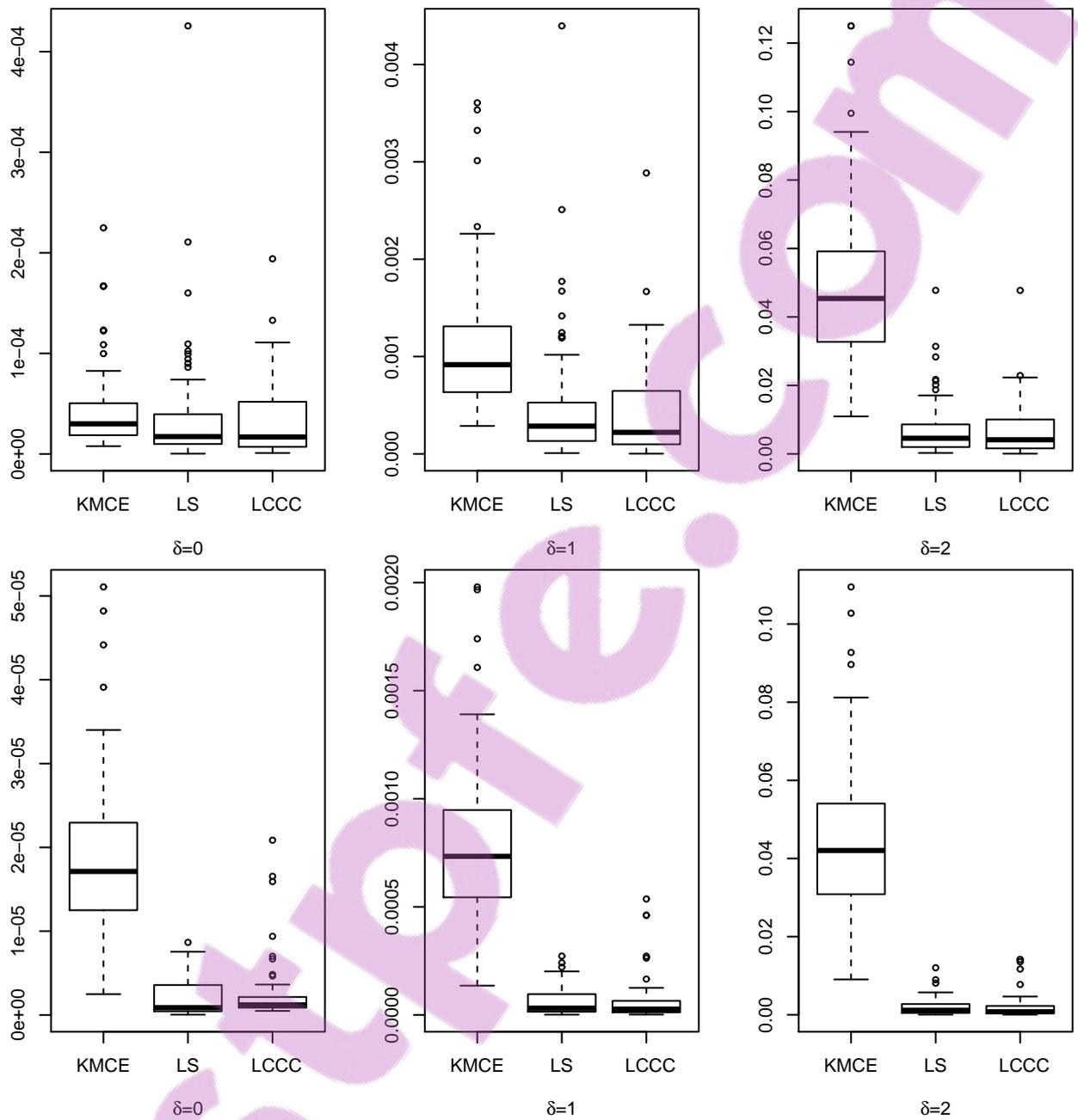


Fig. 7.2: Box plots of the WISE over 100 replications for 95% (upper panels) and 99% (lower panels) tail performances from WB (0.5, 1).

Kupiec Test

Kupiec (1995) developed the proportion of failure (POF) test that can measure the accuracy of a VaR estimation. In the POF test, the number of violations (violations occur when the actual loss exceeds the estimate) from the empirical data are compared with the accepted number of

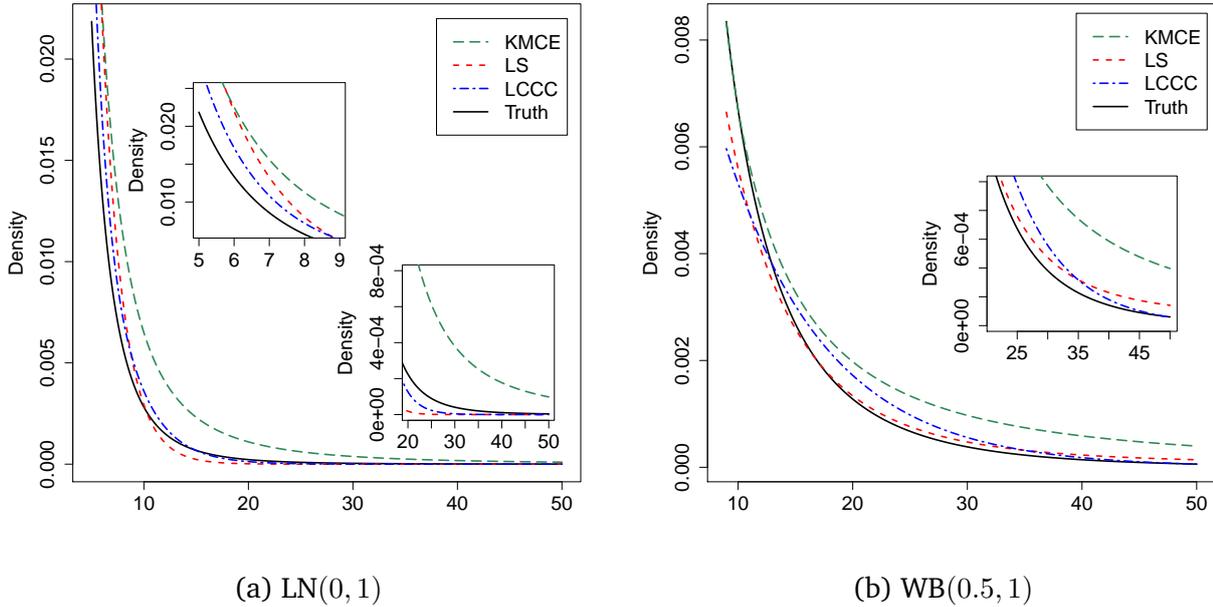


Fig. 7.3: Tail plots among different estimates from different scenarios with sample size 100: LN(0, 1) (left) and Weibull (1.5) (right)

exceedance at a given confidence level α . Denote by n the total number of observations and n_e the number of exceptions provided by the estimate. Under null hypothesis of the model being ‘correct’, the number of exceptions n_e follows a binomial distribution. The null hypothesis for the POF test is

$$H_0 : p = 1 - \alpha.$$

According to Kupiec (1995), the POF test is best conducted as a standard likelihood-ratio (LR) test, and the test statistic is

$$\chi^2 = -2 \log \left(\left(\frac{1-p}{1-\hat{p}} \right)^{n-n_e} \left(\frac{p}{\hat{p}} \right)^{n_e} \right),$$

where $\hat{p} = \frac{n_e}{n}$. Under the null hypothesis that the model is correct, the quantity is asymptotically chi-square distributed with one degree of freedom. If χ^2 is greater than a critical value, the null hypothesis is rejected and the model is deemed as inaccurate for VaR estimation.

Results

Figure 7.4 displays the VaR estimates along with the loss histogram, for each dataset. It shows that the estimates found by the LS and LCCC estimators are both close to those of the empirical distribution. If the histogram is used as a benchmark, then both estimators provide good fits at the tail and differ only slightly. Table 7.2 gives a numerical summary of the estimated VaR’s for each

7.5. Summary

dataset. It shows that for each dataset, there is no evidence to reject any of the estimated VaR's at the 5% significance level according to the Kupiec test.

Tab. 7.2: VaR ($\times 10^{-2}$) estimation and Kupiec's POF test p -values for the financial datasets.

Estimator		MSFT		SPX		NZD/USD	
		95%	99%	95%	99%	95%	99%
LS	VaR	2.379	5.201	1.962	3.712	1.323	2.241
	p -value	0.274	0.717	0.914	0.537	0.138	0.585
LCCC	VaR	2.471	4.983	1.969	3.651	1.291	2.299
	p -value	0.723	0.717	0.973	0.537	0.718	0.988

7.5 Summary

In this chapter, we investigated the heavy tail performance of a few density estimators, in particular the LCCC, LS and KMCE ones. The estimation of VaR has been studied and the Kupiec test has been used.

In the heavy tail performance analysis, two distributions have been considered in the simulation study. The results showed that both the LCCC and LS estimators can generally provide better fits to a heavy tail than the KMCE one. However, the LS estimator has the disadvantage of forcing two tails to be exponential and performs worse than the LCCC estimator. We further apply the LS and LCCC estimators to VaR estimation. The Kupiec test is introduced to measure the accuracy of VaR estimation. The test suggests that there is no evidence to reject the VaR estimates provided by the LS and LCCC estimators.

There are many methods for estimating heavy tails or a VaR. Future study will consider more of these methods, in particular those developed from the extreme value theory.

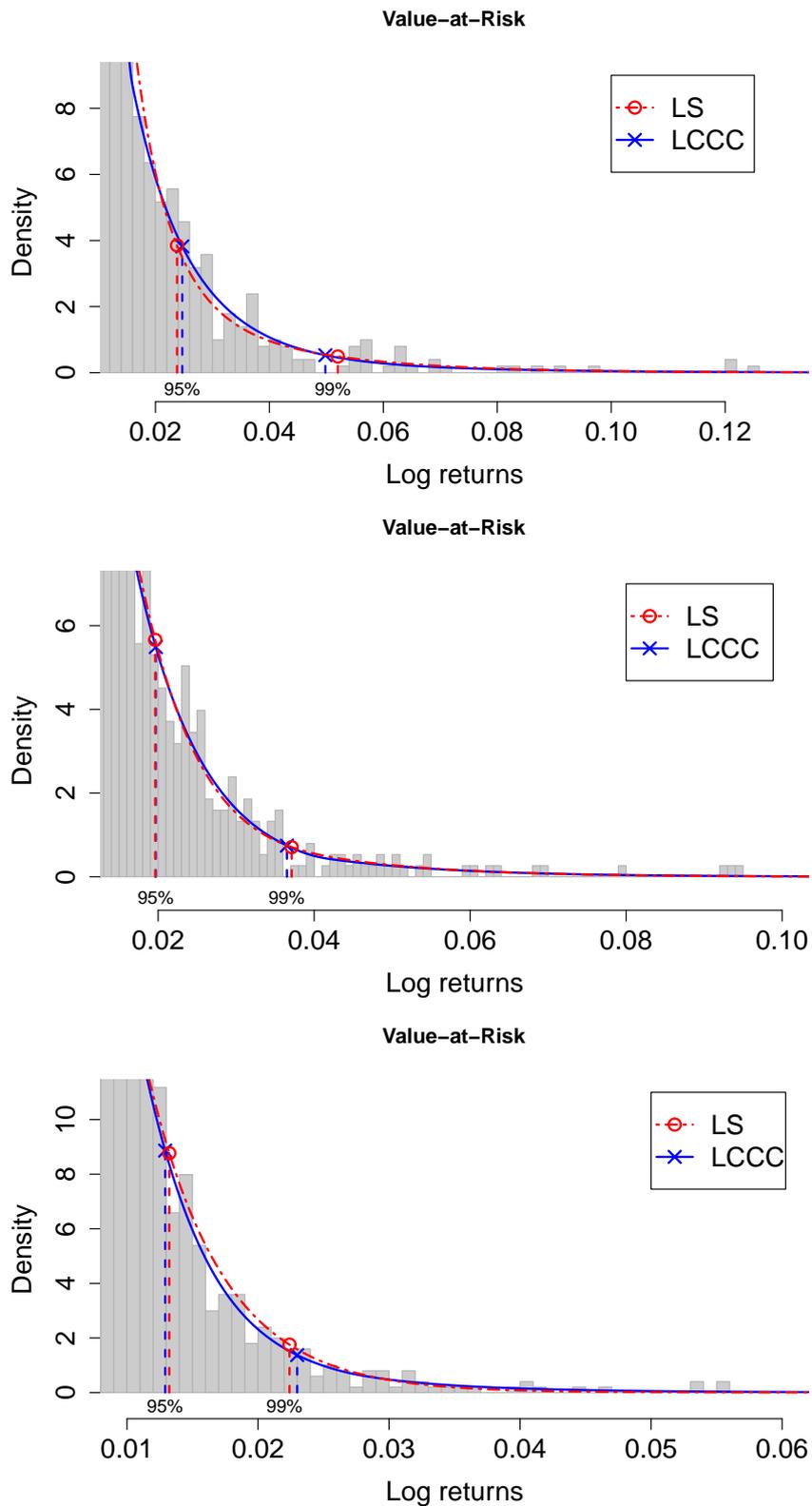


Fig. 7.4: VaR estimation for financial datasets using the LS and LCCC estimators: MSFT (top panel), SPX (middle panel) and NZUS (bottom panel). The circle and cross points are the estimated VaR's from the LS and LCCC estimators, respectively.

Chapter 8

Summary and Future Works

8.1 Summary

This research was motivated by fitting an asset return distribution in finance, which is known to exhibit certain shapes, such as skewness, high peak, heavy tails and unimodality. One of our goals is to let the data speak for themselves as much as possible. This is achieved by using nonparametric methods which make less rigid assumptions about the underlying distribution, as opposed to parametric ones. To avoid issues such as bandwidth or tuning parameter selection in traditional nonparametric approaches, it may be reasonable to make a good use of the prior knowledge for estimating a function by imposing natural qualitative constraints on it. Our main research problem is to find the nonparametric estimation of a density function under shape restrictions. Our methodologies are allocated to the nonparametric maximum likelihood approach. Unimodality is a commonly-used assumption in density estimation, also in financial applications. Unfortunately, the NPMLE of a unimodal density function with mode unknown does not exist. Log-concavity which belongs to the subset of the class of unimodality maybe a good choice for starter, because it has many attractive properties and, most of all the NPMLE of a log-concave density function exists and is unique. Both nonsmooth and smooth NPMLE of a log-concave density function have been proposed. Moreover, in order to estimate the highly heavy-tailed distributions, while preserving unimodality, we impose both heavy-tailedness and unimodality constraints on the underlying density function. As for the computational aspect, the NPMLE of a function has no explicit solution, hence the iterative algorithms have to be developed. Several algorithms are proposed for the proposed estimators.

We reviewed the nonparametric estimation of a density function with or without shape restrictions in Chapter 2. The log-concavity and unimodality with heavy tails are described in more details. Furthermore, algorithms for nonparametric density estimation have also been studied.

In Chapter 3, log-concave density estimation has been studied in great detail. Some properties of the nonparametric maximum likelihood estimation of a log-concave density function have been provided and a new fast algorithm based on CNM (CNMLCD) has been proposed which can efficiently find the NPMLE of a log-concave density function and is superior to other existing algorithms in the literature, including the log barrier, ICMA, ASA and LC algorithms. The convergence of the new algorithm has also been theoretically established.

An extension has been given in Chapter 4, where smoothness has been included. In particular, smooth estimation of a unimodal density function that is also log-concave has been investigated in details. Several new smooth estimators have been proposed and new algorithms for their computation are presented, which extend the CNM method. A combined estimator has also been proposed for its flexibility and convenience from an end user's point of view. Numerical studies that compare the performance of different estimators, including the non-smooth estimator (PL) and smooth estimator (LCDS and PQC), have been done using both simulated and real-world data. The results show that the smooth estimators are generally better than the non-smooth estimator, and our new smooth estimator (PQC) is always better than the LCDS estimator, in terms of the mean integrated squared error, the Kullback-Leibler divergence and the Hellinger risk. In addition, a new test of log-concavity has been developed based on the CNMLCD.

In Chapter 5, we investigated the performance of different methods for the ROC curve estimation. Nonparametric methods based on the log-concave constituent distributions have been studied in great detail. Three ROC curve estimators, the PL-, LCDS- and PQC-based, have been compared using simulated and real-world data. Numerical studies show that the ROC curve estimators based on the smooth estimators are better than the one based on the piecewise linear estimator, and the PQC-based estimator is slightly better than the LCDS-based one.

Heavy-tailed distributions with unimodality have been studied in Chapter 6. Both parametric and nonparametric methods have been investigated. We further proposed a smooth hybrid nonparametric maximum likelihood estimator for this kind of distribution by mixing log-concavity and log-convexity (LCCC). A novel idea is to separately impose the log-concavity shape restriction on the main body and the log-convexity on the tails. The corresponding algorithm has also been developed. Several existing nonparametric methods including the logspline estimator (LS), the unimodal density estimation using regression spline (UMRS), the unimodal density estimation using Bernstein polynomial (UMBP) and the standard kernel (KER) have been included in the numerical

studies for full density estimation. A parametric method, using a skewed- t distribution (ST), has also been included in the real-world study. The results from full density estimation show that the KER introduces spurious features in the tails, the LS estimator is not necessarily unimodal and is forced to have two exponential tails, the UMRS and UMBP have difficulty with estimating heavy tails, the ST can not estimate well the mean and skewness. By contrast, the new LCCC estimator has good fittings in the both center and tails of the distributions. Bootstrap testing for identifying the functional form implied by the LCCC estimator has been developed.

We gave a further investigation on heavy tail estimation in Chapter 7. The kernel density estimator based on the Champernowne distribution (KMCE) has been included in this study. The performance of the estimators, the LS, LCCC and KMCE, have been compared using simulated data. Because of the outstanding performance in the tail domain, we further apply the LS and LCCC estimators to VaR estimation for risk management and utilize the Kupiec test to measure the accuracy of the estimated VaR.

8.2 Future Works

Our main interest in this thesis is to investigate the nonparametric density estimation under shape restrictions. Of course, there are still many issues that can be further investigated in this direction. A number of topics for further studies are listed as follows.

Theoretical Justifications

In this thesis, we have established some theoretical properties of our nonparametric maximum likelihood estimator. Further theoretical results that concern, for example, consistency and efficiency of the estimators can be further developed.

Other Types of Data

In this thesis, we concentrate on exact observations, it should be fairly straightforward to extend the research to other types of data like censored and doubly truncated data. Logspline density estimation for censored data can be found in [Kooperberg and Stone \(1992\)](#). [Dümbgen et al. \(2006\)](#) studied three nonparametric maximum likelihood estimators based on mixed-case interval-censored data, including unrestricted and restricted estimators. One can also see the log-concave density estimation based on the interval-censored in [Dümbgen et al. \(2011\)](#) and [Anderson-Bergman \(2014\)](#). Moreover, some recent works for doubly truncated data have been done by [Shen \(2010\)](#) and [Shen \(2016\)](#).

Multivariate Situation

Our present work has been focused on univariate data analysis. The literature on the multivariate situation is relatively scarce. Computation is also a problem in the multivariate case. The multivariate log-concave density estimation has been investigated by [Cule et al. \(2010\)](#) and [Chen and Samworth \(2013\)](#). The dependence structures for multivariate highly frequency data have been analysed by [Breyman et al. \(2003\)](#) in a non-dynamic setting and by [Dias and Embrechts \(2004\)](#) in a dynamic setting. An expectation-maximization algorithm for maximum likelihood estimation of heavy-tailed multivariate observation can be found in [Øigård et al. \(2005\)](#). In finance, multivariate analysis has a lot of applications; see [Virbickaite et al. \(2015\)](#) and references therein. On one hand, our shape-restricted maximum likelihood estimation method may be extended to the multivariate situation. On the other hand, our new fast algorithm may also be extendable to solve the computational problem.

Other Shape Restrictions

We focus on the density estimation under log-concavity or the combination of log-concavity and log-convexity shape restrictions. It is also worthwhile and relatively easy to apply the presented methods to some other shape restrictions, such as monotonicity and convexity, in both density estimation and regression; see [An \(1998\)](#) and [Yatchew and Härdle \(2006\)](#). Shapes can also be found in utility function in economics, dose-response relationships in medicine, hazard rate in reliability and survival analysis. Moreover, our idea for combining the log-concavity and log-convexity can also be extended to other potential shapes combinations.

Semiparametric Models

In our study, we estimate an asset return distribution without taking into the effect of time, as the behaviour of an asset return in finance is not stable over time. The volatility process such as GARCH and ARCH processes, can be considered into the nonparametric estimator to produce a semiparametric model ([Chen, 2014](#)). Moreover, as we mentioned in [Chapter 2](#), extreme value theory is a powerful and yet fairly robust framework for modelling the tail behaviour. It would be interesting to extend our methods to this area.

Bibliography

- Ait-Sahalia, Y. and J. Duarte (2003). Nonparametric option pricing under shape restrictions. *Journal of Econometrics*, 116, 9–47.
- Alonzo, T. A. and M. S. Pepe (2002). Distribution-free ROC analysis using binary regression techniques. *Biostatistics*, 3, 421–432.
- An, M. Y. (1996). *Log-concave probability distributions: Theory and statistical testing*. Working paper. Centre for Labour Economics, University of Åarhus and Åarhus School of Business.
- An, M. Y. (1998). Logconcavity versus logconvexity: A complete characterization. *Journal of Economic Theory*, 80, 350–369.
- Anderson, T. W. and D. A. Darling (1954). A test of goodness of fit. *Journal of the American Statistical Association*, 49, 765–769.
- Anderson-Bergman, C. (2014). *Semi- and Non-parametric Methods for Interval Censored Data with Shape Constraints*. Ph. D. thesis, University of California, Irvine.
- Ayer, M., H. D. Brunk, G. M. Ewing, W. Reid, E. Silverman, et al. (1955). An empirical distribution function for sampling with incomplete information. *Annals of Mathematical Statistics*, 26, 641–647.
- Bagnoli, M. and T. Bergstrom (2005). Log-concave probability and its applications. *Economic Theory*, 26, 445–469.
- Bai, J. and S. Ng (2005). Tests for skewness, kurtosis, and normality for time series data. *Journal of Business & Economic Statistics*, 23, 49–60.

- Balabdaoui, F., H. Jankowski, M. Pavlides, A. Seregin, and J. Wellner (2011). On the Grenander estimator at zero. *Statistica Sinica*, 21, 873.
- Balabdaoui, F., K. Rufibach, and J. A. Wellner (2009). Limit distribution theory for maximum likelihood estimation of a log-concave density. *Annals of Statistics*, 37, 1299–1331.
- Bao, Y. (2013). On sample skewness and kurtosis. *Econometric Reviews*, 32, 415–448.
- Barlow, R. (1972). *Statistical Inference under Order Restrictions: The Theory and Application of Isotonic Regression*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons.
- Barlow, R. and F. Proschan (1975). *Statistical Theory of Reliability and Life Testing: Probability Models*. International Series in Decision Processes. Holt, Rinehart and Winston.
- Barron, A. R., L. Györfi, and E. C. van der Meulen (1992). Distribution estimation consistent in total variation and in two types of information divergence. *IEEE Transactions on Information Theory*, 38, 1437–1454.
- Bartoszyński, R., B. W. Brown, C. M. McBride, and J. R. Thompson (1981). Some nonparametric techniques for estimating the intensity function of a cancer related nonstationary poisson process. *Annals of Statistics*, 9, 1050–1060.
- Bast Jr, R. C., T. L. Klug, E. S. John, E. Jenison, J. M. Niloff, H. Lazarus, R. S. Berkowitz, T. Leavitt, C. T. Griffiths, and L. Parker (1983). A radioimmunoassay using a monoclonal antibody to monitor the course of epithelial ovarian cancer. *New England Journal of Medicine*, 309, 883–887.
- Bickel, P. J. and J. Fan (1996). Some problems on the estimation of unimodal densities. *Statistica Sinica*, 6, 23–45.
- Birgé, L. (1997). Estimation of unimodal densities without smoothness assumptions. *Annals of Statistics*, 25, 970–981.
- Birke, M. (2009). Shape constrained kernel density estimation. *Journal of Statistical Planning and Inference* 139(8), 2851–2862.
- Birke, M. and H. Dette (2007). Estimating a convex function in nonparametric regression. *Scandinavian Journal of Statistics*, 34, 384–404.
- Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81, 637–654.

- Bobkov, S., M. Madiman, et al. (2011). Concentration of the information in data with log-concave distributions. *Annals of Probability*, 39, 1528–1543.
- Böhning, D. (1985). Numerical estimation of a probability measure. *Journal of Statistical Planning and Inference*, 11, 57–69.
- Bolancé, C., M. Guillén, and J. P. Nielsen (2003). Kernel density estimation of actuarial loss functions. *Insurance: Mathematics and Economics*, 32, 19–36.
- Bolancé, C., M. Guillén, and J. P. Nielsen (2008). Inverse beta transformation in kernel density estimation. *Statistics & Probability Letters*, 78, 1757–1764.
- Bradley, B. O. and M. S. Taqqu (2003). Financial risk and heavy tails. In S. Rachev (Ed.), *Handbook of Heavy-tailed Distributions in Finance*, pp. 35–103. Elsevier, Amsterdam.
- Braun, W. J. and P. Hall (2001). Data sharpening for nonparametric inference subject to constraints. *Journal of Computational and Graphical Statistics*, 10, 786–806.
- Breymann, W., A. Dias, and P. Embrechts (2003). Dependence structures for multivariate high-frequency data in finance. *Quantitative Finance*, 3, 1–14.
- Brooks, S. P. et al. (1998). MCMC convergence diagnosis via multivariate bounds on log-concave densities. *Annals of Statistics*, 26, 398–433.
- Bruffaerts, C., V. Verardi, and C. Vermandele (2014). A generalized boxplot for skewed and heavy-tailed distributions. *Statistics & Probability Letters*, 95, 110–117.
- Buch-Kromann, T. (2009). Comparison of tail performance of the Champernowne transformed kernel density estimator, the generalized Pareto distribution and the g-and-h distribution. *Journal of Operational Risk*, 4, 43–67.
- Buch-Larsen, T., J. P. Nielsen, M. Guillén, and C. Bolancé (2005). Kernel density estimation for heavy-tailed distributions using the Champernowne transformation. *Statistics*, 39, 503–516.
- Buishand, T., L. de Haan, and C. Zhou (2008). On spatial extremes: With application to a rainfall problem. *Annals of Applied Statistics*, 2, 624–642.
- Cai, T. and C. S. Moskowitz (2004). Semi-parametric estimation of the binormal ROC curve for a continuous diagnostic test. *Biostatistics*, 5, 573–586.
- Cao, R., A. Cuevas, and W. G. Manteiga (1994). A comparative study of several smoothing methods in density estimation. *Computational Statistics & Data Analysis*, 17, 153–176.

- Caplin, A. and B. Nalebuff (1991). Aggregation and imperfect competition: On the existence of equilibrium. *Econometrica: Journal of the Econometric Society*, 59, 25–59.
- Champagnat, N., M. Deaconu, A. Lejay, N. Navet, and S. Boukherouaa (2013). An empirical analysis of heavy-tails behavior of financial data: The case for power laws. Technical Report (No. hal-00851429, version 1), Hyper Articles en Ligne.
- Champernowne, D. G. (1953). A model of income distribution. *Economic Journal*, 63, 318–351.
- Charpentier, A. and E. Flachaire (2014). Log-transform kernel density estimation of income distribution. *SSRN Electronic Journal* (doi: 10.2139/ssrn.2514882).
- Chen, S. X. and C. Y. Tang (2005). Nonparametric inference of value-at-risk for dependent financial returns. *Journal of Financial Econometrics*, 3, 227–255.
- Chen, Y. (2013). *Aspects of Shape-constrained Estimation in Statistics*. Ph. D. thesis, University of Cambridge.
- Chen, Y. (2014). Semiparametric time series models with log-concave innovations: Maximum likelihood estimation and its consistency. *Scandinavian Journal of Statistics*, 42, 1–31.
- Chen, Y. and R. J. Samworth (2013). Smoothed log-concave maximum likelihood estimation with applications. *Statistica Sinica*, 23, 1373–1398.
- Cheng, M.-Y., T. Gasser, and P. Hall (1999). Nonparametric density estimation under unimodality and monotonicity constraints. *Journal of Computational and Graphical Statistics*, 8, 1–21.
- Chicago Board Options Exchange (2009). *The CBOE volatility index–VIX*. White Paper.
- Choi, E. and P. Hall (1999). Data sharpening as a prelude to density estimation. *Biometrika*, 86, 941–947.
- Choi, E., P. Hall, and V. Rousson (2000). Data sharpening methods for bias reduction in nonparametric regression. *Annals of Statistics*, 28, 1339–1355.
- Cizeau, P., Y. Liu, M. C. Meyer, C.-K. Peng, and H. Eugene Stanley (1997). Volatility distribution in the S&P500 stock index. *Physica A: Statistical Mechanics and its Applications*, 245, 441–445.
- Clements, A., S. Hurn, and K. Lindsay (2003). Mobius-like mappings and their use in kernel density estimation. *Journal of the American Statistical Association*, 98, 993–1000.
- Collinson, P. (1998). Of bombers, radiologists, and cardiologists: Time to ROC. *Heart*, 80, 215–217.

- Cule, M., R. Samworth, et al. (2010). Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density. *Electronic Journal of Statistics*, 4, 254–270.
- Cule, M., R. Samworth, and M. Stewart (2010). Maximum likelihood estimation of a multidimensional log-concave density. *Journal of the Royal Statistical Society, Series B*, 72, 545–607.
- Curto, J., E. Reis, and J. Esperança (2003). Stable Paretian distributions: An unconditional model for PSI20, DAX and DJIA indexes. *Review of Financial Markets*, 5, 5–17.
- Del Villano, B., S. Brennan, P. Brock, C. Bucher, V. Liu, M. McClure, B. Rake, S. Space, B. Westrick, and H. Schoemaker (1983). Radioimmunoassay for a monoclonal antibody-defined tumor marker, CA 19-9. *Clinical Chemistry*, 29, 549–552.
- Dias, A. and P. Embrechts (2004). Dynamic copula models for multivariate high-frequency data in finance. *Manuscript, ETH Zurich*.
- Dorfman, D. D., K. S. Berbaum, C. E. Metz, R. V. Lenth, J. A. Hanley, and H. A. Dagga (1997). Proper receiver operating characteristic analysis: The bigamma model. *Academic Radiology*, 4, 138–149.
- Dümbgen, L., S. Freitag-Wolf, and G. Jongbloed (2006). Estimating a unimodal distribution from interval-censored data. *Journal of the American Statistical Association*, 101, 1094–1106.
- Dümbgen, L., A. Hüsler, and K. Rufibach (2007). *Active set and EM algorithms for log-concave densities based on complete and censored data*. Technischer Bericht. Institut für mathematische Statistik und Versicherung.
- Dümbgen, L. and K. Rufibach (2009). Maximum likelihood estimation of a log-concave density and its distribution function: Basic properties and uniform consistency. *Bernoulli*, 15, 40–68.
- Dümbgen, L. and K. Rufibach (2011). logcondens: Computations related to univariate log-concave density estimation. *Journal of Statistical Software*, 39, 1–28.
- Dümbgen, L., R. Samworth, D. Schuhmacher, et al. (2011). Approximation by log-concave distributions, with applications to regression. *Annals of Statistics*, 39, 702–730.
- Dutta, K. K., D. F. Babbal, et al. (2002). On measuring skewness and kurtosis in short rate distributions: The case of the US dollar london inter bank offer rates. *Center for Financial Institutions Working Papers 02*, 25.
- Eberlein, E. and U. Keller (1995). Hyperbolic distributions in finance. *Bernoulli*, 281–299.

- Eddy, W. F. (1980). Optimum kernel estimators of the mode. *Annals of Statistics*, 8, 870–882.
- Eggermont, P. and V. LaRiccia (2000). Maximum likelihood estimation of smooth monotone and unimodal densities. *Annals of Statistics*, 28, 922–947.
- Ehlers, R. S. (2015). A study of skewed heavy-tailed distributions as scale mixtures. *American Journal of Mathematical and Management Sciences*, 34, 40–66.
- Embrechts, P., C. Klüppelberg, and T. Mikosch (1997). *Modelling Extremal Events*. Springer Science & Business Media.
- Fama, E. F. (1965). The behavior of stock-market prices. *Journal of Business*, 38, 34–105.
- Fedorov, V. (1972). *Theory of Optimal Experiments*. Series of Probability and Mathematical Statistics. Elsevier Science.
- Fernández, C. and M. F. Steel (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93, 359–371.
- Fletcher, R. (1987). *Practical Methods of Optimization* (2nd ed.). John Wiley & Sons.
- Fougères, A.-L. (1997). Estimation de densités unimodales. *Canadian Journal of Statistics*, 25, 375–387.
- Frencha, C. W. (2003). The Treynor capital asset pricing model. *Journal of Investment Management*, 1, 60–72.
- Fryer, M. (1977). A review of some non-parametric methods of density estimation. *IMA Journal of Applied Mathematics*, 20, 335–354.
- Gencay, R. and F. Selcuk (2006). Overnight borrowing, interest rates and extreme value theory. *European Economic Review*, 50, 547–563.
- Gilks, W. R. and P. Wild (1992). Adaptive rejection sampling for Gibbs sampling. *Journal of the Royal Statistical Society, Series C*, 41, 337–348.
- Gonçalves, L., A. Subtil, M. R. Oliveira, and P. de Zea Bermudez (2014). ROC curve estimation: An overview. *Revstat–Statistical Journal*, 12, 1–20.
- Grenander, U. (1956). On the theory of mortality measurement: Part ii. *Scandinavian Actuarial Journal*, 1956, 125–153.

- Groeneboom, P. and G. Jongbloed (2014). *Nonparametric estimation under shape constraints*, Volume 38. Cambridge University Press.
- Groeneboom, P., G. Jongbloed, and J. A. Wellner (2001). Estimation of a convex function: Characterizations and asymptotic theory. *Annals of Statistics*, 29, 1653–1698.
- Groeneboom, P., G. Jongbloed, and J. A. Wellner (2008). The support reduction algorithm for computing nonparametric function estimates in mixture models. *Scandinavian Journal of Statistics*, 35, 385–399.
- Groeneboom, P. and J. A. Wellner (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*, Volume 19 of *DMV Seminar*. Springer.
- Hall, P. and L.-S. Huang (2002). Unimodal density estimation using kernel methods. *Statistica Sinica*, 12, 965–990.
- Hall, P. and K.-H. Kang (2005). Unimodal kernel density estimation by data sharpening. *Statistica Sinica*, 15, 73–98.
- Hall, P. and B. Presnell (1999). Intentionally biased bootstrap methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61(1), 143–158.
- Hazelton, M. L. (2011). Assessing log-concavity of multivariate densities. *Statistics & Probability Letters*, 81, 121–125.
- Hildreth, C. (1954). Point estimates of ordinates of concave functions. *Journal of the American Statistical Association*, 49, 598–619.
- Hill, B. M. et al. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3, 1163–1174.
- Holan, S. H. and T. S. McElroy (2010). Tail exponent estimation via broadband log density-quantile regression. *Journal of Statistical Planning and Inference*, 140, 3693–3708.
- Horowitz, J. and S. S. Lee (2015). Nonparametric estimation and inference under shape restrictions. Technical report, Centre for Microdata Methods and Practice, Institute for Fiscal Studies, Britain.
- Hsieh, F., B. W. Turnbull, et al. (1996). Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *Annals of Statistics*, 24, 25–40.

- Hughes, G. and B. Bhattacharya (2013). Symmetry properties of bi-normal and bi-gamma receiver operating characteristic curves are described by kullback-leibler divergences. *Entropy*, 15, 1342–1356.
- Ibragimov, I. A. (1956). On the composition of unimodal distributions. *Theory Probability Application*, 1, 255–260.
- Jokiel-Rokita, A. and M. Pulit (2013). Nonparametric estimation of the ROC curve based on smoothed empirical distribution functions. *Statistics and Computing*, 23, 703–712.
- Jones, M. C., J. S. Marron, and S. J. Sheather (1996). A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association*, 91, 401–407.
- Jongbloed, G. (1998). The iterative convex minorant algorithm for nonparametric estimation. *Journal of Computational and Graphical Statistics*, 7, 310–321.
- Jorion, P. (2007). *Value at Risk: The New Benchmark for Managing Financial Risk*. McGraw-Hill New York.
- Karlin, S. (1968). *Total Positivity*. Stanford University Press.
- Kim, Y. S., S. T. Rachev, M. L. Bianchi, and F. J. Fabozzi (2010). Tempered stable and tempered infinitely divisible garch models. *Journal of Banking & Finance*, 34, 2096–2109.
- Klugman, S. A., H. H. Panjer, and G. E. Willmot (2012). *Loss Models: From Data to Decisions*. John Wiley & Sons.
- Koenker, R. and I. Mizera (2010). Quasi-concave density estimation. *Annals of Statistics*, 38, 2998–3027.
- Koo, J.-Y. and C. Kooperberg (2000). Logspline density estimation for binned data. *Statistics & Probability Letters*, 46, 133–147.
- Koo, J.-Y., C. Kooperberg, and J. Park (1999). Logspline density estimation under censoring and truncation. *Scandinavian Journal of Statistics*, 26, 87–105.
- Kooperberg, C. and C. J. Stone (1991). A study of logspline density estimation. *Computational Statistics & Data Analysis*, 12, 327–347.
- Kooperberg, C. and C. J. Stone (1992). Logspline density estimation for censored data. *Journal of Computational and Graphical Statistics*, 1, 301–328.

- Kou, S. (2014). Levy processes in asset pricing. *Wiley StatsRef: Statistics Reference Online* (doi: 10.1002/9781118445112.stat03738).
- Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, 3, 73–84.
- Laffont, J.-J. and J. Tirole (1988). The dynamics of incentive contracts. *Econometrica: Journal of the Econometric Society*, 56, 1153–1175.
- Laird, N. (1978). Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association*, 73, 805–811.
- Lawson, C. L. and R. J. Hanson (1974). *Solving Least Squares Problems*. SIAM.
- Lehmann, E. (1983). *Theory of Point Estimation*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons.
- Lesperance, M. L. and J. D. Kalbfleisch (1992). An algorithm for computing the nonparametric MLE of a mixing distribution. *Journal of the American Statistical Association*, 87, 120–126.
- Liesenfeld, R. and R. Jung (1997). Stochastic volatility models: Conditional normality versus heavy-tailed distributions. *Journal of Applied Econometrics*, 15, 137–160.
- Lindsay, B. (1995). *Mixture Models: Theory, Geometry, and Applications*. Conference Board of the Mathematical Sciences: NSF-CBMS regional conference series in Probability and Statistics. Institute of Mathematical Statistics.
- Lindsay, B. G. et al. (1983). The geometry of mixture likelihoods: A general theory. *Annals of Statistics*, 11, 86–94.
- Liu, Y. and Y. Wang (2015). R package `cnmld`: Maximum likelihood estimation of a log-concave density function (version 1.0-0). <http://cran.r-project.org/package=cnmld>.
- Lloyd, C. J. (1998). Using smoothed receiver operating characteristic curves to summarize and compare diagnostic systems. *Journal of the American Statistical Association*, 93, 1356–1364.
- Lloyd, C. J. and Z. Yong (1999). Kernel estimators of the ROC curve are better than empirical. *Statistics & Probability Letters*, 44, 221–228.
- Maiboroda, R. E. and N. M. Markovich (2004). Estimation of heavy-tailed probability density function with application to web data. *Computational Statistics*, 19, 569–592.

- Mammen, E. (1991). Nonparametric regression under qualitative smoothness assumptions. *Annals of Statistics*, 19, 741–759.
- Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business*, 36, 394–419.
- Manning, W. G., A. Basu, and J. Mullahy (2005). Generalized modeling approaches to risk adjustment of skewed outcomes data. *Journal of Health Economics*, 24, 465–488.
- Markovich, N. M. (2005). Accuracy of transformed kernel density estimates for a heavy-tailed distribution. *Automation and Remote Control*, 66, 217–232.
- Markovich, N. M. (2006). Estimation of heavy-tailed density functions with application to www-traffic. In *2nd Conference on Next Generation Internet Design and Engineering*, pp. 208–215. IEEE.
- Markovitch, N. M. and U. R. Krieger (2000). Nonparametric estimation of long-tailed density functions and its application to the analysis of world wide web traffic. *Performance Evaluation*, 42, 205–222.
- Markovitch, N. M. and U. R. Krieger (2002). The estimation of heavy-tailed probability density functions, their mixtures and quantiles. *Computer Networks*, 40, 459–474.
- Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7, 77–91.
- Markowitz, H. M. and N. Usmen (1996). The likelihood of various stock market return distributions, part 1: Principles of inference. *Journal of Risk and Uncertainty*, 13, 207–219.
- Matthews, S. (1987). Comparing auctions for risk averse buyers: A buyer's point of view. *Econometrica: Journal of the Econometric Society*, 55, 633–646.
- McDonald, J. B. (1984). Some generalized functions for the size distribution of income. *Econometrica: Journal of the Econometric Society*, 52, 647–663.
- Meerschaert, M. M. and H.-P. Scheffler (2003). Nonparametric methods for heavy tailed vector data: A survey with applications from finance and hydrology. *Recent Advances and Trends in Nonparametric Statistics*, 265–279.
- Mengersen, K. L. and R. L. Tweedie (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Annals of Statistics*, 24, 101–121.

- Metz, C. E., B. A. Herman, and J.-H. Shen (1998). Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data. *Statistics in Medicine*, 17, 1033–1053.
- Meyer, M. C. (2008). Inference using shape-restricted regression splines. *Annals of Applied Statistics*, 2, 1013–1033.
- Meyer, M. C. (2012). Nonparametric estimation of a smooth density with shape restrictions. *Statistica Sinica*, 22, 681.
- Meyer, M. C. and M. Woodroffe (2004). Consistent maximum likelihood estimation of a unimodal density using shape restrictions. *Canadian Journal of Statistics*, 32, 85–100.
- Mittnik, S. and S. T. Rachev (1993). Modeling asset returns with alternative stable distributions. *Econometric Reviews*, 12, 261–330.
- Mittnik, S., S. T. Rachev, and M. S. Paolella (1998). Stable Paretian modeling in finance: Some empirical and theoretical aspects. In R. F. Robert Adler and M. Taqqu (Eds.), *A Practical Guide to Heavy Tails*, pp. 79–110. Birkhauser, Boston.
- Müller, S. and K. Rufibach (2009). Smooth tail-index estimation. *Journal of Statistical Computation and Simulation*, 79, 1155–1167.
- Novak, S. (2011). *Extreme Value Methods with Applications to Finance*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.
- Øigård, T. A., A. Hanssen, R. E. Hansen, and F. Godtliebsen (2005). EM-estimation and modeling of heavy-tailed processes with the multivariate normal inverse gaussian distribution. *Signal Processing*, 85, 1655–1673.
- Pal, J. K., M. Woodroffe, and M. C. Meyer (2007). Estimating a Polya frequency function. In W. S. Regina Liu and C.-H. Zhang (Eds.), *Complex Datasets and Inverse Problems: Tomography Networks and Beyond*, Volume 54 of *Lecture Notes–Monograph Series*, pp. 239–249. IMS, Beachwood, OH, USA.
- Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33, 1065–1076.
- Pepe, M. S. and T. Cai (2004). The analysis of placement values for evaluating discriminatory measures. *Biometrics*, 60, 528–535.

- Pitt, I. L. (2010). Estimation of skewness, heavy tails, and music success in a performance rights organization. In *Economic Analysis of Music Copyright*, pp. 105–119. Springer.
- Pivac, J. A. E. J. S. (2006). Parametric forecasting of value at risk using heavy tailed distribution. In *11 th International Conference on Operational Research KOI 2006*, pp. 65–75.
- R Core Team (2015). R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria.
- Rachev, S. T., C. Menn, and F. J. Fabozzi (2005). *Fat-tailed and Skewed Asset Return Distributions: Implications for Risk Management, Portfolio Selection, and Option Pricing*. John Wiley & Sons.
- Rao, B. P. (1969). Estimation of a unimodal density. *Sankhyā: Indian Journal of Statistics, Series A*, 31, 23–36.
- Robertson, T., F. Wright, R. L. Dykstra, and T. Robertson (1988). *Order Restricted Statistical Inference*. Probability and Statistics Series. Wiley New York.
- Rosenberg, J. V. and T. Schuermann (2006). A general approach to integrated risk management with skewed, fat-tailed risks. *Journal of Financial Economics*, 79, 569–614.
- Rubio, F. J. and M. F. Steel (2014). Bayesian modelling of skewness and kurtosis with two-piece scale and shape transformations. *arXiv preprint arXiv:1406.7674*.
- Rufibach, K. (2007). Computing maximum likelihood estimators of a log-concave density function. *Journal of Statistical Computation and Simulation*, 77, 561–574.
- Rufibach, K. (2012). A smooth ROC curve estimator based on log-concave density estimates. *International Journal of Biostatistics*, 8.
- Ryan, D. L. and T. J. Wales (2000). Imposing local concavity in the translog and generalized Leontief cost functions. *Economics Letters*, 67, 253–260.
- Saulo, H., V. Leiva, F. A. Ziegelmann, and C. Marchant (2013). A nonparametric method for estimating asymmetric densities based on skewed Birnbaum–Saunders distributions applied to environmental data. *Stochastic Environmental Research and Risk Assessment*, 27, 1479–1491.
- Saumard, A. and J. A. Wellner (2014). Log-concavity and strong log-concavity: A review. *Statistics Surveys*, 8, 45–114.
- Schuhmacher, D. and L. Dümbgen (2010). Consistency of multivariate log-concave density estimators. *Statistics & Probability Letters*, 80, 376–380.

- Sheather, S. J. et al. (2004). Density estimation. *Statistical Science*, 19, 588–597.
- Shen, P. (2010). Nonparametric analysis of doubly truncated data. *Annals of the Institute of Statistical Mathematics*, 62, 835–853.
- Shen, P. (2016). Analysis of transformation models with doubly truncated data. *Statistical Methodology*, 30, 15–30.
- Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.
- Silverman, B. W. (1981). Using kernel density estimates to investigate multimodality. *Journal of the Royal Statistical Society, Series B (Methodological)*, 43, 97–99.
- Silverman, B. W. (1982). On the estimation of a probability density function by the maximum penalized likelihood method. *Annals of Statistics*, 10, 795–810.
- Simonoff, J. (2012). *Smoothing Methods in Statistics*. Springer Series in Statistics. Springer New York.
- Stone, C. J., M. H. Hansen, C. Kooperberg, Y. K. Truong, et al. (1997). Polynomial splines and their tensor products in extended linear modeling: 1994 Wald memorial lecture. *Annals of Statistics*, 25, 1371–1470.
- Stott, P. A., D. A. Stone, and M. R. Allen (2004). Human contribution to the European heatwave of 2003. *Nature*, 432, 610–614.
- Stoyanov, S. V., S. T. Rachev, B. Racheva-Iotova, and F. J. Fabozzi (2011). Fat-tailed models for risk estimation. *Journal of Portfolio Management*, 37, 107–117.
- Sun, J., E. W. Frees, and M. A. Rosenberg (2008). Heavy-tailed longitudinal data modeling using copulas. *Insurance: Mathematics and Economics*, 42, 817–830.
- Takada, T. et al. (2001). Nonparametric density estimation: A comparative study. *Economics Bulletin*, 3, 1–10.
- Tankov, P. (2004). *Financial Modelling with Jump Processes*. CRC press.
- Terlaky, T. and J. P. Vial (1998). Computing maximum likelihood estimators of convex density functions. *SIAM Journal on Scientific Computing*, 19, 675–694.

- Terrell, G. R. (1990). The maximal smoothing principle in density estimation. *Journal of the American Statistical Association*, 85, 470–477.
- Theodossiou, P. (1998). Financial data and the skewed generalized T distribution. *Management Science*, 44, 1650–1661.
- Turnbull, B. C. and S. K. Ghosh (2014). Unimodal density estimation using Bernstein polynomials. *Computational Statistics & Data Analysis*, 72, 13–29.
- Venturini, S., F. Dominici, and G. Parmigiani (2008). Gamma shape mixtures for heavy-tailed distributions. *Annals of Applied Statistics*, 2, 756–776.
- Virbickaite, A., M. C. Ausín, and P. Galeano (2015). Bayesian inference methods for univariate and multivariate GARCH models: A survey. *Journal of Economic Surveys*, 29, 76–96.
- Walls, W. D. (2005). Modelling heavy tails and skewness in film returns. *Applied Financial Economics*, 15, 1181–1188.
- Walther, G. (2002). Detecting the presence of mixing with multiscale maximum likelihood. *Journal of the American Statistical Association*, 97, 508–513.
- Walther, G. (2009). Inference and modeling with log-concave distributions. *Statistical Science*, 24, 319–327.
- Wand, M. P., J. S. Marron, and D. Ruppert (1991). Transformations in density estimation. *Journal of the American Statistical Association*, 86, 343–353.
- Wand, P. and C. Jones (1994). *Kernel Smoothing*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.
- Wang, X., M. Woodroffe, M. G. Walker, M. Mateo, and E. Olszewski (2005). Estimating dark matter distributions. *Astrophysical Journal*, 626, 145–158.
- Wang, Y. (2007). On fast computation of the non-parametric maximum likelihood estimate of a mixing distribution. *Journal of the Royal Statistical Society, Series B*, 69, 185–198.
- Wang, Y. (2010). Maximum likelihood computation for fitting semiparametric mixture models. *Statistics and Computing*, 20, 75–86.
- Wang, Y., C. L. Lawson, and R. J. Hanson (2015). R package lsei: Solving least squares problems under equality/inequality constraints (version 1.1-1). <http://cran.r-project.org/package=lsei>.

- Wieand, S., M. H. Gail, B. R. James, and K. L. James (1989). A family of nonparametric statistics for comparing diagnostic markers with paired or unpaired data. *Biometrika*, 76, 585–592.
- Wolters, M. (2012). *Methods for Shape-constrained Kernel Density Estimation*. Ph. D. thesis, The School of Graduate and Postdoctoral Studies, University of Western Ontario, Canada.
- Wolters, M. A. (2009). A greedy algorithm for unimodal kernel density estimation by data sharpening. *Journal of Statistical Software*, 47, 1–26.
- Woodroffe, M. and J. Sun (1993). A penalized maximum likelihood estimate of. *Statistica Sinica*, 3, 501–515.
- Wu, C.-F. (1982). Some algorithms for concave and isotonic regression. *Studies in the Management Sciences*, 19, 105–116.
- Yatchew, A. and W. Härdle (2006). Nonparametric state price density estimation using constrained least squares and the bootstrap. *Journal of Econometrics*, 133, 579–599.
- Zhou, X.-H. and J. Harezlak (2002). Comparison of bandwidth selection methods for kernel smoothing of ROC curves. *Statistics in Medicine*, 21, 2045–2055.
- Zhou, X.-H. and H. Lin (2008). Semi-parametric maximum likelihood estimates for ROC curves of continuous-scale tests. *Statistics in Medicine*, 27, 5271–5290.
- Zou, K. H., W. Hall, and D. E. Shapiro (1997). Smooth non-parametric receiver operating characteristic (ROC) curves for continuous diagnostic tests. *Statistics in Medicine*, 16, 2143–2156.
- Zumbach, G. O. (2007). A gentle introduction to the RM2006 methodology. *SSRN Electronic Journal* (doi: 10.2139/ssrn.1420183).
- Zweig, M. H. and G. Campbell (1993). Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, 39, 561–577.