

Computing Confidence Intervals for Log-Concave
Densities

by Mahdis Azadbakhsh

Supervisors: Hanna Jankowski

Xin Gao

York University

August 2012

Contents

1	Introduction	1
1.1	Introduction	1
1.2	Nonparametric estimation of a density function	3
1.2.1	Kernel Smoothing	4
1.3	Maximum likelihood estimators	6
2	Family of log-concave densities	8
2.1	Introduction	8
2.2	Basic definition	8
2.3	Properties of Log-Concave Random Variables	9
2.4	Existence and uniqueness of maximum likelihood estimator . . .	10
2.5	A Computational approach for Univariate Log-Concave Density Estimation	12
3	Confidence Interval	14
3.1	Introduction	14
3.2	Estimating Quantiles of the Limiting Process	15
3.2.1	Simulation Results	16
3.3	Estimating C_2	17
3.4	Bandwidth Selection	18
3.4.1	Least Square Cross Validation	18
3.4.2	The Bandwidth proposed by Duembgen and Rufibach . .	22
3.4.3	Finding a Bandwidth using a reference distribution	23
3.4.4	Kernel Density Method	24

3.5	Consistency	24
4	Simulation Studies	27
4.1	Which Bandwidth is optimal?	27

Abstract

Log-concave density estimation is considered as an alternative for kernel density estimations which does not depend on tuning parameters. Pointwise asymptotic theory has been already developed for the nonparametric maximum likelihood estimator of a log-concave density. Here, the practical aspects of this theory are studied. In order to obtain a confidence interval, estimators of the constants appearing in the limiting distribution are developed and quantiles of the underlying process estimated. We then study the empirical coverage probabilities of pointwise confidence intervals based on these methods. The general methodology can be applied to other shape constrained problems - e.g. k-monotone estimators.

Chapter 1

Introduction

1.1 Introduction

In nonparametric density estimation, kernel smoothing methods are common approaches, introduced by Fix and Hodge(1951) that involve some tuning parameters, such as the order of kernel and bandwidth. An optimal choice of these parameters has not a straightforward solution, since they depend on the unknown underlying density f through the integral of the square of the second derivative of f . Silverman(1982, 1986) and Donoho et al.(1996) did a vast work in this field. To avoid this problem, another approach was formed by imposing some constraints on f , for example, convexity on certain intervals, unimodality and monotonicity. Such shape constraints are sometimes arise from the problem under study directly, (Hampel (1987) or Wang et al. (2005)), or they are at least plausible in many problems. Also, imposing these constraints might improve the quality of the estimator.

In the context of density estimation, Grenander (1956) derived the nonparametric maximum likelihood estimator of a density function that is non-increasing on a half-line. But as the nonparametric MLE does not exist for a unimodal density with unknown mode(Birge (1997)), this result is not useful for estimating such densities. Even with a known mode, the estimator is inconsistent near the mode(Woodroffe and Sun (1993)).

Walther (2002) worked on one kind of shape restricted maximum likelihood

inference, called log-concavity. Although the class of log-concave densities is a subset of the class of the unimodal densities, it contains most of the commonly used parametric distributions. Pal et al.(2006) and Rufibach (2006), independently proved that nonparametric maximum likelihood estimator of a log-concave density, always uniquely exists. Uniform consistency of NPMLE of log-concave densities was shown by Duembgen and Rufibach (2009) for the case $d = 1$ dimension. Cule and Samworth(2010) and Cule , Samworth and Stewart (2010) developed the nonparametric maximum likelihood estimator for multidimensional log-concave densities and studied its limiting behavior.

Silverman(1986), Thompson and Tapia(1990), showed some applications of non-parametric density estimation. Because of the some attractive properties of Log-concave functions, they has been applied in modeling and inference, in both the univariate and the multivariate cases. The family of log-concave distributions contains symmetric and skewed densities. This flexibility and also having subexponential tails and non-decreasing hazard rates are of the properties of this class that widen the area of its applications. (see Karlin (1968) and Barlow and Proschan (1975)). Using the fact that the hazard rate of a log-concave density is automatically monotone, Duembgen and Rufibach (2009) build a non-decreasing estimator of the hazard rate. Utilizing a log-concave estimator leads to an improved performance for some problems in extreme value theory which reported by Muller and Rufibach (2009). Economics(An (1995,1998),Bagnoli and Bergstrom (2005), and Caplin and Nalebuff (1991)), reliability theory(Barlow and Proschan (1975)), sampling and nonparametric Bayesian analysis(see Gilks and Wild (1992), Dellaportas and Smith (1993), and Brooks (1998)) and clustering(Guenther Walther(2008)) are some of the areas which the log-concave distributions are found useful in. Duembgen et al. (2007) expanded EM algorithm to estimate a distribution based on arbitrarily censored data under the assumption of log-concavity. Balabdaoui et al. (2009) investigate the mode of NPMLE \hat{f}_n as an estimator of the mode of true density f . They also provided the pointwise limiting theory of the nonparametric maximum likelihood estimator of this family of densities. Using the limiting theory we developed a computable

confidence interval for estimator, which shows the variability of estimator better than a pointwise estimate.

In the first chapter, nonparametric density estimation, motivation for studying shape constrained densities and nonparametric maximum likelihood estimators are briefly discussed. Definition of log-concave density and some of its useful properties can be found in Chapter 2. In addition, some theorems for existence and uniqueness of NPMLE's are mentioned and a computational method to find these estimators is introduced.

Chapter 3, contains our work to compute a confidence interval for log-concave density. There are constant C_2 and the quantile $C(0)$ which are estimated in this chapter and simulation result for quantile estimation is available. To approximate C_2 , we need to find the optimal bandwidth for kernel smoothing of first and second derivative of log-concave maximum likelihood estimator. For this sake different method of bandwidth selection are studied and also the consistency of the estimator is investigated.

The simulation results for bandwidth selection methods including empirical coverage probabilities of pointwise confidence intervals are shown in Chapter 4.

1.2 Nonparametric estimation of a density function

To estimate a density function directly from data, without some restrictive assumptions, nonparametric approach is helpful. The simplest way to estimate a density is the histogram which needs two parameters: bin width h (or bandwidth) and starting point x_0 of the first bin. The bins are of the form $[x_0 + (i - 1)h, x_0 + ih), i = 1, 2, \dots, m$ and the approximated density at the center of each bin is given by

$$\hat{f}(x) = \frac{1}{n} \frac{\text{Number of observations in the same bin as } x}{h}$$

where n and m are denoted the number of observations and the number of bins, respectively. The choice of bandwidth has a substantial effect on the shape of estimated density. Although simplicity is a benefit of using the histogram, it has several drawbacks such as

- discontinuities which have not arisen from the underlying density and are due to the bin locations
- the density depends on the starting position of the bins
- in high dimensions, a very large sample is needed or else many of the bins would be empty

Therefore, other methods have been developed which improve the estimation.

1.2.1 Kernel Smoothing

Kernel-based methods are of the most popular estimators of density functions. They are smoother than histograms and converge to the true density faster. In other words, the MISE is asymptotically inferior to the kernel estimator, since its convergence rate of MISE (mean integrated square error) of histogram is $O(n^{-2/3})$ compared to the kernel estimators $O(n^{-4/5})$ rate. (see more in Scott(1979)). Kernel estimators remove the dependency on the first points by centering a kernel function at each data point. So two of the problems with histograms are eliminated, but bandwidth problem still remains. Usually a function K is called **kernel** of order r , if it satisfies the following assumptions:

- K is symmetric, i.e. $K(x) = K(-x)$
- $\int_{\mathbb{R}} K(x) dx = 1$
- $\int_{\mathbb{R}} x^i K(x) dx = 0, i = 1, \dots, r - 1$
- $\int_{\mathbb{R}} x^r K(x) dx \neq 0$

Some of the commonly kernels are: Uniform, Triangle, Epanechnikov, Quartic, Triweight and Gaussian. In this work, the Gaussian kernel of zero mean and h^2 variance is used which all its derivatives exist.

$$K(x) = \frac{1}{\sqrt{2\pi h^2}} e^{-\frac{x^2}{2h^2}}$$

$$K'(x) = \left(-\frac{x}{h^2}\right) \left(\frac{1}{\sqrt{2\pi h^2}} e^{-\frac{x^2}{2h^2}}\right)$$

$$K''(x) = \left(\frac{1}{h^2}\right) \left(\frac{x^2}{h^2} - 1\right) \left(\frac{1}{\sqrt{2\pi h^2}} e^{-\frac{x^2}{2h^2}}\right)$$

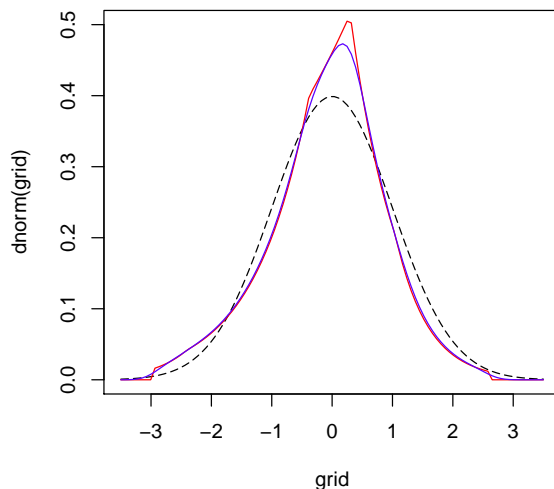


Figure 1.1: The dot black line shows the true standard normal density, red line and blue line show the evaluated MLE at grid points and smoothed MLE, respectively

Given a random sample X_1, \dots, X_n with a continuous, univariate density f . The kernel density estimator is

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i)$$

the estimator for the r -th derivative of the density function $f(x)$ is

$$\hat{f}_n^{(r)}(x) = \frac{1}{n} \sum_{i=1}^n K_h^{(r)}(x - X_i)$$

The value of the bandwidth h has more effect on the quality of kernel smoothed estimate than the shape of kernel K . Therefore it is important to choose the most efficient bandwidth value. Too small or too large values lead to the undersmoothed or oversmoothed estimates (a small bandwidth leading to a higher variance and a large bandwidth causing a higher bias).

So, choosing the optimal smoothing parameter is crucial in kernel estimation. (See Silverman (1986))

1.3 Maximum likelihood estimators

Parametric MLE

Maximum likelihood is one of the methods for parametric estimation and the density can be estimated by $\hat{f}_n(\cdot) = f(\cdot, \hat{\theta}_n)$. Define the likelihood function as

$$L_n(\theta) = \prod_{i=1}^n f(X_i; \theta)$$

and the log-likelihood

$$l_n(\theta) = \frac{1}{n} \sum_{i=1}^n \log f(X_i; \theta) dF_n(x)$$

where F_n denotes the empirical distribution function.

$$\hat{\theta}_n = \max_{\theta \in \Theta} l_n(\theta)$$

Under regularity condition, it may be shown that

$$\|\hat{\theta}_n - \theta_0\|_2 = O(n^{-1/2})$$

Here θ_0 and $\|\cdot\|_2$ denote the true parameter and the Euclidean norm on \mathbb{R} , respectively. Choosing a suitable norm, there would be the same property for $\|\hat{f}_n - f_0\|$.

Nonparametric MLE(NPMLE)

In the absence of a parametric model, there are less assumption about the underlying density and the estimate converges more slowly than the $O(n^{-1/2})$. In this case the maximum likelihood approach is not efficient any more.

The empirical distribution function can be viewed as a nonparametric maximum likelihood estimate (Thompson and Tapia (1990)). The kernel density estimator can considered as a kind of maximum likelihood estimator. The maximum likelihood estimator is a linear function between each two data points (Duemgen et al. (2007)), that is, it is not smoothed. Therefore it is not differentiable every where but in some applications, it is preferred to have a smoothed estimator. kernel density estimator might be written as the smoothed maximum likelihood estimator

$$\tilde{f}(x) = \int_y K_h(x - y) \hat{f}_n(y) dy$$

Here $\hat{f}_n(y)$ denotes the maximum likelihood estimate at point y and $\tilde{f}(x)$ is the smoothed maximum likelihood estimator.

Chapter 2

Family of log-concave densities

2.1 Introduction

Specifying an optimal smoothing parameter or bandwidth was introduced as one of the problems of performing kernel density estimation. So it is useful to find some methods of automatic bandwidth selection.

Imposing some shape constraints on the densities under consideration can result in an explicit solution that does not depend on a tuning parameter. Two constraints monotonicity and unimodality are of most interest in the literature. In fact, monotone densities are included in the class of unimodals, for which the mode is located at either edge of the density's support.

Because of the attractive properties of log-concave functions, it can be used as an alternative for kernel-density estimation.

2.2 Basic definition

Let X be a random variable with distribution function F and Lebesgue density

$$f(x) = \exp(\varphi(x))$$

for some concave function $\varphi : \mathbb{R} \rightarrow [-\infty, \infty)$. In other words, a function f is called log-concave if $\log f$ is concave, i.e.

$$\log f(\lambda x + (1 - \lambda)y) \geq \lambda \log f(x) + (1 - \lambda) \log f(y)$$

for all $\lambda \in (0, 1)$ and all $x, y \in \mathbb{R}$.

Remark: Log-concavity of f on (a, b) is equivalent to each of the following two conditions.

- i. $f'(x)/f(x)$ is monotone decreasing on (a, b)
- ii. If the second derivative of $\log f$ exists, then $(\log f(x))'' < 0$

2.3 Properties of Log-Concave Random Variables

The class of log-concave densities is a large and flexible class with many nice properties.

- Many parametric models, for a certain range of their parameters are log-concave. Normal, uniform, $\text{gamma}(r, \lambda)$ with shape parameter $r \geq 1$, $\text{beta}(a, b)$ for $a \geq 1$ and $b \geq 1$, generalized Pareto, Gumbel, Fréchet, logistic or Laplace, are some of these models. So a log-concave density can be nonparametric alternative for purely parametric models.

Definition A real-valued function f on \mathbb{R} is PF_2 if and only if both conditions hold,

(a) $f(x) \geq 0$

(b) for any two pairs (x_1, y_1) and (x_2, y_2) in \mathbb{R} , where $x_1 < x_2$ and $y_1 < y_2$, the determinant

$$\begin{vmatrix} f(x_1 - y_1) & f(x_1 - y_2) \\ f(x_2 - y_1) & f(x_2 - y_2) \end{vmatrix}$$

is nonnegative.

Each of the following two statements is equivalent to (b).

(b') f is a concave function on \mathbb{R}

(b'') for every fixed $\delta > 0$, the ratio $\frac{f(x+\delta)}{f(x)}$ is decreasing in x , for every x in an interval contained in a support of f .

- Tails of log-concave densities are sub-exponential: i.e. if $X \sim f \in PF_2$, then $E \exp(c|X|) < \infty$ for some $c > 0$.
- For every real valued log-concave random variable X , there exist $\theta_0 > 0$ such

that $M_X(\theta) = E(e^{\theta X}) < \infty$ for all $\theta \in \mathbb{R}$ that $\|\theta\|_2 < \theta_0$, where $\|\cdot\|_2$ is L_2 norm. A proof can be found in Cule's thesis (Theorem 2.6).

- every log concave density is unimodal but it is not necessarily symmetric.
- A log-concave density is unimodal, but the converse is not correct. A function $f : \mathbb{R} \rightarrow \mathbb{R}$ is said to be unimodal if there exist a constant $m \in \mathbb{R}$ such that f is nonincreasing on (m, ∞) and nondecreasing on $(-\infty, m)$.
- A density f on \mathbb{R} is log-concave if and only if its convolution with any unimodal density is unimodal. However the class of unimodal families is not closed under convolution. Log-concavity has been called "strongly unimodality" (Ibragimov, 1956)
- If X and Y are independent, log-concave random variables, then $X+Y$ is log-concave. This is an attractive property of this class which is not held by the class of unimodal random variables.
- Marginal and conditional densities obtained from a log-concave density are log-concave. Inverse is not necessarily true.
- The class of log-concave function is closed under weak limits.
- Unlike the family of log-convex densities, a mixture of log-concave densities is not log-concave, in general.
- Let f be strictly monotonic (increasing or decreasing) on the interval (a, b) . and either $f(a) = 0$ or $f(b) = 0$. Then if f' is log-concave on (a, b) , then $f(x)$ is also log-concave on (a, b) . (see Bagnoli and Bergstrom, 1989 for proof)
- If the density f , is monotone decreasing, then its c.d.f., F , and its left side integral, $G(x) = \int_l^x F(t)dt$, are both log-concave.

2.4 Existence and uniqueness of maximum likelihood estimator

In order to use maximum likelihood in nonparametric situations, imposing some restrictions is necessary to ensure that the density does not get too spiky. For the first time, Grenander (1956) used shape-constrained maximum likelihood for estimating mortality under the assumption of monotonicity. Then some other shape constraints have been studied, such as unimodality (Rao, 1969), convexity (Groeneboom et al., 2001b), k-monotonicity, (Balabdaoui and Wellner, 2007)

and log-concavity (Duembgen and Rufibach, 2008; Walther, 2002). Although, no maximum likelihood estimator exists for some classes such as unimodal densities with unknown mode, (Birge, (1997)), we can use this method by adding some restriction for example log-concavity. The claim that these estimators perform well in practice has been shown by empirical results in different works. Maximum likelihood estimation for log-concave densities in one dimension was introduced by Walther (2002) and further Duembgen and Rufibach (2008) developed this technique.

Let X_1, X_2, \dots, X_n be a random sample of size $n > 1$ from distribution \mathbb{F} . For any log-concave probability density f on \mathbb{R} , the normalized log likelihood function at f is given by

$$l(\varphi) = \int \log f d\mathbb{F}_n = \int \varphi d\mathbb{F}_n \quad (2.1)$$

where \mathbb{F}_n denotes the empirical distribution function of the sample. Let C denote the class of all concave function $\varphi : \mathbb{R} \rightarrow [-\infty, \infty)$. One can show that maximizing $l(\varphi)$ over all function φ satisfying the constraint $\int \exp \varphi(x) dx = 1$ is equivalent to adding a Lagrangian term and maximizing

$$L(\varphi) = \int_{\mathbb{R}} \varphi d\mathbb{F}_n - \int_{\mathbb{R}} \exp \varphi(x) dx \quad (2.2)$$

Hence, the maximum likelihood estimator of φ is

$$\hat{\varphi}_n = \max_{\varphi \text{ concave}} L(\varphi)$$

and the nonparametric log-concave density estimator is $\hat{f}_n = \exp \hat{\varphi}_n$.

Theorem 2.4.1. *Suppose that X_1, X_2, \dots, X_n are iid observations in \mathbb{R} from a log-concave density f and $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ are the corresponding order statistics to the observations. Then, with probability one, there is a unique non-parametric maximum likelihood estimator $\hat{\varphi}_n$ which is linear on all intervals $[X_{(j)}, X_{(j+1)}]$, $1 \leq j < n$. Moreover, $\hat{\varphi}_n = -\infty$ on $\mathbb{R} \setminus [X_1, X_n]$.*

$\hat{\varphi}_n$ is piecewise linear, changes of slope is only possible at data points $X_i \in [X_1, X_n]$, i.e. for a continuous and piecewise linear function $h : [X_1, X_n] \rightarrow \mathbb{R}$,

the set of knots is introduced as

$$S_n(h) = \{t \in (X_1, X_n) : h'(t-) \neq h'(t+)\} \cup \{X_1, X_n\}$$

Knots of $\widehat{\varphi}_n$ only occur at some of the ordered observations $X_{(1)}, X_{(2)}, \dots, X_{(n)}$.

The estimator $\widehat{\varphi}_n$ has the following characterization. For $x \geq X_1$, define the processes

$$\begin{aligned} \widehat{F}_n(x) &= \int_{X_1}^x \exp(\widehat{\varphi}_n(t)) dt, & \widehat{H}_n(x) &= \int_{X_1}^x \widehat{F}_n(t) dt \\ \mathbf{H}_n(x) &= \int_{X_1}^x \mathbb{F}_n(t) dt = \int_{-\infty}^x \mathbb{F}_n(t) dt. \end{aligned}$$

then the concave function $\widehat{\varphi}_n$ is the MLE of the log density φ_0 if and only if,

$$\begin{aligned} \widehat{H}_n(x) &\leq \mathbf{H}_n(x), & \text{for all } x \geq X_1, \\ &= \mathbf{H}_n(x), & \text{if } x \in \widehat{S}_n(\widehat{\varphi}_n) \end{aligned}$$

Groeneboom and Wellner(2001) show that if the true density f_0 is convex on $[0, \infty)$ and that f_0 is twice continuously differentiable in a neighborhood of x_0 with $f_0''(x_0) > 0$, the MLE \widehat{f}_n satisfies

$$n^{2/5}(\widehat{f}_n(x_0) - f_0(x_0)) \rightarrow (24^{-1} f_0^2(x_0) f_0''(x_0))^{1/5} \mathbf{H}''(0) \quad (2.3)$$

where \mathbf{H} is a particular upper envelope of an integrated two-sided Brownian motion $+t^4$.

2.5 A Computational approach for Univariate Log-Concave Density Estimation

In last few years, several algorithms have been proposed to estimate maximum likelihood of a log-concave density. Duembgen and Rufibach(2011) have implemented two of those algorithms, named an iterative convex minorant and an active set algorithm, in an R package, **logcondens**.

This package provides the maximum likelihood estimate (MLE) and smoothed

log-concave density estimator derived from the MLE. Also it is possible to evaluate the estimated density, distribution and quantile functions at arbitrary points and compute the characterizing functions of the MLE. Moreover, the user is enabled to sample from the estimated distribution. There are two datasets available related to log-concave density estimation.

Chapter 3

Confidence Interval

3.1 Introduction

Log-concave estimation has many characteristics similar to convex density estimation problem mentioned in previous chapter. Balabdaoui, Rufibach and Wellner(2009) find that under the following assumptions:

- 1- The density function f_0 belongs to log-concave family.
- 2- $f_0(x_0) > 0$
- 3- The function φ_0 is at least twice continuously differentiable in a neighborhood of x_0
- 4- If $\varphi_0''(x_0) \neq 0$, then $k=2$. otherwise, suppose that k is the smallest integer such that $\varphi_0^{(j)}(x_0) = 0, j = 2, \dots, k - 1$, and $\varphi_0^{(k)}(x_0) \neq 0$ and $\varphi_0^{(k)}$ is continuous in a neighborhood of x_0 .

Then

$$n^{k/(2k+1)}(\widehat{f}_n(x_0) - f_0(x_0)) \rightarrow C_k(x_0, \varphi_0)\mathbf{H}_k''(0)$$

where the constant C_k is given by

$$C_k(x_0, \varphi_0) = \left(\frac{f_0(x_0)^{k+1} |\varphi_0^{(k)}(x_0)|}{(k+2)!} \right)^{1/(2k+1)}$$

Now, suppose that all four assumption hold with $k = 2$.

Let W denote two-sided Brownian motion, starting at 0. That is, $W(t) = W_1(t); t \geq 0$ and $W(t) = W_2(-t); t \leq 0$; where W_1 and W_2 are two independent Brownian

motions with $W_1(0) = W_2(0) = 0$. For $t \in \mathbb{R}$, define:

$$Y_2(t) = \begin{cases} \int_0^t W(s)ds - t^4, & \text{if } t \geq 0, \\ \int_t^0 W(s)ds - t^4, & \text{if } t < 0. \end{cases}$$

\mathbf{H}_2 is called the lower envelope of the process Y_2 if

$\mathbf{H}_2(t) \leq Y_2(t)$, for all $t \in \mathbb{R}$;

\mathbf{H}_2'' is concave;

$\mathbf{H}_2(t) = Y_2(t)$, if the slope of \mathbf{H}_2'' decreases strictly at t .

Based on the limiting distribution of the NPMLE, a confidence interval for the true log-concave the density f_0 is

$$\hat{f}_n(x_0) \pm n^{-2/5} \cdot C_2(x_0, \varphi_0) \mathbb{C}(0)$$

where

$$C_2(x_0, \varphi_0) = \left(\frac{f_0^3(x_0) |\varphi_0''(x_0)|}{24} \right)^{1/5},$$

and the distribution of $\mathbb{C}(0)$ is known and is described in next section.

3.2 Estimating Quantiles of the Limiting Process

According to Balabdaoui, Rufibach and Wellner(2009)

$$n^{2/5} C_2^{-1}(x_0) (\hat{f}_n(x_0) - f_0(x_0)) \rightarrow \mathbf{H}_2''$$

have defined the process \mathbb{C} .

They show that Suppose that C_m denotes the class of concave functions φ , that $\varphi(-m) = \varphi(m) = -12m^2$ The limiting distribution $\mathbb{C}(0)$ at $t = 0$ is the following process:

$$\mathbb{C} = \lim_{m \rightarrow \infty} \min_{\varphi \in C_m} \left\{ \int_{-m}^m \varphi^2(t) dt - 2 \int_{-m}^m \varphi(t) d(W(t) - 4t^3) \right\},$$

Then $\mathbb{C}(t) = \mathbf{H}_2''(t)$, for all $t \in \mathbb{R}$. The constant \mathbb{C} can be seen as the solution

of the continuous time regression problem, where the $Y(t)$ is obtained from the differential equation $dY(t) = 12t^2dt + dW(t)$.

Existence of $\tilde{\mathbb{C}}$ which precedes \mathbb{C} is shown in Groeneboom et al.(2001a). They explain an iterative cubic spline algorithm to simulate from the approximate distribution of \mathbb{C} . Groeneboom et al.(2008) also described an other algorithm based on the support reduction which has many applications and for example has used in Maathuis(2010) and Jankowski et al.(2009). Active set algorithm is the third way, described in Duembgen et al.(2010). Ng and Maechler have implemented this method in the function `conreg`, available in the R package `cobs`, but the constraints $\tilde{\varphi}(-m) = \tilde{\varphi}(m) = -12m^2$ have not been implemented. Groeneboom et al. (2008)and Rufibach (2007, 2010) have introduced some other algorithm which can be useful to deal with this problem.

Method COBS from the "brute force" approach is used for simulation to estimate the quantile. The result can be seen in Table 1. $B = 1000$ samples of $n^{2/5}C_2^{-1}(x_0)(\hat{f}_n(x_0) - f_0(x_0))$ for comparison with the last result.

3.2.1 Simulation Results

Table 1 contains the result of $B = 1000$ simulation of $n = 10000$ data over different values of X_0 . four different distribution were chosen

- (D1). the standard normal distribution,
- (D2). the gamma distribution with shape parameter 3 and rate 1,
- (D3). the beta distribution with both parameters 2,
- (D4). the standard Gumbel distribution.

Comparing with the empirically obtained results, it seems that the chosen method COBS provides sufficient answers as the quantile estimations. .

Table 3.1: Estimated values of $F_{\mathbb{C}(0)}^{-1}(p)$ using COBS, and four simulation using distributions (D1) normal(0, 1), (D2) gamma(3,1), (D3) beta(3,3), and (D4) Gumbel.

p	COBS	(D1)	(D2)	(D3)	(D4)
0.001	-3.5014	-3.6504	-3.2798	-3.4365	-3.4543
0.005	-3.0902	-2.9780	-2.8530	-2.8869	-2.9044
0.010	-2.7925	-2.7106	-2.5708	-2.6470	-2.5991
0.025	-2.4547	-2.3185	-2.1675	-2.2487	-2.2173
0.050	-2.0693	-1.9782	-1.8294	-1.9225	-1.9133
0.100	-1.6675	-1.5836	-1.4797	-1.5441	-1.5616
0.200	-1.1643	-1.0815	-1.0288	-1.0698	-1.0718
0.500	-0.0878	-0.0678	-0.0795	-0.0664	-0.0708
0.800	1.0767	1.0705	1.0023	1.0353	1.0441
0.900	1.7298	1.7150	1.6175	1.6566	1.6541
0.950	2.2855	2.2722	2.1470	2.2038	2.6462
0.975	2.7274	2.7657	2.6070	2.6857	2.5224
0.990	3.2553	3.3377	3.1446	3.2323	3.1627
0.995	3.6527	3.6772	3.4977	3.5944	3.5180
0.999	4.1805	4.3257	4.0823	4.2211	4.1099

3.3 Estimating C_2

To estimate the constants C_2 the approximation of the second derivative of φ_0 is needed. We know that

$$\varphi(x) = \log f(x),$$

and

$$\varphi''(x) = \frac{f(x)f''(x) - f'^2(x)}{f^2(x)}.$$

So, it is necessary to estimate the first and second derivative of the underlying unknown density f .

To estimate the true density $f_0(x_0)$ we use the kernel maximum likelihood

estimator, that is

$$\tilde{f}_{h,n}(x_0) = \int_{\mathbb{R}} K_h(x_0 - x) \hat{f}_n(x) dx$$

where n and h denote the sample size and the bandwidth, respectively. K is the Gaussian kernel with mean zero and standard deviation h and $\hat{f}_n(\cdot)$ is the maximum likelihood estimator.

By differentiation at each point, one can have a pointwise maximum likelihood estimate of the derivatives of the log-concave density

$$\tilde{f}'_{n,h}(x) = \int_y K'_h(x - y) \hat{f}_n(y) dy$$

$$\tilde{f}''_{n,h}(x) = \int_y K''_h(x - y) \hat{f}_n(y) dy$$

The accuracy of the estimation is highly depends on the value of h .

3.4 Bandwidth Selection

This section is devoted to find an optimal bandwidth h for kernel estimation of second derivative of the density function. For this sake, there are several methods such as cross validation, risk estimation, etc. Here we use three different methods to find the optimal smoothing parameter.

3.4.1 Least Square Cross Validation

Cross validation (CV) is a popular method to optimize the bandwidth h . This method tries to find the h that minimizes one kind of error criterion between the true function and the estimated one base on different subsets of observations. In nonparametric density estimation methods, the choice of this criterion can affect the asymptotic results. Devroye and Györfi (1985) show that the L1 norm is the only L_p norm invariant under affine transformation. "For kernel density estimation, the L2 norm has been preferred for its easy decomposition into bias and variance terms." There are two main forms of cross validation: maximum likelihood cross validation (which uses the Kulback-Leibler information as criterion) and least squares cross validation. Here, we have used least square CV as

the first method.

Integrated Square Error(ISE)

Consider the smoothed continuous kernel density estimator $\tilde{f}_{h,n}$,

$$\tilde{f}_{h,n}''(x) = \int_y K_h''(x-y) \hat{f}_n(y) dy$$

where $\hat{f}_n(\cdot)$ is the NPMLE. The Integrated Square Error(ISE) is given by

$$\begin{aligned} ISE(\hat{f}_{h,n}'') &= \int (\tilde{f}_{h,n}'' - f'')(x)^2 dx \\ &= \int (\tilde{f}_{h,n}''(x) - f''(x))^2 dx \\ &= \int \tilde{f}_{h,n}''^2(x) dx - 2 \int \tilde{f}_{h,n}''(x) f''(x) dx + \int f''(x)^2 dx \end{aligned}$$

The last term does not depend on h , so the the criterion that should be minimized is:

$$\begin{aligned} CV(h) &= ISE(\hat{f}_{h,n}'') - \int f''(x)^2 dx \\ &= \int \tilde{f}_{h,n}''^2(x) dx - 2 \int \tilde{f}_{h,n}''(x) f''(x) dx \end{aligned}$$

To estimate numerically, we consider a set of grid points and evaluate the functions in those points. Let x_1, \dots, x_m be the grid points and $\Delta(x_i) = \min_i(x_i - x_{i-1})$. K denotes the Gaussian kernel and $\hat{f}_n(\cdot)$ is the estimated maximum likelihood obtained from the observation of the unknown density. The first term can

be calculated from data

$$\begin{aligned}
\int \tilde{f}_{h,n}''^2(x) dx &= \int_x \left(\int_y \frac{1}{h^3} \cdot K''\left(\frac{y-x}{h}\right) \hat{f}_n(y) dy \right)^2 dx \\
&\cong \int_x \left(\sum_{i=1}^n \frac{1}{h^3} \cdot K''\left(\frac{x_i-x}{h}\right) \hat{f}_n(x_i) \Delta(x_i) \right)^2 dx \\
&= \int_x \sum_{i=1}^n \sum_{j=1}^n \frac{1}{h^6} \cdot K''\left(\frac{x_i-x}{h}\right) \hat{f}_n(x_i) \cdot K''\left(\frac{x_j-x}{h}\right) \hat{f}_n(x_j) \Delta(x_i) \Delta(x_j) dx \\
&= \sum_{i=1}^n \sum_{j=1}^n \frac{1}{h^5} \int_s K''(s) \cdot K''\left(\frac{x_i-x_j}{h} + s\right) \hat{f}_n(x_i) \hat{f}_n(x_j) \Delta(x_i) \Delta(x_j) ds \\
&= \sum_{i=1}^n \sum_{j=1}^n \frac{1}{h^5} \int_s K''(s) \cdot K''\left(\frac{x_j-x_i}{h} - s\right) ds \hat{f}_n(x_i) \hat{f}_n(x_j) \Delta(x_i) \Delta(x_j) \\
&= \sum_{i=1}^n \sum_{j=1}^n \frac{1}{h^5} K'' * K''\left(\frac{x_j-x_i}{h}\right) \hat{f}_n(x_i) \hat{f}_n(x_j) \Delta(x_i) \Delta(x_j)
\end{aligned}$$

In the fourth line the substitution $\frac{x_i-x}{h} = s$ is used and the reason for the equation in the fifth line is that the second derivative of Gaussian kernel is symmetric around 0. The last equation is resulted of the definition of the convolution of a function which is:

$$f * f(x) = \int_R f(x-y) f(y) dy$$

In the second term of $CV(h)$, performing integration by part twice,

$$\begin{aligned}
\int \tilde{f}_{h,n}''(x) f''(x) dx &= \left(\tilde{f}_{h,n}''(x) f'(x) - \tilde{f}_{h,n}'''(x) f(x) \right)_{-\infty}^{\infty} + \int \tilde{f}_{h,n}^{(4)}(x) f(x) dx \\
&= \lim_{b \rightarrow \infty} (\tilde{f}_{h,n}''(b) f'(b) - \tilde{f}_{h,n}'''(b) f(b)) \\
&\quad - \lim_{b \rightarrow -\infty} (\tilde{f}_{h,n}''(b) f'(b) - \tilde{f}_{h,n}'''(b) f(b)) \\
&\quad + \int \tilde{f}_{h,n}^{(4)}(x) f(x) dx \\
&= 0 + E(\tilde{f}_{h,n}^{(4)}(x))
\end{aligned}$$

Let $x_{(1)}, \dots, x_{(n)}$ be ordered data points. By the bounded convergence theorem, and the fact that the rate of decay in exponential functions is much faster than

the growth of power functions, we have

$$\begin{aligned}
\lim_{b \rightarrow \infty} \tilde{f}_{h,n}''(b) &= \lim_{b \rightarrow \infty} \int_{x(1)}^{x(n)} \frac{1}{h^3} \cdot K''\left(\frac{b-y}{h}\right) \hat{f}_n(y) dy \\
&= \lim_{b \rightarrow \infty} \int_{x(1)}^{x(n)} \frac{1}{h^3 \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{b-y}{h}\right)^2} \left[\left(\frac{b-y}{h}\right)^2 - 1\right] \hat{f}_n(y) dy \\
&= 0
\end{aligned}$$

Using this and considering that φ is a concave function, For any fixed m , the first part of the equation can be written

$$\begin{aligned}
\lim_{b \rightarrow \infty} f'(b) \tilde{f}_{h,n}''(b) &= \lim_{b \rightarrow \infty} \lim_{h \rightarrow 0} e^{\varphi(b)} \times \left(\frac{\varphi(b+h) - \varphi(b)}{h}\right) \times \tilde{f}_{h,n}''(b) \\
&\leq \lim_{b \rightarrow \infty} \lim_{h \rightarrow 0} e^{\varphi(b)} \times \left|\frac{\varphi(b+h) - \varphi(b)}{h}\right| \times \tilde{f}_{h,n}''(b) \\
&\leq \lim_{b \rightarrow \infty} e^{\varphi(b)} \times \frac{|\varphi(b+m) - \varphi(b)|}{m} \times \tilde{f}_{h,n}''(b) \\
&\leq \frac{1}{m} \lim_{b \rightarrow \infty} \{e^{\varphi(b)} |\varphi(b+m)| + e^{\varphi(b)} |\varphi(b)|\} \times \tilde{f}_{h,n}''(b) \\
&= 0
\end{aligned}$$

Consider the second part of equation. There is some constant c such that

$$\begin{aligned}
\lim_{b \rightarrow \infty} f(b) \tilde{f}_{h,n}'''(b) &= \lim_{b \rightarrow \infty} e^{\phi(b)} \times \int_{x(1)}^{x(n)} \frac{1}{h^4} \cdot K''' \left(\frac{b-y}{h}\right) \hat{f}_n(y) dy \\
&\leq \lim_{b \rightarrow \infty} e^{-c \cdot b} \times \int_{x(1)}^{x(n)} \frac{1}{h^4 \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{b-y}{h}\right)^2} \left[3\left(\frac{b-y}{h}\right) - \left(\frac{b-y}{h}\right)^3\right] \hat{f}_n(y) dy \\
&= 0
\end{aligned}$$

The inequality arises from concavity of φ and the final answer is obtained similar to the other term . The same result holds for the term when b goes to $-\infty$

For the last term, observe that $\int \tilde{f}_{h,n}^{(4)}(x) f(x) dx = E_Y[\tilde{f}_{h,n}^{(4)}(x)]$ where the expectation should be computed with respect to the additional and independent

observation Y . Let y_1, \dots, y_n denote the observation and x_1, \dots, x_m be the grid points, which maximum likelihood of the underlying density is evaluated at each grid point. To estimate the expectation one can use the leave one out estimate

$$\begin{aligned} E(\widehat{f_{h,n}^{(4)}}(y)) &= \frac{1}{n} \sum_{i=1}^n \tilde{f}_{h,(-i)}^{(4)}(y_i) \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m \frac{1}{h^5} K^{(4)}\left(\frac{x_j - y_i}{h}\right) \hat{f}_{n,(-i)}(x_j) \Delta(x_j) \end{aligned}$$

Hence, we can find an optimal bandwidth h by minimizing $CV(h)$, that is

$$\begin{aligned} CV(h) &= \sum_{i=1}^n \sum_{j=1}^m \frac{1}{h^5} K''\left(\frac{x_j - x_i}{h}\right) \hat{f}_n(x_i) \hat{f}_n(x_j) \Delta(x_i) \Delta(x_j) \\ &\quad - \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m \frac{1}{h^5} K^{(4)}\left(\frac{x_j - y_i}{h}\right) \hat{f}_{n,(-i)}(x_j) \Delta(x_j) \end{aligned} \tag{3.1}$$

Local Cross Validation

Here, we try to minimize criterion at specified points to get the bandwidth h . Let t_0 be an arbitrary point which we are interested to evaluate the local error at,

$$LCV(h) = \sum_{i=1}^n (\tilde{f}_{n,h}''(t_0) - \tilde{f}_{(-i),h}''(t_0))^2 \tag{3.2}$$

Using the previous relation, we get

$$LCV(h) = \sum_{i=1}^n \left(\sum_{j=1}^m K_h''(t_0 - x_j) \hat{f}_n(x_j) \Delta(x_j) - \sum_{k=1}^m K_h''(t_0 - x_k) \hat{f}_{(-i),n}(x_k) \Delta(x_k) \right)^2$$

Simulation studies showed that $LCV(h)$ is decreasing with respect to h , so this is not a useful criterion to find the bandwidth.

3.4.2 The Bandwidth proposed by Duembgen and Rufibach

Duembgen et al(2007) proposed an approach to fit a suitable density to the observation. The smoothed log concave density estimator \hat{f}^* , for some bandwidth

γ

$$\hat{f}_n^*(x) = \int_{-\infty}^{\infty} K_\gamma(x - y) d\hat{F}_n(y)$$

that is, the convolution of a Gaussian kernel with mean 0 and standard deviation γ , where $\gamma^2 = \hat{\sigma}^2 - Var(\hat{F}_n)$.

Let \widehat{F}^* be the distribution function of smoothed density \widehat{f}^* . It is clear that \widehat{F}^* is the distribution function of $X + Y$, where X and Y are independent random variables that $X \sim \widehat{F}$ and $Y \sim N(0, \gamma^2)$

so $Var(\widehat{f}_n^*) = \widehat{\sigma}^2$.

Duembgen et al.(2007) define an explicit computation for $Var(\widehat{F})$.

$$\begin{aligned} Var(\widehat{F}) &= \int_{z_1}^{z_m} (z - \bar{Z})^2 \widehat{f}(x) dx \\ &= \sum_{j=2}^m \Delta(z_j) ((z_{j-1} - \bar{Z})^2 J_{10}(\widehat{\varphi}_{j-1}, \widehat{\varphi}_j) + (z_j - \bar{Z})^2 J_{10}(\widehat{\varphi}_j, \widehat{\varphi}_{j-1}) \\ &\quad - \Delta(z_j)^2 J_{11}(\widehat{\varphi}_{j-1}, \widehat{\varphi}_j)) \end{aligned}$$

Where z_1, \dots, z_m are the support points, $\Delta(z_j) = z_j - z_{j-1}, 2 \leq j \leq m$,

$$\widehat{\varphi} = \max_{\varphi} L(\varphi)$$

which the modified relation

$$L(\varphi) = \frac{1}{m} \sum_{i=1}^m w_i \varphi(z_i) - \sum_{j=1}^{m-1} (z_{j+1} - x_j) J(\varphi_j, \varphi_{j+1})$$

$(w_i)_1^m$ is the vector of positive weights that $\sum_{i=1}^m w_i = 1$. with the auxiliary function

$$J(r, s) = \begin{cases} (\exp(r) - \exp(s))/(r - s) & \text{if } r \neq s, \\ \exp(r) & \text{if } r = s. \end{cases}$$

Here, J_{10} and J_{11} are the partial derivatives of $J(r, s)$ with respect to r , and with respect to r and s , respectively. Duembgen et al.(2010) computed the derivatives in detail. Now, consider the estimator

$$\widehat{\sigma}^2 = n(n-1)^{-1} \sum_{j=1}^m w_j (z_j - \bar{Z})^2$$

As γ usually results in an undersmoothed density estimate, we also tried $\gamma \times n^{0.25}$ as another choice for bandwidth.

3.4.3 Finding a Bandwidth using a reference distribution

Here, we use normal distribution as a reference, i.e. it is assumed that the true density (which is unknown in practice) is normal with mean μ and unknown

standard deviation σ . We need to estimate $|\varphi''|$.

$$\varphi'' = \frac{f(x)f''(x) - f'^2(x)}{f^2(x)} = \frac{\frac{1}{\sigma^2}(\frac{x^2}{\sigma^2} - 1)f^2(x) - \frac{x^2}{\sigma^2}f^2(x)}{f^2(x)}$$

so

$$c_2(x_0, \varphi_0) = \left(\frac{f_0(x_0)^3 |\varphi_0^{(2)}(x_0)|}{4!} \right)^{1/5} = \frac{f_0(x_0)^3}{\sigma^2}$$

We use the smoothed density estimate and the standard deviation of the observation as the approximation of f and σ .

3.4.4 Kernel Density Method

In this method, instead of MLE the regular form of kernel smoothing is used, i.e.

$$\tilde{f}_{h,n}(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - y)$$

Package `ks`, enables us to get the bandwidth and density estimation directly. The function `"kdde"` estimates the kernel density derivatives and function `"hscv"` is used for bandwidth selection; However, it is not always stable for large sample sizes.

3.5 Consistency

Let $\hat{f}_n(\cdot)$, $\tilde{f}_n(\cdot)$ and $f(y)$ be the maximum likelihood estimation of the density, its kernel smoothed version and the true density, respectively.

Theorem 3.5.1. (*Cule and Samworth(2009)*) Suppose that f_n is a sequence of log-concave densities on \mathbb{R} that converges in distribution to some density f . Then for a $a_0 > 0$ and $b_0 \in \mathbb{R}$ such that $f(x) \leq \exp\{-a_0|x| + b_0\}$, there is a $a < a_0$ that

$$\int_{\mathbb{R}} e^{a|x|} |f_n(x) - f(x)| dx \rightarrow 0.$$

Also they show, for every density f with $\int_{\mathbb{R}} |x|f(x)dx < \infty$, $\int_{\mathbb{R}} f \log_+ f < \infty$ and $\text{int}(E) \neq \emptyset$, there exist a unique log concave density f^* such that $f^*(x) \leq \exp\{-a_0|x| + b_0\}$. Then for any $a < a_0$

$$\int_{\mathbb{R}} e^{a|x|} |\hat{f}_n(x) - f^*(x)| dx \rightarrow 0$$

and for continuous f^*

$$\sup_{x \in \mathbb{R}} e^{a|x|} |\widehat{f}_n(x) - f^*(x)| \rightarrow 0$$

Theorem 3.5.2. (Duembgen and Rufibach (2009)) Let \widehat{f}_n be the nonparametric maximum likelihood estimator of f , then \widehat{f}_n is the uniform consistent estimator of f , i.e.

$$\int |\widehat{f}_n(x) - f(x)| dx \rightarrow_p 0$$

Here, we show that for a fixed bandwidth, $\tilde{f}_{h,n}''(\cdot)$ asymptotically converges to of $f_{h,0}''(\cdot)$.

Theorem 3.5.3. Let $\tilde{f}_{h,n}''(\cdot)$ be a sequence of the smoothed log-concave second derivative of maximum likelihood estimators, and $f_{h,0}''(\cdot)$ be the smoothed log-concave second derivative of density f , i.e.

$$f_{h,0}''(x) = \int K_h''(x-y) f_0(y) dy.$$

Then, for every fixed bandwidth h , as $n \rightarrow \infty$

$$|\tilde{f}_{h,n}''(x) - f_{h,0}''(x)| \rightarrow 0$$

Proof.

$$\begin{aligned} |\tilde{f}_{h,n}''(x) - f_{h,0}''(x)| &= \left| \int K_h''(x-y) \widehat{f}_n(y) dy - \int K_h''(x-y) f_0(y) dy \right| \\ &\leq \int |K_h''(x-y)| |\widehat{f}_n(y) - f_0(y)| dy \end{aligned}$$

For $K_h(\cdot)$, the Gaussian kernel with mean 0 and variance h^2 ,

$$\begin{aligned}
|K_h''(x-y)| &= \left| \frac{1}{h^2} \left(\left(\frac{x-y}{h} \right)^2 - 1 \right) K_h(x-y) \right| \\
&= \frac{1}{h^2} \left| \left(\left(\frac{x-y}{h} \right)^2 - 1 \right) K_h(x-y) \right| \\
&= \frac{1}{h^2} |(u^2 - 1)| K(u) \\
&\leq \int_{-\infty}^{\infty} \frac{1}{h^2} |(u^2 - 1)| K(u) du \\
&= \frac{1}{h^2} \left[\int_{-\infty}^{\infty} (u^2 - 1) K(u) du + 2 \int_{-1}^1 (1 - u^2) K(u) du \right] \\
&= \frac{2}{h^2 \sqrt{2\pi}} (e^{1/2} - e^{-1/2})
\end{aligned}$$

In the third line $u = \frac{x-y}{h}$, and $K(u)$ is the the Gaussian kernel with mean 0 and standard deviation 1. Therefore, for a fixed bandwidth h , the term $|K_h''(x-y)|$ is bounded.

Hence, using the result of Theorem from Duembgen and Rufibach(2009)

$$\begin{aligned}
|\tilde{f}_{h,n}''(x) - f_{h,0}''(x)| &\leq \int |K_h''(x-y)| |\hat{f}_n(y) - f_0(y)| dy \\
&\leq \frac{2}{h^2 \sqrt{2\pi}} (e^{1/2} - e^{-1/2}) \int |\hat{f}_n(y) - f_0(y)| dy \\
&\rightarrow 0
\end{aligned}$$

and proof is complete. □

Chapter 4

Simulation Studies

In this chapter some simulation results for finding the optimal bandwidth and estimating C_2 are presented. Simulation was performed with

- B=100 replications, regenerating samples of size $n=1000$
- B=1000 replications, regenerating samples of size $n=200$ and $n=1000$

from four different log-concave distributions

- normal density with mean 0 and standard deviation 1
- gamma density with shape parameter 3 and scale parameter 1
- beta density with parameters 2 and 2
- standard Gumbel density

Also $m=500$ equidistant grid points between -3 and 3 were chosen and relations were evaluated at those points.

4.1 Which Bandwidth is optimal?

In the previous chapter several method was introduced to find the bandwidth. Here, regardless of their asymptotic behavior, it is tried to study and compare their performance in small samples through running some simulations. That is, we computed bandwidth through

- minimizing Integrated Square Error(ISE), which is denoted by *ise*

Simulation showed that this method is extremely slow, so it did not use for approximating in practice.

- the bandwidth proposed by Duembgen and Rufibach, denoted by γ
- $n^{0.25} * \gamma = \gamma^*$
- considering kernel smoothed density, denoted by ks
- using normal density as reference, which is labeled $norm$

To compare the value of C_2 computed using each of the four types of bandwidth, the box plots of difference of estimated C_2 and its true value at each grid point, in L_1 -norm and L_2 -norm, are given in Figure 4.1 and Figure 4.2. Each box plot in Figure 4.3, show the difference between the value of true density and that of the maximum likelihood kernel estimator at grid points for one different bandwidth. As it is expected, bandwidth obtained from $norm$ has the smallest error for normal distribution. For other distribution, it seems that ks , gives the optimal bandwidth compared with the other methods, i.e. its error is small and more stable for different densities.

Figure 4.4 and Figure 4.5 show the empirical coverage of the confidence interval for each method, i.e. the proportion of the 95% confidence intervals for each grid, which contains that grid point. Plots confirm the results of boxplots, that is, empirical coverage of the estimates performed using the bandwidth obtained from ks method, is more than others and more stable in the tails.

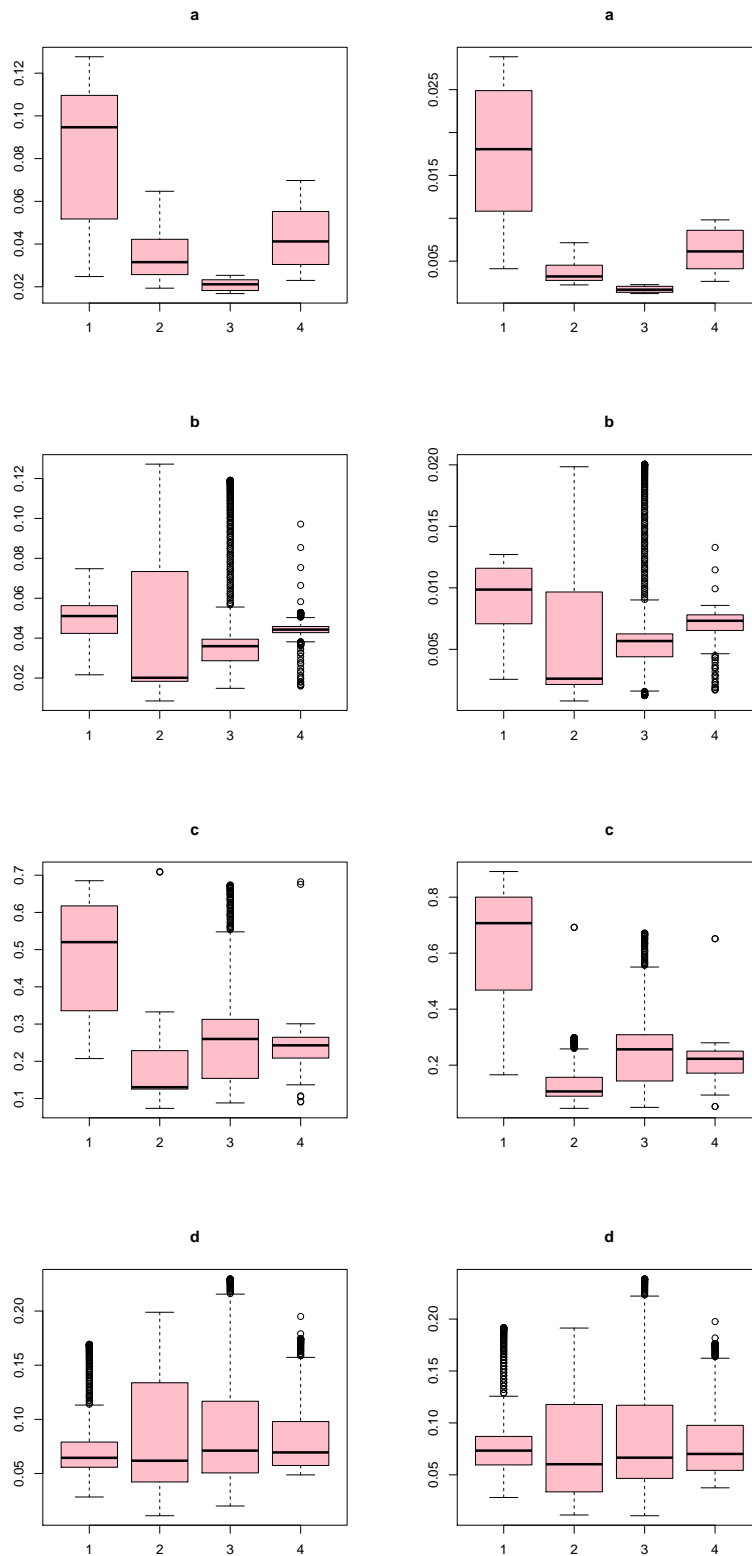


Figure 4.1: Average of difference of true and estimated C_2 in L1 for (a) normal, (b) gamma(3,1), (c) beta(2,2) and (d) standard Gumbel distributions; left to right: γ , γ^* , norm, ks; left column $n=200$, right column $n=1000$, $B=1000$

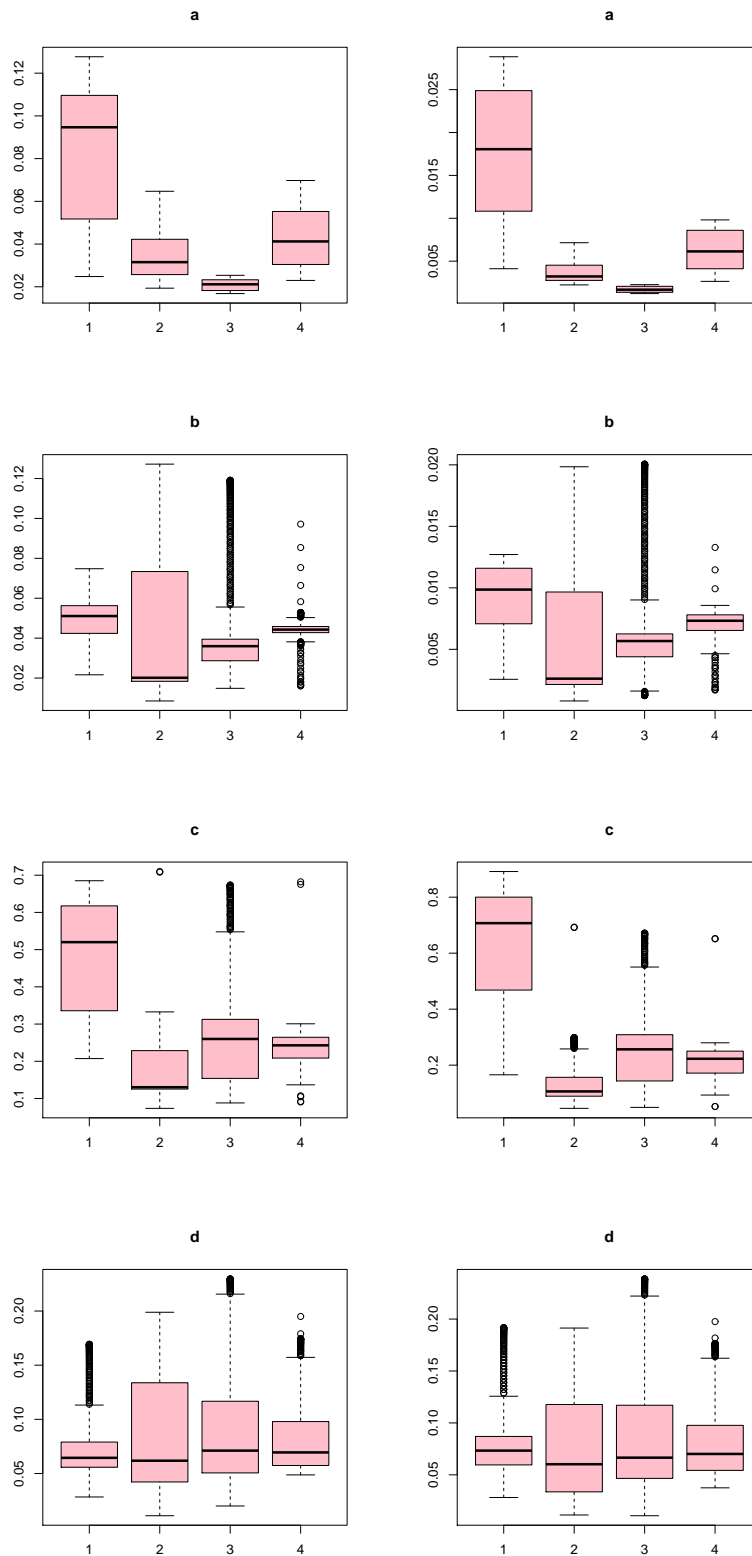


Figure 4.2: Average of difference of true and estimated C_2 in L2 for (a) normal, (b) gamma(3,1), (c) beta(2,2) and (d) standard Gumbel distributions; left to right: γ , γ^* , norm, ks; left column $n=200$, right column $n=1000$, $B=1000$

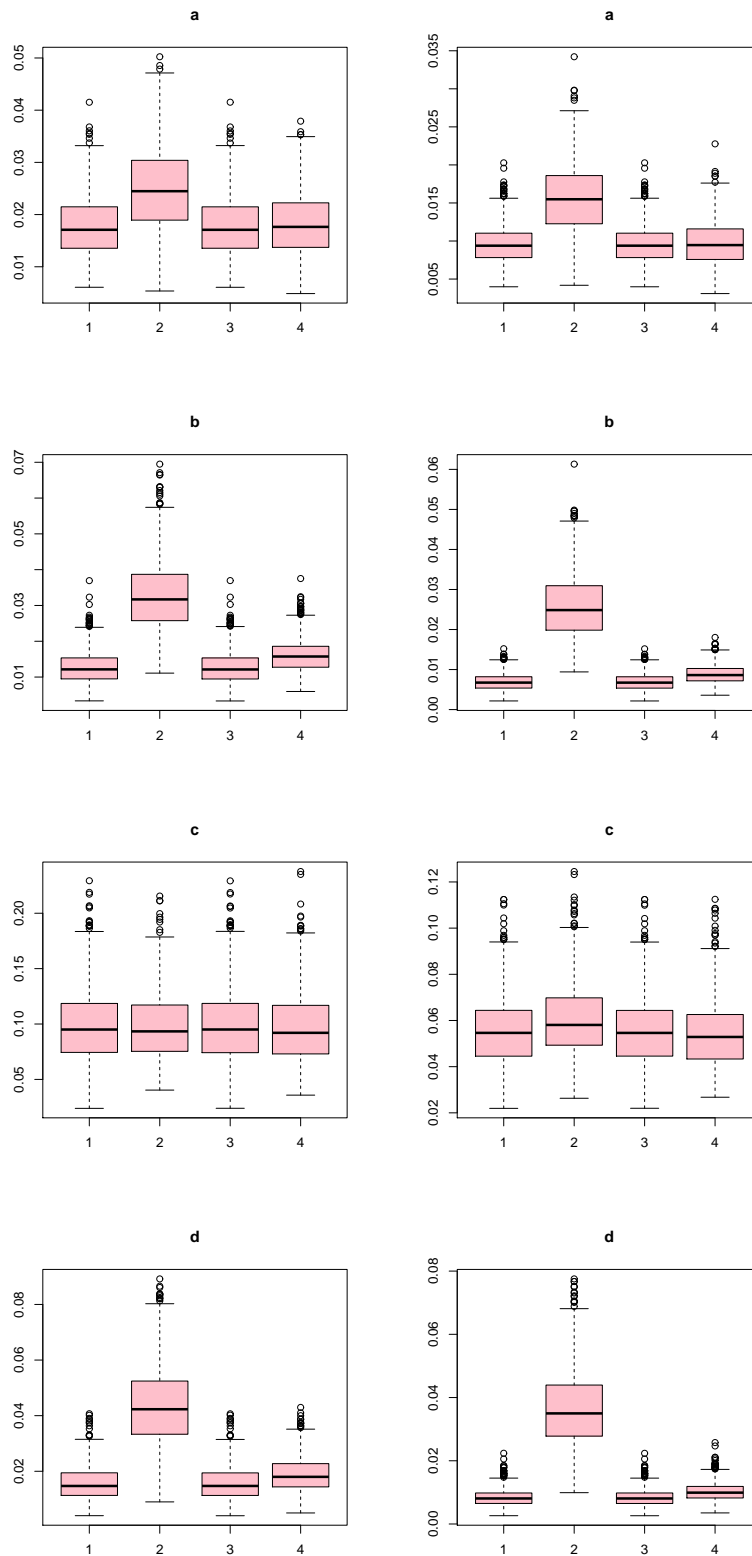


Figure 4.3: Average of difference of true and smoothed density in L1-norm for normal, gamma(3,1), beta(2,2) and standard Gumbel distributions; left to right: γ , γ^* , norm, ks; left column $n=200$, right column $n=1000$, $B=1000$

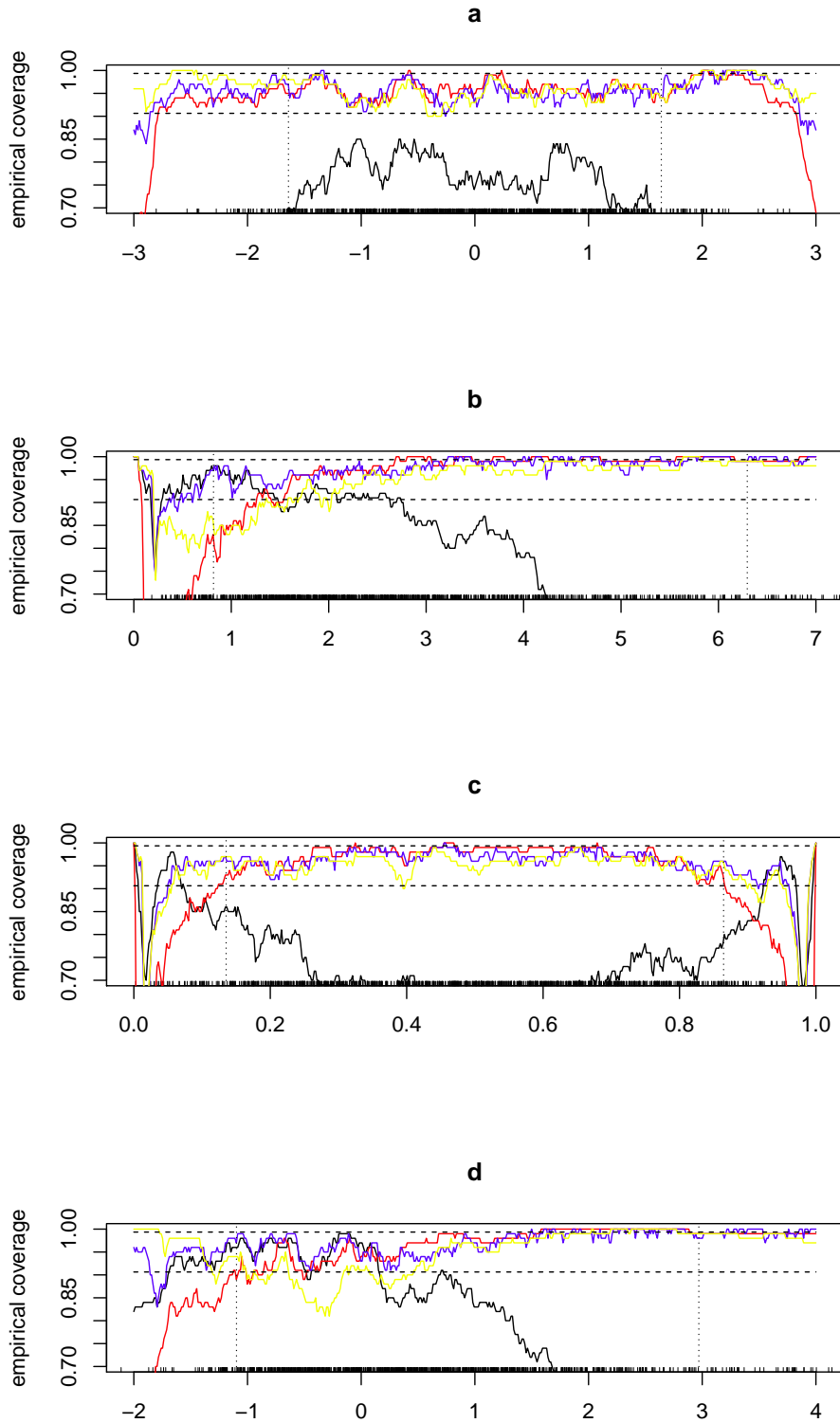


Figure 4.4: empirical coverage for (a) normal (b) gamma(3,1), (c) beta(2,2) and (d) standard Gumbel distributions; red line: norm, blue line: ks, yellow line: γ^* , black line: γ ; $n=1000$; $B=100$

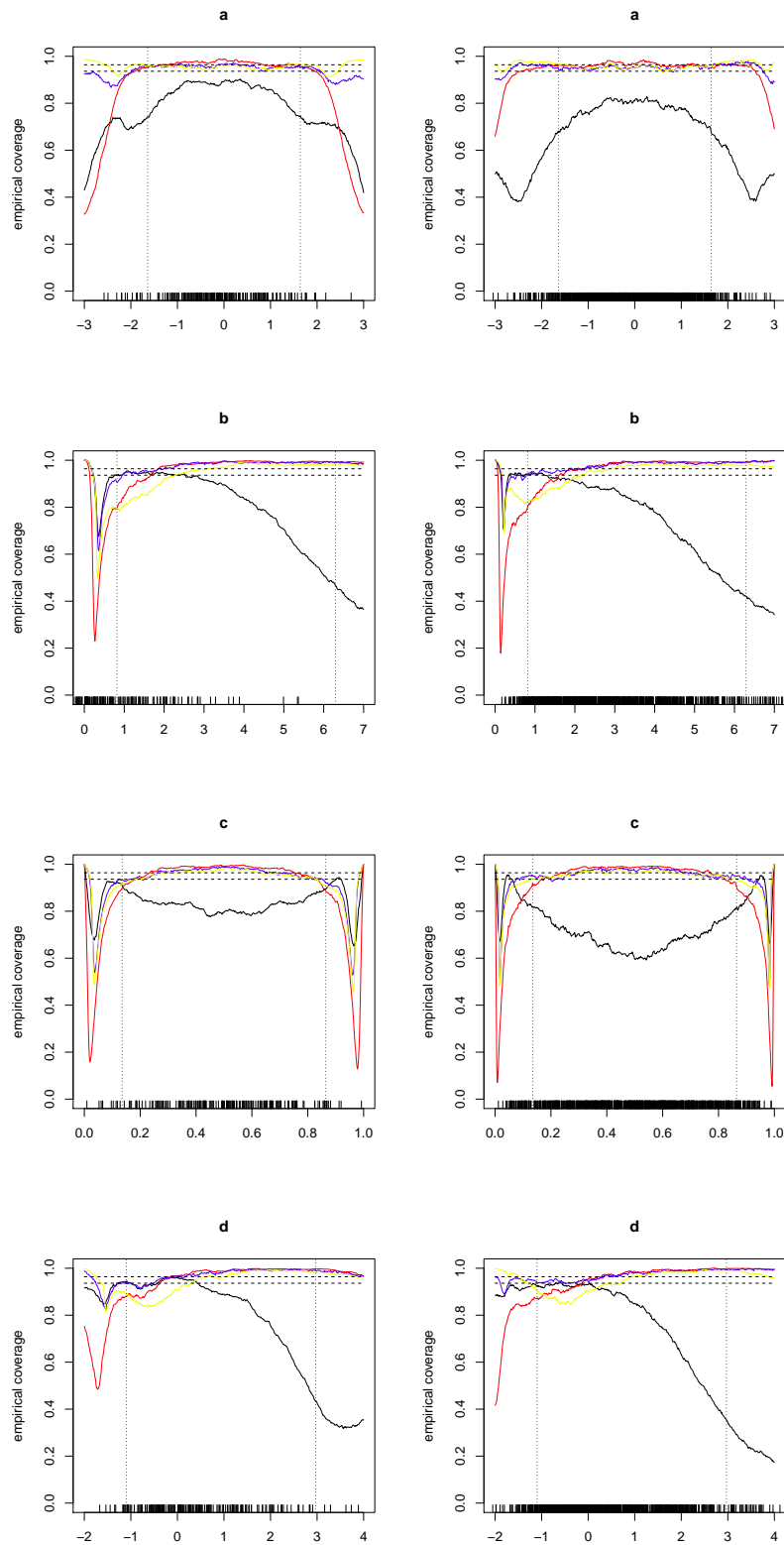


Figure 4.5: empirical coverage for (a) normal (b) gamma(3,1), (c) beta(2,2) and (d) standard Gumbel distributions; red line: norm, blue line: ks, yellow line: γ^* , black line: γ ; left column: $n=200$, right column: $n=1000$; $B=1000$

Bibliography

M. Y. An. *Log-concave probability distributions: theory and statistical testing*. Technical report, Duke University, 1995. M. Y. An. *Log-concavity versus log-convexity: A complete characterization*. *Journal of Economic Theory*, 80:350369, 1998.

M. Bagnoli and T. Bergstrom. *Log-concave probability and its applications*. *Economic Theory*, 26:445469, 2005.

F. Balabdaoui. *Nonparametric estimation of a k -monotone density: a new asymptotic distribution theory*. PhD thesis, University of Washington, 2004.

F. Balabdaoui and J. A. Wellner. *Estimation of a k -monotone density: Limiting distribution theory and the spline connection*. *The Annals of Statistics*, 35:25362564, 2007.

F. Balabdaoui, K. Rufibach, and J. A. Wellner. *Limit distribution theory for maximum likelihood estimation of a log-concave density*. *Annals of Statistics*, 37(3):12991331, 2009.

R. E. Barlow and F. Proschan. *Statistical theory of reliability and life testing*. Holt, Reinhart and Winston, New York, 1975.

O. Barndorff-Nielsen. *Information and exponential families in statistical theory*. Wiley, New Jersey, 1978.

L. Birg. *Estimation of unimodal densities without smoothness assumptions*. *Annals of Statistics*, 25(3):970981, 1997.

S. P. Brooks. *MCMC convergence diagnosis via multivariate bounds on log-concave densities*. *Annals of Statistics*, 26:398433, 1998.

G. Chang and G. Walther. *Clustering with mixtures of log-concave distributions*. *Computational Statistics and Data Analysis*, 51(12):62426251, 2007.

M. L. Cule and L. Duembgen. *On an auxiliary function for log-density estimation*. Technical Report 71, Universitt Bern, 2008.

M. L. Cule, R. J. Samworth, and M. I. Stewart. *Maximum likelihood estimation of a multidimensional log-concave density*. Submitted, 2008.

D. L. Donoho, I. M. Johnstone, G. Kerkycharian, and D. Picard. *Density estimation by wavelet thresholding*. Annals of Statistics, 24(2):508539, 1996.

L. Duembgen and K. Rufibach. *Maximum likelihood estimation of a log-concave density and its distribution function: Basic properties and uniform consistency*. Bernoulli, 15 (1):4068, 2008.

L. Duembgen, R. Samworth, D. Schuhmacher. *Approximation by log-concave distributions, with applications to regression*. The Annals of Statistics, 39(2), 702730, 2011.

L. Duembgen, A. Husler, and K. Rufibach. *Active set and EM algorithms for log-concave densities based on complete and censored data*. Technical report, Universitat Bern, 2007. URL <http://arxiv.org/abs/0709.0334/>.

E. Fix and J. L. Hodges. *Discriminatory analysis nonparametric discrimination: Consistency properties*. Technical Report 4, Project no. 21-29-004, USAF School of Aviation Medicine, Randolph Field, Texas, 1951.

E. Fix and J. L. Hodges. *Discriminatory analysis nonparametric discrimination: Consistency properties*. International Statistical Review, 57:238247, 1989.

P. Groeneboom. *Estimating a monotone density*. In Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Keifer, volume 2, pages 539555, London, 1983. Chapman and Hall.

P. Groeneboom, G. Jongbloed, and J. A. Wellner. *Estimation of a convex function: Characterizations and asymptotic theory*. The Annals of Statistics, 29:16531698, 2001b.

A. I. Ibragimov. *On the composition of unimodal distributions*. Theory of Probability and its Applications, 1(2):255260, 1956.

C.W. Lee. *Subdivisions and triangulations of polytopes*. In J. E. Goodman and J. Rourke, editors, Handbook of discrete and computational geometry, pages

383406. CRC Press, New York, 2004.

F. R. Hampel. *Design, modelling and analysis of some biological datasets*. In C. L. Mallows, editor, *Design, data and analysis: By some friends of Cuthbert Daniel*. Wiley, New Jersey, 1987.

W. Hardle and B. Ronz, editors, *COMPSTAT 2002 Proceedings in Computational Statistics*, pages 575580, Heidelberg, 2002. Physica-Verlag.

S. Mueller and K. Rufibach. Smooth tail index estimation. *Journal of Statistical Computation and Simulation*, 79(9):1155–1167, 2009.

S. Mueller and K. Rufibach. On the max-domain of attraction of distributions with log-concave densities. *Statistics and Probability Letters*, 78(12):14401444, 2007.

B. L. S. Prakasa Rao. *Estimation of a unimodal density*. *Sankhya Series A*, 31:2336, 1969.

M. Rosenblatt. *Remarks on some nonparametric estimates of a density function*. *Annals of Mathematical Statistics*, 27:832837, 1956.

K. Rufibach. *Log-concave density estimation and bump hunting for iid observations*. PhD thesis, University of Bern, 2006.

K. Rufibach. *Computing maximum likelihood estimators of a log-concave density function*. *Journal of Statistical Computation and Simulation*, 77:561574, 2007.

K. Rufibach and L. Duembgen. *logcondens: Estimate a Log-Concave Probability Density from i.i.d. Observations*, 2006. URL <http://CRAN.R-project.org/package=logcondens>. R package version 1.3.2.

D.W. Scott. *On optimal and data-based histograms*. *Biometrika*, 66:60510, 1979.

B. W. Silverman. *Density estimation for statistics and data analysis*. Chapman and Hall, London, 1986.

J. R. Thompson and R. A. Tapia. *Nonparametric function estimation, modeling and simulation*. Society for Industrial and Applied Mathematics, Philadelphia, 1990.

G. Walther. *Detecting the presence of mixing with multiscale maximum like-*

lihood. Journal of the American Statistical Association, 97(458):508513, 2002.

G. Walther. *Inference and modeling with log-concave distributions*. Unpublished manuscript, 2008.

X. Wang, M. Woodroffe, M. Walker, M. Mateo, and E. Olzewski. *Estimating dark matter distributions*. Astrophysics Journal, 626:245158, 2005.

W. H. Wong and X. Shen. *Probability inequalities for likelihood ratios and convergence rates of sieve MLEs*. Annals of Statistics, 23(2):339362, 1995.