Computing confidence intervals for log-concave densities

Mahdis Azadbakhsh, Hanna Jankowski, Xin Gao

Department of Mathematics and Statistics, York University, Toronto, Canada

Abstract

In Balabdaoui, Rufibach, and Wellner (Annals of Statistics, 37, pages 1299-1331, 2009), pointwise asymptotic theory was developed for the nonparametric maximum likelihood estimator of a log-concave density. Here, the practical aspects of their results are explored. Namely, the theory is used to develop pointwise confidence intervals for the true log-concave density. To do this, the quantiles of the limiting process are estimated and various ways of estimating the nuisance parameter appearing in the limit are studied. The finite sample size behaviour of these estimated confidence intervals is then studied via a simulation study of the empirical coverage probabilities.

Keywords: nonparametric density estimation, log-concave, maximum likelihood, confidence interval

1. Introduction

The nonparametric maximum likelihood estimator (MLE) of a log-concave density has received much attention in the statistics literature of late. It has been studied, for example, in Walther (2002); Dümbgen and Rufibach (2009); Dümbgen and Rufibach (2011); Chang and Walther (2007); Chen and Samworth (2013); Cule et al. (2010); Cule and Samworth (2010). For an overview, we recommend the review article of Walther (2009). The appeal of this estimator is that, unlike a kernel-density approach, it does not require a choice of bandwidth. Indeed, the log-concave MLE is not only fully automatic, but also automatically locally adaptive. Furthermore, the simulations in Chen and Samworth (2013, page 12-13) show that the log-concave MLE outperforms the kernel-density estimator for larger sample sizes, when the true density is log-concave. For smaller sample sizes, an (automatic) smoothed version of the MLE continues to have improved

performance over the kernel-density estimator. (Chen and Samworth (2013) consider the density on $\mathbb{R}^d$ with $d = 2, 3$.)

Here, we focus on the MLE of a log-concave density on $\mathbb{R}$. That is, let f_0 denote a log-concave density on $\mathbb{R}$ and suppose that we observe $X_1, \ldots, X_n$ independent and identically distributed samples from f_0. Let $\mathcal{F}$ denote the class of log-concave densities on $\mathbb{R}$. Then the nonparametric MLE of a log-concave density on $\mathbb{R}$ is defined as

$$\widehat{f_n} \;=\; \text{argmax}_{f \in \mathcal{F}} \sum_{i=1}^{n} \log f(X_i).$$

Dümbgen and Rufibach (2009) show that this estimator exists and is unique, and also study its consistency. Additional results on consistency can also be found in Pal et al. (2007); Cule and Samworth (2010). The estimator may be calculated using the active set algorithm, and this has been implemented in the R package **logcondens** (Dümbgen and Rufibach, 2006; Dümbgen and Rufibach, 2011). Pointwise asymptotic theory for $\widehat{f_n}$ was developed in Balabdaoui et al. (2009).

Suppose that the true density f_0 is log-concave with $f_0(x_0) > 0$ and $\varphi_0 = \log f_0$ is twice continuously differentiable in a neighbourhood of x_0 with $\varphi_0''(x_0) \neq 0$. One of the main results of Balabdaoui et al. (2009) is that

$$n^{2/5}\left(\widehat{f_n}(x_0) - f_0(x_0)\right) \;\Rightarrow\; \left(\frac{f_0^3(x_0)|\varphi_0^{(2)}(x_0)|}{4!}\right)^{1/5} \mathbb{C}(0), \tag{1}$$

where the distribution of $\mathbb{C}(0)$ is known (here, we describe it in Section 2). For a fixed f_0, define

$$c_2(x) \;=\; \left(\frac{f_0^3(x)|\varphi_0^{(2)}(x)|}{4!}\right)^{1/5}. \tag{2}$$

If $c_2(x)$ is known, and q_α denotes the quantile such that $P(\mathbb{C}(0) \leq q_\alpha) = \alpha$, then the result in (1) implies that

$$\left(\widehat{f_n}(x_0) - \frac{c_2(x_0)}{n^{2/5}}\, q_{1-\alpha/2},\; \widehat{f_n}(x_0) - \frac{c_2(x_0)}{n^{2/5}}\, q_{\alpha/2}\right) \tag{3}$$

forms an asymptotically correct $100(1-\alpha)\%$ confidence interval for $f_0(x_0)$. The main goal of this paper is to provide estimators for the quantiles q_α and $c_2(x_0)$ so

that the confidence intervals (3) may be implemented in practice, and to assess the quality of this procedure.

In Section 2 we describe the process $\mathbb{C}(0)$ and provide its quantile estimates based on simulations (the simulations are detailed in Appendix A). In Section 3 we consider estimation of the constant c_2, and in Section 4 we use simulations to understand the empirical performance of the estimated confidence intervals (3). The methods presented here have been implemented in the R package **logcondens** (Dümbgen and Rufibach, 2006).

2. Quantiles of the limiting process

Let $\mathbb{B}(t)$, $t \in \mathbb{R}$ denote a two-sided Brownian motion. That is, $\mathbb{B}(t) = \mathbb{B}_1(t), t \geq 0$ and $\mathbb{B}(t) = \mathbb{B}_2(-t), t \leq 0$, where $\mathbb{B}_1, \mathbb{B}_2$ are two independent Brownian motions with $\mathbb{B}_1(0) = \mathbb{B}_2(0) = 0$. Let

$$\mathbb{Y}(t) = \begin{cases} \int_0^t \mathbb{B}(s)ds - t^4, & t \geq 0 \\ \int_t^0 \mathbb{B}(s)ds - t^4, & t < 0, \end{cases}$$

and let $\mathbb{H}$ be the almost surely unique process such that

1. $\mathbb{H}(t) \leq \mathbb{Y}(t)$ for all $t \in \mathbb{R}$,
2. $\mathbb{H}''(t)$ is concave,
3. $\mathbb{H}(t) = \mathbb{Y}(t)$ if the slope of $\mathbb{H}''(t)$ is strictly decreasing at t.

The process $\mathbb{H}$ thus defined exists and is unique (Balabdaoui et al., 2009, Theorem 2.1). Let $\mathbb{C}(t) = \mathbb{H}''(t)$ for all $t \in \mathbb{R}$, then the quantity of interest, $\mathbb{C}(0)$, is simply $\mathbb{C}(t)$ evaluated at $t = 0$.

The process $\mathbb{H}$, or rather its close relative, was first shown to exist in Groeneboom et al. (2001a), and we refer to Appendix A for further details. Using their approach, one could show that $\mathbb{C}(t) = \lim_{m\to\infty} \mathbb{C}_m(t)$, where $\mathbb{C}_m$ is defined as:

$$\mathbb{C}_m \quad = \quad \operatorname{argmin}_{\varphi \in C_m} \left\{ \int_{-m}^{m} \varphi^2(t)dt - 2 \int_{-m}^{m} \varphi(t) d\left(\mathbb{B}(t) - 4t^3\right) \right\},$$

where C_m denotes the class of concave functions with the restriction that $\varphi(-m) = \varphi(m) = -12m^2$. Thus, we can think of $\mathbb{C}(t)$ as the concave regression on the function $-12t^2$ plus white noise.

Our interest here is limited to the value of $\mathbb{C}(t)$ at $t = 0$, and below we present some observed properties based on $n = 100\,000$ independent samples. Details on the algorithm used to generate these samples is given in Appendix A. Figure 1

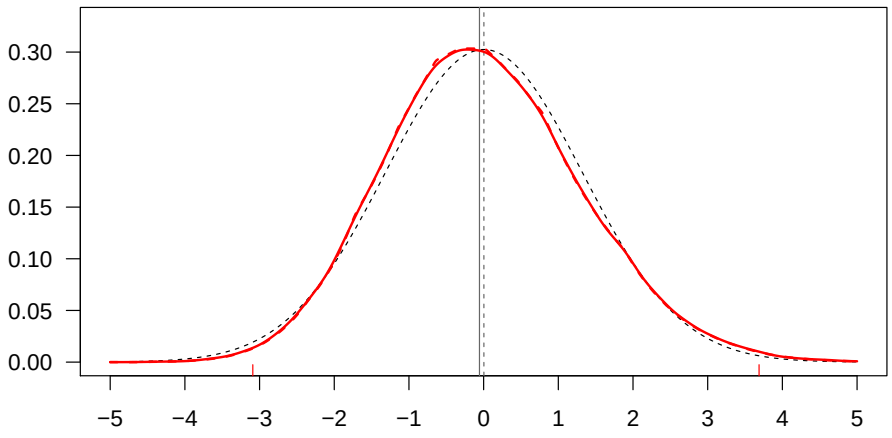

Figure 1: Estimates of $f_{\mathbb{C}(0)}$: log-concave MLE (dashed red line) and kernel density estimator (solid red line) from a sample size of $n = 100\,000$. The dashed black line is the Gaussian density with mean zero and variance equal to that of the data, and has been added for reference. The vertical lines denote the observed mean (dashed line) and median (solid line) of the data. The observed range of the data was $(-4.55, 6.49)$, while 99% of the data fell into the interval $(-3.09, 3.69)$, the latter are marked on the plot in red.

shows the estimate of the density of $\mathbb{C}(0)$. Visually, two things are immediately striking: first, the density appears to be asymmetric (right-skewed), and second, the density appears to be log-concave. We address these two questions in Sections 2.1 and 2.2.

Table 1: Moment estimates for $\mathbb{C}(0)$

	estimate	error
mean	0.0036	(0.0042)
median	-0.0557	(0.0029)
variance	1.7395	(0.0078)
$E[\mathbb{C}^2(0)]$	1.7395	(0.0079)
$E[\mathbb{C}^3(0)]$	0.599	(0.0294)
$E[\mathbb{C}^4(0)]$	9.3086	(0.1053)
$E[\mathbb{C}^5(0)]$	9.9032	(0.4689)

Moment (plus median) estimates of $\mathbb{C}(0)$ are given in Table 1 while quantile estimates are given in Table 2. Table 2 gives four different quantile estimates based on

(A). the empirical distribution function,
(B). the kernel density estimate,
(C). the log-concave MLE,
(D). the normal approximation.

The last column of the table gives standard errors of the values in column (A), (see Shorack and Wellner, 1986, Example 1, page 639). The Gaussian approximation is given for reference only. Our simulations indicate that $E[\mathbb{C}(0)] = 0$, as expected.

Table 2: Estimated values of $F^{-1}_{\mathbb{C}(0)}(p)$

p	(A)	(B)	(C)	(D)	error
0.001	-3.6442	-3.6569	-3.6349	-4.0722	(0.0017)
0.005	-3.0905	-3.1108	-3.0878	-3.3937	(0.0019)
0.010	-2.8172	-2.8358	-2.8178	-3.0647	(0.0020)
0.025	-2.4157	-2.4327	-2.4157	-2.5815	(0.0021)
0.050	-2.0574	-2.0737	-2.0617	-2.1659	(0.0023)
0.100	-1.6440	-1.6537	-1.6457	-1.6867	(0.0024)
0.200	-1.1184	-1.1236	-1.1186	-1.1065	(0.0026)
0.500	-0.0557	-0.0565	-0.0575	0.0036	(0.0029)
0.800	1.0948	1.0986	1.0916	1.1136	(0.0029)
0.900	1.7421	1.7467	1.7387	1.6938	(0.0027)
0.950	2.2653	2.2767	2.2657	2.1730	(0.0026)
0.975	2.7536	2.7618	2.7498	2.5886	(0.0025)
0.990	3.3193	3.3328	3.3188	3.0718	(0.0024)
0.995	3.6881	3.7059	3.6959	3.4009	(0.0022)
0.999	4.5140	4.5270	4.4939	4.0793	(0.0020)

2.1. Is the density of $\mathbb{C}(0)$ *log-concave?*

Let $\widehat{g}_n$ denote the nonparametric MLE of a decreasing density on $\mathbb{R}_+$, that is, the Grenander estimator. If the true decreasing density, g_0, satisfies $g_0(x_0) > 0$

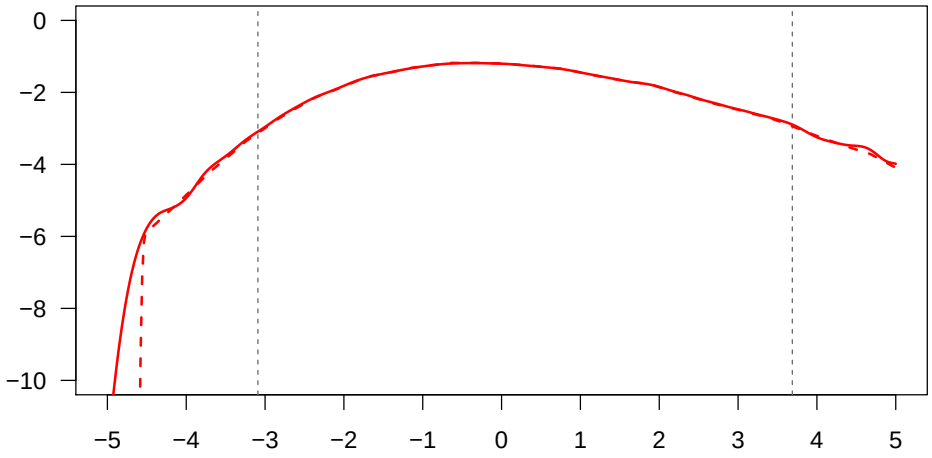

Figure 2: Estimates of $\log(f_{\mathbb{C}(0)})(x) + 0.5\, c_0\, x^2$: log-concave MLE (dashed) and kernel density estimator (solid) from a sample size of $n = 100\,000$ with $c_0 = 0.25$. The vertical line denotes the mean of the data. The observed range of the data was $(-4.55, 6.49)$, while 99% of the data fell into the interval $(-3.09, 3.69)$, the latter are marked on the plot as dashed vertical lines.

and $g'(x_0) < 0$, then it was established in Prakasa Rao (1969) (see also Groeneboom (1985)) that

$$n^{1/3}\,(\widehat{g}_n(x_0) - g_0(x_0)) \;\;\Rightarrow\;\; 2\left(\frac{g_0(x_0)|g_0'(x_0)|}{2}\right)^{1/3} \mathbb{Z},$$

where the density of $\mathbb{Z}$ is known as Chernoff's distribution. Moreover, one can show that $\mathbb{Z} = \operatorname{argmax}\{\mathbb{B}(t) - t^2\}$. Chernoff's distribution arises as the limit in many monotone problems, including estimation of the mode (Chernoff, 1964; Balabdaoui and Wellner, 2012), and the form of the density of $\mathbb{Z}$ was determined in Groeneboom (1989). Computation of $f_{\mathbb{Z}}$ was considered in Groeneboom and Wellner (2001), where exact quantiles were also calculated. More recently, Balabdaoui and Wellner (2012) established that Chernoff's density is log-concave, and conjecture that it is also strongly log-concave. A density f is strongly log-concave if it satisfies $(-\log f)''(x) \geq \delta > 0$ for all x in the support of f. Equivalently, f is strongly log-concave if $(\log f)(x) + c_0\, x^2/2$ is concave for some $c_0 > 0$.

Figure 1 shows kernel density and log-concave MLE estimates of the density of $\mathbb{C}(0)$, and the two appear quite close. The Gaussian density with mean and variance equal to that of the data has also been added for reference in the plot. To

test if the density of $\mathbb{C}(0)$ is log-concave we preformed the "trace test" developed in Chen and Samworth (2013). The test attained a p-value of 1.00 (to minimize computational time we chose $B = 100$). Furthermore, in Figure 2 we show plots of $\log \widehat{f_n} + x^2/8$ (i.e. $c_0 = 0.25$) where $\widehat{f_n}$ is either the kernel density estimate or the log-concave MLE. The result appears to be concave, except in the far tails, and therefore it seems plausible that the density $f_{\mathbb{C}(0)}$ is also strongly log-concave.

2.2. Is the density of $\mathbb{C}(0)$ *symmetric about zero?*

It is well-known that Chernoff's density is symmetric, see e.g. Groeneboom and Wellner (2001, page 395). One might think, at first, that $\mathbb{C}(0)$ is also symmetric about zero, and indeed this was stated in Balabdaoui et al. (2009, page 1307) and was also conjectured by the authors of this work. However, the assumption of symmetry appears to be quite clearly violated by the fitted densities, see Figure 1. A comparison of the mean and median in Table 1 (along with their standard errors) appears to affirm this statement. The following result shows that a different version of symmetry does hold for both $\mathbb{H}(t)$ and $\mathbb{C}(t)$.

Proposition 2.1. *The processes* $\mathbb{H}(t)$ *and* $\mathbb{C}(t)$ *are symmetric about zero in the sense that* $\mathbb{H}(t) \stackrel{d}{=} \mathbb{H}(-t)$, $\mathbb{H}'(t) \stackrel{d}{=} -\mathbb{H}(-t)$, *and* $\mathbb{C}(t) \stackrel{d}{=} \mathbb{C}(-t)$. *It follows that the density of* $\mathbb{H}'(0)$ *is symmetric about zero.*

To study the problem further, we consider the (simpler) regression:

$$\begin{aligned}
\mathbb{G}_1 &= \operatorname{argmin}_g \left\{ \sum_{k=-5}^{5} \left(g(k) - (k^2 + \varepsilon_k) \right)^2 ; g \text{ convex} \right\} \\
\mathbb{G}_2 &= \operatorname{argmin}_g \left\{ \sum_{k=-5}^{5} \left(g(k) - (-k^2 + \varepsilon_k) \right)^2 ; g \text{ concave} \right\},
\end{aligned}$$

where ε_k, $k = -5, \ldots, 5$ are IID standard normal. Using the function `conreg` in the package **cobs**, we simulate $n = 100\,000$ samples of $\mathbb{G}_1(0)$ and $\mathbb{G}_2(0)$, and the resulting densities are shown in Figure 4. (Note that our sample size is defined as follows: in each sample we observe the *entire* function $\varepsilon_{-5}, \ldots, \varepsilon_5$ exactly once.) The observed moments of both densities were $\widehat{\mu}_{\mathbb{G}_1(0)} = -0.0013, \widehat{\mu}_{\mathbb{G}_2(0)} = -0.0021, \widehat{\sigma}_{\mathbb{G}_1(0)} = 0.839, \widehat{\sigma}_{\mathbb{G}_2(0)} = 0.838$. The median of $\mathbb{G}_1(0)$ (resp. $\mathbb{G}_2(0)$) was 0.0445 (resp. -0.0472). The confidence interval for $\mu_{\mathbb{G}_1(0)}$ (resp. $\mu_{\mathbb{G}_2(0)}$) is $(-0.007, 0.004)$ (resp. $(-0.007, 0.003)$).

The plots and estimates above indicate that $\mathbb{G}_1(0)$ is slightly left-skewed, while $\mathbb{G}_2(0)$ is slightly right-skewed. In Figure 3, 25 samples from $\mathbb{G}_1$ are shown, along

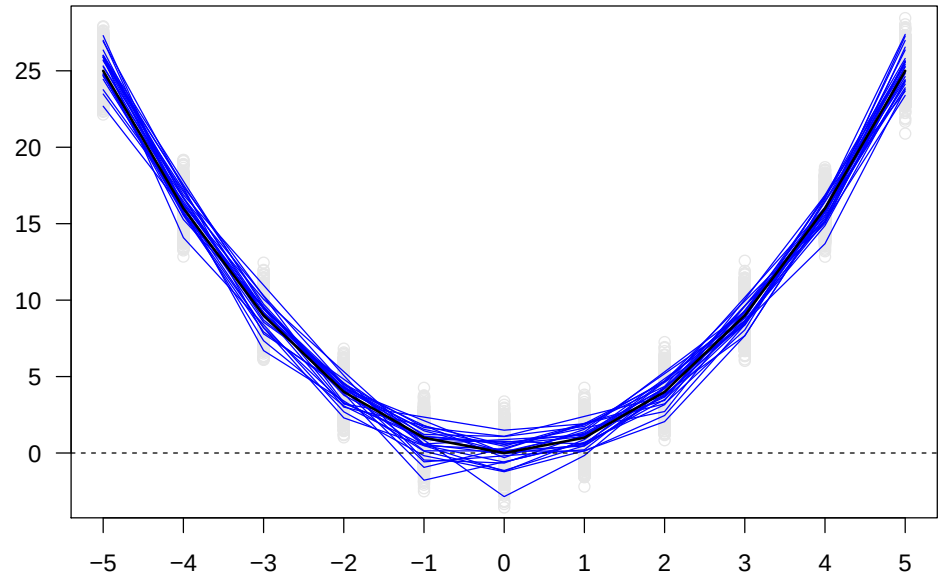

Figure 3: 25 samples of $\mathbb{G}_1$ (blue), the function k^2 is shown in black along with samples of the errors (light gray).

with the underlying function $k^2 + \varepsilon_k$. Since $\mathbb{G}_1$ must be convex, the values at $\mathbb{G}_1(0)$ sometimes get "pulled down" to accommodate the overall fit. The reverse occurs for $\mathbb{G}_2$, which must be concave.

3. Estimating the constant $c_2(x_0)$

To estimate the constant c_2, we need to estimate the second derivative of $\varphi_0 = \log f_0$. An estimator of φ_0 is provided directly by the MLE: $\widehat{\varphi}_n = \log \widehat{f}_n$. However, this is piecewise linear and therefore the second derivative is equal either to zero or to $-\infty$. Therefore, an alternative approach is required. Here, we consider three main methods. The first two are based on the fact that

$$\varphi_0''(x) \quad = \quad \frac{f_0(x) f_0''(x) - \left(f_0'(x)\right)^2}{f_0^2(x)},$$

which allows us to write

$$c_2(x) \quad = \quad \left(\frac{f_0(x)|\psi_2(x)|}{4!}\right)^{1/5}, \tag{4}$$

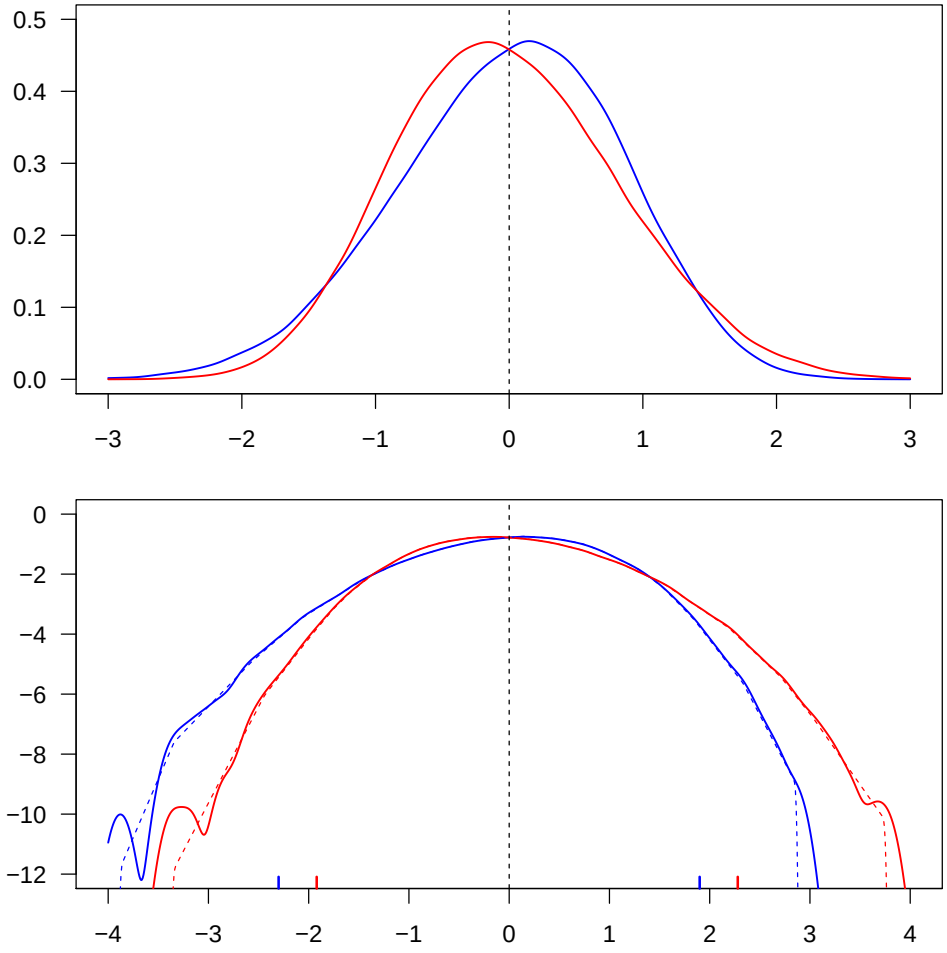

Figure 4: Kernel density estimates of the densities of $\mathbb{G}_1(0)$ (convex, blue) and $\mathbb{G}_2(0)$ (concave, red) shown in the top plot and on the log-densities are shown in the bottom plot. The dashed lines correspond to the log-concave MLE estimates. 99% of the convex (resp. concave) data fell into the region $(-2.30, 1.90)$ (resp. $(-1.92, 2.28)$) which is marked in blue (resp. red) in the plot.

where $\psi_2(x) = f_0(x)f_0''(x) - \left(f_0'(x)\right)^2$.

The three methods which we consider are:

1. Smoothed MLE approach: The smoothed log-concave MLE was introduced in Dümbgen and Rufibach (2009) and is defined as

$$\widetilde{f}_n(x) \quad = \quad \int \phi_{\widehat{\gamma}_n}(x-y)\widehat{f}_n(y)dy, \tag{5}$$

   where ϕ_γ denotes the normal density with mean zero and standard deviation γ. The value $\widehat{\gamma}_n$ is defined by $\widehat{\gamma}_n^2 = \widehat{\sigma}_n^2 - \mathrm{var}_{\widehat{f}_n}(X)$, where $\widehat{\sigma}_n^2$ is the observed variance of the data and $\mathrm{var}_{\widehat{f}_n}(X)$ is the variance of the random variable X with density $\widehat{f}_n$. It is known that $\widehat{\gamma}_n^2 > 0$ and with this choice $\mathrm{var}_{\widetilde{f}_n}(X) = \widehat{\sigma}_n^2$. It was shown in Chen and Samworth (2013) that $\widetilde{f}_n$ is also a consistent estimator of f_0. An exact calculation of $\widetilde{f}_n$ is provided in the R package **logcondens** (Dümbgen and Rufibach, 2006; Dümbgen and Rufibach, 2011). Unlike the MLE, the smoothed estimator is directly differentiable, and our first approach is to use plug in estimates $\widetilde{f}_n$, $\widetilde{f}_n'$ and $\widetilde{f}_n''$ in estimating $c_2(x)$. We denote this estimator as $\widehat{c}_{n,2}^{smle}(x)$.

2. Kernel-density estimation: Here, we use kernel density estimators of the zero, first, and second derivatives, $\overline{f}_n$, $\overline{f}_n'$ and $\overline{f}_n''$ and plug these in directly into $c_2(x)$ to obtain the estimate. Various methods exist to obtain the bandwidth, and we refer the reader to Härdle (1991, Chapter 4) or Härdle et al. (2004, Section 3.3). In our implementation, we used the R package **ks** (Duong, 2012, 2007).

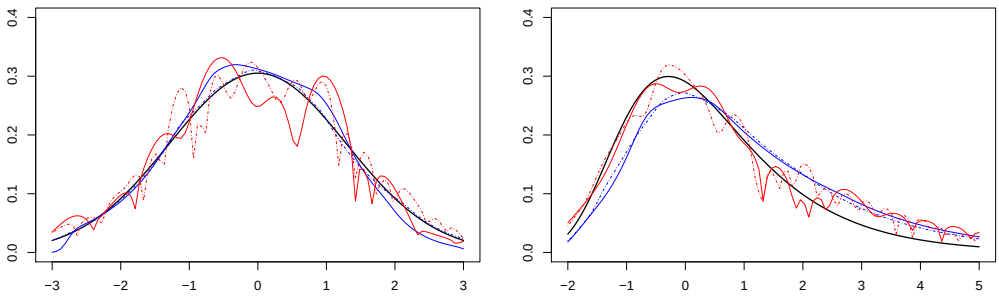

Figure 5: Two examples of c_2 estimation. The estimator $\widehat{c}_{n,2}^{nrd}$ is shown in blue and the estimator $\widehat{c}_{n,2}^{kde}$ with bandwidth selected using `hns` is shown in red. The sample sized are $n = 1000$ (solid lines) and $n = 10000$ (dashed lines). The true density is standard normal on the left and standard Gumbel on the right. The true c_2 is shown in black.

We denote this estimator as $\widehat{c}_{n,2}^{kde}(x)$. With this package, the available bandwidth selection methods are

(i) `hscv`: bandwidth selected using smoothed cross-validation as introduced in Jones et al. (1991),

(ii) `hlscv`: bandwidth selected using least squares cross-validation (Bowman, 1984; Rudemo, 1982),

(iii) `hpi`: the plug-in bandwidth selector of Wand and Jones (1994),

(iv) `hns`: asymptotically optimal bandwidth assuming a normal density (Härdle et al., 2004, page 52).

3. Normal reference distribution: Lastly, we considered the approach of estimating φ_0'' by assuming that the true density is Gaussian. In this case, we find that $\varphi_0''(x) = -\sigma^{-2}$, which is straightforward to implement. That is, in this case we estimate c_2 as

$$\widehat{c}_{n,2}^{nrd}(x) \quad = \quad \left(\frac{\widetilde{f}_n^3(x)}{4!\ \widehat{\sigma}_n^2} \right)^{1/5}.$$

In addition to these methods, we tried adhoc and cross-validation methods of choosing γ in the smoothed MLE approach above. However, these did not perform as well and/or were too computationally intensive to report here. In Figure 6 we show box plots of the squared L_2 error of the various estimators $\widehat{c}_{n,2}$ of the true c_2. That is, we report values of

$$d_2(\widehat{c}_{n,2}, c_2) \quad = \quad \int_I (\widehat{c}_{n,2}(x) - c_2(x))^2 dx, \tag{6}$$

which is approximated using a Riemann sum with $m = 500$ summands. We considered two samples sizes, $n = 100$ and $n = 1000$, and the following densities.

(a). the standard normal density with $I = [-3, 3]$,

(b). the gamma density with shape parameter $\alpha = 3$ and rate parameter $\lambda = 1$ with $I = [0, 8]$,

(d). the beta density with parameters $\alpha = \beta = 3$ with $I = [0, 1]$,

(d). the standard Gumbel density with $I = [-2, 5]$.

In Figure 6, the estimator $\widehat{c}_{2,n}^{smle}$ is denoted as `DR`, $\widehat{c}_{2,n}^{nrd}$ is denoted as `nrd`, and the estimator $\widehat{c}_{2,n}^{kde}$ is denoted by its bandwidth selection method.

Of the three estimators of c_2 we consider, only the kernel density estimator is known to be consistent (Baird, 2012). The simple normal reference estimator is

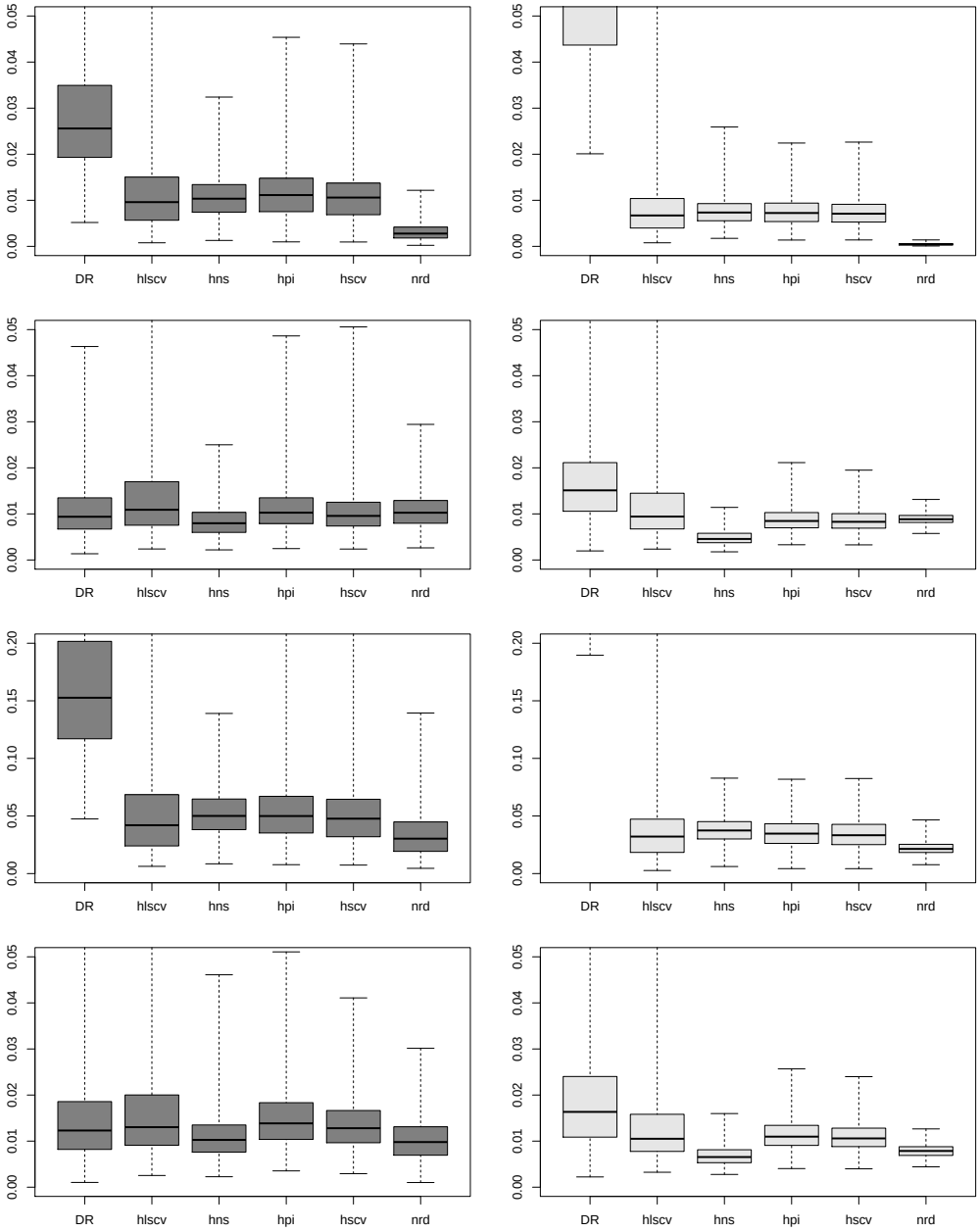

Figure 6: Monte Carlo comparison of the various estimators of c_2. From top to bottom the distributions are (a)-(d), as indicated in the text. The sample size is $n = 100$ (dark grey) and $n = 1000$ (light grey). The error is measured as in (6) with a Monte Carlo sample of $B = 1000$.

also consistent, but only when the true density is Gaussian. The smoothed MLE estimator of Dümbgen and Rufibach (2009) is a consistent estimator of the true density f_0 (Chen and Samworth, 2013), however, simulations from Figure 6 indicate that $\widehat{c}_{n,2}^{smle}$ is not a consistent estimator of c_2 in general. From Figure 6 we see that the the estimator $\widehat{c}_{n,2}^{nrd}$ has by far the best performance for the normal density, as expected. Although it is not a consistent estimator for the other densities, it also performs relatively well, probably because of its relative smoothness (see Figure 5). For even larger sample sizes we would expect that the bias will take over and the estimator $\widehat{c}_{n,2}^{nrd}$ will begin to diverge for non-normal densities.

4. A simulation study of confidence intervals

Finally, we use the estimated quantiles from Section 2 and c_2 estimators from Section 3 to estimate the confidence intervals (3). The performance of these is tested via simulations of empirical coverage probabilities.

We consider two sample sizes: $n = 100$ and $n = 1000$, and the same four densities as in Section 3. In Figures 7 and 8, the densities are (from top to bottom in both figures):

(a). the standard normal density
(b). the gamma density with shape $\alpha = 3$ and rate $\lambda = 1$
(d). the beta density with parameters $\alpha = \beta = 3$
(d). the standard Gumbel density.

All results reported are based on a Monte Carlo sample of $B = 1000$. We chose the most popular confidence interval with coverage of 95%, and the results are reported in Figure 7 for $n = 100$ and in Figure 8 for $n = 1000$. Each figure shows the observed empirical coverage for the following methods:

(i). "oracle" confidence interval with the true c_2 used in (3). This is included for comparison, and is shown in black in all plots.
(ii). confidence interval with c_2 in (3) estimated by $\widehat{c}_{n,2}^{smle}$. This is shown in light gray in all plots, and performed so poorly that it is sometimes not visible.
(iii). confidence interval with c_2 in (3) estimated by $\widehat{c}_{n,2}^{kde}$ with bandwidth estimated by `hns`. This is shown in blue in all plots.
(iv). confidence interval with c_2 in (3) estimated by $\widehat{c}_{n,2}^{nrd}$. This is shown in red in all plots.
(v). ECDF-bootstrap percentile confidence interval (see Remark 4.1 for details). This is shown in green in all plots, and is based on re-sampling from the empirical distribution.

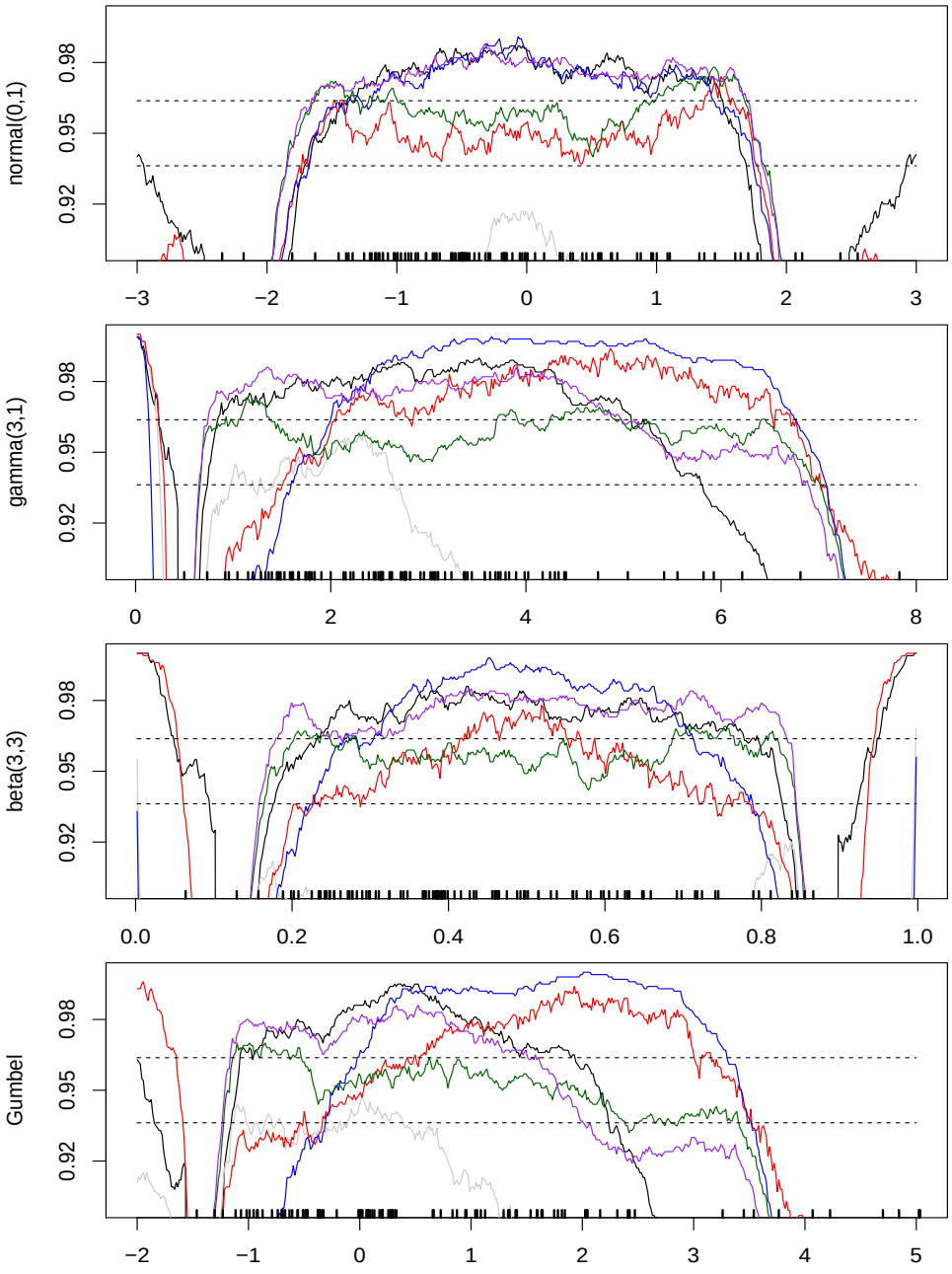

Figure 7: Monte Carlo comparison of 95% confidence intervals for $n = 100$. Four methods using the asymptotic results are shown as follows: red (normal reference method), grey (DR bandwidth), blue (kernel density approach), and black (oracle c_2). The two bootstrap methods are shown in green (ECDF) and purple (NPML). For reference, each plot also shows a sample of size $n = 100$ from the sampling distribution.

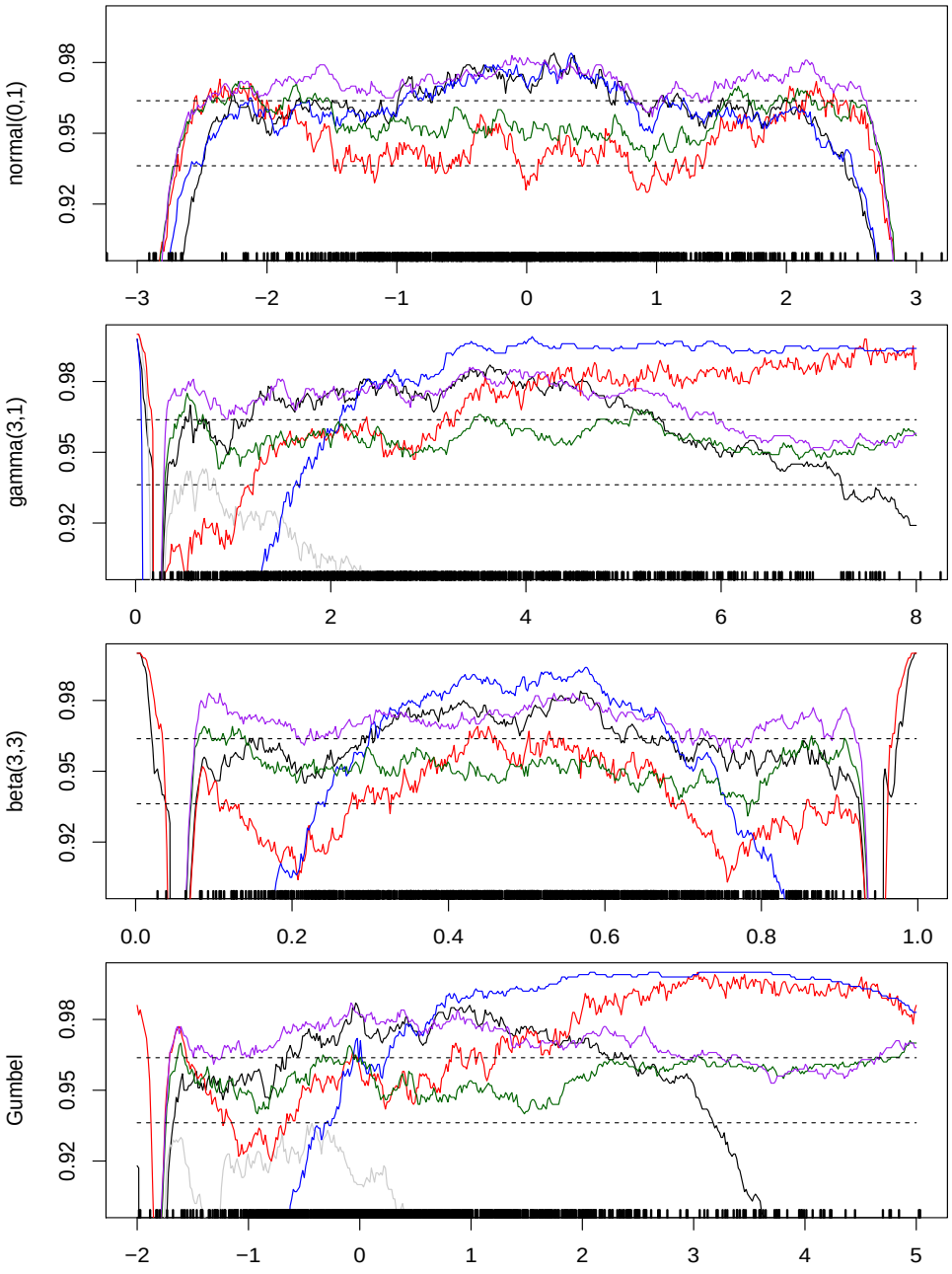

Figure 8: Monte Carlo comparison of 95% confidence intervals for $n = 1000$. Four methods using the asymptotic results are shown as follows: red (normal reference method), grey (DR bandwidth), blue (kernel density approach), and black (oracle c_2). The two bootstrap methods are shown in green (ECDF) and purple (NPML). For reference, each plot also shows a sample of size $n = 1000$ from the sampling distribution.

(vi). NPML-bootstrap percentile confidence interval (see Remark 4.1 for details). This is shown in purple in all plots, and is based on re-sampling from the fitted nonparametric MLE.

In each plot, the horizontal lines indicate the bounds $0.95 \pm 2 * \sqrt{0.95 * 0.05/B}$ where $B = 1000$ is the size of the Monte Carlo sample. Each plot also shows a sample data set of the same sample size as in the simulations (e.g., the top panel in Figure 7 also shows a sample data set of size $n = 100$ from a standard Gaussian distribution). This is done for reference.

Remark 4.1 (Details of bootstrap methods used)**.** *Let $X_1, \ldots, X_n$ denote an IID sample from a density f_0, and let $\widehat{f_n}$ denote its log-concave MLE. For $j = 1, \ldots, \kappa = 500$, let $X^*_{1j}, \ldots, X^*_{nj}$ denote an IID sample from either the empirical distribution $\mathbb{F}_n$ (for the ECDF-bootstrap in (v)), or the fitted density $\widehat{f_n}$ (for the NPML-bootstrap in (vi)). For a fixed x_0, let $\widehat{f^*_{(\kappa\alpha)}}(x_0)$ denote the α quantile of the bootstrap sample. The 95% confidence intervals reported are calculated as $(\widehat{f^*_{(0.025\kappa)}}(x_0), \widehat{f^*_{(0.975\kappa)}}(x_0))$, for each x_0. This is called the "bootstrap percentile interval" in Wasserman (2006, page 34). Notably, we also tried the "bootstrap pivotal confidence interval" (Wasserman, 2006, page 33), calculated as*

$$(2\widehat{f_n}(x_0) - \widehat{f^*_{(0.975\kappa)}}(x_0), 2\widehat{f_n}(x_0) - \widehat{f^*_{(0.025\kappa)}}(x_0)),$$

but it did not perform well: The mean empirical coverage varied between 70% and 90%, and often the results would not be visible in the figure plots. We therefore do not report these results in Figures 7 and 8.

Upon examining the results in Figures 7 and 8, it is immediately clear that the method using the $\widehat{c}^{smle}_{n,2}$ (shown in grey) does not work well. Of the other methods based on the asymptotic theory, (iii) & (iv), it seems that $\widehat{c}^{nrd}_{n,2}$ (shown in red) performs best, although the quality of its performance is far from uniform. The NPML-bootstrap percentile interval (shown in purple) appears to perform most closely to the oracle c_2 method (shown in black). The ECDF-bootstrap percentile interval (shown in green) appears to perform best of all. All methods are sensitive to the sample size - that is, they work well only in a likely range for the data (e.g. consider the first panel of Figure 7 - here, the methods work well only for a range of about $(-1.75, 1.75)$). This behaviour is to be expected. Notably, the ECDF-bootstrap percentile interval is least sensitive in this regard. We note also that the points where $\varphi''(x_0) = 0$ (i.e. the modes) will not satisfy the conditions required for (1) to hold.

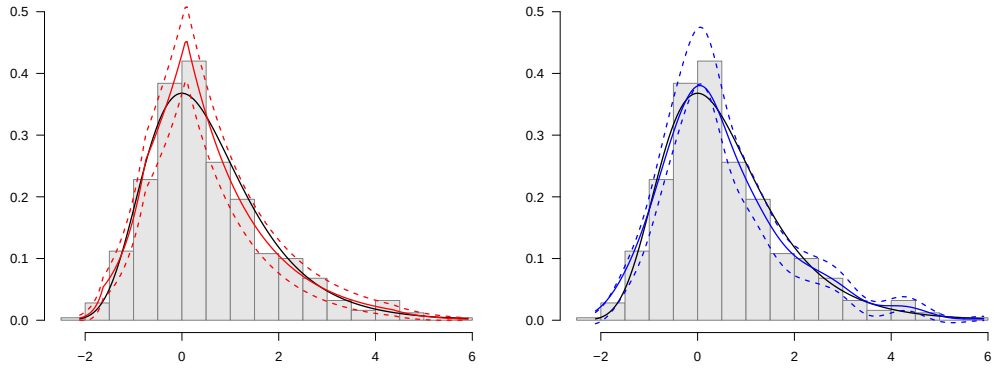

Figure 9: 95% confidence intervals based on maximum likelihood (left, using method (iv)) and kernel density (right) estimators for a Gumbel data set with sample size $n = 500$.

5. Examples

5.1. A simulated example

To compare the log-concave MLE to the kernel density estimator, we simulated a data set of size $n = 500$ from a standard Gumbel density. The results (along with 95% confidence intervals) are shown in Figure 9. The confidence interval based on the kernel density estimator was calculated as in Härdle et al. (2004, Section 2.3, pages 61-62) for a bandwidth calculated based on the normal reference distribution `hns`, as this satisfies $h = cn^{-1/5}$ for some constant c which depends on the data. The confidence interval for the log-concave MLE is calculated based on method (iv) as described on page 13. Both confidence intervals have width proportional to $n^{-2/5}$, and visually the calculated confidence intervals appear quite similar. However, the confidence interval based on the kernel density estimator is not centred at the estimator, due to the bias correction term.

5.2. A reliability data set

In Dümbgen and Rufibach (2009), the authors use the log-concave MLE to analyze the reliability of a certain device. This data set has been made available as part of the package **logcondens**. In Figure 10 we show the estimator $\widehat{f_n}$, along with a 95% confidence interval based on methods (iv) and (v) as described on page 13. Method (iv) uses the asymptotic theory, while method (v) is the ECDF-bootstrap. Visually, the confidence interval provides much needed information on the variability of $\widehat{f_n}$.

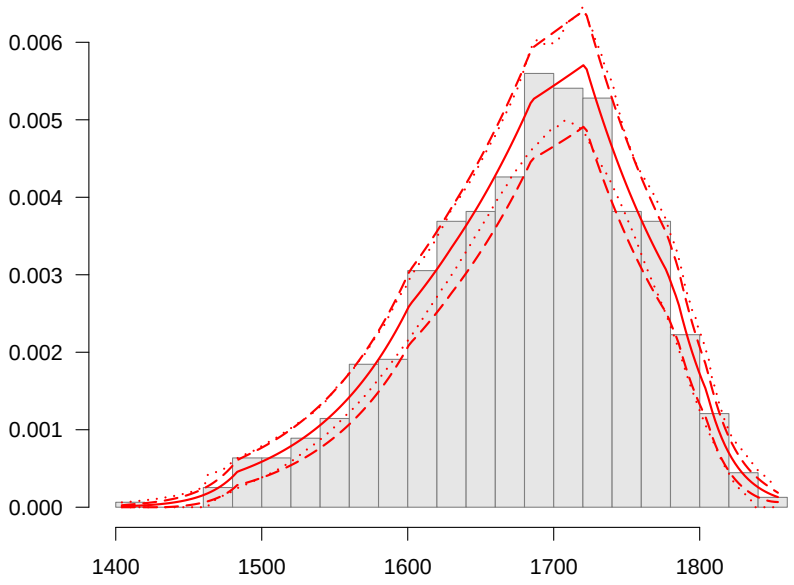

Figure 10: 95% confidence intervals in the reliability data set: method (iv) based on asymptotics is shown as the dashed line and the ECDF-bootstrap, method (v), is shown as the dotted line; the MLE is shown as the solid line.

6. Discussion

The confidence intervals we present here are confidence intervals for the true density at a point, and assume that the true density satisfies the conditions of the result in Balabdaoui et al. (2009). The limiting distribution at a point x_0 depends on the location x_0 only in the constant $c_2(x_0)$. The latter depends on the second derivative of $\log f_0(x_0)$ and is quite difficult to estimate. Also, it is quite likely that for a fixed sample size n, the quality of the approximation of the distribution by the asymptotic distribution depends on the point x_0. Indeed, even for a fixed x_0, the quality of this approximation is not well-understood. All of these factors combine to make the problem of inference for a log-concave density based on (1) relatively difficult.

A similar problem was investigated in Banerjee and Wellner (2005) for the current status model with censoring. Banerjee and Wellner (2005) compare two methods of calculating confidence intervals in that setting: one based on the likelihood ratio test of Banerjee and Wellner (2001), and the second based on the

pointwise asymptotics for their problem (Prakasa Rao, 1969), and similar to those of Balabdaoui et al. (2009). They do not appear to investigate the quality of the approaches for various locations x_0. The benefit of the likelihood ratio approach of Banerjee and Wellner (2001) is that its asymptotic distribution is free of nuisance parameters, and is therefore easier to estimate. Indeed, the simulations in Banerjee and Wellner (2005) show that the likelihood ratio approach outperforms confidence intervals based on pointwise asymptotics for current status data. At this time, we are not aware of any such likelihood ratio results for the nonparametric estimator based on the log-concave density.

The bootstrap approach has been considered extensively for the Grenander estimator and other cube-root asymptotics (Kosorok, 2008; Sen et al., 2010; Léger and MacGibbon, 2006; Politis et al., 1999). To our best knowledge, no such examination has yet been done for the log-concave setting, or $n^{-2/5}$ asymptotics. The very preliminary results on the bootstrap presented here indicate that such work would be of great practical interest: indeed, the ECDF-bootstrap percentile confidence interval worked best in the cases we considered here. The benefit of the bootstrap is that, like the likelihood ratio approach, it does not require the estimation of nuisance parameters. However, due to the current lack of theory backing its use, it is difficult to fully recommend the bootstrap at this time.

The work presented here was designed to study the practical benefits of the asymptotic theory developed in Balabdaoui et al. (2009). Our simulation study allows practitioners to gain some understanding of both the advantages and limitations of the resulting methodology. The approximate confidence intervals we study here have reasonable performance, but do sometimes overcover and/or undercover. Overall, performance is not ideal, but we believe that it is useful for practitioners in that it provides an approximate visual representation of the variability of the log-concave MLE. Of the methods considered, the ECDF-bootstrap (method (v)) has the best performance. Notably, the empirical performance of the NPML-bootstrap (method (vi)) follows closely that of the oracle method (method (i)). The oracle method uses the true value of $c_2(x_0)$, which in practice would be unknown. As pointed out by one of the reviewers, the bias in the coverage of the oracle method is therefore attributable to the discrepancy between the actual sampling distribution (of the re-scaled error) and that of the limiting distribution. It is quite interesting that the NPML-bootstrap follows this discrepancy, whereas the ECDF-bootstrap does not.

The various methods based on (3) have been implemented in the package **logcondens** (Dümbgen and Rufibach, 2006) in the function `logConCI`. In practice, the smoothed MLE defined in (5) is preferable to the MLE, particularly for smaller

sample sizes. However, we are not aware of asymptotic results for this estimator at this time, beyond the consistency results of Chen and Samworth (2013). The simulated data from Section 2, sample code from Section 5, as well as other sample code is available online at
`www.math.yorku.ca/~hkj/Research/AJGsupplement.zip`.

7. Acknowledgements

We thank Jane Heffernan for allowing us to use her server `immune` to run all of our simulations. We thank Piet Groeneboom for sharing the ICS algorithm with us, and for taking the time to prepare the algorithm for us. We thank both Piet Groeneboom and Jon Wellner for interesting discussion on symmetry. Lastly, we thank Kaspar Rufibach for including our work in the package **logcondens**.

Appendix A. Simulating from the limiting distribution $\mathbb{C}(0)$

In this section we give the details on the algorithm used to simulate the quantile estimates of $\mathbb{C}(0)$ given in Section 2. As mentioned previously, the distribution of $\mathbb{C}(0)$ is closely related to the asymptotic distribution which shows up in the estimation of a convex decreasing density on $\mathbb{R}_+$, a problem which was studied in Groeneboom et al. (2001a,b). To define the latter, let $\mathbb{B}(t)$, $t \in \mathbb{R}$ denote a two-sided Brownian motion, as before, and let

$$\widetilde{\mathbb{Y}}(t) = \begin{cases} \int_0^t \mathbb{B}(s)ds + t^4, & t \geq 0, \\ \int_t^0 \mathbb{B}(s)ds + t^4, & t < 0. \end{cases}$$

Define $\widetilde{\mathbb{H}}$ to be the almost surely unique process such that

1. $\widetilde{\mathbb{H}}(t) \geq \widetilde{\mathbb{Y}}(t)$ for all $t \in \mathbb{R}$,
2. $\widetilde{\mathbb{H}}''(t)$ is convex,
3. $\widetilde{\mathbb{H}}(t) = \widetilde{\mathbb{Y}}(t)$ if the slope of $\widetilde{\mathbb{H}}''(t)$ is strictly decreasing at t.

It was shown in Groeneboom et al. (2001a) that the process $\widetilde{\mathbb{H}}$ exists and is unique. In fact, Balabdaoui et al. (2009) show existence of their limiting process simply by noting that $\mathbb{H} = -\widetilde{\mathbb{H}}$ in distribution. Groeneboom et al. (2001a) show that $\widetilde{\mathbb{C}}(t) = \widetilde{\mathbb{H}}''(t) = \lim_{m\to\infty} \widetilde{\mathbb{C}}_m(t)$, where $\widetilde{\mathbb{C}}_m$ is defined as:

$$\widetilde{\mathbb{C}}_m \quad = \quad \operatorname{argmin}_{\varphi\in\widetilde{C}_m} \left\{ \int_{-m}^{m} \varphi^2(t)dt - 2\int_{-m}^{m} \varphi(t)d\left(\mathbb{B}(t) - 4t^3\right) \right\},$$

Table A.3: Two-sides Kologorov-Smirnov test p-values for simulations with different values of m and g. The settings are reported as m/g; that is, $5/1000$ indicates that $m = 5$ and $g = 1000$.

p-value	5/1000	5/2000	5/5000	10/1000
5/2000	0.5361			
5/5000	0.9996	0.2705		
10/1000	0.9802	0.3582	0.9011	
15/1000	0.9557	0.6815	0.8623	0.9541

where $\widetilde{C}_m$ denotes the class of convex functions with the restriction that $\varphi(-m) = \varphi(m) = 12m^2$. As Groeneboom et al. (2001a, Corollary 2.4) point out, other restrictions, beyond that of $\varphi(-m) = \varphi(m) = 12m^2$, may also be used.

Since $\widetilde{\mathbb{C}}_m$ is a solution to a class of optimization problems, various algorithms may be applied to find a solution in practice. A nice review of these, cast in the context of mixture models, is given in Groeneboom et al. (2008, 2004), where the support reduction algorithm is also introduced. The support reduction algorithm is a close relative of the iterative cubic spline algorithm, see Groeneboom et al. (2004, page 2, line -7), Groeneboom et al. (2008, page 390, line -15). The iterative cubic spline (ICS) algorithm is the algorithm discussed in Groeneboom et al. (2001a, Section 3) to simulate from the approximate distribution of $\widetilde{\mathbb{C}}_m$. Both algorithms (support reduction and ICS) use some approximation on a grid (Groeneboom et al. (2008), Groeneboom et al. (2001a, page 1645, line -6)) in the solution. A third alternative to solving these types of problems is the active set algorithm, described for example in Dümbgen et al. (2010). The latter has been implemented in the function `conreg` (which finds the least squares convex/concave regression on a grid), available in the R package **cobs**, Ng and Maechler (2011). The function `conreg`, however, does not implement the constrained least squares problem ($\widetilde{\varphi}(m) = \widetilde{\varphi}(-m) = 12m^2$).

Here, to simulate samples of $\mathbb{C}(0)$ (equivalently, $-\widetilde{\mathbb{C}}(0)$), we employ the function `conreg` from the R package **cobs**, to simulate samples from

$$\mathbb{C}_m^* \quad = \quad \text{argmin}_{\varphi \in C_m^*} \left\{ \int_{-m}^{m} \varphi^2(t)dt - 2 \int_{-m}^{m} \varphi(t) d\left(\mathbb{B}(t) + 4t^3\right) \right\},$$

where C_m^* is the class of concave functions defined on $[-m, m]$. Our approach, thus, does not use the constrained least squares solutions suggested by Groeneboom et al. (2001a). All integrals were approximated on a grid of size $2mg + 1$.

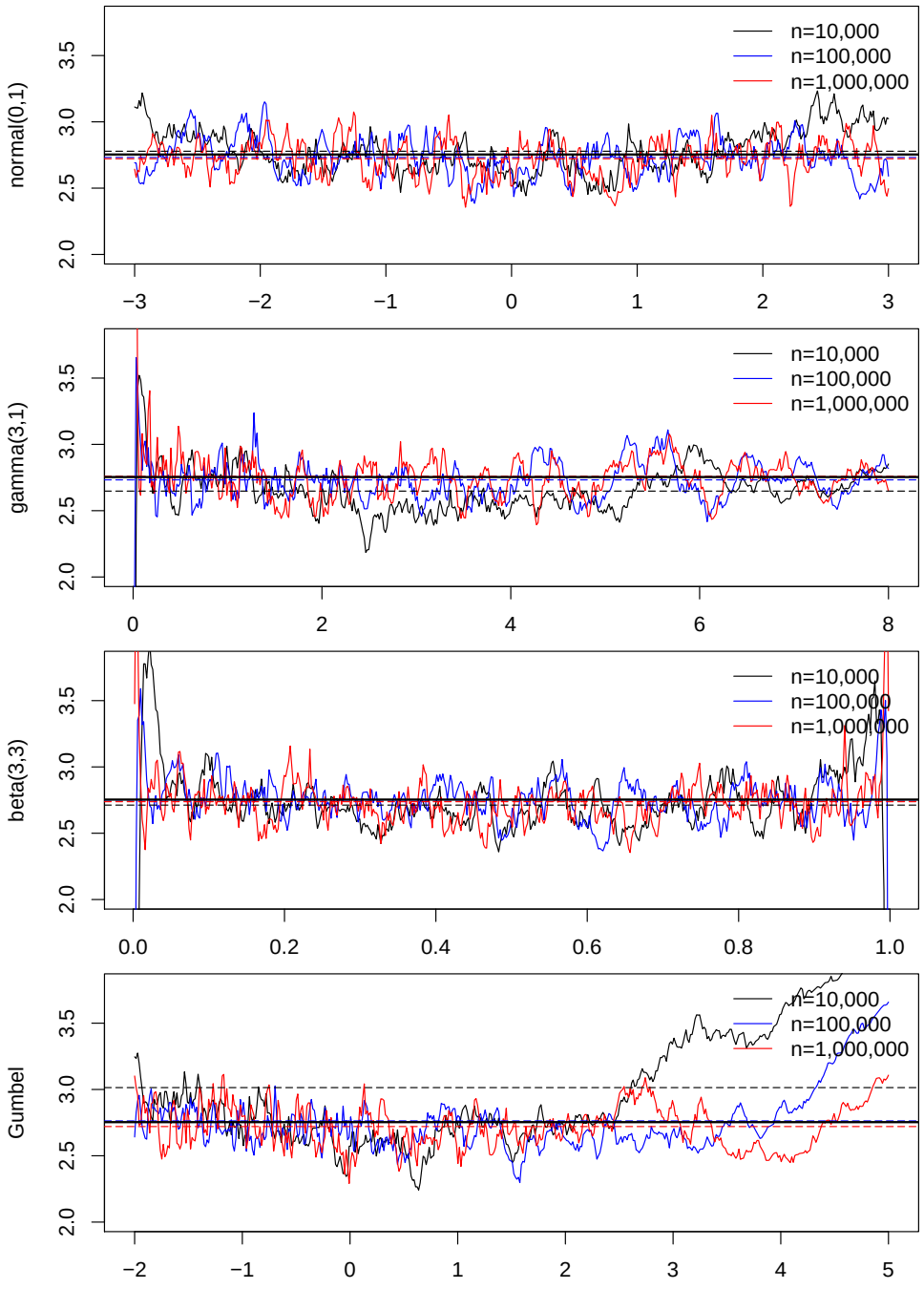

Figure A.11: Comparison of 97.5% quantiles for simulations from distributions (D1)-(D4) for samples sizes $n = 10\,000, 100\,000, 1\,000\,000$ and the 97.5% quantile of $\mathbb{C}(0)$ (shown as horizontal black line). Means for the distributions (D1)-(D4) are shown as the dashed horizontal lines.

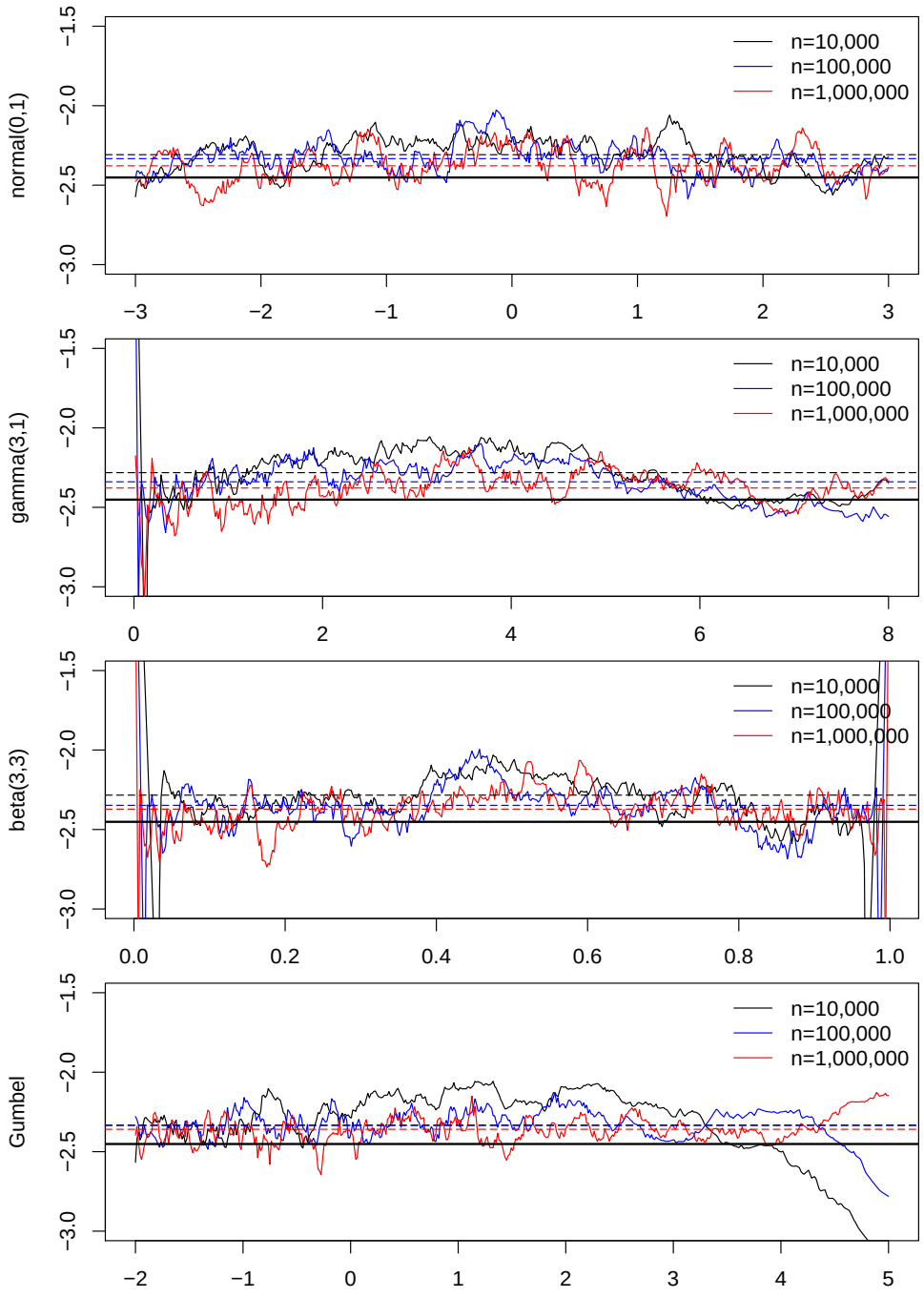

Figure A.12: Comparison of 2.5% quantiles for simulations from distributions (D1)-(D4) for samples sizes $n = 10\,000, 100\,000, 1\,000\,000$ and the 2.5% quantile of $\mathbb{C}(0)$ (shown as horizontal black line). Means for the distributions (D1)-(D4) are shown as the dashed horizontal lines.

Table A.4: Estimated values of $F^{-1}_{\mathbb{C}(0)}(p)$ and four simulations using distributions (D1) normal(0,1), (D2) gamma(3,1), (D3) beta(3,3), and (D4) Gumbel for sample size $n = 10\,000$.

p	$\mathbb{C}(0)$	(D1)	(D2)	(D3)	(D4)
0.001	-3.6442	-3.6426	-3.502	-3.4945	-3.9236
0.005	-3.0905	-2.9407	-2.8997	-2.9071	-3.0015
0.010	-2.8172	-2.6953	-2.6407	-2.6546	-2.7257
0.025	-2.4157	-2.3083	-2.2818	-2.2838	-2.3331
0.050	-2.0574	-1.9578	-1.9473	-1.9648	-2.0083
0.100	-1.644	-1.5679	-1.5656	-1.5919	-1.5815
0.200	-1.1184	-1.0734	-1.0752	-1.1068	-1.0541
0.500	-0.0557	-0.0588	-0.0885	-0.104	0.0154
0.800	1.0948	1.0769	1.0232	1.0215	1.1737
0.900	1.7421	1.7197	1.651	1.6582	1.8421
0.950	2.2653	2.2714	2.1901	2.2145	2.4325
0.975	2.7536	2.7782	2.6477	2.7122	3.014
0.990	3.3193	3.3776	3.1755	3.2975	3.6574
0.995	3.6881	3.795	3.549	3.688	4.1097
0.999	4.514	4.881	4.2587	4.3994	5.1261

Comparisons of various choices of m and g are given in Table A.3, and the p-values reported indicate that no statistical difference exists between the selected options. In the reported quantile results in this paper, we used $m = 5$ and $g = 1000$. For these settings, the simulations of 100 000 samples took roughly 10 days to complete, while for larger m/g the simulation times were significantly longer.

To assess the quality of these estimates, we compared our quantiles to sample quantiles of

$$n^{2/5} c_2^{-1}(x_0) \left(\widehat{f_n}(x_0) - f_0(x_0) \right),$$

using the known values of $f_0(x_0)$ and $c_2(x_0)$, for four different distributions. We chose a sample size of $n = 10\,000, 100\,000$ and $n = 1\,000\,000$, as well as various values of x_0. The four choices of f_0 were

(D1). the normal with mean zero and variance 1,

(D2). the gamma with shape parameter 3 and rate 1,

(D3). the beta with both parameters equal to 2,

(D4). the standard Gumbel distribution.

Table A.5: Estimated values of $F^{-1}_{\mathbb{C}(0)}(p)$ and four simulations using distributions (D1) normal(0,1), (D2) gamma(3,1), (D3) beta(3,3), and (D4) Gumbel for sample size $n = 100\,000$.

p	$\mathbb{C}(0)$	(D1)	(D2)	(D3)	(D4)
0.001	-3.6442	-3.5547	-3.5786	-3.5583	-3.6052
0.005	-3.0905	-2.9683	-2.9806	-2.977	-2.973
0.010	-2.8172	-2.7194	-2.7331	-2.7177	-2.7191
0.025	-2.4157	-2.3333	-2.3393	-2.3479	-2.3351
0.050	-2.0574	-2.004	-2.0062	-2.0233	-1.9799
0.100	-1.644	-1.6081	-1.6062	-1.6159	-1.5743
0.200	-1.1184	-1.104	-1.1078	-1.1133	-1.0692
0.500	-0.0557	-0.072	-0.0922	-0.0857	-0.0508
0.800	1.0948	1.0729	1.0427	1.0473	1.099
0.900	1.7421	1.7001	1.6876	1.6931	1.7232
0.950	2.2653	2.2439	2.2407	2.2493	2.2752
0.975	2.7536	2.7316	2.7324	2.7505	2.7631
0.990	3.3193	3.3249	3.3256	3.3635	3.3263
0.995	3.6881	3.763	3.7211	3.8326	3.7367
0.999	4.514	4.7569	4.6067	4.9339	4.8466

For each distribution we calculated the quantile based on $B = 1000$ simulations and averaged this over the various x_0 chosen. The results are shown in Tables A.4, A.5, and A.6. The results in the tables report averages over different values of x_0. For a more detailed view, we also report results for specific x_0 for quantiles corresponding to $p = 0.025, 0.975$ in Figures A.11 and A.12. The figures in particular seem to indicate convergence, albeit rather slow, to the estimated quantiles. This is particularly evident for $p = 0.025$ (Figure A.12). In the tables, the only worrisome results are for the smallest quantiles at $p = 0.001, 0.999$ for distributions (D2) and (D4) in Table A.6. We conjecture that these are due to numerical accuracy issues, as these are only observable at the highest sample size.

Remark Appendix A.1. *We did try using the ICS algorithm to generate samples from $\widetilde{\mathbb{C}}_m$ (the program for which was generously given to us by Piet Groeneboom), but we found that this approach underestimated the quantiles as compared to the empirical results of Table A.4- A.6, and therefore did not perform well in empirical coverage simulations for finite sample sizes. Notably, however, the ICS algorithm was much faster than the algorithm currently implemented in* `conreg`.

Table A.6: Estimated values of $F^{-1}_{\mathbb{C}(0)}(p)$ and four simulations using distributions (D1) normal(0,1), (D2) gamma(3,1), (D3) beta(3,3), and (D4) Gumbel for sample size $n = 1\,000\,000$.

p	$\mathbb{C}(0)$	(D1)	(D2)	(D3)	(D4)
0.001	-3.6442	-3.6907	-32.4047	-3.7050	-43.9143
0.005	-3.0905	-3.0447	-3.2235	-3.0511	-3.1146
0.010	-2.8172	-2.7692	-2.8519	-2.7793	-2.7945
0.025	-2.4157	-2.3782	-2.3784	-2.3718	-2.3591
0.050	-2.0574	-2.0270	-2.0395	-2.0279	-1.9982
0.100	-1.644	-1.6175	–1.6346	-1.6265	-1.596
0.200	-1.1184	-1.1040	-1.1100	-1.1145	-1.0874
0.500	-0.0557	-0.0619	-0.0787	-0.0733	-0.0541
0.800	1.0948	1.0722	1.0795	1.0691	1.077
0.900	1.7421	1.7075	1.7226	1.7111	1.7181
0.950	2.2653	2.2540	2.2745	2.2664	2.2570
0.975	2.7536	2.7477	2.7796	2.7647	2.7475
0.990	3.3193	3.3334	3.4223	3.3382	3.3313
0.995	3.6881	3.7390	4.0490	3.7505	3.8067
0.999	4.514	4.6840	55.5786	4.8066	47.1243

Appendix A.1. Discussion

The simulation results presented here are based on an ad-hoc approximation approach, as described above. Although the empirical evidence presented here indicates that this approach is reasonable, a more theoretical way of calculating the quantiles of $\mathbb{C}(0)$ would be highly desirable. Such an approach is available for Chernoff's distribution (see Groeneboom and Wellner, 2001) thanks to the incredible work in Groeneboom (1989). Groeneboom and Wellner (2001) also provide some history on the estimation of quantiles of Chernoff's distribution.

Appendix B. Proofs

Proof of Proposition 2.1. The proof of this is straightforward, and is included only for completeness. Note that $\mathbb{H}(t)$ (the invelope of $\mathbb{Y}(t)$) is uniquely characterized by

1. $\mathbb{H}(t) \leq \mathbb{Y}(t)$ for all $t \in \mathbb{R}$,
2. $\mathbb{H}''(t)$ is concave,

3. $\mathbb{H}(t) = \mathbb{Y}(t)$ if the slope of $\mathbb{H}''(t)$ is strictly decreasing at t.

To prove the result it is enough to note that $\mathbb{Y}(-t)$ has the same distribution as $\mathbb{Y}(t)$, and that $\mathbb{H}(-t)$ is the (unique!) invelope of $\mathbb{Y}(-t)$. Taking derivatives yields the results for $\mathbb{H}'(t), \mathbb{C}(t)$. □

References

Baird, M., 2012. Cross validation bandwidth selection for derivatives of various dimensional densities: a Monte Carlo study of kernel order and CV criteria. Tech. rep., UCLA.

Balabdaoui, F., Rufibach, K., Wellner, J. A., 2009. Limit distribution theory for maximum likelihood estimation of a log-concave density. Ann. Statist. 37 (3), 1299–1331.

Balabdaoui, F., Wellner, J. A., 2012. Chernoff's density is log-concave. Tech. rep., University of Washington, available at arXiv:1203.0828.

Banerjee, M., Wellner, J. A., 2001. Likelihood ratio tests for monotone functions. Ann. Statist. 29 (6), 1699–1731.

Banerjee, M., Wellner, J. A., 2005. Confidence intervals for current status data. Scand. J. Statist. 32 (3), 405–424.

Bowman, A. W., 1984. An alternative method of cross-validation for the smoothing of density estimates. Biometrika 71 (2), 353–360.

Chang, G., Walther, G., 2007. Clustering with mixtures of log-concave distributions. Comput. Statist. Data Anal. 51, 6242–6251.

Chen, Y., Samworth, R. J., 2013. Smoothed log-concave maximum likelihood estimation with applications. Statistica Sinica. 23, 1373–1398.

Chernoff, H., 1964. Estimation of the mode. Ann. Inst. Statist. Math. 16, 31–41.

Cule, M., Samworth, R., 2010. Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density. Electronic J. Stat. 4, 254–270.

Cule, M., Samworth, R., Stewart, M., 2010. Maximum likelihood estimation of a multidimensional log-concave density. J. R. Stat. Soc. Ser. B Stat. Methodol. 72 (5), 545–607.

Dümbgen, L., Hüsler, A., Rufibach, K., 2010. Active set and EM algorithms for log-concave densities based on complete and censored data. Tech. rep., University of Bern, available at arXiv:0707.4643.

Dümbgen, L., Rufibach, K., 2006. logcondens: Estimate a Log-Concave Probability Density from iid Observations. R package version 2.0-7.

Dümbgen, L., Rufibach, K., 2009. Maximum likelihood estimation of a log-concave density and its distribution function. Bernoulli 15, 40–68.

Dümbgen, L., Rufibach, K., 2011. logcondens: Computations related to univariate log-concave density estimation. Journal of Statistical Software 39, 1–28.

Duong, T., 2007. ks: Kernel density estimation and kernel discriminant analysis for multivariate data in R. Journal of Statistical Software 21, 1–16.

Duong, T., 2012. ks: Kernel smoothing. R package version 1.8-9.

Groeneboom, P., 1985. Estimating a monotone density. In: Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer, Vol. II (Berkeley, Calif., 1983). Wadsworth Statist./Probab. Ser. Wadsworth, Belmont, CA, pp. 539–555.

Groeneboom, P., 1989. Brownian motion with a parabolic drift and Airy functions. Probab. Theory Related Fields 81 (1), 79–109.

Groeneboom, P., Jongbloed, G., Wellner, J. A., 2001a. A canonical process for estimation of convex functions: the "invelope" of integrated Brownian motion $+t^4$. Ann. Statist. 29 (6), 1620–1652.

Groeneboom, P., Jongbloed, G., Wellner, J. A., 2001b. Estimation of a convex function: characterizations and asymptotic theory. Ann. Statist. 29 (6), 1653–1698.

Groeneboom, P., Jongbloed, G., Wellner, J. A., 2004. The support reduction algorithm for computing nonparametric function estimates in mixture models. Available on arXiv: arXiv:math/0405511v1 [math.ST].

Groeneboom, P., Jongbloed, G., Wellner, J. A., 2008. The support reduction algorithm for computing non-parametric function estimates in mixture models. Scand. J. Statist. 35 (3), 385–399.

Groeneboom, P., Wellner, J. A., 2001. Computing Chernoff's distribution. J. Comput. Graph. Statist. 10 (2), 388–400.

Härdle, W., 1991. Smoothing techniques. Springer Series in Statistics. Springer-Verlag, New York, with implementation in S.

Härdle, W., Müller, M., Sperlich, S., Werwatz, A., 2004. Nonparametric and semiparametric models. Springer Series in Statistics. Springer-Verlag, New York.

Jones, M. C., Marron, J. S., Park, B. U., 1991. A simple root *n* bandwidth selector. Ann. Statist. 19 (4), 1919–1932.

Kosorok, M. R., 2008. Bootstrapping in Grenander estimator. In: Beyond parametrics in interdisciplinary research: Festschrift in honor of Professor Pranab K. Sen. Vol. 1 of Inst. Math. Stat. Collect. Inst. Math. Statist., Beachwood, OH, pp. 282–292.

Léger, C., MacGibbon, B., 2006. On the bootstrap in cube root asymptotics. Canad. J. Statist. 34 (1), 29–44.

Ng, P. T., Maechler, M., 2011. cobs: COBS – Constrained B-splines (Sparse matrix based). R package version 1.2-2.

Pal, J. K., Woodroofe, M., Meyer, M., 2007. Estimating a Polya frequency function$_2$. In: Complex datasets and inverse problems. Vol. 54 of IMS Lecture Notes Monogr. Ser. Inst. Math. Statist., Beachwood, OH, pp. 239–249.

Politis, D. N., Romano, J. P., Wolf, M., 1999. Subsampling. Springer Series in Statistics. Springer-Verlag, New York.

Prakasa Rao, B. L. S., 1969. Estimation of a unimodal density. Sankhya Series A 31, 23–36.

Rudemo, M., 1982. Empirical choice of histograms and kernel density estimators. Scand. J. Statist. 9 (2), 65–78.

Sen, B., Banerjee, M., Woodroofe, M., 2010. Inconsistency of bootstrap: the Grenander estimator. Ann. Statist. 38 (4), 1953–1977.

Shorack, G. R., Wellner, J. A., 1986. Empirical processes with applications to statistics. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York.

Walther, G., 2002. Detecting the presence of mixing with multiscale maximum likelihood. J. Amer. Statist. Assoc. 97 (458), 508–513.

Walther, G., 2009. Inference and modeling with log-concave distributions. Statist. Sci. 24 (3), 319–327.

Wand, M. P., Jones, M. C., 1994. Multivariate plug-in bandwidth selection. Comput. Statist. 9 (2), 97–116.

Wasserman, L., 2006. All of nonparametric statistics. Springer Texts in Statistics. Springer, New York.