Bi-Log-Concave Distribution Functions[*]

Lutz Dümbgen[1], Petro Kolesnyk[1] and Ralf A. Wilke[2,3]
([1]University of Bern, [2]Copenhagen Business School, [3] University of Strasbourg)

October 2016

Abstract

Nonparametric statistics for distribution functions F or densities $f = F'$ under qualitative shape constraints constitutes an interesting alternative to classical parametric or entirely nonparametric approaches. We contribute to this area by considering a new shape constraint: F is said to be bi-log-concave, if both $\log F$ and $\log(1 - F)$ are concave. Many commonly considered distributions are compatible with this constraint. For instance, any c.d.f. F with log-concave density $f = F'$ is bi-log-concave. But in contrast to log-concavity of f, bi-log-concavity of F allows for multimodal densities. We provide various characterisations. It is shown that combining any nonparametric confidence band for F with the new shape constraint leads to substantial improvements, particularly in the tails. To pinpoint this, we show that these confidence bands imply non-trivial confidence bounds for arbitrary moments and the moment generating function of F.

AMS subject classifications.   62G15, 62G20, 62G30.

Key words.   hazard, honest confidence region, moment generating function, moments, reverse hazard, shape constraint.

1   Introduction

In nonparametric statistics one is often interested in estimators or confidence regions for curves such as densities or regression functions. Estimation of such curves is typically an ill-posed problem and requires additional assumptions. Interesting alternatives to smoothness assumptions are qualitative constraints such as monotonicity or concavity.

Estimation of a distribution function F based on independent, identically distributed random variables $X_1, X_2, \ldots, X_n$ with c.d.f. F is common practice and does not require restrictive assumptions. But nontrivial confidence regions for certain functionals of F such as the mean do not exist without substantial additional constraints (cf. Bahadur and Savage, 1956).

A growing literature on density estimation under shape constraints considers the family of log-concave densities. These are probability densities f on $\mathbb{R}^d$ such that $\log f : \mathbb{R}^d \to [-\infty, \infty)$ is a concave function. For more details see Bagnoli and Bergstrom (2005), Cule et al. (2010), Dümbgen and Rufibach (2009, 2011), Walther (2009), Seregin and Wellner (2010), Dümbgen et al. (2011) and the references cited therein. Most efforts in these papers are devoted to point estimation. Schuhmacher et al. (2011) obtain a nonparametric confidence region by combining the log-concavity constraint and a standard Kolmogorov-Smirnov confidence region. But its explicit computation is difficult, and this is one motivation to search for alternative shape constraints in terms of the distribution function F directly.

While many popular densities are log-concave, this constraint can be too restrictive in applications with a multimodal density. In the present paper we consider a model with a new and weaker constraint on the distribution function:

Definition (Bi-log-concavity). A distribution function F on the real line is called *bi-log-concave* if both $\log F$ and $\log(1 - F)$ are concave functions from $\mathbb{R}$ to $[-\infty, 0]$.

Many distribution functions satisfy this constraint. In particular, when F has a log-concave density $f = F'$, it is bi-log-concave (Bagnoli and Bergstrom, 2005). But indeed, bi-log-concavity of F is a much weaker constraint. As shown later, F may have a density with an arbitrary number of modes. Thus, we consider estimation of distributions under shape constraints for a wider family of distributions.

The remainder of this paper is organised as follows: In Section 2 we present characterisations of bi-log-concavity and explicit bounds for F and its density $f = F'$. In Section 3 we describe exact (conservative) confidence bands for F. They are constructed by combining the bi-log-concavity constraint with standard confidence bands for F such as, for instance, the Kolmogorov-Smirnov band or Owen's (1995) band. A numerical example with the distribution of CEO salaries (Woolridge 2000) illustrates the usefulness of the proposed method. The benefits of adding the shape constraint are pinpointed in Section 4. It is shown that combining a reasonable confidence band with the new shape constraint leads to non-trivial honest confidence bounds for various quantities related to F. These include its density, hazard function and reverse hazard function, its moment generating function and arbitrary moments. All proofs are deferred to Section 5.

2 Bi-log-concave distribution functions

In what follows we call a distribution function F *non-degenerate* if the set

$$J(F) \ := \ \{x \in \mathbb{R} : 0 < F(x) < 1\}$$

is nonvoid. Notice that in the case of $J(F) = \emptyset$ the distribution function F would correspond to the Dirac measure δ_m at some point $m \in \mathbb{R}$, i.e. $F(x) = 1_{[x \geq m]}$.

Our first theorem provides three alternative characterisations of bi-log-concavity which are expressed by different constraints for F and its derivatives.

Theorem 1. *For a non-degenerate distribution function F the following four statements are equivalent:*

(i) F is bi-log-concave;

(ii) F is continuous on $\mathbb{R}$ and differentiable on $J(F)$ with derivative $f = F'$ such that

$$F(x+t) \begin{cases} \leq \ F(x) \exp\left(\dfrac{f(x)}{F(x)} t\right) \\ \geq \ 1 - (1 - F(x)) \exp\left(-\dfrac{f(x)}{1 - F(x)} t\right) \end{cases} \tag{1}$$

for arbitrary $x \in J(F)$ and $t \in \mathbb{R}$.

(iii) F is continuous on $\mathbb{R}$ and differentiable on $J(F)$ with derivative $f = F'$ such that the hazard function $f/(1 - F)$ is non-decreasing and the reverse hazard function f/F is non-increasing on $J(F)$.

(iv) F is continuous on $\mathbb{R}$ and differentiable on $J(F)$ with bounded and strictly positive derivative $f = F'$. Furthermore, f is locally Lipschitz-continuous on $J(F)$ with L^1-derivative $f' = F''$ satisfying

$$\frac{-f^2}{1 - F} \ \leq \ f' \ \leq \ \frac{f^2}{F}. \tag{2}$$

The set of all distribution functions F with the properties stated in Theorem 1 is denoted as $\mathcal{F}_{\mathrm{blc}}$. The inequalities (2) in statement (iv) can be reformulated as follows: $\log f$ is locally Lipschitz-continuous on $J(F)$ with L^1-derivative $(\log f)'$ satisfying

$$(\log(1 - F))' \ \leq \ (\log f)' \ \leq \ (\log F)'.$$

Recall that the L^1-derivative of a function h on an open interval $J \subset \mathbb{R}$ is a locally integrable function h' on J such that $h(y) - h(x) = \int_x^y h'(t)\, dt$ for all $x, y \in J$.

Example (Bi-modal density). Consider the mixture $2^{-1}\mathcal{N}(-\delta, 1) + 2^{-1}\mathcal{N}(\delta, 1)$ with $\delta > 0$. It can be verified numerically that the corresponding c.d.f. F is bi-log-concave for $\delta \leq 1.34$ but not for $\delta \geq 1.35$. This distribution has a bi-modal density for $\delta = 1.34$. The corresponding c.d.f. F is shown in Figure 1(a), together with the functions $1 + \log F \leq F \leq -\log(1 - F)$, the inequalities following from $\log(1+y) \leq y$ for arbitrary $y \geq -1$. Bi-log-concavity means that the lower bound $1 + \log F$ is concave while the upper bound $-\log(1 - F)$ is convex. Figures 1-2 illustrate the various characterisations of the bi-log-concavity constraint as given in Theorem 1. In particular, Figure 1(b) shows the bounds from part (ii) for one particular point $x \in J(F)$. Figure 2(a) shows the density f together with the hazard function $f/(1 - F)$ and the reverse hazard function f/F. It is apparent that the latter two satisfy the monotonicity properties of part (iii). Figure 2(b) contains the derivative f' together with the bounds $-f^2/(1 - F)$ and f^2/F as given in part (iv).

While the previous example illustrates bi-log-concavity for a bi-modal density, the next example considers a multi modal density.

Example (k-modal density). For any integer $k > 0$ and $a \in (0, 1)$,

$$f(x) := 1_{[0 < x < 1]}(1 + a\sin(2\pi k x))$$

defines a probability density with k local maxima. The corresponding c.d.f. is given by $F(x) = x + a(1 - \cos(2\pi k x))/(2\pi k)$ for $x \in [0, 1]$. One can easily deduce from Theorem 1 (iv) that F is bi-log-concave if a is sufficiently small.

Remark. For $F \in \mathcal{F}_{\mathrm{blc}}$, its moment-generating function is finite in a neighborhood of 0. In particular, we show in Section 5 that

$$\left\{t \in \mathbb{R} : \int e^{tx}\, F(dx) < \infty\right\} = \left(-T_1(F), T_2(F)\right) \tag{3}$$

with

$$T_1(F) := \sup_{x \in J(F)} \frac{f(x)}{F(x)} \begin{cases} > 0, \\ = \infty & \text{if } \inf(J(F)) > -\infty, \end{cases}$$

$$T_2(F) := \sup_{x \in J(F)} \frac{f(x)}{1 - F(x)} \begin{cases} > 0, \\ = \infty & \text{if } \sup(J(F)) < \infty. \end{cases}$$

3 Confidence bands

A confidence band for $F \in \mathcal{F}_{\mathrm{blc}}$ may be constructed by intersecting a standard confidence band for a (continuous) distribution function with this class $\mathcal{F}_{\mathrm{blc}}$.
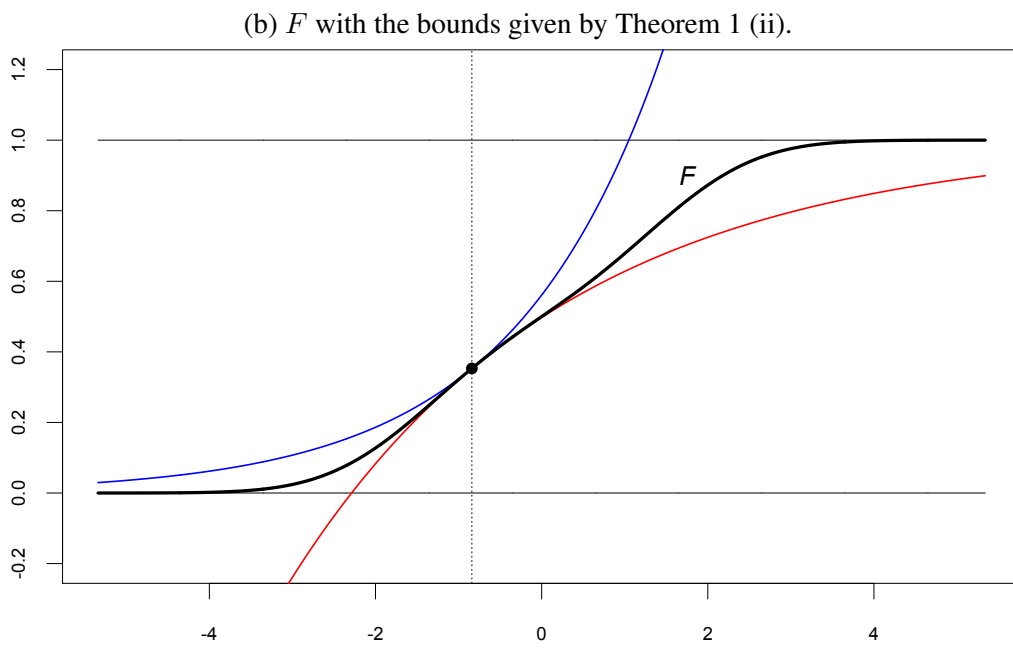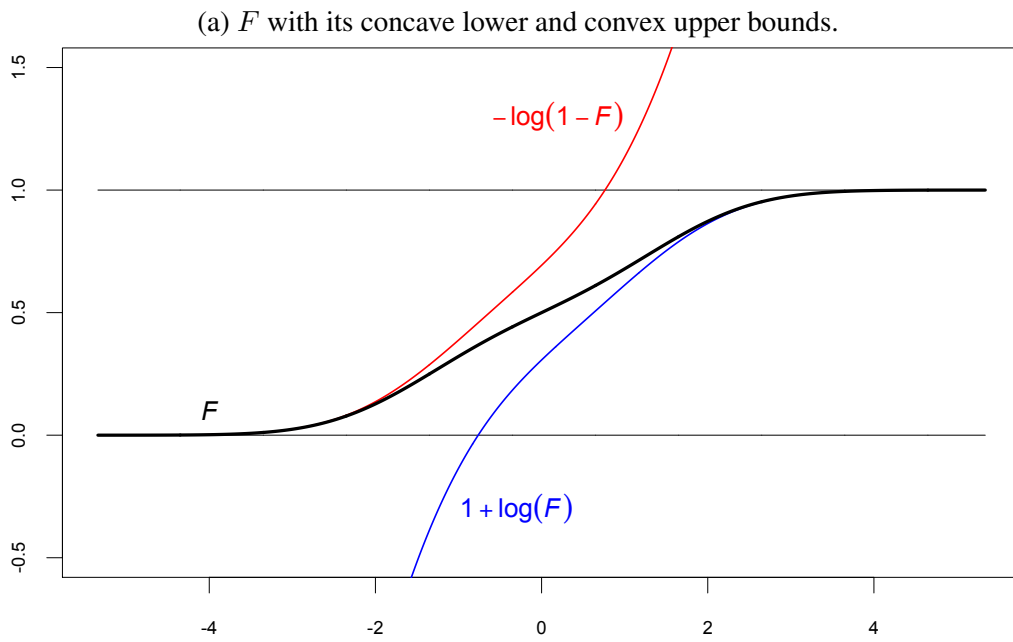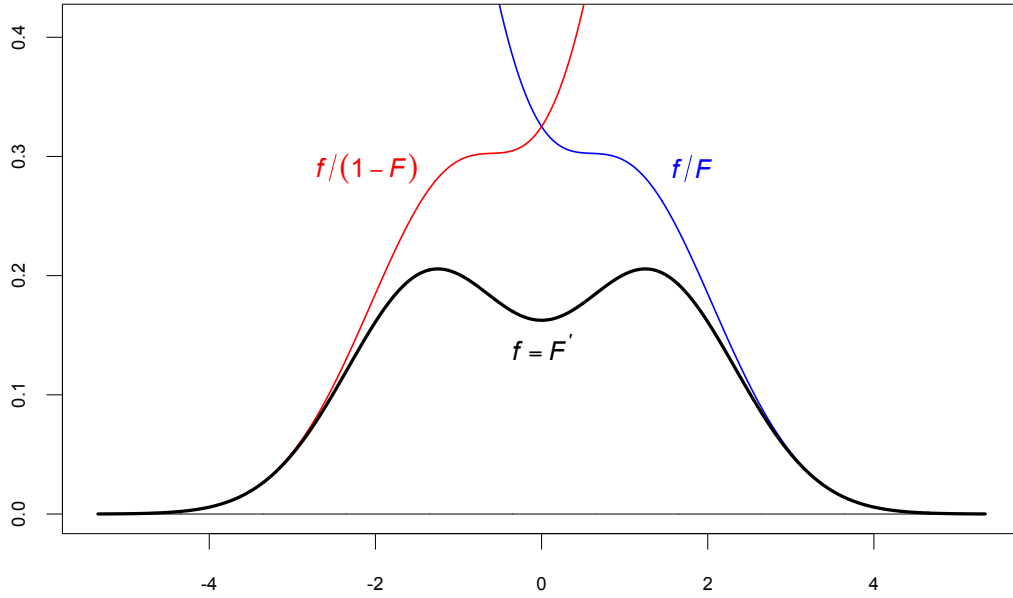
(a) F with its concave lower and convex upper bounds.

(b) F with the bounds given by Theorem 1 (ii).

Figure 1: A bi-log-concave F with its bounds.

(a) f with monotonic hazard and reversed hazard function as given by Theorem 1 (iii).
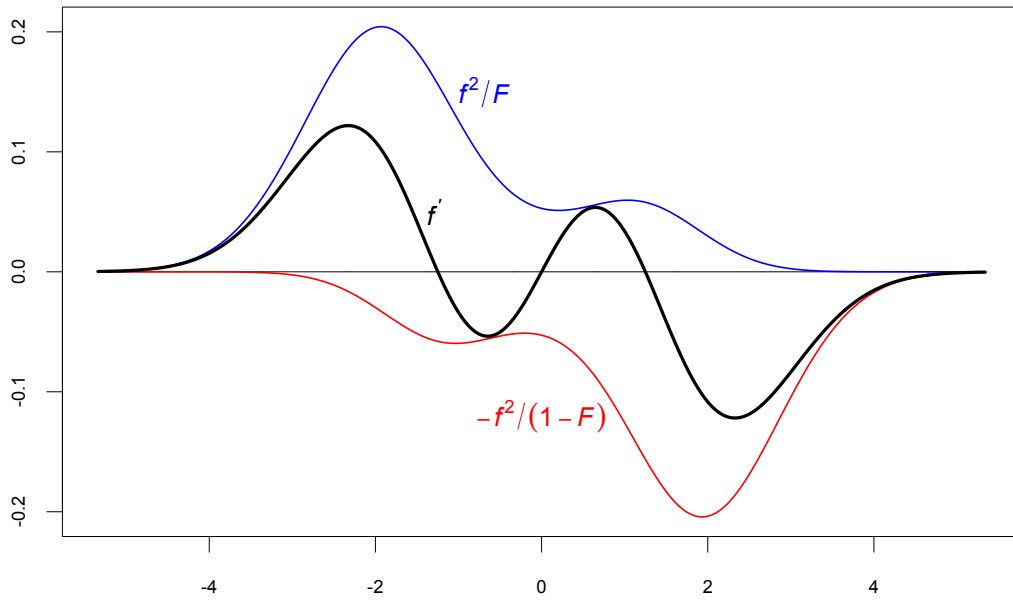

(b) f' with its bounds as given by Theorem 1 (iv).

Figure 2: Different characterisations of a bi-log-concave F.

Unconstrained nonparametric confidence bands. Let $X_1, \ldots, X_n$ be independent random variables with continuous distribution function F. In what follows let (L_n, U_n) be a $(1 - \alpha)$-confidence band for F with $0 < \alpha \leq 0.5$. This means, $L_n = L_{n,\alpha}(\cdot \mid X_1, \ldots, X_n) < 1$ and $U_n = U_{n,\alpha}(\cdot \mid X_1, \ldots, X_n) > 0$ are data-driven non-decreasing functions on the real line such that $L_n \leq U_n$ pointwise and

$$P\big(L_n(x) \leq F(x) \leq U_n(x) \text{ for all } x \in \mathbb{R}\big) \;=\; 1 - \alpha.$$

Example (Kolmogorov-Smirnov band). A standard example for (L_n, U_n) is given by

$$\big[L_n(x), U_n(x)\big] \;:=\; \Big[\widehat{F}_n(x) - \frac{\kappa^{\mathrm{KS}}_{\alpha,n}}{\sqrt{n}}, \widehat{F}_n(x) + \frac{\kappa^{\mathrm{KS}}_{\alpha,n}}{\sqrt{n}}\Big] \cap [0, 1],$$

where

$$\widehat{F}_n(x) \;:=\; \frac{1}{n}\sum_{i=1}^n 1_{[X_i \leq x]},$$

and $\kappa^{\mathrm{KS}}_{n,\alpha}$ denotes the $(1 - \alpha)$-quantile of $\sup_{x \in \mathbb{R}} n^{1/2}\big|\widehat{F}(x) - F(x)\big|$; cf. Shorack and Wellner (1986). Notice also that $\kappa^{\mathrm{KS}}_{n,\alpha} \leq \sqrt{\log(2/\alpha)/2}$ by Massart's (1990) inequality.

Example (Weighted Kolmogorov-Smirnov band). Let $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ denote the order statistics of $X_1, X_2, \ldots, X_n$ and $U_{(i)} := F(X_{(i)})$. It is well known that $U_{(1)} < U_{(2)} < \cdots < U_{(n)}$ are distributed like the order statistics of n independent random variables with uniform distribution on $[0, 1]$. By noting that $\mathbb{E}(U_{(i)}) = t_i := i/(n+1)$ for $1 \leq i \leq n$, and using empirical process theory, one can show that for any $\gamma \in [0, 1/2)$, the random variable

$$\sqrt{n}\max_{i=1,2,\ldots,n}\frac{|U_{(i)} - t_i|}{(t_i(1 - t_i))^\gamma} \tag{4}$$

converges in distribution to $\sup_{t \in (0,1)}(t(1 - t))^{-\gamma}|B(t)| < \infty$ as $n \to \infty$, where B is standard Brownian bridge. In particular, the $(1-\alpha)$-quantile $\kappa^{\mathrm{WKS}}_{n,\alpha}$ of the test statistic (4) satisfies $\kappa^{\mathrm{WKS}}_{n,\alpha} = O(1)$. Inverting this test leads to the $(1 - \alpha)$-confidence band (L_n, U_n) for F with

$$\big[L_n(x), U_n(x)\big] \;=\; \Big[t_i - \frac{\kappa^{\mathrm{WKS}}_{n,\alpha}}{\sqrt{n}}(t_i(1 - t_i))^\gamma, t_{i+1} + \frac{\kappa^{\mathrm{WKS}}_{n,\alpha}}{\sqrt{n}}(t_{i+1}(1 - t_{i+1}))^\gamma\Big] \cap [0, 1]$$

for $i \in \{0, 1, \ldots, n\}$ and $x \in [X_{(i)}, X_{(i+1)})$. Here $X_{(0)} := -\infty$ and $X_{(n+1)} := \infty$.

Example (Owen's band refined). Another confidence band which may be viewed as a refinement of Owen's (1995) method has been proposed recently by Dümbgen and Wellner (2014). Let

$$K(\widehat{p}, p) \;:=\; \widehat{p}\log\frac{\widehat{p}}{p} + (1 - \widehat{p})\log\frac{1 - \widehat{p}}{1 - p}$$

for $p, \widehat{p} \in [0, 1]$ with the usual conventions that $0 \log(\cdot) := 0$ and $a \log(a/0) := \infty$ for $a > 0$. Furthermore, for $t \in (0, 1)$ let

$$C(t) := \log(1 + \mathrm{logit}(t)^2/2)/2 \quad \text{and} \quad D(t) := \log(1 + C(t)^2/2)/2.$$

Then for any fixed $\nu > 2$,

$$\max_{j=1,2,\ldots,n} \big((n+1)K(t_j, U_{(j)}) - C(t_j) - \nu D(t_j)\big) \tag{5}$$

converges in distribution to

$$\sup_{t \in (0,1)} \left(\frac{B(t)^2}{t(1-t)} - C(t) - \nu D(t)\right) < \infty.$$

In particular, the $(1 - \alpha)$-quantile $\kappa_{n,\alpha}^{\mathrm{ODW}}$ of the test statistic (5) is bounded as $n \to \infty$. Inverting this test leads to the following confidence band (L_n, U_n):

$$L_n(x) := 0 \quad \text{for } x < X_{(1)},$$

$$L_n(x) := \min\{p \in (0, t_j] : K(t_j, p) \le \gamma_n(t_j)\} \quad \text{for } 1 \le j \le n, X_{(j)} \le x < X_{(j+1)},$$

$$U_n(x) := \max\{p \in [t_j, 1) : K(t_j, p) \le \gamma_n(t_j)\} \quad \text{for } 1 \le j \le n, X_{(j-1)} \le x < X_{(j)},$$

$$U_n(x) := 1 \quad \text{for } x \ge X_{(n)},$$

where

$$\gamma_n(t) := \frac{C(t) + \nu D(t) + \kappa_{n,\alpha}^{\mathrm{ODW}}}{n+1}.$$

Confidence bands for a bi-log-concave F. Now suppose that F belongs to $\mathcal{F}_{\mathrm{blc}}$. Under this assumption, a $(1 - \alpha)$-confidence band (L_n, U_n) for F may be refined as follows:

$$L_n^o(x) := \inf\{G(x) : G \in \mathcal{F}_{\mathrm{blc}}, L_n \le G \le U_n\},$$

$$U_n^o(x) := \sup\{G(x) : G \in \mathcal{F}_{\mathrm{blc}}, L_n \le G \le U_n\}.$$

It may happen that no bi-log-concave distribution function fits into the band (L_n, U_n). In this case we set $L_n^o \equiv 1$ and $U_n^o \equiv 0$ and conclude with confidence $1 - \alpha$ that $F \notin \mathcal{F}_{\mathrm{blc}}$. But in the case of $F \in \mathcal{F}_{\mathrm{blc}}$ this happens with probability at most α. Indeed, the construction of (L_n^o, U_n^o) implies that

$$\mathbb{P}(L_n^o \le F \le U_n^o) = \mathbb{P}(L_n \le F \le U_n) \quad \text{if } F \in \mathcal{F}_{\mathrm{blc}}.$$

The following algorithm is used to determine the refined band (L_n^o, U_n^o). An essential ingredient is a procedure $\mathrm{ConcInt}(\cdot, \cdot)$ (concave interior). Given any finite set $\mathcal{T} = \{t_0, t_1, \ldots, t_m\}$

of real numbers $t_0 < t_1 < \cdots < t_m$ and any pair (ℓ, u) of functions $\ell, u : \mathcal{T} \to [-\infty, \infty)$ with $\ell < u$ pointwise and $\ell(t) > -\infty$ for at least two different points $t \in \mathcal{T}$, this procedure computes the pair (ℓ^o, u^o), where

$$\ell^o(x) := \inf\{g(x) : g \text{ concave on } \mathbb{R},\ \ell \le g \le u \text{ on } \mathcal{T}\},$$

$$u^o(x) := \sup\{g(x) : g \text{ concave on } \mathbb{R},\ \ell \le g \le u \text{ on } \mathcal{T}\}.$$

This is a standard and solvable problem. On the one hand, ℓ^o is the smallest concave majorant of ℓ on $\mathcal{T}$ which may be computed via a suitable version of the pool-adjacent-violators algorithm (Robertson et al. 1988). Indeed, there exist indices $0 \le j(0) < j(1) < \cdots < j(b) \le m$ such that

$$\ell^o \begin{cases} \equiv -\infty & \text{on } \mathbb{R} \setminus [t_{j(0)}, t_{j(b)}], \\ \text{is linear on } [t_{j(a-1)}, t_{j(a)}] \text{ for } 1 \le a \le b, \\ \text{changes slope at } t_{j(a)} \text{ if } 1 \le a < b. \end{cases}$$

Having computed ℓ^o, we can check whether $\ell^o \le u$ on $\mathcal{T}$. If this is not the case, there is no concave function fitting in between ℓ and u, and the procedure returns a corresponding error message. Otherwise the value of $u^o(x)$ equals

$$\min\left\{u(s) + \frac{u(s) - \ell^o(r)}{s - r}(x - s) : r \in \mathcal{T}_o, s \in \mathcal{T},\ r < s \le x \text{ or } x \le s < r\right\},$$

where $\mathcal{T}_o = \{t_{j(0)}, t_{j(1)}, \ldots, t_{j(b)}\}$. To maximise $g(x)$ over all concave functions g such that $\ell \le g \le u$, we may assume without loss of generality that for fixed x and a given value y of $g(x)$, the function g is the smallest concave function such that $g \ge \ell_o$ and $g(x) = y$. But the latter function is piecewise linear with changes in slope at x and some points in $\mathcal{T}_o$. Moreover, if y is chosen as large as possible, $g(s)$ has to be equal to $u(s)$ for at least one point $s \in \mathcal{T}$.

Figure 3 illustrates this procedure for a $\mathcal{T}$ that consists of of 21 points. It shows two (parallel) functions ℓ and u evaluated at all points in $\mathcal{T}$, indicated by bullets and interpolating dashed lines. In addition the plot shows the resulting functions ℓ^o and u^o on $\mathcal{T} \cup (-\infty, t_0) \cup (t_m, \infty)$, which are displayed as interpolating solid lines.

In our context, $\mathcal{T}$ is chosen as a fine grid of points such that $t_0 < X_{(1)}$ and $t_m > X_{(n)}$ and $\{X_1, X_2, \ldots, X_n\} \subset \mathcal{T}$. Table 1 contains pseudo-code for our algorithm to compute (L_n^o, U_n^o). We tacitly assume that whenever $\mathrm{ConcInt}(\cdot, \cdot)$ returns an error message, the whole algorithm stops and reports the fact that there is no $G \in \mathcal{F}_{\mathrm{blc}}$ satisfying $L_n \le G \le U_n$.

The next step is to show that our proposed new band (L_n^o, U_n^o) has some desirable properties under rather weak conditions on (L_n, U_n). In particular, both L_n^o and U_n^o are Lipschitz-continuous on $\mathbb{R}$, unless $\inf\{x \in \mathbb{R} : L_n(x) > 0\} \ge \sup\{x \in \mathbb{R} : U_n(x) < 1\}$. Moreover,
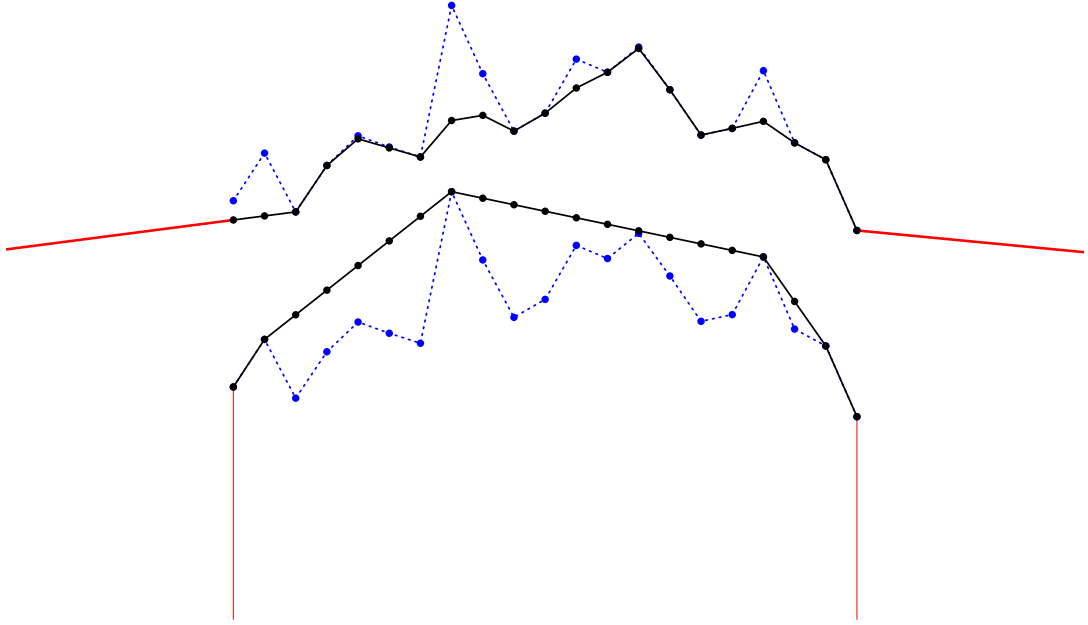
Figure 3: Graphical illustration of the procedure $\mathrm{ConcInt}(\cdot, \cdot)$.

$$(L_n^o, U_n^o) \leftarrow (L_n, U_n)$$
$$(\ell_o, u_o) \leftarrow \mathrm{ConcInt}\big(\log(L_n^o), \log(U_n^o)\big)$$
$$(\widetilde{L}_n^o, \widetilde{U}_n^o) \leftarrow \big(\exp(\ell^o), \exp(u^o)\big)$$
$$(\ell_o, u_o) \leftarrow \mathrm{ConcInt}\big(\log(1 - \widetilde{U}_n^o), \log(1 - \widetilde{L}_n^o)\big)$$
$$(\widetilde{L}_n^o, \widetilde{U}_n^o) \leftarrow \big(1 - \exp(u^o), 1 - \exp(\ell^o)\big)$$

while $(\widetilde{L}_n^o, \widetilde{U}_n^o) \neq (L_n^o, U_n^o)$ do
$$\quad (L_n^o, U_n^o) \leftarrow (\widetilde{L}_n^o, \widetilde{U}_n^o)$$
$$\quad (\ell_o, u_o) \leftarrow \mathrm{ConcInt}\big(\log(L_n^o), \log(U_n^o)\big)$$
$$\quad (\widetilde{L}_n^o, \widetilde{U}_n^o) \leftarrow \big(\exp(\ell^o), \exp(u^o)\big)$$
$$\quad (\ell_o, u_o) \leftarrow \mathrm{ConcInt}\big(\log(1 - \widetilde{U}_n^o), \log(1 - \widetilde{L}_n^o)\big)$$
$$\quad (\widetilde{L}_n^o, \widetilde{U}_n^o) \leftarrow \big(1 - \exp(u^o), 1 - \exp(\ell^o)\big)$$
end while

Table 1: Pseudocode for the computation of (L_n^o, U_n^o).

if $\lim_{x \to \infty} L_n(x) > \lim_{x \to -\infty} U_n(x)$, then $U_n^o(x)$ converges exponentially fast to 0 as $x \to -\infty$ while $L_n^o(x)$ converges exponentially fast to 1 as $x \to \infty$. These properties are implied by the following lemma.

Lemma 2. *For real numbers $a < b$ and $0 < r < s < 1$ define*

$$\gamma_1 := \frac{\log(s/r)}{b - a} \quad and \quad \gamma_2 := \frac{\log\big((1 - r)/(1 - s)\big)}{b - a}.$$

(i) If $L_n(a) \geq r$ and $U_n(b) \leq s$, then L_n^o and U_n^o are Lipschitz-continuous on $\mathbb{R}$ with Lipschitz constant $\max\{\gamma_1, \gamma_2\}$.

(ii) If $U_n(a) \leq r$ and $L_n(b) \geq s$, then

$$U_n^o(x) \leq r \exp(\gamma_1(x - a)) \quad for\ x \leq a$$

and

$$1 - L_n^o(x) \leq (1 - s) \exp(-\gamma_2(x - b)) \quad for\ x \geq b.$$

3.1 A numerical example

We illustrate our methods with a data set from Woolridge (2000). It contains for $n = 177$ randomly chosen companies in the U.S. the annual salaries of their CEOs in 1990, rounded to multiples of 1000 USD. Since it is not clear to us how the rounding has been done, we assume that an observation $Y_{i,\mathrm{raw}} \in \mathbb{N}$ corresponds to an unobserved true salary Y_i within $(Y_{i,\mathrm{raw}} - 1, Y_{i,\mathrm{raw}} + 1)$, and we consider $Y_1, Y_2, \ldots, Y_n$ to be a random sample from a distribution function G on $(0, \infty)$. Salary distributions are well-known to be heavily right-skewed with heavy right tails. A standard model is that $Y \sim G$ has the same distribution as 10^X for some Gaussian random variable X, see Kleiber and Kotz (2003). We assume that the distribution function $F(x) := G(10^x)$ of $X_i := \log_{10}(Y_i)$ is bi-log-concave. More specifically, we compute an unrestricted confidence band (L_n, U_n), where L_n is computed with $\big(\log_{10}(Y_{i,\mathrm{raw}} + 1)\big)_{i=1}^n$ and U_n with $\big(\log_{10}(Y_{i,\mathrm{raw}} - 1)\big)_{i=1}^n$.

Figure 4(a) shows the Kolmogorov-Smirnov 95%-confidence bands for F, without (black lines) and with (blue lines) the restriction of bi-log-concavity. Figure 4(b) shows the confidence bands based on the weighted Kolmogorov-Smirnov 95%-confidence band, where $\gamma = 0.4$. The corresponding quantiles have been estimated in $2 \cdot 10^6$ Monte Carlo simulations. In both cases the shape constraint yields a substantial gain of precision. Notice also that the bounds in Fig-
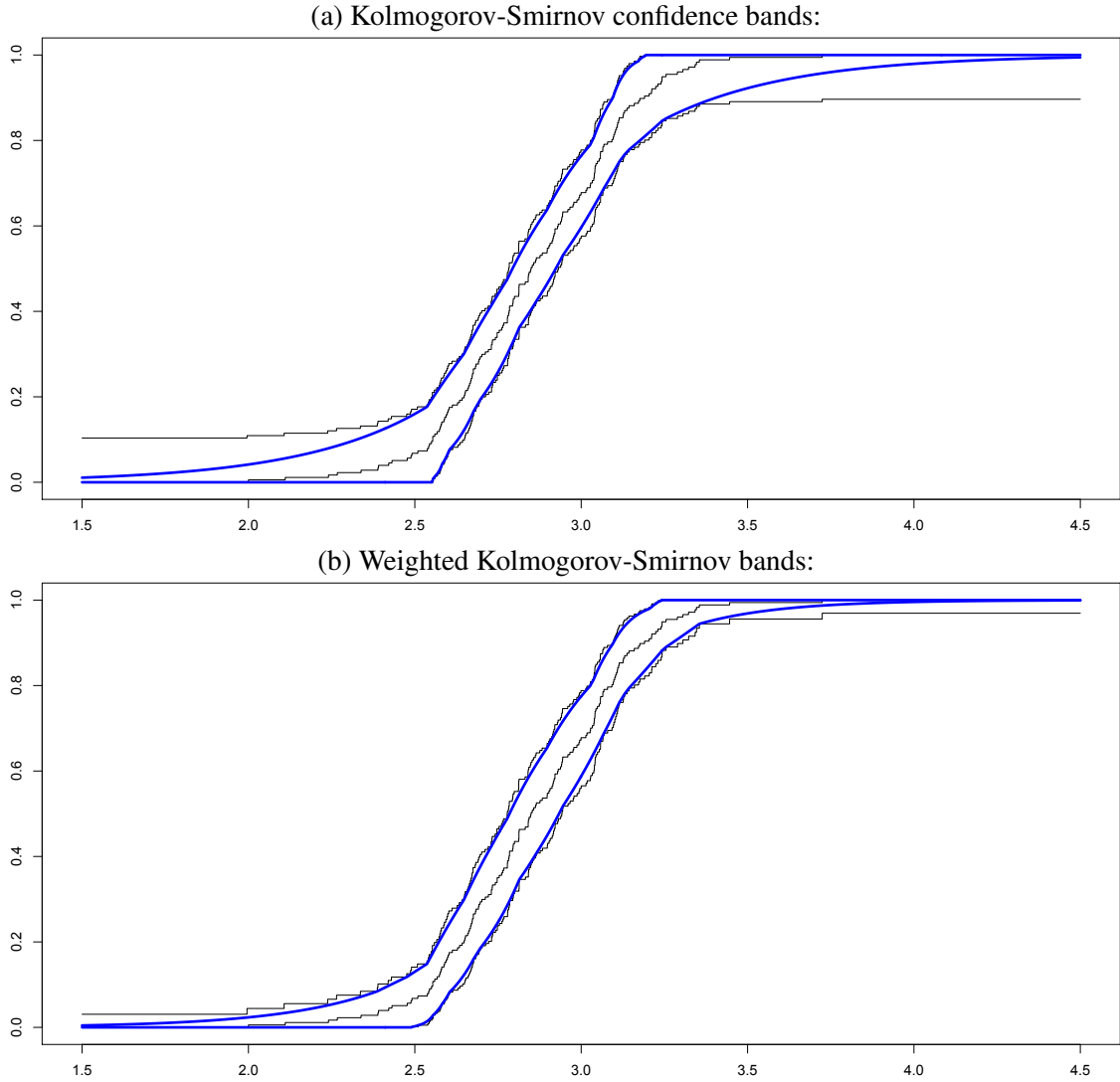
Figure 4: Estimated distribution function with unconstrained and constrained confidence bands for CEO salaries.

ure 4(b) are tighter in the tails but slightly wider in the central part than those in Figure 4(a), for the unconstrained band as well as for the band with shape constraint.

4 Consistency properties

In this section we study the asymptotic behaviour of the proposed confidence band (L_n^o, U_n^o) when $F \in \mathcal{F}_{\mathrm{blc}}$. Our goal is to pinpoint the benefits of utilising the shape constraint of bi-log-concavity. All asymptotic statements refer to $n \to \infty$ while F is fixed.

We start with rather general consistency results for (L_n^o, U_n^o). Recall that we set $L_n^o \equiv 1$ and

$U_n^o \equiv 0$ in the case of no $G \in \mathcal{F}_{\mathrm{blc}}$ fitting in between L_n and U_n, concluding with confidence $1 - \alpha$ that $F \notin \mathcal{F}_{\mathrm{blc}}$. The supremum norm of a function $h : \mathbb{R} \to \mathbb{R}$ is denoted by $\|h\|_\infty = \sup_{x \in \mathbb{R}} |h(x)|$, and for $K \subset \mathbb{R}$ we write $\|h\|_{K,\infty} := \sup_{x \in K} |h(x)|$.

Theorem 3. *Suppose that the original confidence band (L_n, U_n) is consistent in the sense that for any fixed $x \in \mathbb{R}$, both $L_n(x)$ and $U_n(x)$ tend to $F(x)$ in probability.*

(i) *Suppose that $F \notin \mathcal{F}_{\mathrm{blc}}$. Then $\mathbb{P}(L_n^o \le U_n^o) \to 0$.*

(ii) *Suppose that $F \in \mathcal{F}_{\mathrm{blc}}$. Then $\mathbb{P}(L_n^o \le U_n^o) \ge 1 - \alpha$, and*

$$\sup_{G \in \mathcal{F}_{\mathrm{blc}} : L_n \le G \le U_n} \|G - F\|_\infty \to_p 0,$$

where $\sup(\emptyset) := 0$. Moreover, for any compact interval $K \subset J(F)$,

$$\sup_{G \in \mathcal{F}_{\mathrm{blc}} : L_n \le G \le U_n} \|h_G - h_F\|_{K,\infty} \to_p 0,$$

where h_G stands for any of the three functions G', $\log(G)'$ and $\log(1 - G)'$. Finally, for any fixed $x_1 \in J(F)$ and $b_1 < f(x_1)/F(x_1)$,

$$\mathbb{P}\big(U_n^o(x) \le U_n(x') \exp(b_1(x - x')) \text{ for } x \le x' \le x_1\big) \to 1,$$

while for any fixed $x_2 \in J(F)$ and $b_2 < f(x_2)/(1 - F(x_2))$,

$$\mathbb{P}\big(1 - L_n^o(x) \le (1 - L_n(x')) \exp(-b_2(x - x')) \text{ for } x \ge x' \ge x_2\big) \to 1.$$

A direct consequence of Theorem 3 are consistent confidence bounds for functionals $\int \phi \, dF$ of F with well-behaved integrands $\phi : \mathbb{R} \to \mathbb{R}$:

Corollary 4. *Suppose that the original confidence band (L_n, U_n) is consistent, and let $F \in \mathcal{F}_{\mathrm{blc}}$. Let $\phi : \mathbb{R} \to \mathbb{R}$ be absolutely continuous with a derivative ϕ' satisfying the following constraint: For constants $a \in \mathbb{R}$ and $0 \le b_1 < T_1(F)$, $0 \le b_2 < T_2(F)$,*

$$|\phi'(x)| \le \exp(a + b_1 x^- + b_2 x^+)$$

with $x^{\pm} := \max\{\pm x, 0\}$. Then

$$\sup_{G : L_n^o \le G \le U_n^o} \left| \int \phi \, dG - \int \phi \, dF \right| \to_p 0.$$

The previous supremum is meant over all distribution functions G within the confidence band (L_n^o, U_n^o), which is larger than the supremum over all distribution functions $G \in \mathcal{F}_{\mathrm{blc}}$ between L_n

and U_n. Corollary 4 applies to $\phi(x) := e^{tx}$ with $-T_1(F) < t < T_2(F)$. Indeed, the proof of (3) implies the following explicit formulae in the case $L_n^o \leq U_n^o$:

$$\inf_{G\,:\,L_n^o \leq G \leq U_n^o} \int e^{tx} G(dx) = \begin{cases} \int_{\mathbb{R}} te^{tx}(1 - U_n^o(x))\,dx & \text{if } t > 0, \\ \int_{\mathbb{R}} |t| e^{tx} L_n^o(x)\,dx & \text{if } t < 0, \end{cases}$$

$$\sup_{G\,:\,L_n^o \leq G \leq U_n^o} \int e^{tx}\, G(dx) = \begin{cases} \int_{\mathbb{R}} te^{tx}(1 - L_n^o(x))\,dx & \text{if } t > 0, \\ \int_{\mathbb{R}} |t| e^{tx} U_n^o(x)\,dx & \text{if } t < 0. \end{cases}$$

Now we refine Corollary 4 by providing rates of convergence, assuming that the original confidence band (L_n, U_n) satisfies the following property:

Condition (*)   For certain constants $\gamma \in [0, 1/2)$ and $\kappa, \lambda > 0$,

$$\max\{\widehat{F}_n - L_n, U_n - \widehat{F}_n\} \leq \kappa n^{-1/2}(\widehat{F}_n(1 - \widehat{F}_n))^{\gamma}$$

on the interval $\{\lambda n^{-1/(2-2\gamma)} \leq \widehat{F}_n \leq 1 - \lambda n^{-1/(2-2\gamma)}\}$.

Obviously this condition is satisfied with $\gamma = 0$ in the case of the Kolmogorov-Smirnov band. For the weighted Kolmogorov-Smirnov band it is satisfied with the given value of $\gamma \in [0, 1/2)$. In the refined version of Owen's band, it is satisfied for *any* fixed number $\gamma \in (0, 1/2)$.

Theorem 5.   *Suppose that $F \in \mathcal{F}_{\mathrm{blc}}$, and let (L_n, U_n) satisfy Condition (*). Let $\phi : \mathbb{R} \to \mathbb{R}$ be absolutely continuous.*

(i) Suppose that $|\phi'(x)| = O(|x|^{k-1})$ as $|x| \to \infty$ for some number $k \geq 1$. Then

$$\sup_{G\,:\,L_n^o \leq G \leq U_n^o}\left| \int \phi\,dG - \int \phi\,dF \right| = \begin{cases} O_p\big(n^{-1/2}(\log n)^k\big) & \text{if } \gamma = 0, \\ O_p\big(n^{-1/2}\big) & \text{if } \gamma > 0. \end{cases}$$

(ii) Suppose that ϕ satisfies the conditions in Corollary 4. Then

$$\sup_{G\,:\,L_n^o \leq G \leq U_n^o}\left| \int \phi\,dG - \int \phi\,dF \right| = O_p(n^{-\beta}) \tag{6}$$

for any exponent $\beta \in (0, 1/2]$ such that

$$\beta < \frac{1 - \max\{b_1/T_1(F), b_2/T_2(F)\}}{2(1 - \gamma)}.$$

The additional factor $(\log n)^k$ in part (i) cannot be avoided. To verify this we consider $\phi(x) = x^k$ and the distribution function F of a standard exponential random variable X, i.e. $F(x) =$

$1 - e^{-x}$ for $x \geq 0$. Further let F_n be the conditional distribution function of X, given that $X \leq x_n := (\log n)/2 - \log c$ with a fixed $c > 0$. Then both F and F_n are bi-log-concave, $\|F_n - F\|_\infty = e^{-x_n} = cn^{-1/2}$, but

$$\begin{aligned}
\int \phi \, d(F_n - F) &= \mathbb{E}(X^k) - \mathbb{E}(X^k \mid X \leq x_n) \\
&= \mathbb{P}(X > x_n)\big(\mathbb{E}(X^k \mid X > x_n) - \mathbb{E}(X^k \mid X \leq x_n)\big) \\
&\geq \mathbb{P}(X > x_n)\big(x_n^k - \mathbb{E}(X^k)/\mathbb{P}(X \leq x_n)\big) \\
&= 2^{-k} c n^{-1/2} (\log n)^k (1 + o(1)).
\end{aligned}$$

Consequently, if we use the Kolmogorov-Smirnov confidence band, the asymptotic probability of $n^{1/2}\|\widehat{F}_n - F\|_\infty \leq \kappa_{n,\alpha}^{\mathrm{KS}} - c$ is strictly positive, provided that $0 < c < \lim_{n \to \infty} \kappa_{n,\alpha}^{\mathrm{KS}}$. But then F_n satisfies $n^{1/2}\|F_n - \widehat{F}_n\|_\infty \leq \kappa_{n,\alpha}^{\mathrm{KS}}$, so $L_n^o \leq F_n \leq U_n^o$, and the k-th moments of F and F_n differ by $2^{-k} c n^{-1/2} (\log n)^k (1 + o(1))$.

If (L_n^o, U_n^o) is constructed with the refined version of Owen's confidence band, we may choose γ arbitrarily close to $1/2$, so the term $2(1 - \gamma)$ is arbitrarily close to 1. Thus (6) holds for any exponent $\beta \in (0, 1/2]$ such that

$$\beta \; < \; 1 - \max\{b_1/T_1(F), b_2/T_2(F)\}.$$

In particular,

$$\sup_{G \,:\, L_n^o \leq G \leq U_n^o} \left| \int e^{tx} \, G(dx) - \int e^{tx} \, F(dx) \right| \; = \; O_p(n^{-1/2})$$

whenever $-T_1(F)/2 < t < T_2(F)/2$.

5 Proofs

When proving Theorem 1 we assume that the reader is acquainted with the following facts about concave functions that we summarise in two lemmas:

Lemma 6. *Suppose that $h : \mathbb{R} \to [-\infty, +\infty)$ is a concave function. Then it satisfies the following properties:*

(i) h is continuous on the interior of $\{h > -\infty\} := \{x \in \mathbb{R} : h(x) > -\infty\}$.

(ii) For each interior point x of $\{h > -\infty\}$, the left- and right-sided derivatives $h'(x-)$ and $h'(x+)$ exist in $\mathbb{R}$ and satisfy $h'(x-) \geq h'(x+)$. Moreover, $h(x\pm)$ is non-decreasing in x.

(iii) For each interior point x of $\{h > -\infty\}$ and $a \in [h'(x+), h'(x-)]$,

$$h(x + t) \leq h(x) + at \quad \text{for all } t \in \mathbb{R}.$$

Lemma 7. *Let h be a real-valued function on an open interval $J \subset \mathbb{R}$, and let $[a, b] \in [-\infty, \infty]$. Then the following two statements are equivalent:*

(i) For arbitrary different $x, y \in J$,

$$\frac{h(y) - h(x)}{y - x} \in [a, b].$$

(ii) For arbitrary $x \in J$,

$$\liminf_{y \to x} \frac{h(y) - h(x)}{y - x} \geq a \quad \text{and} \quad \limsup_{y \to x} \frac{h(y) - h(x)}{y - x} \leq b.$$

In the case of $[a, b] = [0, \infty]$ or $[a, b] = [-\infty, 0]$, part (i) is equivalent to h being non-decreasing or non-increasing, respectively. In the case of $[a, b] \subset \mathbb{R}$, part (i) is equivalent to h having an L^1-derivative h' on J with values in $[a, b]$.

Lemma 7 follows essentially from a bisection argument and the following observation: For points $r < s < t$ in J,

$$\frac{h(t) - h(r)}{t - r} = \alpha \frac{h(s) - h(r)}{s - r} + (1 - \alpha) \frac{h(t) - h(s)}{t - s}$$

with $\alpha := (s - r)/(t - r) \in (0, 1)$. In particular,

$$\frac{h(t) - h(r)}{t - r} \begin{cases} \geq \min\Big\{ \dfrac{h(s) - h(r)}{s - r}, \dfrac{h(t) - h(s)}{t - s} \Big\}, \\ \leq \max\Big\{ \dfrac{h(s) - h(r)}{s - r}, \dfrac{h(t) - h(s)}{t - s} \Big\}. \end{cases}$$

Proof of Theorem 1. Equivalence of (i-iv) will be verified in four steps.

Proof of (i) $\Rightarrow$ (ii). Suppose that F is bi-log-concave. Since $\log F$ is concave, it follows from Lemma 6 that F is continuous on (a, ∞), where $a := \inf\{F > 0\}$. Furthermore, concavity of $\log(1 - F)$ implies that F is continuous on $(-\infty, b)$ with $b := \sup\{F < 1\} \geq a$. But $a < b$, because otherwise F would be degenerate. Hence F is continuous on $\mathbb{R}$. In particular, $J(F)$ is the open and nonvoid interval (a, b).

Concavity of $h := \log F$ implies that for $a < x < b$ its left- and right-sided derivatives $h'(x-), h'(x+)$ exist in $\mathbb{R}$ and satisfy $h'(x-) \geq h'(x+)$. But then

$$F'(x\pm) = \lim_{t \to 0, \pm t > 0} \frac{\exp(h(x + t)) - \exp(h(x))}{t} = F(x)h'(x\pm)$$

16

exist in $\mathbb{R}$, too, and satisfy the inequality

$$F'(x-) \geq F'(x+).$$

Using similar reasoning one can deduce from concavity of $h := \log(1 - F)$ that

$$-F'(x-) = (1-F)'(x-) \geq (1-F)'(x+) = -F'(x+),$$

so that $F'(x-) = F'(x+)$. This proves differentiability of F on $J(F)$.

Finally, the inequalities (1) follow directly from the last part of Lemma 6, applied to $h = \log F$ and $h = \log(1 - F)$.

Proof of (ii) $\Rightarrow$ (iii). Suppose that F is continuous on $\mathbb{R}$, differentiable on $J(F)$ with derivative $f = F'$ and satisfies the inequalities (1). This implies that $h := f/F$ is non-increasing and $\widetilde{h} := f/(1-F)$ is non-decreasing on $J(F)$. For if $x, y \in J(F)$ with $x < y$, then by (1),

$$\begin{aligned}
\log F(x) &\leq \log F(y) + h(y)(x-y) \\
&\leq \log F(x) + h(x)(y-x) + h(y)(x-y) \\
&= \log F(x) + \big(h(x) - h(y)\big)(y-x)
\end{aligned}$$

and

$$\begin{aligned}
\log(1 - F(x)) &\leq \log(1 - F(y)) - \widetilde{h}(y)(x-y) \\
&\leq \log(1 - F(x)) - \widetilde{h}(x)(y-x) - \widetilde{h}(y)(x-y) \\
&= \log(1 - F(x)) + \big(\widetilde{h}(y) - \widetilde{h}(x)\big)(y-x),
\end{aligned}$$

whence $h(x) \geq h(y)$ and $\widetilde{h}(x) \leq \widetilde{h}(y)$.

Proof of (iii) $\Rightarrow$ (iv). Suppose that F satisfies the conditions in part (iii). First of all this implies $f > 0$ on $J(F)$. Suppose $f(x_o) = 0$ for some $x_o \in J(F)$. Then isotonicity of $\widetilde{h} = f/(1-F)$ implies $f(x) = 0$ for $x \leq x_o$, and antitonicity of $h = f/F$ implies $f(x) = 0$ for $x \geq x_o$. Hence F would be constant on $J(F)$, which violates that F is a continuous distribution function on $\mathbb{R}$. Another consequence of these monotonicity properties is boundedness of f on $J(F)$: If we fix any $x_o \in J(F)$, then for any other point $x \in J(F)$,

$$f(x) = \begin{cases} F(x)h(x) \leq h(x_o) & \text{if } x \geq x_o, \\ (1 - F(x))\widetilde{h}(x) \leq \widetilde{h}(x_o) & \text{if } x \leq x_o. \end{cases}$$

Finally, local Lipschitz-continuity of f may be verified via Lemma 7: Let $c, d \in J(F)$ with $c < d$. For arbitrary different $x, y \in (c, d)$,

$$
\begin{aligned}
\frac{f(y) - f(x)}{y - x} &= \frac{F(y)h(y) - F(x)h(x)}{y - x} \\
&= h(y)\frac{F(y) - F(x)}{y - x} + F(x)\frac{h(y) - h(x)}{y - x} \\
&\leq h(c)\frac{F(y) - F(x)}{y - x} \\
&\leq h(c)\frac{\exp(h(x)(y - x)) - 1}{y - x}F(x) \\
&\to h(c)h(x)F(x) \leq h(c)^2 F(d)
\end{aligned}
$$

as $y \to x$. Hence

$$
\limsup_{y \to x} \frac{f(y) - f(x)}{y - x} \leq h(c)^2 F(d) \quad \text{for all } x \in (c, d). \tag{7}
$$

Analogously, one can show that

$$
\liminf_{y \to x} \frac{f(y) - f(x)}{y - x} \geq -\widetilde{h}(d)^2(1 - F(c)) \quad \text{for all } x \in (c, d). \tag{8}
$$

In particular, f is Lipschitz-continuous on (c, d) with Lipschitz-constant

$$
\max\big\{h(c)^2 F(d), \widetilde{h}(d)^2(1 - F(c))\big\}.
$$

This proves local Lipschitz-continuity of f on $J(F)$. In particular, f is absolutely continuous with L^1-derivative f'. This means, f' is a locally integrable function on $J(F)$ such that

$$
f(y) - f(x) = \int_x^y f'(t)\, dt \quad \text{for all } x, y \in J(F),
$$

and f' may be chosen such that

$$
f'(x) \in \left[\liminf_{y \to x} \frac{f(y) - f(x)}{y - x}, \limsup_{y \to x} \frac{f(y) - f(x)}{y - x}\right]
$$

for any $x \in J(F)$. But for $c, d \in J(F)$ with $c < x < d$, the latter interval is contained in

$$
\left[-\widetilde{h}(d)^2(1 - F(c)), h(c)^2 F(d)\right] = \left[\frac{-f(d)^2(1 - F(c))}{(1 - F(d))^2}, \frac{f(c)^2 F(d)}{F(c)^2}\right]
$$

according to (7) and (8). Since F and f are continuous, letting $c, d \to x$ implies (2).

Proof of (iv) $\Rightarrow$ (i). One can easily verify that a continuous distribution function F is bi-log-concave if, and only if, $\log F$ and $\log(1 - F)$ are concave on $J(F)$. Hence (i) is a consequence of (iii), and it suffices to show that (iv) implies (iii).

According to Lemma 7, h is non-increasing on $J(F)$ if, and only if,

$$\limsup_{y \to x} \frac{h(y) - h(x)}{y - x} \le 0$$

for any $x \in J(F)$. To verify this, let $y \in J(F) \setminus \{x\}$ and set $r := \min(x, y)$, $s := \max(x, y)$. Then it follows from (2) and from continuity of f that

$$
\begin{aligned}
\frac{h(y) - h(x)}{y - x} &= \frac{f(y)/F(y) - f(x)/F(x)}{y - x} \\
&= \frac{1}{F(y)} \frac{f(y) - f(x)}{y - x} - \frac{f(x)}{F(x)F(y)} \frac{F(y) - F(x)}{y - s} \\
&= \frac{1}{F(y)(s - r)} \int_r^s f'(t)\,dt - \frac{f(x)}{F(x)F(y)(s - r)} \int_r^s f(t)\,dt \\
&\le \frac{1}{F(y)(s - r)} \int_r^s \frac{f(t)^2}{F(t)}\,dt - \frac{f(x)}{F(x)F(y)(s - r)} \int_r^s f(t)\,dt \\
&\to \frac{f(x)^2}{F(x)^2} - \frac{f(x)^2}{F(x)^2} = 0
\end{aligned}
$$

as $y \to x$.

Analogously one can show that $\widetilde{h}$ is non-decreasing on $J(F)$. $\qquad\square$

Proof of (3). For any fixed $x_o \in J(F)$, monotonicity of $f/F = \log(F)'$ implies that for $x \in J(F)$, $x < x_o$,

$$\frac{f}{F}(x) \ge \frac{\log F(x_o) - \log F(x)}{x - x_o}.$$

Since $\log F(x) \to -\infty$ as $x \to \inf(J(F))$, this inequality implies that

$$T_1(F) = \sup_{x \in J(F)} \frac{f}{F}(x) = \lim_{x \to \inf(J(F))} \frac{f}{F}(x) \begin{cases} > 0, \\ = \infty \quad \text{if } \inf(J(F)) > -\infty. \end{cases}$$

Analogously one can show that

$$T_2(F) = \sup_{x \in J(F)} \frac{f}{1 - F}(x) = \lim_{x \to \sup(J(F))} \frac{f}{1 - F}(x) \begin{cases} > 0, \\ = \infty \quad \text{if } \sup(J(F)) < \infty. \end{cases}$$

For symmetry reasons it suffices to show that $\int e^{tx} F(dx)$ is finite for $t \in (0, T_2(F))$ and infinite for $t \ge T_2(F)$. Notice that for $t > 0$, Fubini's theorem yields

$$
\begin{aligned}
\int e^{tx} F(dx) &= \int \int 1_{[z \le x]} t e^{tz}\,dz\, F(dx) \\
&= t \int e^{tz}(1 - F(z))\,dz \\
&= t \int \exp\big(tz + \log(1 - F(z))\big)\,dz.
\end{aligned}
$$

In the case of $m := \sup(J(F)) < \infty$, the previous integral is smaller than $e^{tm} < \infty$ for $t < \infty = T_2(F)$. In the case of $m = \infty$, notice that $tz + \log(1 - F(z))$ is concave in $z \in \mathbb{R}$ with limit $-\infty$ as $z \to -\infty$. Thus the integral $\int e^{tx} F(dx)$ is finite if, and only if,

$$\lim_{z \to \infty} \frac{d}{dz}\big(tz + \log(1 - F(z))\big) \;=\; \lim_{z \to \infty}\Big(t - \frac{f(z)}{1 - F(z)}\Big) \;=\; t - T_2(F)$$

is strictly negative, which is equivalent to $t < T_2(F)$. $\qquad\square$

Proof of Lemma 2. The assertions are trivial if $L_n^o \equiv 1$ and $U_n^o \equiv 0$, meaning that no $G \in \mathcal{F}_{\mathrm{blc}}$ fits in between L_n and U_n. Otherwise let $G \in \mathcal{F}_{\mathrm{blc}}$ such that $L_n \le G \le U_n$.

For part (i) it suffices to show that for any $x \in J(G)$ the density $g = G'$ satisfies the inequality $g(x) \le \max\{\gamma_1, \gamma_2\}$. This is equivalent to Lipschitz-continuity of G with the latter constant, and this property carries over to the pointwise infimum L_n^o and supremum U_n^o. For $x \ge b$ it follows from concavity of $\log G$ and $G(a) \ge r$, $G(b) \le s$ that

$$g(x) \;\le\; \frac{g}{G}(x) \;\le\; \frac{g}{G}(b) \;\le\; \frac{\log G(b) - \log G(a)}{b - a} \;\le\; \frac{\log s - \log r}{b - a} \;=\; \gamma_1.$$

Similarly, convexity of $-\log(1 - G)$ and the inequalities $G(a) \ge r$, $G(b) \le s$ imply

$$g(x) \;\le\; \frac{g}{1 - G}(x) \;\le\; \frac{g}{1 - G}(a) \;\le\; \frac{-\log(1 - G(b)) + \log(1 - G(a))}{b - a} \;\le\; \gamma_2$$

for $x \le a$. For $a < x < b$ we obtain the two inequalities

$$g(x) \;=\; G(x)\frac{g}{G}(x) \;\le\; G(x)\frac{\log G(x) - \log r}{x - a}$$

and

$$g(x) \;=\; (1 - G(x))\frac{g}{1 - G}(x) \;\le\; (1 - G(x))\frac{\log(1 - G(x)) - \log(1 - s)}{b - x}.$$

The former inequality times $x - a$ plus the latter inequality times $b - x$ yields

$$g(x) \;\le\; \frac{G(x)\log(G(x)/r) + (1 - G(x))\log\big((1 - G(x))/(1 - s)\big)}{b - a}.$$

But $h(y) := y\log(y/r) + (1 - y)\log((1 - y)/(1 - s))$ is easily shown to be convex in $y \in (0, 1)$, so

$$g(x) \;\le\; \max_{y = r, s} h(y) \;=\; \max\{\gamma_1, \gamma_2\}.$$

As to part (ii), it suffices to show that $G(x) \le G(a)\exp(\gamma_1(x - a))$ for $x \le a$ and $G(x) \ge 1 - (1 - G(b))\exp(-\gamma_2(x - b))$ for $x \ge b$. We know from Theorem 1 (ii) that this is true with

$(g/G)(a)$ and $(g/(1-G))(b)$ in place of γ_1 and γ_2, respectively. But it follows from $G(a) \leq r$, $G(b) \geq s$ and concavity of $\log G$ that

$$\frac{g}{G}(a) \geq \frac{\log G(b) - \log G(a)}{b - a} \geq \frac{\log s - \log r}{b - a} = \gamma_1,$$

while convexity of $-\log(1-G)$ yields that $(g/(1-G))(b) \geq \gamma_2$. $\qquad\square$

Proof of Theorem 3. Suppose that $F \notin \mathcal{F}_{\mathrm{blc}}$. This means, either $\log F$ or $\log(1-F)$ or both are not concave. When $\log F$ is not concave there exist real numbers $x_0 < x_1 < x_2$ such that $\log F(x_1) < (1-\lambda)\log F(x_0) + \lambda \log F(x_2)$, where $\lambda := (x_1 - x_0)/(x_2 - x_0) \in (0,1)$. Then with probability tending to one, $\log U_n(x_1) < (1-\lambda)\log L_n(x_0) + \lambda \log L_n(x_2)$, whence no log-concave distribution function fits between L_n and U_n. Analogous arguments apply in the case of $\log(1-F)$ violating concavity.

Now suppose that $F \in \mathcal{F}_{\mathrm{blc}}$. Obviously, $\mathbb{P}(L_n^o \leq U_n^o) \geq \mathbb{P}(L_n \leq F \leq U_n) \geq 1 - \alpha$. Since L_n and U_n are assumed to be non-decreasing, and since F is continuous, a standard argument shows that pointwise convergence implies uniform convergence in probability, i.e. $\|L_n - F\|_\infty \to_p 0$ and $\|U_n - F\|_\infty \to_p 0$. This implies that

$$\sup_{G \in \mathcal{F}_{\mathrm{blc}} : L_n \leq G \leq U_n} \|G - F\|_\infty \leq \|L_n - F\|_\infty + \|U_n - F\|_\infty \to_p 0, \tag{9}$$

because $L_n \leq L_n^o \leq U_n^o \leq U_n$ in the case of $L_n^o \leq U_n^o$.

Now let K be a compact subset of $J(F)$, and let $h_G := \log(G)'$ for $G \in \mathcal{F}_{\mathrm{blc}}$. Since $h_F = f/F$ is continuous and non-increasing on $J(F)$, for any fixed $\varepsilon > 0$ there exist points $a_0 < a_1 < \cdots < a_m < a_{m+1}$ in $J(F)$ such that $K \subset [a_1, a_m]$ and

$$0 \leq h_F(a_{i-1}) - h_F(a_i) \leq \varepsilon \quad \text{for } 1 \leq i \leq m+1.$$

For $G \in \mathcal{F}_{\mathrm{blc}}$ with $L_n \leq G \leq U_n$, for any $x \in K$ it follows from monotonicity of h_F and h_G that

$$
\begin{aligned}
\sup_{x \in K} \big(h_G(x) - h_F(x)\big) &\leq \max_{i=1,\ldots,m-1} \big(h_G(a_i) - h_F(a_{i+1})\big) \\
&\leq \max_{i=1,\ldots,m-1} \left(\frac{\log G(a_i) - \log G(a_{i-1})}{a_i - a_{i-1}} - h_F(a_{i+1}) \right) \\
&\leq \max_{i=1,\ldots,m-1} \left(\frac{\log U_n(a_i) - \log L_n(a_{i-1})}{a_i - a_{i-1}} - h_F(a_{i+1}) \right) \\
&= \max_{i=1,\ldots,m-1} \left(\frac{\log F(a_i) - \log F(a_{i-1})}{a_i - a_{i-1}} - h_F(a_{i+1}) \right) + o_p(1) \\
&\leq \max_{i=1,\ldots,m-1} \big(h_F(a_{i-1}) - h_F(a_{i+1})\big) + o_p(1) \\
&\leq 2\varepsilon + o_p(1).
\end{aligned}
$$

Analogously,

$$\sup_{x \in K} \big(h_F(x) - h_G(x) \big) \leq \max_{i=1,\dots,m-1} \big(h_F(a_i) - h_F(a_{i+2}) \big) + o_p(1)$$
$$\leq 2\varepsilon + o_p(1).$$

Since $\varepsilon > 0$ is arbitrarily small, this shows that

$$\sup_{G \in \mathcal{F}_{\mathrm{blc}} \,:\, L_n \leq G \leq U_N} \big\| \log(G)' - \log(F)' \big\|_{K,\infty} = o_p(1). \tag{10}$$

Using similar reasoning one can show that

$$\sup_{G \in \mathcal{F}_{\mathrm{blc}} \,:\, L_n \leq G \leq U_N} \big\| \log(1-G)' - \log(1-F)' \big\|_{K,\infty} = o_p(1).$$

Moreover, since $G' = \log(G)'\, G$, it follows from (9) and (10) that

$$\sup_{G \in \mathcal{F}_{\mathrm{blc}} \,:\, L_n \leq G \leq U_N} \big\| G' - F' \big\|_{K,\infty} = o_p(1).$$

Finally, let $x_1 < \sup(J(F))$ and $b_1 < f(x_1)/F(x_1)$. As in the proof of Lemma 2 (ii) one may argue that for any fixed $x_1' > x_1$, $x_1' \in J(F)$,

$$U_n^o(x) \leq U_n(x') \exp\Big(\frac{\log L_n(x_1') - \log U_n(x_1)}{x_1' - x_1} (x - x') \Big)$$

for all $x \leq x' \leq x_1$. But

$$\frac{\log L_n(x_1') - \log U_n(x_1)}{x_1' - x_1} \to_p \frac{\log F(x_1') - \log F(x_1)}{x_1' - x_1} > b_1$$

if $x_1 \leq \inf(J(F))$ or x_1' is sufficiently close to $x_1 \in J(F)$. This shows that asymptotically with probability one,

$$U_n^o(x) \leq U_n(x') \exp(b_1(x - x'))$$

for all $x \leq x' \leq x_1$. Analogously, one can prove the claim about $1 - L_n^o$ on halflines $[x_2, \infty)$, $x_2 > \inf(J(F))$. $\qquad\qquad \square$

Proof of Corollary 4. Without loss of generality let $0 \in J(F)$; otherwise we could shift the coordinate system suitably and adjust the constant a in our bound for $|\phi'|$. Notice that for any $z \in \mathbb{R}$,

$$\phi(z) - \phi(0) = \int_{-\infty}^{\infty} \big(1_{[0 \leq x < z]} - 1_{[z \leq x < 0]} \big) \phi'(x)\, dx,$$

so by Fubini's theorem,

$$\int \phi\, dG = \phi(0) + \int_{\mathbb{R}} \phi'(x) \big(1_{[x \geq 0]} - G(x) \big)\, dx,$$

provided that

$$\int |\phi'(x)| \big| 1_{[x \geq 0]} - G(x) \big| \, dx \ < \ \infty. \tag{11}$$

By assumption, for arbitrary numbers $b'_1 \in (0, T_1(F))$ and $b'_2 \in (0, T_2(F))$ there exist points $x_1, x_2 \in J(F)$ with $x_1 \leq 0 \leq x_2$ and

$$f(x_1)/F(x_1) \ > \ b'_1, \quad f(x_2)/(1 - F(x_2)) \ > \ b'_2.$$

Then it follows from Theorem 3 (ii) that asymptotically with probability one,

$$U_n^o(x) \ \leq \ U_n(x') \exp(b'_1(x - x')) \quad \text{for } x \leq x' \leq x_1 \tag{12}$$

and

$$1 - L_n^o(x) \ \leq \ (1 - L_n(x')) \exp(-b'_2(x - x')) \quad \text{for } x \geq x' \geq x_2. \tag{13}$$

If we choose $b'_1 > b_1$ and $b'_2 > b_2$, the inequalities (12) and (13) imply (11) for arbitrary distribution functions G with $L_n^o \leq G \leq U_n^o$. More precisely, for any fixed $c \geq 0$ and $\delta := \min\{b'_1 - b_1, b'_2 - b_2\} > 0$,

$$\int_{-\infty}^{x_1-c} |\phi'(x)| U_n^o(x) \, dx \ \leq \ U_n(x_1) \int_{-\infty}^{x_1-c} \exp(a - b_1 x + b'_1(x - x_1)) \, dx$$

$$\leq \ U_n(x_1) \exp(a - b_1 x_1 - \delta c) \int_{-\infty}^{0} \exp(\delta y) \, dy$$

$$= \ \frac{U_n(x_1) \exp(a - b_1 x_1 - \delta c)}{\delta}$$

and

$$\int_{x_2+c}^{\infty} |\phi'(x)| (1 - L_n^o(x)) \, dx \ \leq \ \frac{(1 - L_n(x_1)) \exp(a + b_2 x_2 - \delta c)}{\delta}.$$

The same inequalities hold if L_n, U_n, L_n^o and U_n^o are all replaced by F. Thus

$$\sup_{G : L_n^o \leq G \leq U_n^o} \left| \int \phi \, dG - \int \phi \, dF \right| \ = \ \sup_{G : L_n^o \leq G \leq U_n^o} \left| \int_{-\infty}^{\infty} \phi'(x)(F - G)(x) \, dx \right| \tag{14}$$

is not larger than

$$\sup_{G : L_n^o \leq G \leq U_n^o} \|G - F\|_{\infty} \int_{x_1-c}^{x_2+c} |\phi'(x)| \, dx$$

$$+ \int_{-\infty}^{x_1-c} |\phi'(x)| (U_n^o + F)(x) \, dx + \int_{x_2+c}^{\infty} |\phi'(x)| (2 - L_n^o - F)(x) \, dx$$

$$\leq \ \frac{2F(x_1) \exp(a - b_1 x_1 - \delta c)}{\delta} + \frac{2(1 - F(x_2)) \exp(a + b_2 x_2 - \delta c)}{\delta} + o_p(1).$$

But the limit on the right hand side becomes arbitrarily small for sufficiently large $c > 0$. $\qquad \square$

23

Proof of Theorem 5. It follows from standard results about the empirical process on the real line that for any fixed $\varepsilon \in (0,1)$ there exists a constant $\kappa_\varepsilon > 0$ such that with probability at least $1 - \varepsilon$,

$$|\widehat{F}_n - F| \leq \kappa_\varepsilon n^{-1/2}(F(1 - F))^\gamma$$

on $\mathbb{R}$. Let us assume that the previous inequalities hold and that $L_n^o \leq U_n^o$.

For a constant $\lambda_\varepsilon > 0$ to be specified later it follows from $\lambda_\varepsilon n^{-1/(2-2\gamma)} \leq F \leq 1 - \lambda_\varepsilon n^{-1/(2-2\gamma)}$ that

$$\widehat{F}_n \geq \Big(1 - \frac{|\widehat{F}_n - F|}{F}\Big)F \geq (1 - \kappa_\varepsilon \lambda_\varepsilon^{\gamma-1})\lambda_\varepsilon n^{-1/(2-2\gamma)} = (\lambda_\varepsilon - \kappa_\varepsilon \lambda_\varepsilon^\gamma)n^{-1/(2-2\gamma)}$$

and

$$1 - \widehat{F}_n \geq (\lambda_\varepsilon - \mu_\varepsilon \lambda_\varepsilon^\gamma)n^{-1/(2-2\gamma)}.$$

Thus we choose λ_ε sufficiently large such that the number $\lambda_\varepsilon - \kappa_\varepsilon \lambda_\varepsilon^\gamma$ exceeds λ. Then the interval

$$J_n := \{\lambda_\varepsilon n^{-1/(2-2\gamma)} \leq F \leq 1 - \lambda_\varepsilon n^{-1/(2-2\gamma)}\}$$

is a subset of $\{\lambda n^{-1/(2-2\gamma)} \leq \widehat{F}_n \leq 1 - \lambda n^{-1/(2-2\gamma)}\}$. On this interval J_n,

$$\frac{\widehat{F}_n(1 - \widehat{F}_n)}{F(1 - F)} \leq \max\Big\{\frac{\widehat{F}_n}{F}, \frac{1 - \widehat{F}_n}{1 - F}\Big\} \leq 1 + \frac{|\widehat{F}_n - F|}{\min(F, 1 - F)} \leq 1 + \kappa_\varepsilon \lambda_\varepsilon^{\gamma-1},$$

and for any function h with $L_n \leq h \leq U_n$,

$$\frac{|h - F|}{(F(1 - F))^\gamma} \leq \frac{|h - \widehat{F}_n|}{(\widehat{F}_n(1 - \widehat{F}_n))^\gamma}\Big(\frac{\widehat{F}_n(1 - \widehat{F}_n)}{F(1 - F)}\Big)^\gamma + \frac{|\widehat{F}_n - F|}{(F(1 - F))^\gamma} \leq \nu_\varepsilon n^{-1/2} \qquad (15)$$

with $\nu_\varepsilon := \kappa(1 + \kappa_\varepsilon \lambda_\varepsilon^{\gamma-1})^\gamma + \kappa_\varepsilon$. In particular, the boundaries L_n and U_n themselves satisfy (15) on J_n.

Again we assume without loss of generality that $0 \in J(F)$. For arbitrary fixed numbers $b_1' \in (0, T_1(F))$ and $b_2' \in (0, T_2(F))$ we choose points $x_1, x_2 \in J(F)$ with $x_1 < 0 < x_2$ such that $f(x_1)/F(x_1) > b_1'$ and $f(x_2)/(1 - F(x_2)) > b_2'$. For sufficiently large n, $[x_1, x_2] \subset J_n$, and we may even assume that (12) and (13) are satisfied, too. Writing $J_n = [x_{n1}, x_{n2}]$, we can deduce from (14) and (15) that

$$\sup_{G : L_n^o \leq G \leq U_n^o}\Big|\int \phi\, d(G - F)\Big| \leq \nu_\varepsilon n^{-1/2}\int_{x_{n1}}^{x_{n2}} |\phi'(x)| F(x)^\gamma(1 - F(x))^\gamma\, dx$$

$$+ \int_{-\infty}^{x_{n1}} |\phi'(x)|(F + U_n^o)(x)\, dx$$

$$+ \int_{x_{n2}}^\infty |\phi'(x)|(2 - F - L_n^o)(x)\, dx.$$

24

Notice that

$$F(x) \leq F(x_1) \exp(b'_1(x - x_1)) \quad \text{for } x \leq x_1,$$

$$1 - F(x) \leq (1 - F(x_2)) \exp(-b'_2(x - x_2)) \quad \text{for } x \geq x_2.$$

In particular, for $x = x_{n1}, x_{n2}$ it follows from these inequalities and $F(x_{n1}) = 1 - F(x_{n2}) = \lambda_\varepsilon n^{-1/(2-2\gamma)}$ that

$$x_{n1} \geq O(1) - \frac{\log n}{b'_1(2 - 2\gamma)} \quad \text{and} \quad x_{n2} \leq O(1) + \frac{\log n}{b'_2(2 - 2\gamma)}. \tag{16}$$

Notice also that by (12), (13) and (15),

$$(F + U_n^o)(x) \leq (F + U_n^o)(x_{n1}) \exp(b'_1(x - x_{n1}))$$

$$\leq \omega_\varepsilon n^{-1/(2-2\gamma)} \exp(b'_1(x - x_{n1})) \quad \text{for } x \leq x_{n1},$$

$$(2 - F - L_n^o)(x) \leq \omega_\varepsilon n^{-1/(2-2\gamma)} \exp(-b'_2(x - x_{n2})) \quad \text{for } x \geq x_{n2},$$

where $\omega_\varepsilon := \lambda_\varepsilon + \nu_\varepsilon \lambda_\varepsilon^\gamma$. These considerations show that

$$\sup_{G : L_n^o \leq G \leq U_n^o} \left| \int \phi \, d(G - F) \right| \leq I_{n0} + I_{n1} + I'_{n1} + I_{n2} + I'_{n2}$$

with

$$I_{n0} := \nu_\varepsilon n^{-1/2} \int_{x_1}^{x_2} |\phi'(x)| \, dx = O(n^{-1/2}),$$

$$I_{n1} := \nu_\varepsilon n^{-1/2} \int_{x_{n1}}^{x_1} |\phi'(x)| F(x)^\gamma \, dx = O\left(n^{-1/2} \int_{x_{n1}}^{x_1} |\phi'(x)| e^{\gamma b'_1 x} \, dx\right),$$

$$I'_{n1} := \int_{-\infty}^{x_{n1}} |\phi'(x)| (F + U_n^o)(x) \, dx = O\left(n^{-1/(2-2\gamma)} \int_{-\infty}^{x_{n1}} |\phi'(x)| e^{b'_1(x - x_{n1})} \, dx\right),$$

$$I_{n2} := \nu_\varepsilon n^{-1/2} \int_{x_2}^{x_{n2}} |\phi'(x)| (1 - F(x))^\gamma \, dx = O\left(n^{-1/2} \int_{x_2}^{x_{n2}} |\phi'(x)| e^{-\gamma b'_2 x} \, dx\right),$$

$$I'_{n2} := \int_{x_{n2}}^{\infty} |\phi'(x)| (2 - F - L_n^o)(x) \, dx = O\left(n^{-1/(2-2\gamma)} \int_{x_{n2}}^{\infty} |\phi'(x)| e^{-b'_2(x - x_{n2})} \, dx\right).$$

As to part (i), suppose that $|\phi'(x)| \leq a(1 + |x|^{k-1})$ for arbitrary $x \in \mathbb{R}$ and some constant $a > 0$. Then both I_{n1} and I_{n2} are of order

$$O\left(n^{-1/2} \int_0^{O(\log n)} (1 + s^{k-1}) \exp(-\gamma b' s) \, ds\right) = \begin{cases} O(n^{-1/2}) & \text{if } \gamma > 0, \\ O(n^{-1/2}(\log n)^k) & \text{if } \gamma = 0, \end{cases}$$

where $b' := \min\{b'_1, b'_2\} > 0$. Moreover, both I'_{n1} and I'_{n2} are of order

$$O\left(n^{-1/(2-2\gamma)} \int_0^\infty O((\log n)^{k-1} + s^{k-1}) e^{-b' s} \, ds\right) = O(n^{-1/(2-2\gamma)}(\log n)^{k-1})$$

$$= \begin{cases} o(n^{-1/2}) & \text{if } \gamma > 0, \\ O(n^{-1/2}(\log n)^{k-1}) & \text{if } \gamma = 0. \end{cases}$$

25

This proves the assertion in part (i).

For functions ϕ as in part (ii), let $b_1' > b_1$ and $b_2' > b_2$ such that $b_1 \neq \gamma b_1'$ and $b_2 \neq \gamma b_2'$. Then

$$I_{n1} = O\left(n^{-1/2} \int_0^{O(1)+(\log n)/(b_1'(2-2\gamma))} \exp((b_1 - \gamma b_1')s)\,ds\right) = O(n^{-\beta_1})$$

with

$$\beta_1 := \frac{1}{2} - \frac{(b_1 - \gamma b_1')^+}{b_1'(2-2\gamma)} = \frac{1 - \gamma - (b_1/b_1' - \gamma)^+}{2(1-\gamma)} = \frac{1 - \max(b_1/b_1', \gamma)}{2(1-\gamma)},$$

and

$$I_{n2} = O(n^{-\beta_2}) \quad \text{with} \quad \beta_2 := \frac{1 - \max(b_2/b_2', \gamma)}{2(1-\gamma)}.$$

Furthermore,

$$\begin{aligned}
I_{n1}' &= O\left(n^{-1/(2-2\gamma)} \int_{-\infty}^{x_{n1}} \exp(-b_1 x + b_1'(x - x_{n1}))\,dx\right) \\
&= O\left(n^{-1/(2-2\gamma)} \exp(-b_1 x_{n1}) \int_0^\infty \exp(-(b_1' - b_1)s)\,ds\right) \\
&= O\left(n^{-1/(2-2\gamma)} \exp(-b_1 x_{n1})\right) = O\left(n^{-(1-b_1/b_1')/(2-2\gamma)}\right) = O(n^{-\beta_1})
\end{aligned}$$

and

$$I_{n2}' = O(n^{-\beta_2}).$$

This proves the assertion in part (ii). If $\widetilde{\gamma} := \max\{b_1/T_1(F), b_2/T_2(F)\} < \gamma$, we may choose b_1' and b_2' such that $b_1/b_1', b_2/b_2' < \gamma$, resulting in $\beta_1 = \beta_2 = 1/2$. If $\widetilde{\gamma} \geq \gamma$, the exponents β_1, β_2 are strictly smaller than but arbitrarily close to $(1 - \widetilde{\gamma})/(2(1-\gamma))$. $\qquad\square$

References

[1] BAGNOLI, M. and BERGSTROM, T. (2005). Log-concave probability and its applications. *Econ. Theory* **26**, 445-469.

[2] BAHADUR, R.R. and SAVAGE, L.J. (1956). The nonexistence of certain statistical procedures in nonparametric problems. *Ann. Math. Statist.* **27**, 1115-1122.

[3] BERK, R.H. and JONES, D.H. (1979). Goodness-of-fit statistics that dominate the Kolmogorov statistics. *Z. Wahrsch. Verw. Gebiete* **47**, 47-59.

[4] CULE, M.L., SAMWORTH, R.J. and STEWART, M.I. (2010). Maximum likelihood estimation of a multidimensional log-concave density (with discussion). *J. Roy. Statist. Soc., Ser. B* **72**, 545-600.

[5] DÜMBGEN, L. (1998). New goodness-of-fit tests and their application to nonparametric confidence sets. *Ann. Statist.* **26(1)**, 288-314.

[6] DÜMBGEN, L. and RUFIBACH, K. (2009). Maximum likelihood estimation of a log-concave density and its distribution function: basic properties and uniform consistency. *Bernoulli* **15(1)**, 40-68.

[7] DÜMBGEN, L. and RUFIBACH, K. (2011). logcondens: Computations related to univariate log-concave density estimation. *J. Statist. Software* **39(6)**.

[8] DÜMBGEN, L., SAMWORTH, R.J. and SCHUHMACHER, D. (2011). Approximation by log-concave distributions, with applications to regression. *Ann. Statist.* **39(2)**, 702-730.

[9] DÜMBGEN, L. and WELLNER, J.A. (2014). Confidence bands for distribution functions: A new look at the law of the iterated logarithm. Preprint.

[10] KLEIBER, C. and KOTZ, S. (2003). *Statistical size distributions in economics and actuarial sciences.* Wiley Series in Probability and Statistics.

[11] MASSART, P. (1990). The tight constant in the Dvoretzky-Kiefer-Wolfowitz Inequality. *Ann. Probab.* **18**, 1269-1283.

[12] OWEN, A.B. (1995). Nonparametric likelihood confidence bands for a distribution function. *J. Amer. Statist. Assoc.* **90(430)**, 516-521.

[13] ROBERTSON, T., WRIGHT, F.T. and DYKSTRA, R.L. (1988). *Order restricted statistical inference*, Wiley & Sons.

[14] SCHUHMACHER, D., HÜSLER, A. and DÜMBGEN, L. (2011). Log-concave distributions as a nearly parametric model. *Statist. Risk Model.* **28(3)**, 277-295.

[15] SEREGIN, A. and WELLNER, J.A. (2010). Nonparametric estimation of multivariate convex-transformed densities. *Ann. Statist.* **38(6)**, 3751-3781.

[16] SHORACK, G.R. and WELLNER, J.A. (1986). *Empirical processes with applications to statistics.* Wiley, New York.

[17] WALTHER, G. (2009). Inference and modeling with log-concave distributions. *Statist. Sci.* **24**, 319-327.

[18] WOOLRIDGE, J.M. (2000). CEOSAL2 (Instructional Stata data sets for Econometrics). Boston College Department of Economics.