Bi-Log-Concave Distribution and Regression Functions

Inauguraldissertation

der Philosophisch-naturwissenschaftlichen Fakultät

der Universität Bern

vorgelegt von

Petro Kolesnyk

von Kiew, Ukraine

Leiter der Arbeit:

Prof. Dr. L. Dümbgen

Institut für mathematische Statistik und Versicherungslehre

Bi-Log-Concave Distribution and Regression Functions

Inauguraldissertation
der Philosophisch-naturwissenschaftlichen Fakultät
der Universität Bern

vorgelegt von

Petro Kolesnyk
von Kiew, Ukraine

Leiter der Arbeit:
Prof. Dr. L. Dümbgen
Institut für mathematische Statistik und Versicherungslehre

Von der Philosophisch-naturwissenschaftlichen Fakultät angenommen.

Bern, 10. 11. 2016

Der Dekan:
Prof. Dr. G. Colangelo

Acknowledgements

First of all, I would like to thank Lutz Dümbgen for his top-notch support throughout my doctoral project. His enthusiastic attitude and patience made it a big pleasure to work together.

A special thanks to Ralf Wilke for his valuable inputs.

I am also grateful to Prof. Dr. Richard Samworth for reviewing my thesis.

Many thanks to David Ginsbourger, Ilya Molchanov, Riccardo Gatto, Michael Schmutz, Johanna Ziegel, Benjamin Baumgartner, Niki Zumbrunnen, Dominic and Heike Schuhmacher, Alain Hauser, Felix Nagel and all other members of IMSV for helpful and interesting discussions, after-lunch coffee and other pleasant breaks and activities.

This work was supported by the research group FOR916 of the Swiss National Science Foundation (SNF), the Deutsche Forschungsgemeinschaft (DFG) and by IMSV.

Contents

Overview

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. An interesting alternative to smoothness assumptions and data smoothing methods (cf. Silverman (1986)) are qualitative constraints, e.g. monotonicity, concavity or log-concavity. On the other hand, estimation of a distribution function based on independent, identically distributed random variables $X_1, X_2, \ldots, X_n$ with cumulative distribution function F is a common practice and does not require restrictive assumptions. But non-trivial confidence regions for certain functionals of F such as the mean do not exist without substantial additional constraints (cf. Bahadur and Savage (1956)).

In density estimation, a particular constraint which attracted considerable attention recently is log-concavity. That means, one estimates a probability density f on $\mathbb{R}^d$ under the constraint that $\log f : \mathbb{R}^d \to [-\infty, \infty)$ is a concave function. The research topic of the present dissertation was motivated and inspired by results on log-concavity and related constraints that were obtained in particular by Bagnoli and Bergstrom (2005), Cule et al. (2010), Dümbgen and Rufibach (2009, 2011), Walther (2009), Seregin and Wellner (2010), Dümbgen, Samworth et al. (2011). While all these papers are focusing on point estimation, Schuhmacher et al. (2011) show that combining the log-concavity constraint on density and a standard Kolmogorov-Smirnov confidence region yields non-trivial confidence sets for the moments of an unknown distribution. 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. Besides that, while many popular densities are log-concave, this constraint can be too restrictive in applications with multimodal densities. Therefore in the present dissertation a new and weaker constraint on distribution function is introduced:

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 commonly used parametric distributions satisfy this constraint. In particular, if F has a log-concave density $f = F'$, then F is bi-log-concave, according to Bagnoli and Bergstrom (2005). However bi-log-concavity of F is a much weaker constraint. In particular, F may have a density with an arbitrarily large number of modes. Thus we consider estimation of distribution functions under shape constraints for a wider family of distributions. And the bi-log-concavity constraint on F can be considered as a tool for extrapolation of the class of unimodal distributions to the class of multimodal ones.

In Chapter 1 we present characterizations of bi-log-concavity and explicit bounds for F and its density's derivative $f' = F''$. Examples of distributions with bi-log-concave distribution function are provided, and properties of such class of distributions are established together with relations to log-concave as well as to the increasing hazard rate distributions. Besides that, we describe exact (conservative) confidence bounds for F. These bounds are constructed by combining the bi-log-concavity constraint with the standard confidence bands for F such as the Kolmogorov-Smirnov band, Owen's (1995) band and a refinement of the latter one introduced by Dümbgen and Wellner (2014). Examples of such bands for simulated and real data demonstrate the benefits of adding the shape constraint. 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. We also prove consistency properties for those quantities as well as for some functionals of F in the case of combined bands. Under some additional property the corresponding rates of convergence were derived. In particular, in the case of the refined version of Owen's band combined with the bi-log-concavity constraint, the root-n-consistency is achieved for the moments of F and its moment generating function.

In the context of binary regression, bi-log-concavity provides a natural extension of standard logistic regression. In Chapter 2 a nonparametric maximum-likelihood estimator is developed in this context. It is proven that such an estimator is consistent in a certain sense and its rates of convergence are derived. An algorithm for computing this estimator was developed and implemented in the statistical software environment R. It's full description together with numerical examples is provided. Notice that for the estimation procedure no tuning parameters, such as a bandwidth, are necessary. Besides that, the advantages of the bi-log-concave estimator can be seen in the case of rather small samples. For large samples any smoothing estimator would provide desirable and indistinguishable results as the number of observations is large enough. All proofs are deferred to Chapter 3.

1. Bi-log-concave Distribution Functions

In this chapter we introduce a new constraint on a distribution function F – bi-log-concavity. Characterizations of the new constraint and explicit bounds for F and its density $f = F'$ are obtained. Most of the commonly used parametric distributions satisfy this bi-log-concavity constraint, in particular, Gaussian, exponential, Gamma and Beta for certain parameters. Various properties of distributions with bi-log-concave distribution function are derived and illustrated with examples. In particular, connections to log-concave and increasing hazard rate distribution classes are studied. We also 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 the Kolmogorov-Smirnov band, Owen's (1995) band and a refinement of the latter one introduced by Dümbgen and Wellner (2014). A numerical example with the distribution of CEO salaries (Woolridge (2000)) illustrates the usefulness of the proposed method. The benefits of adding the new shape constraint are pinpointed in Section 1.7, where the main results on consistency properties are presented. In particular, it is shown that the new confidence bands based on the Kolmogorov-Smirnov and the refined Owen's statistics imply nontrivial honest confidence bounds for arbitrary moments of F and its moment generating function. Namely, root-n-consistency is achieved for these quantities when combining the latter band with the bi-log-concavity constraint. Notice that honest confidence bounds are understood in the sense that they need to achieve asymptotically correct coverage for all possible model parameters, that is, be valid uniformly in F (see Li (1989)). Some parts of this chapter are also presented in Kolesnyk, Dümbgen et al. (2016).

1.1. Bi-log-concave Functions

We consider functions $F : \mathbb{R} \to [0, 1]$.

Definition 1.1. *A $[0, 1]$-valued function F is called non-degenerate if the set*

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

is non-void and contains at least two different points.

If F is a cumulative distribution function, then $J(F) = \emptyset$ would be equivalent to $F(x) = 1_{\{x \geq m\}}$ for some $m \in \mathbb{R}$. That means, F would correspond to the Dirac measure δ_m at some point m.

If $J(F)$ is non-void and contains just one point then F would have a jump discontinuity at this isolated point.

Now we define a new shape-constraint:

Definition 1.2 (Bi-log-concavity). *A non-degenerate, $[0,1]$-valued 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 $[-\infty, 0]$.*

Notation 1.3. *In what follows let $\mathcal{F}_{\mathrm{blc}}$ be the class of all bi-log-concave functions on $\mathbb{R}$.*

The constraint of bi-log-concavity is natural in many situations. For example, recall that a log-concave distribution is a distribution with log-concave density (see, for example, Walther (2009) for the review and Ibragimov (1956) and Prekopa (1973) for the seminal results on log-concave distributions). Notice that log-concavity of a density function implies log-concavity of the corresponding cumulative distribution function (see Prekopa (1973) or Bagnoli and Bergstrom (2005) for a simpler proof). Thus the following straightforward result provides an important link between distributions with bi-log-concave c.d.f. and the class of log-concave distributions:

Lemma 1.4. *Any cumulative distribution function F with log-concave density $f = F'$ is bi-log-concave.*

Proof is an immediate corollary from Theorems 1 and 3 in Bagnoli and Bergstrom (2005), and from Theorem 5 in Prekopa (1973).

Notation 1.5. *In what follows let $\mathcal{F}_{\mathrm{blcd}}$ denote a class of distributions with bi-log-concave cumulative distribution functions on $\mathbb{R}$.*

Corollary 1.6. *The class of log-concave distributions is embedded into $\mathcal{F}_{\mathrm{blcd}}$.*

One should notice, however, that bi-log-concavity of F alone is a much weaker constraint. As one can see from subsequent examples, F may have a density with an arbitrarily large number of modes.

Table 1.1 contains examples of well-known distributions with bi-log-concave cumulative distribution functions. They all have a log-concave density function and therefore bi-log-concavity of the c.d.f. follows straightforwardly from the results of Bagnoli and Bergstrom (2005) and An (1996) on log-concave distributions.

Examples of distributions without bi-log-concave c.d.f. are log-normal distributions, power laws on $(0, 1)$ with c.d.f. $F(x) = x^c$ for $0 < c < 1$, Pareto distributions, Weibull distributions with parameters $0 < a < 1$ and $b > 0$, Arc-sine, Student's t, Cauchy and F distributions, Beta distributions with at least one of its

Distribution	Support	c.d.f. $F(x)$
Uniform	$[a, b]$	$(x-a)/(b-a)$
Normal	$(-\infty, +\infty)$	$*$
Logistic	$(-\infty, +\infty)$	$(1 + e^{-(x-\mu)/s})^{-1}$
Exponential ($\lambda > 0$)	$[0, +\infty)$	$1 - e^{-\lambda x}$
Gamma ($a \geq 1$, $b > 0$)	$(0, +\infty)$	$*$
Beta ($a \geq 1$, $b \geq 1$)	$[0, 1]$	$*$
Weibull ($k \geq 1$, $\lambda > 0$)	$[0, +\infty)$	$1 - e^{-(\frac{x}{\lambda})^k}$
Gumbel (log-Weibull) $\mu \in \mathbb{R}$, $\beta > 0$)	$(-\infty, +\infty)$	$e^{-e^{-(x-\mu)/\beta}}$
Laplace (Double Exponential, $\mu \in \mathbb{R}$, $\beta > 0$)	$(-\infty, +\infty)$	$\frac{1}{2}e^{(x-\mu)/\beta}$ if $x < \mu$ $1 - \frac{1}{2}e^{-(x-\mu)/\beta}$ if $x \geq \mu$
Power Laws ($c \geq 1$)	$(0, 1]$	x^c
Chi-Squared ($c \geq 2$)	$[0, \infty)$	$*$
Chi ($c \geq 1$)	$[0, \infty)$	$*$

Table 1.1.: Examples of the distributions with bi-log-concave c.d.f. (distribution functions marked $*$ lack a closed-form representation)

parameters smaller than 1 and Gamma distributions with shape parameters $a < 1$. These facts follow immediately from Bagnoli and Bergstrom (2005) and Patel et al. (1976).

The first theorem provides three alternative and seemingly stronger characterizations of bi-log-concavity.

Theorem 1.7. *For a non-degenerate 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

$$F(x+t) \begin{cases} \leq \; F(x)\exp\Big(\dfrac{f(x)}{F(x)}\,t\Big) \\ \geq \; 1 - (1 - F(x))\exp\Big(-\dfrac{f(x)}{1-F(x)}\,t\Big) \end{cases} \tag{1.1}$$

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 derivative $f = F'$. Furthermore, f is locally Lipschitz-continuous on $J(F)$ with L_1-derivative

$f' = F''$ satisfying

$$\frac{-f^2}{1-F} \;\leq\; f' \;\leq\; \frac{f^2}{F}. \tag{1.2}$$

Remark 1.8. *Inequalities (1.2) in Theorem 1.7, (iv) can be reformulated as follows: $\log f$ is locally Lipschitz-continuous on $J(F)$ with L_1-derivative $(\log f)'$ satisfying*

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

We recall that the L_1-derivative of a function g on an open interval $J \subset \mathbb{R}$ is a locally integrable function g' on J such that $g(y) - g(x) = \int_x^y g'(t)\,dt$ for all $x, y \in J$. Local Lipschitz-continuity of g on J means that for every $x \in J$ there exists a neighbourhood $E \subset J$ such that g is Lipschitz-continuous on E.

Remark 1.9. *In this chapter mainly cumulative distribution functions will be considered. If F is such a function then condition (iv) of Theorem 1.7 can be replaced by the following condition*

(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

$$\frac{-f^2}{1-F} \;\leq\; f' \;\leq\; \frac{f^2}{F}. \tag{1.3}$$

Indeed, condition (iii) of Theorem 1.7 implies that $f > 0$ on $J(F)$. For if $f(x_o) = 0$ for some $x_o \in J(F)$, then isotonicity of $\widetilde{h} := f/(1-F)$ would imply that $f(x) = 0$ for $x \leq x_o$, and antitonicity of $h := f/F$ would yield $f(x) = 0$ for $x \geq x_o$. Hence F would be constant on $J(F)$, a violation of F being a continuous distribution function on $\mathbb{R}$. Recall that a function $g : \Omega \to \mathbb{R}$ (where $\Omega \subset \mathbb{R}$) is called isotonic if it preserves the order, i.e. for any x and y in Ω such that $x > y$ it implies $g(x) \geq g(y)$, and $x < y$ implies $g(x) \leq g(y)$. If the order is reversed, i.e. for any x and y in Ω such that $x > y$ it follows that $g(x) \leq g(y)$ and $x < y$ yields $g(x) \geq g(y)$, then function g is called antitonic.

Remark 1.10. *The proof (i) $\Rightarrow$ (ii) of the Theorem 1.7 shows that the set $J(F)$ is convex as $J(F) \equiv (a, b)$ for some numbers a and b such that $-\infty \leq a < b \leq \infty$.*

Remark 1.11. *Notice that $f' = F''$ may be discontinuous on $J(F)$, as the following example shows:*

Example 1.12. *Consider Laplace distribution with $F(x)$ defined on $\mathbb{R}$ as*

$$F(x) = \begin{cases} \frac{1}{2}e^{x} & \text{if } x < 0, \\ 1 - \frac{1}{2}e^{-x} & \text{if } x \geq 0. \end{cases}$$

Notice that $J(F) = \mathbb{R}$. The density is given as $f(x) = \frac{1}{2}e^{-|x|}$ and its L_1-derivative

$$f'(x) = -\operatorname{sign}(x)\frac{1}{2}e^{-|x|}$$

is discontinuous at 0.

Remark 1.13. *A bi-log-concave function on the real line is either a cumulative distribution function or a survival function or a constant function (see Proposition 1.33).*

Corollary 1.14. *For* $F \in \mathcal{F}_{\text{blc}}$ *and* $[a,b] \subset J(F)$ *one has* $F \in C^1(J(F))$ *and, moreover,* $F\,|_{[a,b]} \in \mathscr{H}^{2,2}([a,b])$*, where* $\mathscr{H}^{k,\alpha}([a,b])$ *is the Hölder class of functions on* $[a,b]$ *having continuous derivatives up to order* $k-1$ *with* $k-1$*-th derivative being Lipschitz-continuous of order* $\alpha - 1$.

Indeed, condition (iv) of Theorem 1.7 implies that the derivative f of a bi-log-concave function F is continuous on $J(F)$. Besides that, the proof of inclusion (iii)⇒(iv) yields (global) Lipschitz-continuity of f on the interval $[a,b] \subset J(F)$.

Example 1.15 (Bi-modal density)**.** Consider the mixture $2^{-1}\mathcal{N}(-\delta,1)+2^{-1}\mathcal{N}(\delta,1)$ with $\delta > 0$. It can be easily numerically verified that the corresponding c.d.f. F is bi-log-concave for $\delta \le 1.34$ but not for $\delta \ge 1.35$. This distribution has a bi-modal density for $\delta = 1.34$. The corresponding c.d.f. F is shown in Figure 1.1(a), together with the functions $1 + \log F \le F \le -\log(1-F)$, the inequalities following from $\log(1+y) \le y$ for arbitrary $y \ge -1$. Bi-log-concavity means that the lower bound $1+\log F$ is concave while the upper bound $-\log(1-F)$ is convex. Figures 1.1-1.2 illustrate the various characterisations of the bi-log-concavity constraint as given in Theorem 1.7. In particular, Figure 1.1(b) shows the bounds from part (ii) for one particular point $x \in J(F)$. Figure 1.2(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 1.2(b) contains the derivative f' together with the bounds $-f^2/(1-F)$ and f^2/F as given in part (iv).

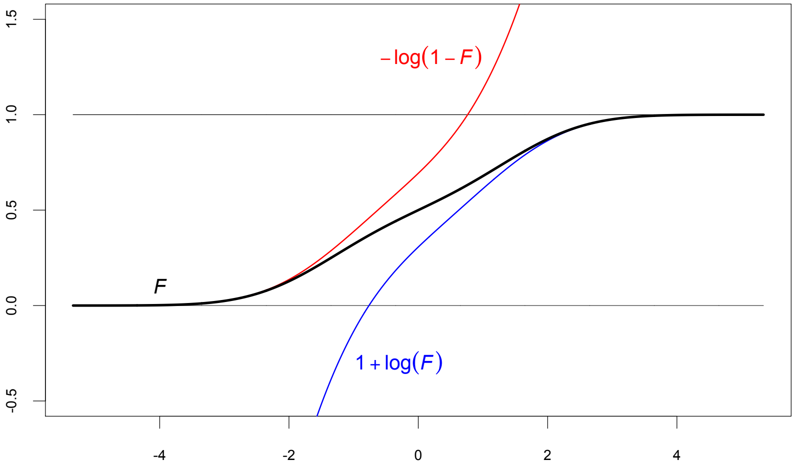
Remark 1.16. *This example shows that the density of the distribution with bi-log-concave c.d.f. (and, in particular, with log-concave c.d.f.) may be non-log-concave since the log-concave density is necessarily unimodal (see, e.g. An (1996), Proposition* 2*).*

Example 1.17 (k-modal density)**.** As it was seen in the previous example, bi-log-concavity allows for distributions with more than one mode. In fact, for any integer $k > 0$ and $a \in (0,1)$,

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

defines a probability density with k local maxima. The corresponding c.d.f. is given by $F(x) = x + a(1-\cos(2\pi kx))/(2\pi k)$ for $x \in [0,1]$, and one can easily deduce from Theorem 1.7 (iii) that it is bi-log-concave if a is sufficiently small.

(a) F with its concave lower and convex upper bounds.



(b) F with the bounds given by Theorem 1.7 (ii).

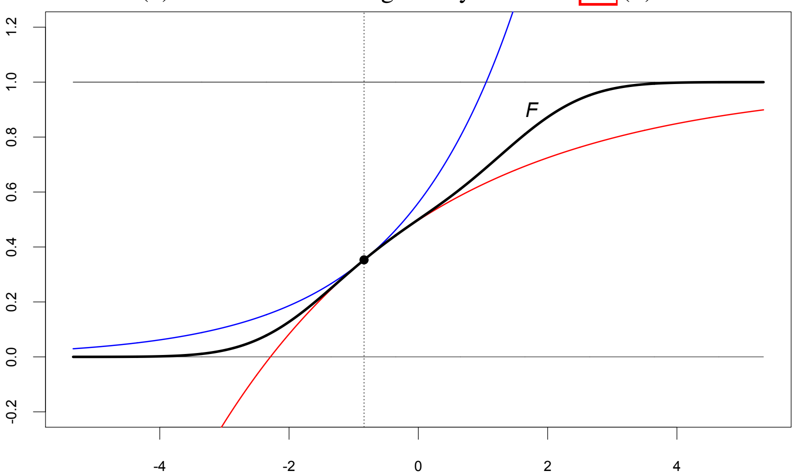
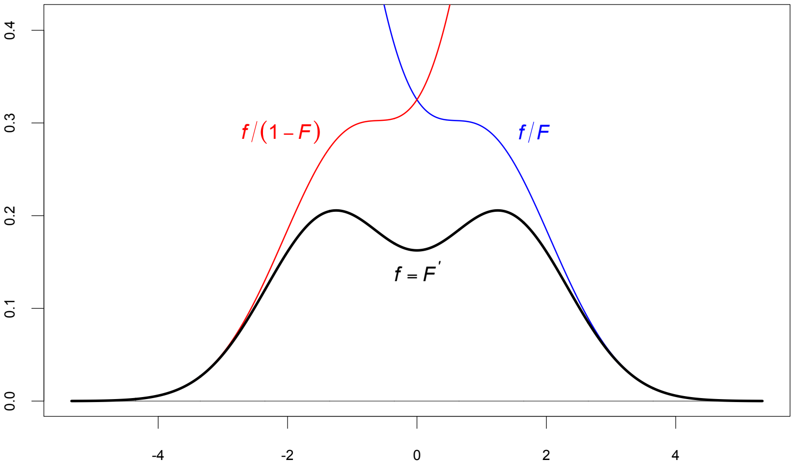


Figure 1.1.: Example of *bi-modal density*: bi-log-concave F with its bounds.

(a) $f = F'$ with monotonic hazard and reverse hazard as given by Theorem 1.7 (iii).



(b) f' with its bounds as given by Theorem 1.7 (iv).

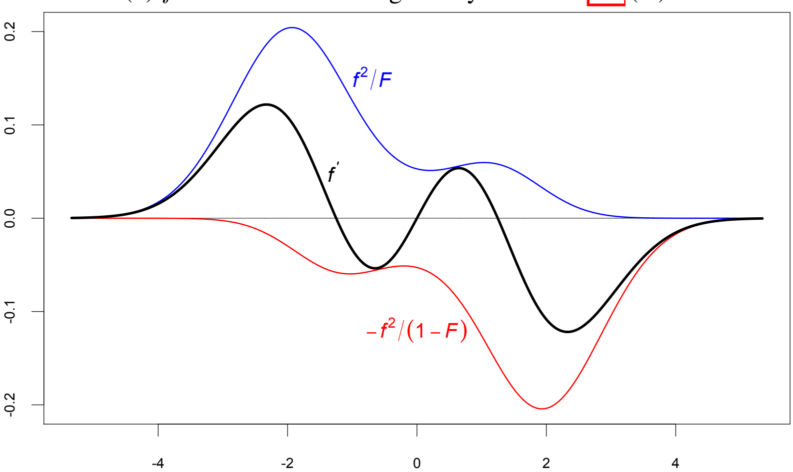


Figure 1.2.: Example of *bi-modal density*: different characterisations of bi-log-concave F.

Example 1.18 (Moment-generating function)**.** *The moment-generating function of $F \in \mathcal{F}_{\text{blcd}}$ is finite in a neighbourhood of* 0*. Precisely, it will be shown in Chapter 3 that*

Proposition 1.19. *If $F \in \mathcal{F}_{\text{blcd}}$ then*

$$\left\{ t \in \mathbb{R} : \int_{\mathbb{R}} e^{tx} \, F(dx) < \infty \right\} \;=\; \big(-T_1(F), T_2(F)\big) \tag{1.4}$$

with

$$\begin{aligned} T_1(F) \;&:=\; \sup_{x \in J(F)} \frac{f(x)}{F(x)} \begin{cases} > 0, & \\ = \infty & \text{if } \inf(J(F)) > -\infty, \end{cases} \\ T_2(F) \;&:=\; \sup_{x \in J(F)} \frac{f(x)}{1 - F(x)} \begin{cases} > 0, & \\ = \infty & \text{if } \sup(J(F)) < \infty. \end{cases} \end{aligned}$$

Corollary 1.20. *All the moments of a distribution with bi-log-concave c.d.f. exist and it is a light-tailed distribution.*

1.2. Relation to Log-concave Distributions

Recall that any cumulative distribution function F with log-concave density $f = F'$ is bi-log-concave (see Lemma 1.4 and its corollary). Therefore a distribution with bi-log-concave c.d.f. $F(x)$ and log-concave density function possesses all well-known properties of the class of log-concave distributions (cf. Prekopa (1973), An (1996), An (1998), Bagnoli and Bergstrom (2005), Walther (2009) and Kotz et al. (2006)). Namely, the class of the distributions with bi-log-concave c.d.f. $F(x)$ and log-concave density function $f(x)$ is closed under

- affine transformations (of a random variable)
- convolutions
- ”truncations” (conditioning on intervals (a, b) with $F(b) - F(a) > 0$)
- taking weak limits of distributions (i.e. considering convergence in distribution of the corresponding measures, - see Cule and Samworth (2010), Proposition 2): $F_n \xrightarrow{\text{w}} F$ implies $F \in \mathcal{F}_{\text{blcd}}$ and log-concavity of f if $F_n \in \mathcal{F}_{\text{blcd}}$ and f_n is log-concave for every n; moreover, $f_n \to f$ almost everywhere.

The property of ”truncated” distributions is a consequence from Theorem 9 in Bagnoli and Bergstrom (2005).

Remark 1.21. *Notice that the product of two bi-log-concave functions is not in general bi-log-concave. An example is the product of distribution function*

$$F_1(x) := \begin{cases} \dfrac{1+\cos(x)}{2} & \textit{for } x \in (-\pi, 0), \\ 0 & \textit{for } x \leq -\pi, \\ 1 & \textit{for } x \geq 0, \end{cases}$$

and distribution function $F_2(x)$ of the uniform distribution $U([-2\pi, -\pi/2])$. The product $F_1(x)F_2(x)$ is log-concave as the product of two log-concave functions; but $1 - F_1(x)F_2(x)$ fails to satisfy log-concavity constraint for all $x \geq -\pi/2$.

1.3. Further Properties of Bi-log-concave Distribution Functions

1.3.1. Affine transformations and truncations

The class $\mathcal{F}_{\mathrm{blcd}}$ is invariant under the following set of operations, namely:

- affine transformations
- ”truncations” (conditioning on intervals (a, b) with $F(b) - F(a) > 0$)

The first property follows from the following fact (see e.g. Rockafellar (1970)): if the mapping $h : \mathbb{R} \to [-\infty, +\infty)$ is concave then so is the mapping $\widetilde{h} : x \mapsto h(a + bx)$. Indeed, if $X \sim F \in \mathcal{F}_{\mathrm{blcd}}$ and $Y := a + bX$, then $Y \sim G(y) := F(b^{-1}(y - a))$ and therefore $\log G(y) = \log F(b^{-1}(y - a))$ and $\log(1 - G(y)) = \log(1 - F(b^{-1}(y - a)))$ are both concave.

The property of ”truncations” holds indeed, since for the ”truncated” cumulative distribution function

$$\widetilde{F}(x) := \frac{F(x) - F(a)}{F(b) - F(a)}$$

with $(a, b) \cap J(F) \neq \emptyset$ (w.l.o.g. $(a, b) \subset J(F)$) and for $x \in (a, b)$ one can write

$$(\log \widetilde{F}(x))' = \frac{f(x)}{F(x) - F(a)} = \frac{f(x)}{F(x)\left(1 - \frac{F(a)}{F(x)}\right)}$$

and

$$(\log(1 - \widetilde{F}(x)))' = -\frac{f(x)}{F(b) - F(x)} = -\frac{f(x)}{(1 - F(x))\left(1 - \frac{1-F(b)}{1-F(x)}\right)}.$$

Both functions are decreasing. Indeed, since $f(x)$ is strictly positive on $J(F)$ (Theorem 1.7, (iv)′) $F(x)$ is increasing and $1 - F(x)$ is decreasing. Hence the function $g_1(x) := (1 - F(a)/F(x))^{-1}$ is decreasing and the function

$$g_2(x) := \frac{1}{1 - \frac{1-F(b)}{1-F(x)}}$$

is increasing. Besides that, functions $f(x)/F(x) =: h(x)$ and $f(x)/(1-F(x)) =: \widetilde{h}(x)$ are strictly positive non-increasing and non-decreasing, respectively (Theorem 1.7, (iii) and (iv)′). Therefore

$$g_1(x)h(x) = (\log \widetilde{F}(x))'$$

and

$$g_2(x)\widetilde{h}(x) = (\log(1 - \widetilde{F}(x)))'$$

are both decreasing. Indeed, $(g_1(x)h(x))' = g_1'(x)h(x) + g_1(x)h'(x)$, where $h(x)$, $g_1(x) > 0$, $g_1'(x) < 0$ and $h'(x) \leq 0$. Hence $(g_1(x)h(x))' < 0$. Consider $(g_2(x)\widetilde{h}(x))' = g_2'(x)\widetilde{h}(x) + g_2(x)\widetilde{h}'(x)$, where $\widetilde{h}(x)$, $g_2(x) > 0$, $g_2'(x) > 0$ and $\widetilde{h}'(x) \geq 0$. Hence $(g_2(x)\widetilde{h}(x))' > 0$. Therefore $(\log \widetilde{F}(x))'' < 0$ and $(\log(1 - \widetilde{F}(x)))'' < 0$, that is $\widetilde{F} \in \mathcal{F}_{\text{blcd}}$.

Remark 1.22. *Notice that neither $\mathcal{F}_{\text{blc}}$ nor $\mathcal{F}_{\text{blcd}}$ is a convex set, and a mixture distribution of two random variables with bi-log-concave distribution functions may have distribution function which is not bi-log-concave. As an example, consider convex combination of c.d.f. F_1 of the uniform distribution $U([-2,-1])$ and c.d.f. F_2 of distribution $U([0,1])$. Then the mixture distribution function $F := (F_1 + F_2)/2$ is such that $F \equiv 0.5$ on $[-1, 0]$. That is, derivative f of F isn't strictly positive on $J(F)$ as it should be according to Remark 1.9 of Theorem 1.7. Besides, F is not differentiable at points -1 and 0 from $J(F)$. It again contradicts to Theorem 1.7. Thus F can not be bi-log-concave.*

1.3.2. Weak convergence

Let us now consider the weak convergence of bi-log-concave distribution functions. Here weak convergence of a sequence of distribution functions to some limit means pointwise convergence on the set of all continuity points of this limit (see, e.g. Durrett (2010), p. 97). Recall also that the supremum (or uniform) norm of a function $h : \mathbb{R} \to \mathbb{R}$ is defined and denoted by $\|h\|_\infty = \sup_{x\in\mathbb{R}} |h(x)|$ (see, e.g., van der Vaart (1998)). For a compact $K \subset \mathbb{R}$ we will write $\|h\|_{K,\infty} := \sup_{x\in K} |h(x)|$.

Lemma 1.23. *The weak limit F of a sequence of bi-log-concave distribution functions F_n is either a degenerate function or bi-log-concave function and*

$$\begin{aligned} ||F_n - F||_{\infty,[a,b]} &\to_p 0, \\ ||f_n - f||_{\infty,[a,b]} &\to_p 0, \\ ||f||_{\infty,[a,b]} &= O_p(1), \\ ||f_n||_{\infty,[a,b]} &= O_p(1), \\ ||f'||_{\infty,[a,b]} &= O_p(1), \\ ||f'_n||_{\infty,[a,b]} &= O_p(1) \end{aligned}$$

as $n \to \infty$, where F, $F_n : \mathbb{R} \to [0,1]$ and f, f_n are corresponding densities and $J(F) \cap J(F_n) \cap [a,b] \neq \emptyset$ for each n and $[a,b] \subset \mathbb{R}$.

We refer to the properties $||f'_n||_\infty = O_p(1)$ and $||f'||_\infty = O_p(1)$ as the uniform boundedness on $[a,b]$. Lemma 1.23 yields the following

Corollary 1.24. *Under the setting and conditions of Lemma 1.23 the following holds true:*

$$\begin{aligned} ||F_n - F||_{q,[a,b]} &\to_p 0, \\ ||f_n - f||_{q,[a,b]} &\to_p 0, \\ f, f_n \in L_q([a,b]), & \\ f', f'_n \in L_q([a,b]) & \end{aligned}$$

for any $q \in [1,\infty]$.

1.3.3. Hazard function and tail behaviour

The following result provides a set of useful inequalities with hazard and reverse hazard functions:

Lemma 1.25. *Suppose that $F \in \mathcal{F}_{\mathrm{blcd}}$ with density $f = F'$ on $J(F)$. For arbitrary $x_1, x_2 \in J(F)$ with $x_1 < x_2$,*

$$\frac{f(x_2)}{F(x_2)} \leq \frac{\log\left[F(x_2)/F(x_1)\right]}{x_2 - x_1} \leq \frac{f(x_1)}{F(x_1)}$$

and

$$\frac{f(x_1)}{1 - F(x_1)} \leq \frac{\log\left[(1 - F(x_1))/(1 - F(x_2))\right]}{x_2 - x_1} \leq \frac{f(x_2)}{1 - F(x_2)}.$$

Now we consider tail behaviour of bi-log-concave distribution functions. The first result gives the bounds for the tails.

Since bi-log-concave distribution function F is also log-concave, the following upper bounds on the survival function of log-concave distribution function (cf. Kotz et al. (2006)) are valid also for the class of distributions with bi-log-concave distribution function:

Proposition 1.26.

$$1 - F(t) \begin{cases} \leq 1, & \text{if } t \leq \mu, \\ \leq 1 - e^{-1} \left(\frac{t}{t-\mu}\right)^{\frac{t}{\mu}-1}, & \text{if } t > \mu, \end{cases}$$

where μ is the mean of the distribution function F; the upper bound is sharp. The sharp lower bound on $1 - F(t)$ is 0.

We conclude this section by showing how bi-log-concavity constraint impacts the behaviour of the mean excess function and, respectively, the behaviour of the tail of distribution. Namely, the following fact holds true:

Lemma 1.27. *If distribution function $F_X(x)$ is bi-log-concave, then the mean excess function $E(X-u|X>u)$ is non-increasing in $u \in \{x \in \mathbb{R} : F(x) < 1\}$, and the function $E(X-u|X<u)$ is non-decreasing in $u \in \{x \in \mathbb{R} : F(x) > 0\}$.*

Indeed, concavity of $\log(1-F)$ implies that the mean excess function $E(X-u|X>u)$ is non-increasing in $u \in \{x \in \mathbb{R} : F(x) < 1\}$. This follows essentially from the following well-known representation:

$$E(X-u|X>u) = \int_0^\infty P(X-u>t|X>u)dt =$$

$$\int_0^\infty \exp(\log(1-F(u+t)) - \log(1-F(u)))dt.$$

Since for a concave function $h : \mathbb{R} \to [-\infty, \infty)$ it follows (cf. Rockafellar (1970), Groeneboom et al. (2001)) that the function $\{h > -\infty\} \ni y \mapsto h(y+s) - h(y)$ is non-increasing for any fixed $s > 0$, the last expression will be non-increasing function in u due to the bi-log-concavity of F. By symmetry, concavity of $\log(F)$ implies that $E(X-u|X<u)$ is non-decreasing function in $u \in \{x \in \mathbb{R} : F(x) > 0\}$.

Such behaviour of the mean excess function, as it is well-known from the extreme value theory, implies the following

Corollary 1.28. *A distribution with bi-log-concave c.d.f. is a light-tailed distribution.*

1.4. Bi-log-concave Distribution Function Implies Increasing Hazard Rate Distribution

Theorem 1.7, (iii) implies that distributions with bi-log-concave c.d.f. F belong to the class of increasing hazard rate (IHR) distributions, i.e. hazard function $f/(1-F)$ is non-decreasing. Therefore $\mathcal{F}_{\text{blcd}}$ possesses all the properties of IHR distributions (see, e.g. Kotz et al. (2006) or Barlow and Proschan (1965) (pp. $9-39$), Barlow and Proschan (1975) (pp. $52-126$) and Johnson and Kotz (1970) (pp. $284-287$)), in particular:

1. If two independent random variables X_1 and X_2 have bi-log-concave distribution functions, then $X_1 + X_2$ has non-decreasing hazard rate function with hazard rate
$$\lambda(x) \leq \min\left(\frac{f_1(x)}{1-F_1(x)}, \frac{f_2(x)}{1-F_2(x)}\right),$$
where $f_1(x)$, $f_2(x)$ and $F_1(x)$, $F_2(x)$ are the densities and the distribution functions of X_1 and X_2, respectively.

2. Assume that a random variable X has bi-log-concave distribution function. If $X_1, X_2, \ldots, X_n$ are n independent observations of X, then each of the order statistics
$$X_{(1)} < X_{(2)} < \ldots < X_{(n)}$$
has IHR distribution.

3. Let m_r be r-th moment of $F \in \mathcal{F}_{\text{blcd}}$ and $F(0) = 0$; then
$$m_r \begin{cases} \leq \Gamma(r+1)m_1^r, & \text{if } r \geq 1, \\ \geq \Gamma(r+1)m_1^r, & \text{if } 0 \leq r \leq 1. \end{cases}$$

The following result shows how bi-log-concavity constraint impacts the behaviour of the tails of distribution function F, reiterating the fact that a distribution with bi-log-concave c.d.f. is a light-tailed distribution.

Proposition 1.29. *If $F(\xi_p) = p$ (i.e., ξ_p is $100p$-th percentile), then*
$$1 - F(x) \leq \exp(-\alpha x), \text{ if } x \geq \xi_p,$$
where $\alpha = -(\log(1-p))/\xi_p$.

That is, bi-log-concave distribution function decays in the tails at least as fast as the exponential function.

Corollary 1.30. $\mathcal{F}_{\text{blcd}} |_{(-\infty,b]} \subset L_p((-\infty,b])$ *for $p \geq 1$ and any $b < \infty$.*

1.5. Confidence Bands

A confidence band for $F \in \mathcal{F}_{\text{blcd}}$ may be constructed by intersecting a standard confidence band for a (continuous) distribution function with this class $\mathcal{F}_{\text{blcd}}$.

Unconstrained nonparametric confidence bands. Let $X_1, X_2, \ldots, X_n$ be independent random variables with continuous distribution function F. Let (L_n, U_n) be an exact $(1-\alpha)$-confidence band for F, where $0 < \alpha \le 0.5$. That means, $L_n = L_{n,\alpha}(\cdot \mid X_1, \ldots, X_n)$ and $U_n = U_{n,\alpha}(\cdot \mid X_1, \ldots, X_n)$ are data-driven non-decreasing functions on the real line such that $L_n \le U_n$ pointwise and

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

A standard example is given by

Kolmogorov-Smirnov band. Consider

$$L_n(x) \ := \ \max\big(\widehat{F}_n(x) - \kappa_{n,\alpha}^{\text{KS}}, 0\big) \quad \text{and} \quad U_n(x) \ := \ \min\big(\widehat{F}_n(x) + \kappa_{n,\alpha}^{\text{KS}}, 1\big),$$

where

$$\widehat{F}_n(x) \ := \ \frac{1}{n} \sum_{i=1}^n 1\{X_i \le x\}$$

is the empirical distribution function and $\kappa_{n,\alpha}^{\text{KS}}$ denotes the $(1-\alpha)$-quantile of $\sup_{x\in\mathbb{R}} \big|\widehat{F}_n(x) - F(x)\big|$ (cf. Shorack and Wellner (1986)). With the order statistics $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ of observations $X_1, X_2, \ldots, X_n$ and $U_{(i)} := F(X_{(i)})$ one may write

$$\sup_{x\in\mathbb{R}} \big|\widehat{F}_n(x) - F(x)\big| \ = \ \max_{i=1,\ldots,n} \max\big(i/n - U_{(i)}, \, U_{(i)} - (i-1)/n\big),$$

and for $i \in \{0, 1, \ldots, n\}$ and $x \in [X_{(i)}, X_{(i+1)})$,

$$\big[L_n(x), U_n(x)\big] \ = \ \Big[\max\big(i/n - \kappa_{n,\alpha}^{\text{KS}}, 0\big), \, \min\big(i/n + \kappa_{n,\alpha}^{\text{KS}}, 1\big)\Big],$$

where $X_{(0)} := -\infty$ and $X_{(n+1)} := \infty$. Notice that $U_{(1)} < U_{(2)} < \cdots < U_{(n)}$ are distributed like the order statistics of n independent random variables with uniform distribution on $[0,1]$. Notice also that $\kappa_{n,\alpha}^{\text{KS}} \le \sqrt{\log(2/\alpha)/2n}$ by Massart's inequality (Massart (1990)).

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 $1 \le i \le n$, and using empirical process theory (see, e.g., Dümbgen (2010)), one can show that for any $\gamma \in [0, 1/2)$, the random variable

$$\sqrt{n} \max_{i=1,2,\ldots,n} \frac{|U_{(i)} - t_i|}{(t_i(1-t_i))^\gamma} \tag{1.5}$$

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. Let $\kappa_{n,\alpha}^{\mathrm{WKS}}$ denote the $(1-\alpha)$-quantile of the test statistic (1.5). It follows that $\kappa_{n,\alpha}^{\mathrm{WKS}} = O(1)$. Inverting this test leads to the $(1-\alpha)$-confidence band (L_n, U_n) for F with

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

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

Owen's band. Another possible confidence band was introduced by Owen (1995). He proposed to invert the goodness-of-fit test of Berk and Jones (1979) which is based on the following test statistic

$$\sup_{x\in\mathbb{R}} K(\widehat{F}_n(x), F(x))$$

with

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

for $p, \widehat{p} \in [0,1]$ and the usual conventions that $0\log(\cdot) := 0$ and $a\log(a/0) := \infty$ for $a > 0$. Notice that $K(\widehat{p}, \cdot) : [0,1] \to [0,\infty]$ is continuous and strictly convex function with minimal value $K(\widehat{p}, \widehat{p}) = 0$, and $K(\widehat{p}, 0) = \infty$ if $\widehat{p} > 0$, $K(\widehat{p}, 1) = \infty$ if $\widehat{p} < 1$. Now we define $L_n(x), U_n(x)$ implicitly via

$$[L_n(x), U_n(x)] \;=\; \big\{p \in [0,1] : K(\widehat{F}(x), p) \le \kappa_{n,\alpha}^{\mathrm{BJO}}\big\},$$

where $\kappa_{n,\alpha}^{\mathrm{BJO}}$ is the $(1-\alpha)$-quantile of $\sup_{x\in\mathbb{R}} K(\widehat{F}_n(x), F(x))$, which has the same distribution as

$$\max_{i=1,2,\ldots,n} \max\big(K(i/n, U_{(i)}), K((i-1)/n, U_{(i)})\big).$$

For $i \in \{0, 1, \ldots, n\}$ and $x \in [X_{(i)}, X_{(i+1)})$,

$$L_n(x) \;=\; \min\big\{p \in [0,1] : K(i/n, p) \le \kappa_{n,\alpha}^{\mathrm{BJO}}\big\},$$

$$U_n(x) \;=\; \max\big\{p \in [0,1] : K(i/n, p) \le \kappa_{n,\alpha}^{\mathrm{BJO}}\big\}.$$

For $i = 0$, the latter interval equals $\Big[0, 1 - \exp(-\kappa_{n,\alpha}^{\mathrm{BJO}})\Big]$, while for $i = n$, it is given by $\big[\exp(-\kappa_{n,\alpha}^{\mathrm{BJO}}), 1\big]$. In case of $1 \le i < n$, it is a compact subinterval of $(0,1)$ with interior point i/n.

As pointed out by Jager and Wellner (2007), one can generalize the Berk-Jones test statistic considerably, replacing the special function $K(\cdot,\cdot)$ with a more general type of function. In all examples considered by Jager and Wellner (2007), the $(1-\alpha)$-quantile $\kappa_{n,\alpha}^{\mathrm{BJO}}$ of $\sup_{x\in\mathbb{R}} K(\widehat{F}_n(x), F(x))$ satisfies

$$\kappa_{n,\alpha}^{\mathrm{BJO}} \;=\; \frac{(1+o(1))\log\log n}{n} \quad \text{as } n \to \infty.$$

Owen's band refined. Dümbgen and Wellner (2014) refined the method of Owen (1995) and considered the following test statistic

$$T_n^{ODW}(F) = \max_{j=1,2,\ldots,n} \big((n+1)K(t_j, F(X_{(j)})) - C(t_j) - \nu D(t_j)\big)$$

with $t_j := j/(n+1)$ and

$$\begin{aligned} C(t) &:= \log\log\big(e/(4t(1-t))\big), \\ D(t) &:= \log(1 + C^2(t)), \end{aligned}$$

while $\nu > 1$ is an arbitrary fixed number. Then for any fixed $\nu > 2$,

$$\max_{j=1,2,\ldots,n} \big((n+1)K(t_j, U_{(j)}) - C(t_j) - \nu D(t_j)\big) \tag{1.6}$$

converges in distribution to

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

In particular, the $(1-\alpha)$-quantile $\kappa_{n,\alpha}^{\mathrm{ODW}}$ of the test statistic (1.6) is bounded as $n \to \infty$. Inverting this test leads to the following confidence band (L_n, U_n):

$$\begin{aligned} L_n(x) &:= 0 \quad \text{for } x < X_{(1)}, \\ L_n(x) &:= \min\big\{p \in (0, t_j] : K(t_j, p) \le \gamma_n(t_j)(1 + o(1))\big\} \\ &\qquad \text{for } 1 \le j \le n, X_{(j)} \le x < X_{(j+1)}, \\ U_n(x) &:= \max\big\{p \in [t_j, 1) : K(t_j, p) \le \gamma_n(t_j)(1 + o(1))\big\} \\ &\qquad \text{for } 1 \le j \le n, X_{(j-1)} \le x < X_{(j)}, \\ U_n(x) &:= 1 \quad \text{for } x \ge X_{(n)}, \end{aligned}$$

where

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

and $\nu > 1$ as before.

The following lemma shows how close (under some additional conditions) two c.d.f.'s from the refined Owen's band are, quantifying the proximity in terms of bounds on the absolute value and the supremum norm of their difference.

Lemma 1.31. *Let F and G be two distribution functions such that $T_n^{ODW}(H) \le \kappa_{n,\alpha}^{ODW}$ for $H = F, G$. Let $\delta_n = c_n n^{-1} \log\log n$ for some $c_n > 0$ such that $c_n \to \infty$ but $\delta_n \to 0$. Then uniformly on $\{\delta_n \le F \le 1 - \delta_n\}$*

$$|F - G| \le \sqrt{(8 + o(1))F(1-F)\gamma_n(F)}.$$

Moreover,

$$||F - G||_\infty = O\left(n^{-1/2}\right).$$

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

$$\begin{aligned} L_n^o(x) &:= \inf\bigl\{G(x) : G \in \mathcal{F}_{\rm blcd}, L_n \le G \le U_n\bigr\}, \\ U_n^o(x) &:= \sup\bigl\{G(x) : G \in \mathcal{F}_{\rm blcd}, L_n \le G \le U_n\bigr\}. \end{aligned}$$

Notice that 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 $F \notin \mathcal{F}_{\rm blcd}$. But in the case of $F \in \mathcal{F}_{\rm blcd}$ this happens with probability at most α. Indeed, the construction of (L_n^o, U_n^o) implies that

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

Thus the coverage probability of the original confidence band and the shape-constrained one coincide, if F itself satisfies the shape constraint. On the other hand, if F does not satisfy the shape-constraint, the coverage probability converges to zero, see Theorem 1.57 (i).

We present two algorithms for computing the confidence band for a bi-log-concave F. The first one is based on bi-log-concavity property only. The second algorithm is a version of the first one which utilizes the so called concave interior procedure $\mathrm{ConcInt}(\cdot,\cdot)$ (see also Dümbgen, Kolesnyk et al. (2016)) which facilitates the computation and makes the practical implementation more straightforward. Besides that, derivations of the algorithms give some useful results about bi-log-concave functions in general.

Algorithm 1. To compute and analyze the refined confidence band, we utilize various inequalities. Consider arbitrary real numbers $x_1 < x_2$ and $0 \le y_1 < y_2 \le 1$. In the case of $y_1 > 0$ there exists a unique exponential function F_1 such that $F_1(x_i) = y_i$ for $i = 1, 2$. Namely,

$$F_1(x) = F_1(x \,|\, x_1, y_1, x_2, y_2) := y_1 \exp\Bigl(\frac{x - x_1}{x_2 - x_1} \log\Bigl(\frac{y_2}{y_1}\Bigr)\Bigr). \tag{1.7}$$

In the case of $y_2 < 1$ there exists a unique "co-exponential" function F_2, i.e. $1 - F_2$ is an exponential function, such that $F_2(x_i) = y_i$ for $i = 1, 2$. Namely,

$$F_2(x) = F_2(x \,|\, x_1, y_1, x_2, y_2) := 1 - (1 - y_1) \exp\Bigl(\frac{x - x_1}{x_2 - x_1} \log\Bigl(\frac{1 - y_2}{1 - y_1}\Bigr)\Bigr). \tag{1.8}$$

Figure 1.3 shows these interpolating functions F_1, F_2 in the case of $x_1 = 0, x_2 = 2$ and $y_1 = 0.2$, $y_2 = 0.5$. Now we are ready to state the key inequalities:

Lemma 1.32. *Let F be non-degenerate function on $\mathbb{R}$ and $x_1 < x_2$, $y_1, y_2 \in [0,1]$. In the case of* $\log F$ *being concave and* $\min(y_1, y_2) > 0$,

$$\begin{cases} F \ge F_1 \text{ on } [x_1, x_2] & \text{if } F(x_1) \ge y_1, F(x_2) \ge y_2, \\ F \le F_1 \text{ on } (-\infty, x_1] & \text{if } F(x_1) \le y_1, F(x_2) \ge y_2, \\ F \le F_1 \text{ on } [x_2, \infty) & \text{if } F(x_1) \ge y_1, F(x_2) \le y_2, \end{cases}$$

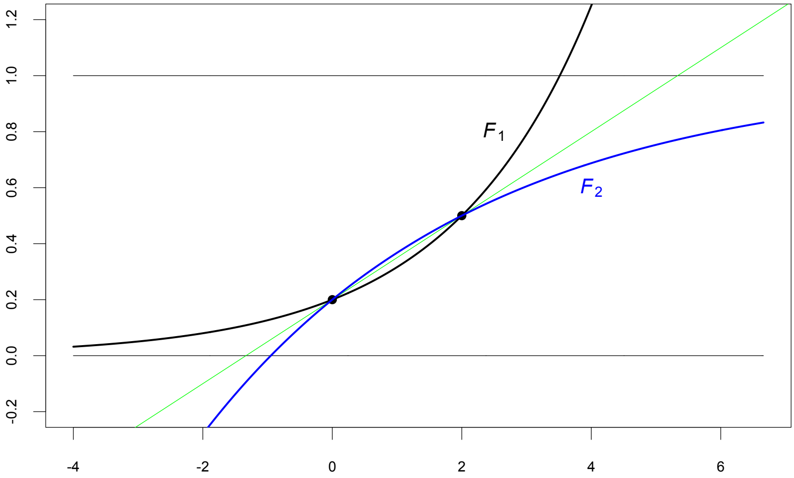


Figure 1.3.: The interpolating functions F_1 and F_2 in (1.7) and (1.8).

with $F_1 = F_1(\cdot \,|\, x_1, y_1, x_2, y_2)$ *as in* (1.7). *Similarly in the case of* $\log(1-F)$ *being concave and* $\max(y_1, y_2) < 1$,

$$\begin{cases} F \leq F_2 \text{ on } [x_1, x_2] & \text{if } F(x_1) \leq y_1, F(x_2) \leq y_2, \\ F \geq F_2 \text{ on } (-\infty, x_1] & \text{if } F(x_1) \geq y_1, F(x_2) \leq y_2, \\ F \geq F_2 \text{ on } [x_2, \infty) & \text{if } F(x_1) \leq y_1, F(x_2) \geq y_2. \end{cases}$$

with $F_2 = F_2(\cdot \,|\, x_1, y_1, x_2, y_2)$ *as in* (1.8).

This lemma is also essential for establishing the following

Proposition 1.33. *A bi-log-concave function* μ *on the real line is either a cumulative distribution function or a survival function or a constant function.*

Notice that any continuous cumulative distribution function G satisfying $L_n \leq G \leq U_n$ also satisfies

$$\begin{aligned} \ell_{ni} := L_n(X_{(i)}+) \;&\leq\; G \quad \text{on } [X_{(i)}, \infty), \\ u_{ni} := U_n(X_{(i)}-) \;&\geq\; G \quad \text{on } (-\infty, X_{(i)}]. \end{aligned}$$

Under the additional constraint that $G \in \mathcal{F}_{\mathrm{blcd}}$, we may untilize Lemma 1.32. Namely, for all integers $1 \leq i < j \leq n$ and arbitrary real numbers $x \leq X_{(i)} \leq y \leq$

$X_{(j)} \le z$:

$$G(x) \begin{cases} \ge F_2(x \,|\, X_{(i)}, \ell_{ni}, X_{(j)}, u_{nj}) & \text{if } u_{nj} < 1, \\ \le F_1(x \,|\, X_{(i)}, u_{ni}, X_{(j)}, \ell_{nj}) & \text{if } u_{ni} < \ell_{nj}, \end{cases}$$

$$G(y) \begin{cases} \ge F_1(y \,|\, X_{(i)}, \ell_{ni}, X_{(j)}, \ell_{nj}) & \text{if } \ell_{ni} > 0, \\ \le F_2(y \,|\, X_{(i)}, u_{ni}, X_{(j)}, u_{nj}) & \text{if } u_{nj} < 1, \end{cases}$$

$$G(z) \begin{cases} \ge F_2(z \,|\, X_{(i)}, u_{ni}, X_{(j)}, \ell_{nj}) & \text{if } u_{ni} < \ell_{nj}, \\ \le F_1(z \,|\, X_{(i)}, \ell_{ni}, X_{(j)}, u_{nj}) & \text{if } \ell_{ni} > 0. \end{cases}$$

Combining all these inequalities amounts to a pointwise maximum $L_n^{(1)}$ of $O(n^2)$ simple functions and a pointwise minimum $U_n^{(1)}$ of another $O(n^2)$ simple functions. Thereafter we choose a fine grid of points $t_1, t_2, \ldots, t_m$ in $[X_{(1)}, X_{(n)}]$. Then for $k = 1, 2, 3, \ldots$ we replace $L_n^{(k)}(x)$ with the maximum $L_n^{(k+1)}(x)$ of 0 and the following $O(m^2)$ numbers:

$$F_1(x \,|\, t_i, L_n^{(k)}(t_i), t_j, L_n^{(k)}(t_j)) \text{ for } i, j \text{ with } t_i < x < t_j, 0 < L_n^{(k)}(t_i) < L_n^{(k)}(t_j),$$
$$F_2(x \,|\, t_i, L_n^{(k)}(t_i), t_j, U_n^{(k)}(t_j)) \text{ for } i, j \text{ with } x < t_i < t_j, L_n^{(k)}(t_i) > 0,$$
$$F_2(x \,|\, t_i, L_n^{(k)}(t_i), t_j, U_n^{(k)}(t_j)) \text{ for } i, j \text{ with } t_i < t_j < x, U_n^{(k)}(t_i) < L_n^{(k)}(t_j).$$

Likewise, $U_n^{(k)}(x)$ is replaced with the minimum $U_n^{(k+1)}(x)$ of 1 and the following $O(m^2)$ numbers:

$$F_2(x \,|\, t_i, U_n^{(k)}(t_i), t_j, U_n^{(k)}(t_j)) \text{ for } i, j \text{ with } t_i < x < t_j, U_n^{(k)}(t_i) < U_n^{(k)}(t_j) < 1,$$
$$F_1(x \,|\, t_i, U_n^{(k)}(t_i), t_j, L_n^{(k)}(t_j)) \text{ for } i, j \text{ with } x < t_i < t_j, U_n^{(k)}(t_i) < L_n^{(k)}(t_j),$$
$$F_1(x \,|\, t_i, U_n^{(k)}(t_i), t_j, L_n^{(k)}(t_j)) \text{ for } i, j \text{ with } t_i < t_j < x, U_n^{(k)}(t_j) < 1.$$

After a finite number k_o of iterations there won't be any change of our bounds $L_n^{(k)}$ and $U_n^{(k)}$ on the set $\mathcal{X} = (-\infty, X_{(1)}) \cup \{t_1, t_2, \ldots, t_m\} \cup (X_{(n)}, \infty)$, and we stop with $(L_n^{(k_o)}, U_n^{(k_o)})$ as a continuous approximation for (L_n^o, U_n^o), i.e. $L_n^{(k_o)} \le L_n^o$ and $U_n^{(k_o)} \ge U_n^o$.

Algorithm 2. An essential ingredient for this algorithm is, as it was already mentioned above, 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 a pair (ℓ^o, u^o), where

$$\begin{aligned} \ell^o(x) &:= \inf\bigl\{g(x) : g \text{ concave on } \mathbb{R},\ \ell \le g \le u \text{ on } \mathcal{T}\bigr\}, \\ u^o(x) &:= \sup\bigl\{g(x) : g \text{ concave on } \mathbb{R},\ \ell \le g \le u \text{ on } \mathcal{T}\bigr\}. \end{aligned}$$

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

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

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 1.4 illustrates this procedure for a $\mathcal{T}$ that consists 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, Figure 1.4 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 1.2 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 $G \in \mathcal{F}_{\text{blcd}}$ satisfying $L_n \le G \le U_n$.

The next lemmas imply 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, 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 $x \to \infty$.

Lemma 1.34. *Suppose that* $\inf\{x \in \mathbb{R} : L_n(x) > 0\} < \sup\{x \in \mathbb{R} : U_n(x) < 1\}$. *Then both L_n^o and U_n^o are Lipschitz-continuous on $\mathbb{R}$.*

Lemma 1.35. *For real numbers $a < b$ and $0 < r < s < 1$ define*

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

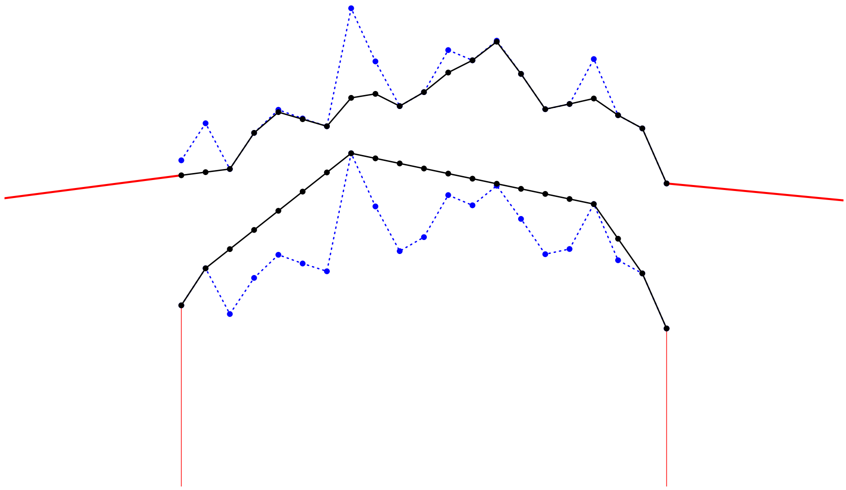


Figure 1.4.: Graphical illustration of the procedure $\mathrm{ConcInt}(\cdot,\cdot)$.

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

Table 1.2.: Pseudocode for the computation of (L_n^o, U_n^o).

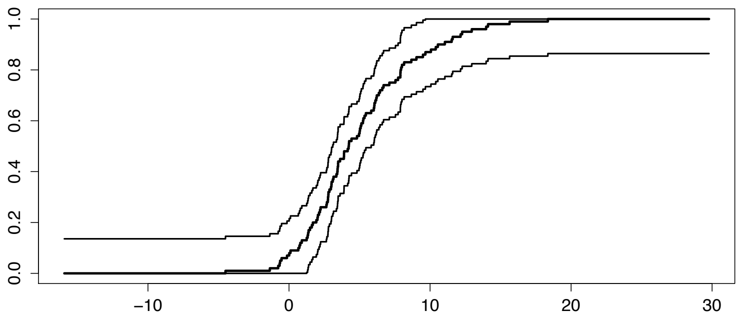


Figure 1.5.: Kolmogorov-Smirnov band (logistic distribution).

(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*

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

and

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

1.6. Numerical Examples

Effectiveness of bi-log-concave shape constraint can be demonstrated when combining it with the standard nonparametric confidence bands such as Kolmogorov-Smirnov and Owen's band. These new shape-constrained confidence regions are substantially more stringent and, due to the bi-log-concavity, gain substantial improvements in either of the central and the tail regions. We illustrate our results by providing corresponding numerical examples for the simulated data.

For the following examples a sample of $n = 100$ data points $\{X_1, X_2, \dots, X_n\}$ was simulated from the logistic distribution with the location parameter 5 and the scale parameter 2. Simulations were performed in the statistical computing environment **R** (see R Development Core Team (2016)).

Example 1.36. In this example the standard Kolmogorov-Smirnov band for F at 0.95-confidence level was constructed (see Figure 1.5).

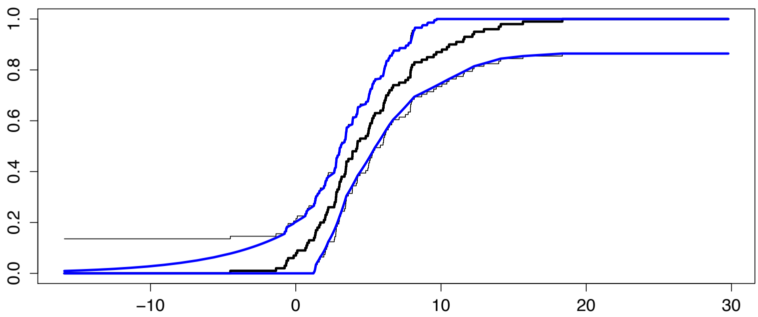


Figure 1.6.: Kolmogorov-Smirnov band for F having concave $\log F$ (logistic distribution).

Example 1.37. Kolmogorov-Smirnov band at 0.95-confidence level for log-concave F was constructed (Figure 1.6; the new band is shown in blue colour, and previous, the standard Kolmogorov-Smirnov band was plotted in black). Due to the concavity of the function $\log F$, the resulting confidence band has more stringent lower bound.

Example 1.38. We constructed Kolmogorov-Smirnov band at 0.95-confidence level for the case when $1 - F$ is log-concave. Due to the concavity of the function $\log(1 - F)$, the resulting confidence band has more stringent upper bound (Figure 1.7).

Example 1.39. Kolmogorov-Smirnov band at 0.95-confidence level for bi-log-concave F (Figure 1.8) was constructed. As a result of the bi-log-concavity, there are more stringent lower and upper bounds comparing to the standard Kolmogorov-Smirnov band.

Example 1.40. In this example we constructed Kolmogorov-Smirnov band with $\alpha = 0.95$ for bi-log-concave F (Figure 1.9) in order to demonstrate the possibility of getting a proxy for the point estimator.

Example 1.41. In this example the standard Owen's band for F at 0.95-confidence level was constructed (see Figure 1.10). As we can see, there are substantial improvements comparing to the Kolmogorov-Smirnov band, especially in the tail regions.

Example 1.42. Owen's band at 0.95-confidence level for log-concave F was constructed (Figure 1.11; the new band is shown in blue colour, and previous, the stan-

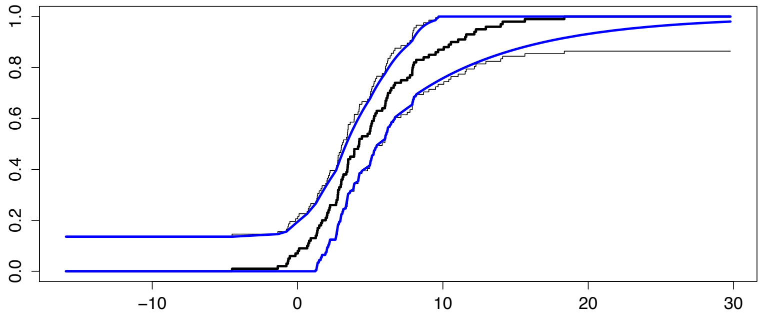


Figure 1.7.: Kolmogorov-Smirnov band for F having concave $\log(1-F)$ (logistic distribution).

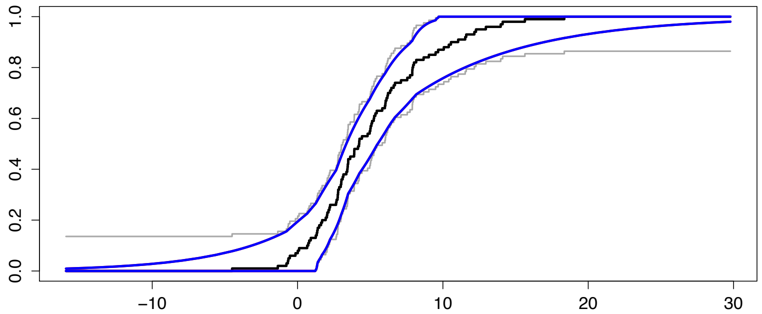


Figure 1.8.: Kolmogorov-Smirnov band for bi-log-concave F (logistic distribution).

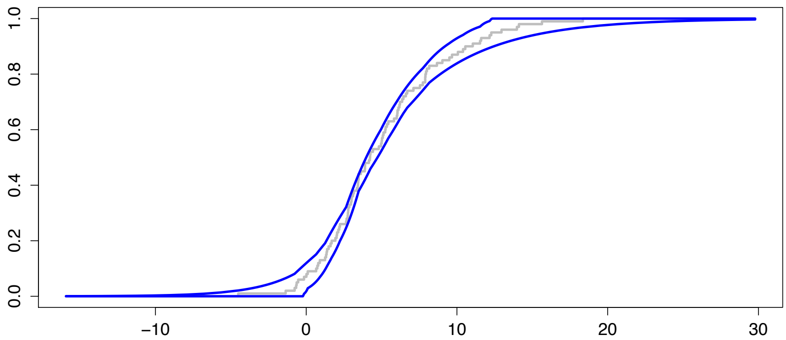


Figure 1.9.: Kolmogorov-Smirnov band for bi-log-concave F with $\alpha = 0.95$ (logistic distribution).

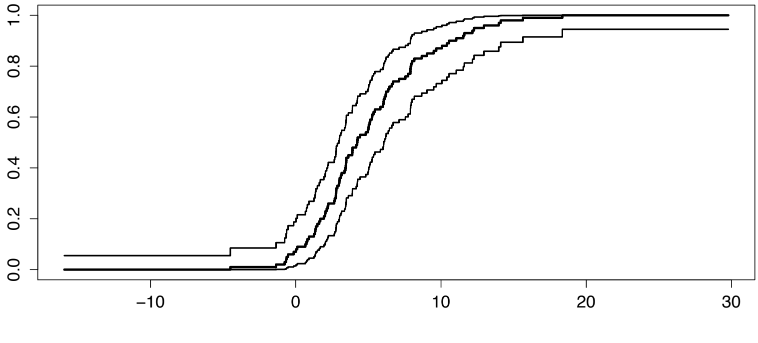


Figure 1.10.: Owen's band (logistic distribution).

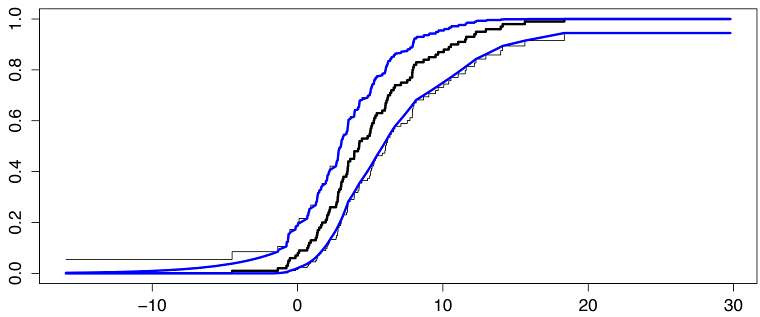


Figure 1.11.: Owen's band for F having concave $\log F$ (logistic distribution).

dard Owen's band was plotted in black). Due to the concavity of the function $\log F$, the resulting confidence band has more stringent lower bound.

Example 1.43. We constructed Owen's band at 0.95-confidence level for the case when $1 - F$ is log-concave. Due to the concavity of the function $\log(1 - F)$, the resulting confidence band has more stringent upper bound (Figure 1.12).

Example 1.44. Here we constructed Owen's band at 0.95-confidence level for bi-log-concave F (Figure 1.13). As we can see, there are substantial improvements comparing to the standard Owen's band, especially in the tail regions.

Example 1.45. In this example we constructed Owen's band with $\alpha = 0.95$ for bi-log-concave F (Figure 1.14) in order to demonstrate the possibility of getting a proxy for the point estimator.

In order to study the impact of bi-log-concavity constraint on the skewed distribution, we simulated a sample of $n = 100$ data points $\{X_1, X_2, \dots, X_n\}$ from the Gamma distribution with the shape parameter 2 and the scale parameter 0.5. Simulations were performed in the statistical computing environment **R** (see R Development Core Team (2016)).

Example 1.46. In this example the standard Kolmogorov-Smirnov band for F at 0.95-confidence level was constructed (see Figure 1.15).

Example 1.47. Kolmogorov-Smirnov band at 0.95-confidence level for log-concave F was constructed (Figure 1.16; the new band is shown in blue colour, and previous, the standard Kolmogorov-Smirnov band was plotted in black). Due to the concavity of the function $\log F$, the resulting confidence band has more stringent lower bound.

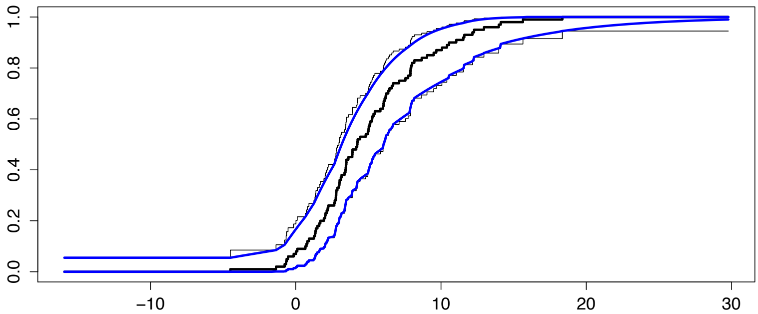


Figure 1.12.: Owen's band for F having concave $\log(1 - F)$ (logistic distribution).

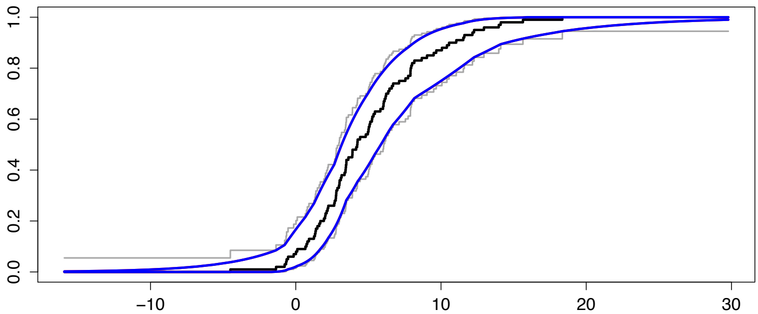


Figure 1.13.: Owen's band for bi-log-concave F (logistic distribution).

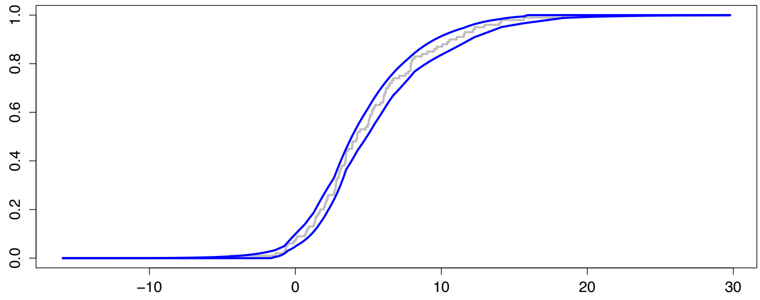


Figure 1.14.: Owen's band for bi-log-concave F with $\alpha = 0.95$ (logistic distribution).

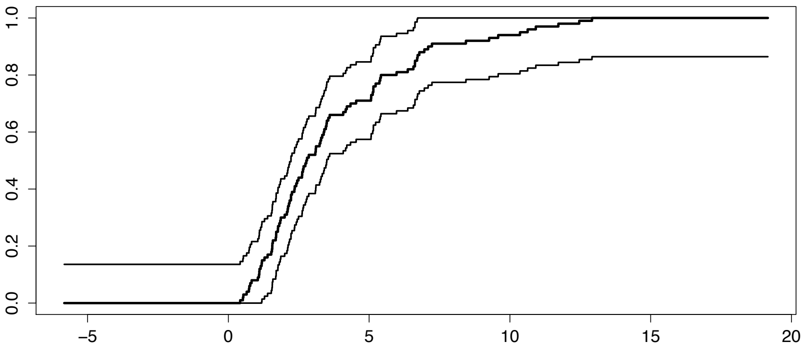


Figure 1.15.: Kolmogorov-Smirnov band (Gamma distribution).

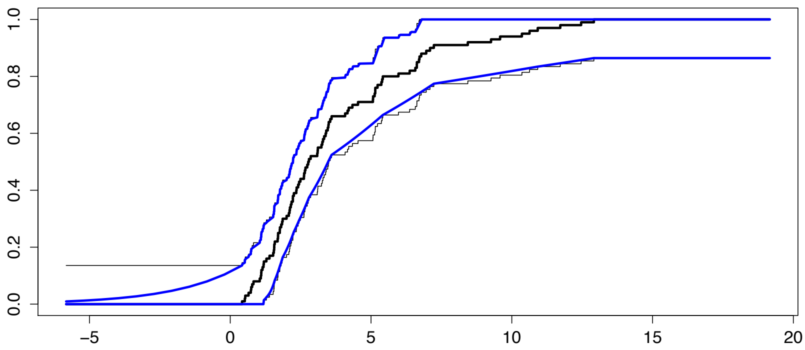


Figure 1.16.: Kolmogorov-Smirnov band for F having concave $\log F$ (Gamma distribution).

Example 1.48. We constructed Kolmogorov-Smirnov band at 0.95-confidence level for the case when $1 - F$ is log-concave. Due to the concavity of the function $\log(1 - F)$, the resulting confidence band has more stringent upper bound (Figure 1.17).

Example 1.49. Kolmogorov-Smirnov band at 0.95-confidence level for bi-log-concave F (Figure 1.18) was constructed. As a result of the bi-log-concavity, there are more stringent lower and upper bounds comparing to the standard Kolmogorov-Smirnov band.

Example 1.50. In this example we constructed Kolmogorov-Smirnov band with $\alpha = 0.80$ for bi-log-concave F (Figure 1.19) in order to demonstrate the possibility of getting a proxy for the point estimator.

Example 1.51. In this example the standard Owen's band for F at 0.95-confidence level was constructed (see Figure 1.20). As we can see, there are substantial improvements comparing to the Kolmogorov-Smirnov band, especially in the tail regions.

Example 1.52. Owen's band at 0.95-confidence level for log-concave F was constructed (Figure 1.21; the new band is shown in blue colour, and previous, the standard Owen's band was plotted in black). Due to the concavity of the function $\log F$, the resulting confidence band has more stringent lower bound.

Example 1.53. We constructed Owen's band at 0.95-confidence level for the case when $1 - F$ is log-concave. Due to the concavity of the function $\log(1 - F)$, the resulting confidence band has more stringent upper bound (Figure 1.22).

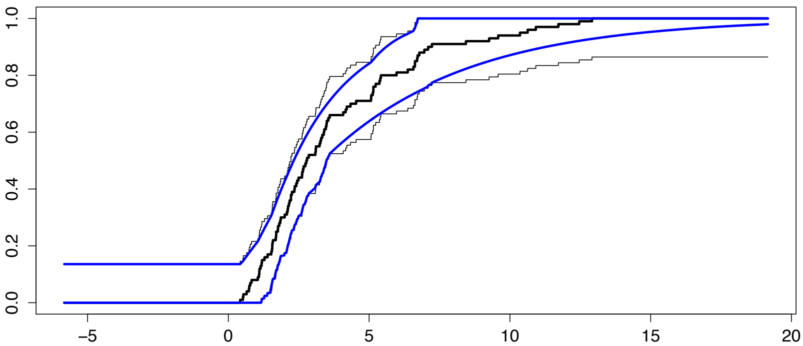


Figure 1.17.: Kolmogorov-Smirnov band for F having concave $\log(1-F)$ (Gamma distribution).

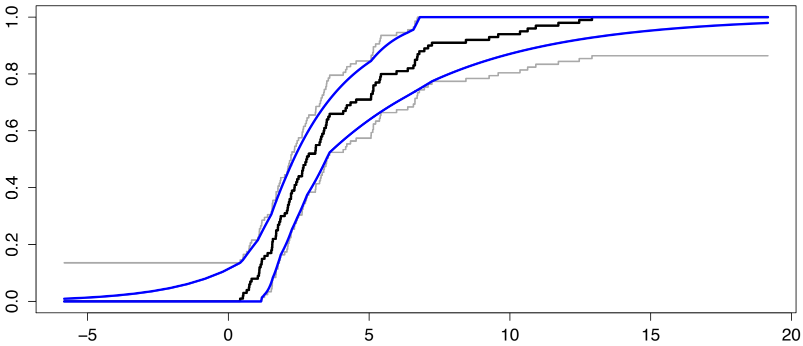


Figure 1.18.: Kolmogorov-Smirnov band for bi-log-concave F (Gamma distribution).

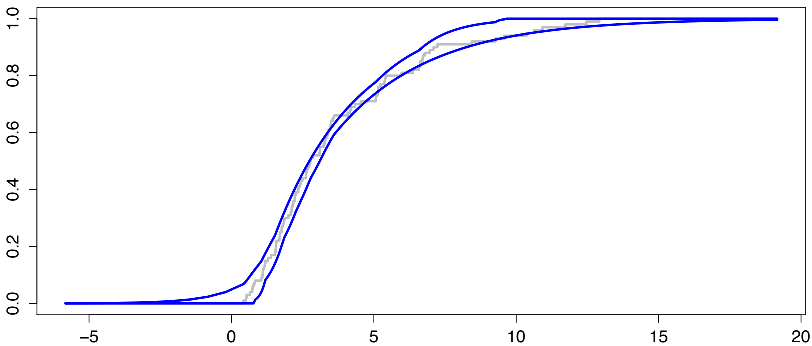


Figure 1.19.: Kolmogorov-Smirnov band for bi-log-concave F with $\alpha = 0.80$ (Gamma distribution).

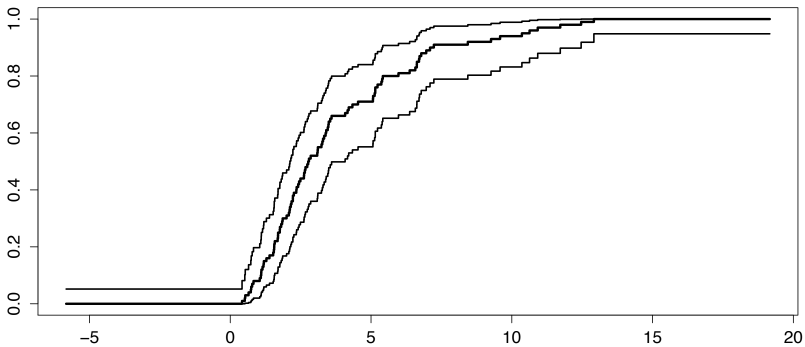


Figure 1.20.: Owen's band (Gamma distribution).

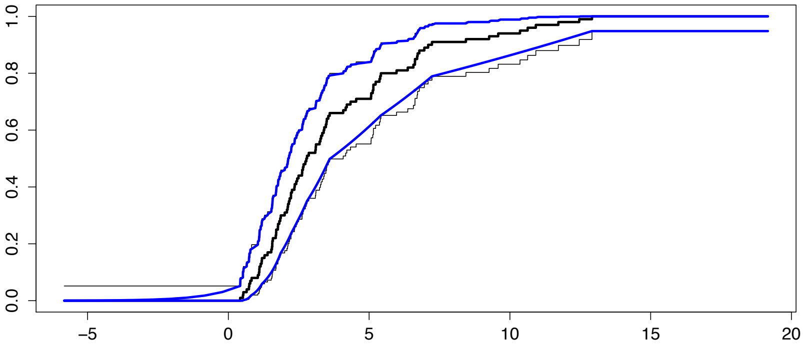


Figure 1.21.: Owen's band for F having concave $\log F$ (Gamma distribution).

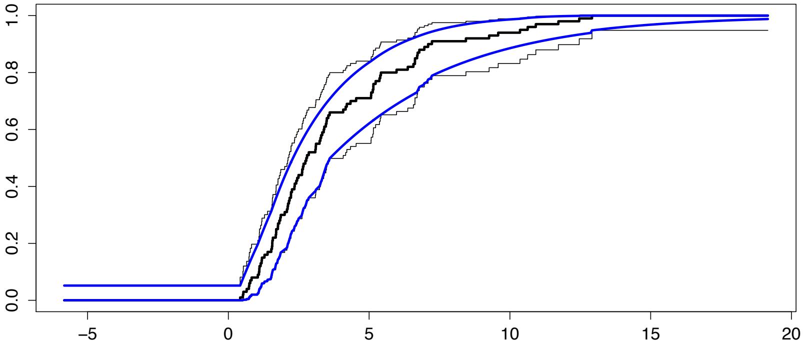


Figure 1.22.: Owen's band for F having concave $\log(1 - F)$ (Gamma distribution).

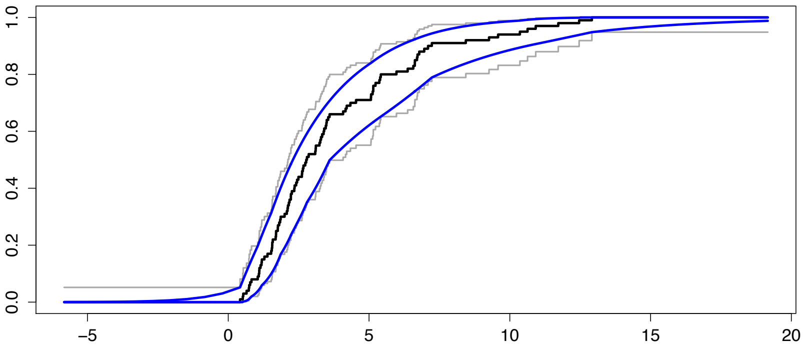


Figure 1.23.: Owen's band for bi-log-concave F (Gamma distribution).

Example 1.54. Here we constructed Owen's band at 0.95-confidence level for bi-log-concave F (Figure 1.23). As we can see, there are substantial improvements comparing to the standard Owen's band, especially in the tail regions.

Example 1.55. Here we constructed Owen's band with $\alpha = 0.80$ for bi-log-concave F (Figure 1.24) in order to demonstrate the possibility of getting a proxy for the point estimator.

Example 1.56. In this example we illustrate our methods with a data set from Wooldridge (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 $Y_{i,\text{raw}} \in \mathbb{N}$ corresponds to an unobserved true salary Y_i within $(Y_{i,\text{raw}} - 1, Y_{i,\text{raw}} + 1)$, 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 $\big(\log_{10}(Y_{i,\text{raw}} + 1)\big)_{i=1}^n$ and U_n with $\big(\log_{10}(Y_{i,\text{raw}} - 1)\big)_{i=1}^n$.

Figure 1.25(a) shows the Kolmogorov-Smirnov 95%-confidence bands for F, without (black lines) and with (blue lines) the restriction of bi-log-concavity. Figure 1.25(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

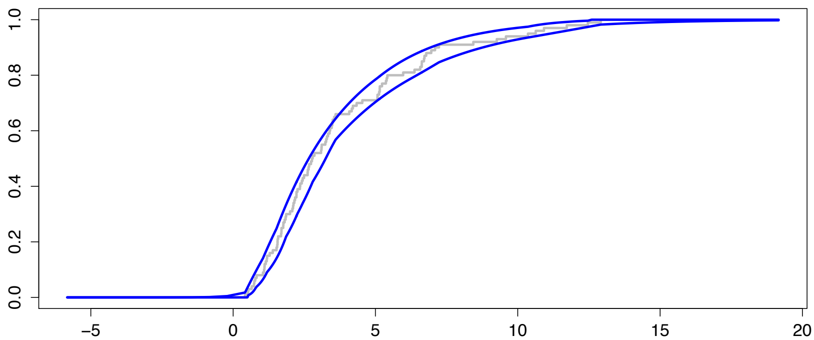


Figure 1.24.: Owen's band for bi-log-concave F with $\alpha = 0.80$ (Gamma distribution).

a substantial gain of precision. Notice also that the bounds in Figure 1.25(b) are tighter in the tails but slightly wider in the central part than those in Figure 1.25(a), for the unconstrained band as well as for the band with shape constraint.

1.7. Consistency Properties

In this section we study the asymptotic behaviour of the proposed confidence band (L_n^o, U_n^o) when $F \in \mathcal{F}_{\rm blcd}$. We illustrate and quantify the benefits of this shape constraint which leads to the confidence band (L_n^o, U_n^o) in place of (L_n, U_n). 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 $G \in \mathcal{F}_{\rm blcd}$ fitting in between L_n and U_n, concluding with confidence $1-\alpha$ that $F \notin \mathcal{F}_{\rm blcd}$.

First of all we are stating the following

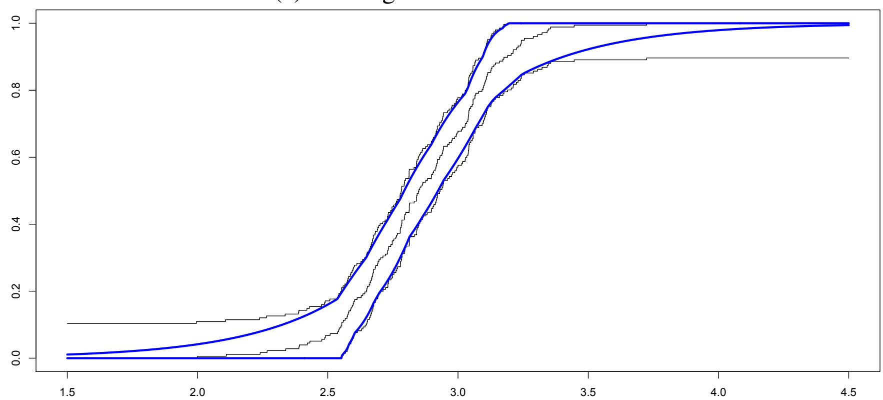
Theorem 1.57. *Suppose that the original confidence band (L_n, U_n) is consistent in the sense that for any fixed $x \in \mathbb{R}$, both $L_n(x)$ and $U_n(x)$ tend to $F(x)$ in probability.*

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

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

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

(a) Kolmogorov-Smirnov bands.



(b) Weighted Kolmogorov-Smirnov bands.

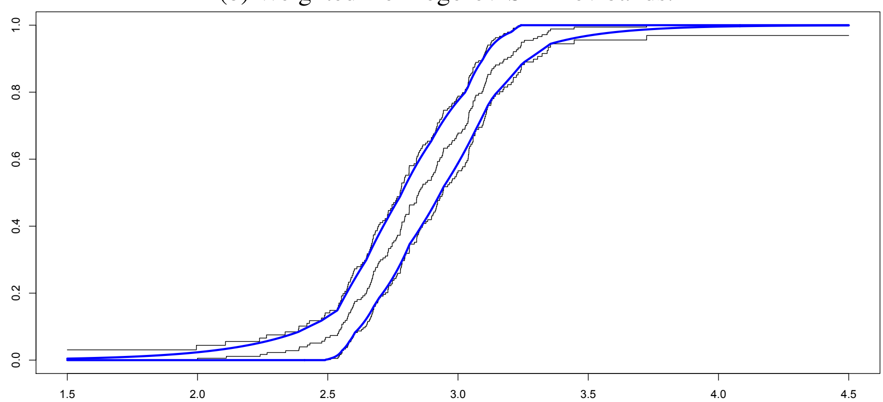


Figure 1.25.: Estimated distribution function with unconstrained and constrained confidence bands for CEO salaries.

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

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

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

Corollary 1.58. *Suppose that the original confidence band* (L_n, U_n) *is consistent, and let* $F \in \mathcal{F}_{\mathrm{blcd}}$. *Let* $\phi : \mathbb{R} \to \mathbb{R}$ *be absolutely continuous function 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*

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

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 $G \in \mathcal{F}_{\mathrm{blcd}}$ between L_n and U_n. Corollary 1.58 applies to $\phi(x) := e^{tx}$ with $-T_1(F) < t < T_2(F)$. Indeed, the proof of Proposition 1.19 implies the following explicit formulae in the case $L_n^o \le U_n^o$:

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

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

Now we refine Corollary 1.58 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 $\kappa, \lambda > 0$,

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

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

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 1.59. *Suppose that $F \in \mathcal{F}_{\rm blcd}$, 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 $k \ge 1$. Then*

$$\sup_{G \,:\, L_n^o \le G \le U_n^o} \Bigl| \int \phi\, dG - \int \phi\, dF \Bigr| \;=\; \begin{cases} O_p\bigl(n^{-1/2}(\log n)^k\bigr) & \text{if } \gamma = 0, \\ O_p(n^{-1/2}) & \text{if } \gamma > 0. \end{cases}$$

(ii) *Suppose that ϕ satisfies the conditions in Corollary 1.58. Then*

$$\sup_{G \,:\, L_n^o \le G \le U_n^o} \Bigl| \int \phi\, dG - \int \phi\, dF \Bigr| \;=\; O_p(n^{-\beta}) \qquad (1.9)$$

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

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

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 $x \ge 0$. Further let F_n be the conditional distribution function of X, given that $X \le x_n := (\log n)/2 - \log c$ 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

$$\begin{aligned} \int \phi\, d(F_n - F) \;&=\; \mathbb{E}(X^k) - \mathbb{E}(X^k \,|\, X \le x_n) \\ &=\; \mathbb{P}(X > x_n)\bigl(\mathbb{E}(X^k \,|\, X > x_n) - \mathbb{E}(X^k \,|\, X \le x_n)\bigr) \\ &\ge\; \mathbb{P}(X > x_n)\bigl(x_n^k - \mathbb{E}(X^k)/\,\mathbb{P}(X \le x_n)\bigr) \\ &=\; 2^{-k} c n^{-1/2} (\log n)^k (1 + o(1)). \end{aligned}$$

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

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. Hence (1.9) holds for any exponent $\beta \in (0, 1/2]$ such that

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

In particular,

$$\sup_{G \,:\, L_n^o \le G \le U_n^o} \Bigl| \int e^{tx}\, G(dx) - \int e^{tx}\, F(dx) \Bigr| \ = \ O_p(n^{-1/2})$$

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

Thus we have shown that bi-log-concavity shape constraint leads to rather accurate confidence bounds for arbitrary moments

$$\mu_k(F) \ := \ \int_{\mathbb{R}} x^k\, F(dx)$$

where $\mathbb{N} \ni k \ge 1$ provided that the original confidence band (L_n, U_n) satisfies **Condition (*)**.

Theorem 1.59 (i) has the following two corollaries for the case of the Kolmogorov-Smirnov and the refined Owen's band (L_n, U_n).

Corollary 1.60. *Suppose that $F \in \mathcal{F}_{\rm blcd}$, and let (L_n^o, U_n^o) be defined with the Kolmogorov-Smirnov band (L_n, U_n) at $1-\alpha$ confidence level for fixed α. Then*

$$\sup_{\widetilde{F} \,:\, L_n^o \le \widetilde{F} \le U_n^o} \bigl| \mu_k(\widetilde{F}) - \mu_k(F) \bigr| \ = \ O_p\bigl((\log n)^k/\sqrt{n}\bigr)$$

for any integer $k \ge 1$.

Corollary 1.61. *Suppose that $F \in \mathcal{F}_{\rm blcd}$, and let (L_n^o, U_n^o) be defined with the refined Owen's band (L_n, U_n) at $1-\alpha$ confidence level for fixed α. Then*

$$\sup_{\widetilde{F} \,:\, L_n^o \le \widetilde{F} \le U_n^o} \bigl| \mu_k(\widetilde{F}) - \mu_k(F) \bigr| \ = \ O_p\bigl(1/\sqrt{n}\bigr)$$

for any integer $k \ge 1$.

2. Bi-log-concave Regression Functions

We consider observation pairs $(X, Y) \in \mathbb{R} \times \{0,1\}$ such that X is fixed or random, and Y is a random variable with unknown mean function μ given by

$$\mu(x) \;=\; \mathbb{E}(Y \,|\, X = x) \;=\; \mathbb{P}(Y \,|\, X = x)$$

In standard logistic regression one assumes that

$$\mu(x) \;=\; \ell(a + bx)$$

with unknown parameters $a, b \in \mathbb{R}$, where $\ell : \mathbb{R} \to (0,1)$ is the logistic function (or, the inverse logit function) given by

$$\ell(z) \;:=\; \frac{e^z}{1+e^z} \;=\; \frac{1}{e^{-z}+1},$$

which is log-concave. We will replace this parametric assumption by a shape constraint, namely *bi-log-concavity* of μ. That means, we assume that

$$\log(\mu), \log(1-\mu) : \mathbb{R} \to [-\infty, 0] \text{ are concave functions.} \tag{2.1}$$

This shape-constraint has been introduced in the previous chapter.

Nonparametric maximum likelihood estimators and their consistency with applications to the shape-constrained setting have been widely studied (see, e.g., Dümbgen et al. (2006) and van de Geer (1993)). In particular, special attention was devoted to the consistency of concave regression (see Dümbgen et al. (2004)). Motivated by those results, in this chapter we devise a maximum likelihood estimator of μ under constraint (2.1), based on independent observations (X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n) in $\mathbb{R} \times \{0,1\}$ such that $\mathbb{E}(Y_i \,|\, X_i = x) = \mu(x)$ for $1 \le i \le n$ and $x \in \mathbb{R}$. We show that such an estimator of a binary bi-log-concave regression is consistent in a certain sense, in particular, it is Hellinger consistent. We also derive the rates of convergence under certain conditions and develop explicit algorithm to perform the estimation. In order to illustrate the estimation procedure and to compare the results to the logistic and isotonic regression estimators, we provide numerical examples for simulated data.

2.1. Introduction

In the case of $\mu(x) = \ell(a + bx)$ one can easily deduce from $\ell' = \ell(1 - \ell)$ that Condition (iii) of Theorem 1.7 is satisfied. Hence bi-log-concavity of μ provides a nonparametric extension of standard logistic regression.

For algorithmic and conceptual reasons it is convenient to reparametrize the regression function μ via the logit-transform. To this end we extend the logistic function to a continuous and bijective function $\ell : [-\infty, +\infty] \to [0, 1]$ via $\ell(-\infty) := 0$ and $\ell(+\infty) = 1$. The corresponding inverse function is denoted as $\mathrm{logit} : [0, 1] \to [-\infty, +\infty]$ and given by

$$\mathrm{logit}(u) \ := \ \log\Bigl(\frac{u}{1-u}\Bigr) \ \text{ for } 0 < u < 1,$$

and $\mathrm{logit}(0) := -\infty$, $\mathrm{logit}(1) := +\infty$. Now we consider the function

$$\theta \ := \ \mathrm{logit}(\mu)$$

from $\mathbb{R}$ into $[-\infty, +\infty]$, and $J(\theta) = \{x \in \mathbb{R} : \theta(x) \in \mathbb{R}\}$. One can easily rephrase conditions (i), (iii) and (iv) of Theorem 1.7 in terms of $\theta = \mathrm{logit}(\mu)$:

Lemma 2.1. *If $J(\theta)$ is non-empty, the following three statements are equivalent:*
(i) *$\mu = \ell(\theta)$ satisfies (2.1).*
(ii) *θ is continuous on $\mathbb{R}$, differentiable on $J(\theta)$, and its derivative θ' satisfies the following two conditions on $J(\theta)$:*

$$\frac{\theta'}{1 + \exp(\theta)} \ \textit{is non-increasing}, \qquad \frac{\theta'}{1 + \exp(-\theta)} \ \textit{is non-decreasing}.$$

(iii) *θ is continuous on $\mathbb{R}$, differentiable on $J(\theta)$, and its derivative θ' is locally Lipschitz-continuous on $J(\theta)$ with L_1-derivative ℓ'' satisfying*

$$-(\theta')^2\ell(-\theta) \ \le \ \theta'' \ \le \ (\theta')^2\ell(\theta).$$

This follows essentially from continuity of the logistic and logit function and the fact that $\ell' = \ell(1 - \ell)$ on $\mathbb{R}$. Hence on $J(\theta)$,

$$\mu' \ = \ \ell'(\theta)\theta' \ = \ \mu(1-\mu)\theta'$$

and thus

$$\frac{\mu'}{\mu} \ = \ (1-\mu)\theta' \ = \ \frac{\theta'}{1+\exp(\theta)} \quad \text{and} \quad \frac{\mu'}{1-\mu} \ = \ \mu\theta' \ = \ \frac{\theta'}{1+\exp(-\theta)}.$$

Moreover,

$$\begin{aligned} \mu'' \ &= \ (1-2\mu)\mu'\,\theta' + \ell'(\theta)\,\theta'' \\ &= \ \ell'(\theta)\bigl((1-2\mu)\,(\theta')^2 + \theta''\bigr), \end{aligned}$$

and

$$\frac{(\mu')^2}{1-\mu} = \frac{(\ell'(\theta))^2\,(\theta')^2}{1-\mu} = \ell'(\theta)\mu\,(\theta')^2,$$
$$\frac{(\mu')^2}{\mu} = \frac{(\ell'(\theta))^2\,(\theta')^2}{\mu} = \ell'(\theta)(1-\mu)\,(\theta')^2,$$

Hence the inequalities in Condition (iv) of Theorem 1.7 are equivalent to

$$\theta'' \in \big[(\mu-1)\,(\theta')^2, \mu\,(\theta')^2\big] = \big[-\ell(-\theta)\,(\theta')^2, \ell(\theta)\,(\theta')^2\big].$$

2.2. The Target Functionals

We will determine a function $\widehat{\mu} : \mathbb{R} \to [0,1]$ which minimizes the conditional negative log-likelihood given the covariables X_1, ..., X_n:

$$L(\mu) := -\sum_{i=1}^{n}\big((1-Y_i)\log(1-\mu(X_i)) + Y_i \log\mu(X_i)\big).$$

To obtain an approximate solution, we choose a fine grid of $N \gg n$ points $t_1 < t_2 < \cdots < t_N$ such that $\{X_1, X_2, \ldots, X_n\} \subset \{t_1, t_2, \ldots, t_N\}$ and $t_1 \ll \min_i X_i$, $t_N \gg \max_i X_i$. Hereinafter without loss of generality we assume that observations are already ordered, i.e. $X_1 = X_{(1)}$, ..., $X_n = X_{(n)}$. Then bi-log-concavity constraints (2.1) can be approximated on the grid $\{t_1, t_2, \ldots, t_N\}$ as follows:

$$\frac{\log(\mu(t_2)) - \log(\mu(t_1))}{t_2 - t_1} \geq \frac{\log(\mu(t_3)) - \log(\mu(t_2))}{t_3 - t_2} \geq$$
$$\cdots \geq \frac{\log(\mu(t_N)) - \log(\mu(t_{N-1}))}{t_N - t_{N-1}}$$

simultaneously with

$$\frac{\log(1-\mu(t_2)) - \log(1-\mu(t_1))}{t_2 - t_1} \geq \frac{1-\log(\mu(t_3)) - \log(1-\mu(t_2))}{t_3 - t_2} \geq$$
$$\cdots \geq \frac{\log(1-\mu(t_N)) - \log(1-\mu(t_{N-1}))}{t_N - t_{N-1}}$$

According to the reparametrization, $\mu(t_j) = \ell(\theta_j)$, therefore these inequalities can be re-written in the following form:

$$\log(1-\ell(\theta_j)) - \lambda_{j\ell}\log(1-\ell(\theta_{j-1})) - \lambda_{jr}\log(1-\ell(\theta_{j+1})) \geq 0, \qquad (2.2)$$

$$\log\ell(\theta_j) - \lambda_{j\ell}\log\ell(\theta_{j-1}) - \lambda_{jr}\log\ell(\theta_{j+1}) \geq 0, \qquad (2.3)$$

where

$$\lambda_{j\ell} := \frac{t_{j+1} - t_j}{t_{j+1} - t_{j-1}},$$
$$\lambda_{jr} := \frac{t_j - t_{j-1}}{t_{j+1} - t_{j-1}}.$$

Inequalities (2.2) and (2.3) are the reparameterized bi-log-concavity constraints.

For $j \in \{1, 2, \ldots, N\}$ and $y \in \{0, 1\}$ we define

$$w_{jy} := \sum_{i=1}^{n} 1_{\{X_i = t_j, Y_i = y\}},$$

and for a vector $\boldsymbol{\theta} \in \mathbb{R}^N$ we define our target functional

$$L_\beta(\boldsymbol{\theta}) := L(\boldsymbol{\theta}) + \beta R(\boldsymbol{\theta})$$

with a small number $\beta > 0$. Here

$$L(\boldsymbol{\theta}) := -\sum_{j=1}^{N} \bigl(w_{j0} \log(1 - \ell(\theta_j)) + w_{j1} \log \ell(\theta_j)\bigr),$$
$$R(\boldsymbol{\theta}) := -\sum_{j=2}^{N-1} \bigl(\log(s_{j0}(\boldsymbol{\theta})) + \log(s_{j1}(\boldsymbol{\theta}))\bigr)$$

with

$$s_{j0}(\boldsymbol{\theta}) := \log(1 - \ell(\theta_j)) - \lambda_{j\ell} \log(1 - \ell(\theta_{j-1})) - \lambda_{jr} \log(1 - \ell(\theta_{j+1})),$$
$$s_{j1}(\boldsymbol{\theta}) := \log \ell(\theta_j) - \lambda_{j\ell} \log \ell(\theta_{j-1}) - \lambda_{jr} \log \ell(\theta_{j+1}).$$

Function $R(\boldsymbol{\theta})$ is called a logarithmic barrier penalty. Throughout this chapter, the convention $\log a := -\infty$ for $a \leq 0$ is used. Thus $\boldsymbol{\theta}$ represents $(\theta(t_j))_{j=1}^N$, $L(\boldsymbol{\theta})$ is a proxy for $L(\mu)$ with $\mu = \ell(\theta)$, and $R(\boldsymbol{\theta})$ forces $s_{jy}(\boldsymbol{\theta})$ to be strictly positive for $2 \leq j < N$ and $y \in \{0, 1\}$.

We will minimize $L_\beta(\boldsymbol{\theta})$ over all $\boldsymbol{\theta} \in \mathbb{R}^N$ such that $L_\beta(\boldsymbol{\theta}) > -\infty$ utilizing the Newton-Raphson method with the step size correction based on the Armijo condition (all the details are deferred to Section 2.4). As a starting point we propose to use $\boldsymbol{\theta}^{(0)} = (\widehat{a} + \widehat{b} t_j)_{j=1}^N$, where $(\widehat{a}, \widehat{b})$ is the maximum likelihood estimator (MLE) for (a, b) in the standard logistic regression model.

2.3. Consistency of the Bi-log-concave Estimator

In this section we provide consistency properties of the MLE $\widehat{\mu}$ in the following triangular observation scheme. For $n = 2, 3, 4, \ldots$ we observe independent random pairs (X_{n1}, Y_{n1}), (X_{n2}, Y_{n2}), ..., (X_{nn}, Y_{nn}) in $\mathbb{R} \times \{0, 1\}$ with fixed or random values X_{ni} and random variables Y_{ni} such that

$$\mathbb{P}(Y_{ni} = 1 \mid X_{ni} = x) = \mu(x)$$

for any $x \in \mathbb{R}$.

2.3.1. A General Consistency Result

Let us consider the X_{ni} as fixed numbers such that $X_{n1} \leq X_{n2} \leq ... \leq X_{nn}$. (In random design settings we may condition on the X_{ni} and sort the observations so that X_{ni} is non-decreasing in i.)

Concerning μ, we only assume that it belongs to a family $\mathscr{F} \subset \mathscr{F}_{mon}$, where

$$\mathscr{F}_{mon} := \{F : \mathbb{R} \to [0, 1] : F \text{ non-decreasing or non-increasing } \}.$$

Further we assume that for any $n \geq 2$,

$$\{(F(X_{ni}))_{i=1}^n : F \in \mathscr{F}\}$$

is a closed subset of $[0, 1]^n$. This property implies that there exists a maximizer $\widehat{\mu}_n \in \mathscr{F}$ of

$$L_n(F) := \frac{1}{n} \sum_{i=1}^n (Y_{ni} \log(F(X_{ni})) + (1 - Y_{ni}) \log(1 - F(X_{ni})))$$

over all $F \in \mathscr{F}$, because

$$[0, 1]^n \ni (p_i)_{i=1}^n \mapsto \frac{1}{n} \sum_{i=1}^n (Y_{ni} \log p_{ni} + (1 - Y_{ni}) \log(1 - p_{ni})) \in [-\infty, 0]$$

is continuous.

Depending on the model $\mathscr{F}$, the MLE $\widehat{\mu}_n$ may be non-unique. But it is consistent in the following sense:

Theorem 2.2. *If $\mu \in \mathscr{F} \subset \mathscr{F}_{mon}$, and if $\{(F(X_{ni}))_{i=1}^n : F \in \mathscr{F}\}$ is a closed subset of $[0, 1]^n$, then $\widehat{\mu}_n$ exists almost surely, and*

$$\mathbb{E}\left(\frac{1}{n} \sum_{i=1}^n (\widehat{\mu}_n(X_{ni}) - \mu(X_{ni}))^2\right) \leq 8n^{-1/2}.$$

Remark 2.3. *In the case of $\mathscr{F} = \mathfrak{F}_{\mathrm{blc}} \cup \{0\} \cup \{1\}$, the set*

$$\mathbb{F}_n := \{(F(X_{ni}))_{i=1}^n : F \in \mathscr{F}\}$$

is a closed subset of $[0, 1]^n$.

2.3.2. Hellinger Consistency in a Special Setting

We consider the special case when $X_{n1}, X_{n2}, ..., X_{nn}$ are the order statistics of i.i.d. $X_1, ..., X_n$. Assume that $X_i \sim Q$, where Q is a probability measure supported by the compact $[a, b] \subset J(\mu)$ such that $a, b \in \text{supp}(Q) \subset [a, b]$ (recall that $J(\mu) = \{x \in \mathbb{R} : 0 < \mu(x) < 1\}$). We also define $\nu := Q \otimes \delta$, where δ is the counting measure on $\{0, 1\}$.

Let $\mathscr{P} := \{P_\mu, \mu \in \mathfrak{F}_{\text{blc}}\}$ be a family of probability measures on some measurable space. We assume that $\mathscr{P}$ is dominated by a σ-finite measure ν. Consider i.i.d. observations $(X_i, Y_i) =: Z_i$, $i = 1, ..., n$ from a probability measure $P_o := P_{\mu_o}$, where $\mu_o(x)$ is a true bi-log-concave regression function. Thus Y_i takes value either 0 or 1. Notice that we consider probability measures P_μ and P_o that correspond to μ and μ_o when they are distribution functions, and to $1 - \mu$ and $1 - \mu_o$ when μ and μ_o are survival functions, respectively. We exclude from this treatment the last remaining and trivial (with respect to the consistency and the rates of convergence of bi-log-concave regression estimator) case of μ and μ_o being constant functions (see Proposition 1.33).

The density in our model is given by

$$f_\mu(x, y) := \frac{dP_\mu}{d\nu} = \mu(x) 1_{\{y=1\}}(y) + (1 - \mu(x)) 1_{\{y=0\}}(y) = \mu(x)^y (1 - \mu(x))^{1-y},$$

where $x \in [a, b]$ and $y \in \{0, 1\}$. Then

$$\frac{dP_\mu}{d\nu}(\cdot, y) = \begin{cases} \mu & \text{if } \;\; y = 1, \\ 1 - \mu & \text{if } y = 0. \end{cases}$$

Recall that the Hellinger distance (see, e.g. Le Cam and Lo Yang (2000) and Groeneboom and Jongbloed (2014)) is the distance between probability measures P_μ and P_o or, equivalently, between the corresponding densities f_μ and $f_o := f_{\mu_o}$:

$$\begin{aligned} H(P_\mu, P_o) \equiv \quad H(f_\mu, f_o) \quad &:= \sqrt{\frac{1}{2} \int (\sqrt{f_\mu} - \sqrt{f_o})^2 d\nu} \\ &= \sqrt{1 - \frac{1}{2} \int \sqrt{f_\mu f_o} d\nu}. \end{aligned}$$

Notice, however, that we will follow the notation of van de Geer (1993), namely:

$$h(f_\mu, f_o) \equiv h(P_\mu, P_o) := H(f_\mu, f_o).$$

By definition of the Hellinger distance, for our setting it follows that:

$$
\begin{aligned}
h(f_\mu, f_o) &= \sqrt{\frac{1}{2}\int(\sqrt{f_\mu}-\sqrt{f_o})^2 d\nu} \\
&= \sqrt{\frac{1}{2}\int_a^b \left(\int_{\{0,1\}}(\sqrt{f_\mu}-\sqrt{f_o})^2 d(\delta_o+\delta_1)\right) dQ} \\
&= \sqrt{\frac{1}{2}\int_a^b ((\sqrt{1-\mu}-\sqrt{1-\mu_o})^2+(\sqrt{\mu}-\sqrt{\mu_o})^2)dQ} \\
&= \sqrt{1-\int_a^b (\sqrt{(1-\mu)(1-\mu_o)}+\sqrt{\mu\mu_o})dQ},
\end{aligned}
$$

where δ_o and δ_1 are the Dirac measures at points $\{0\}$ and $\{1\}$, respectively. Now we are ready to introduce the following

Definition 2.4. *The Hellinger distance $h(\cdot,\cdot)$ between μ and $\widehat{\mu}_n \equiv \widehat{\mu}$ is defined as*

$$
\begin{aligned}
h(\widehat{\mu}_n, \mu) &:= \sqrt{\frac{1}{2}\int_a^b \left(\left(\sqrt{1-\mu}-\sqrt{1-\widehat{\mu}_n}\right)^2+\left(\sqrt{\mu}-\sqrt{\widehat{\mu}_n}\right)^2\right) dQ} \\
&= \sqrt{1-\int_a^b \left(\sqrt{(1-\mu)(1-\widehat{\mu}_n)}+\sqrt{\mu\widehat{\mu}_n}\right) dQ}.
\end{aligned}
$$

In this setting the rate of convergence of the estimators $\widehat{\mu}_n$ on the compact $[a, b]$ is obtained via the following

Theorem 2.5. *Let $\mu \in \mathcal{F}_{\mathrm{blc}}$; then*

$$h(\widehat{\mu}_n, \mu) = O_p\left(\frac{1}{n^{2/5}}\right).$$

Notice that this rate of convergence is the optimal rate and $\widehat{\mu}_n$ are asymptotically optimal (on $[a, b]$) with respect to L_1-norm, i.e. the total variation distance (see Stone (1982), p. 1042, Eggermont and LaRiccia (2009), p. 19, and Appendix A). Namely, the following result holds true:

Corollary 2.6.

$$||\widehat{\mu}_n - \mu||_{1,Q} = O_p\left(\frac{1}{n^{2/5}}\right).$$

This is the optimal convergence rate and $\widehat{\mu}_n$ are asymptotically optimal estimators on the compact $[a, b]$.

2.4. Algorithms

Taylor expansions of second order. For $\theta, h \in \mathbb{R}$, as $h \to 0$,

$$\begin{aligned}
\log \ell(\theta + h) &= \log\Big(\ell(\theta) + \ell'(\theta)h + \ell''(\theta)\frac{h^2}{2} + O(h^3)\Big) \\
&= \log \ell(\theta) + \log\Big(1 + \frac{\ell'(\theta)}{\ell(\theta)}h + \frac{\ell''(\theta)}{\ell(\theta)}\frac{h^2}{2} + O(h^3)\Big) \\
&= \log \ell(\theta) + \frac{\ell'(\theta)}{\ell(\theta)}h + \frac{\ell''(\theta)}{\ell(\theta)}\frac{h^2}{2} - \frac{\ell'(\theta)^2}{\ell(\theta)^2}\frac{h^2}{2} + O(h^3) \\
&= \log \ell(\theta) + \ell(-\theta)h - \ell'(\theta)\frac{h^2}{2} + O(h^3)
\end{aligned}$$

Here we utilized the formulae $\ell' = \ell(1-\ell)$, $\ell'' = (1-2\ell)\ell'$ and $1 - \ell = \ell(-\,\cdot)$. The latter equation yields

$$\begin{aligned}
\log(1 - \ell(\theta + h)) &= \log \ell(-\theta - h) \\
&= \log \ell(-\theta) - (1 - \ell(-\theta))h - \ell(-\theta)(1 - \ell(-\theta))\frac{h^2}{2} + O(h^3) \\
&= \log(1 - \ell(\theta)) - \ell(\theta)h - \ell'(\theta)\frac{h^2}{2} + O(h^3).
\end{aligned}$$

To determine the Newton step, we need various second order Taylor expansions. First of all, our expansions of $\log \ell$ and $\log(1-\ell)$ imply that for arbitrary $\boldsymbol{\theta}, \boldsymbol{v} \in \mathbb{R}^N$,

$$L(\boldsymbol{\theta} + \boldsymbol{v}) = L(\boldsymbol{\theta}) + \nabla L(\boldsymbol{\theta})^\top \boldsymbol{v} + \frac{1}{2}\boldsymbol{v}^\top D^2 L(\boldsymbol{\theta})\boldsymbol{v} + O(\|\boldsymbol{v}\|^3)$$

as $\boldsymbol{v} \to \mathbf{0}$, where

$$\begin{aligned}
\nabla L(\boldsymbol{\theta}) &= \big(w_{j0}\ell(\theta_j) - w_{j1}\ell(-\theta_j)\big)_{j=1}^N \\
&= \Big(\frac{w_{j0}e^{\theta_j} - w_{j1}}{1 + e^{\theta_j}}\Big)_{j=1}^N = \Big(\frac{w_{j0} - w_{j1}e^{-\theta_j}}{1 + e^{-\theta_j}}\Big)_{j=1}^N, \\
D^2 L(\boldsymbol{\theta}) &= \mathrm{diag}\Big(\big((w_{j0} + w_{j1})\ell'(\theta_j)\big)_{j=1}^N\Big) = \mathrm{diag}\Big(\Big(\frac{w_{j0} + w_{j1}}{e^{\theta_j} + e^{-\theta_j} + 2}\Big)_{j=1}^N\Big).
\end{aligned}$$

Moreover, for $2 \le j < N$ and $y \in \{0, 1\}$,

$$s_{jy}(\boldsymbol{\theta} + \boldsymbol{v}) = s_{jy}(\boldsymbol{\theta}) + \boldsymbol{\gamma}_{jy}^\top \boldsymbol{v} + \frac{1}{2}\boldsymbol{v}^\top \mathrm{diag}(\boldsymbol{\eta}_j)\boldsymbol{v} + O(\|\boldsymbol{v}\|^3)$$

as $\boldsymbol{v} \to \mathbf{0}$, where

$$\begin{aligned}
\boldsymbol{\gamma}_{j0} &:= \Big(\underbrace{0,\dots,0}_{j-2}, \quad \lambda_{j\ell}\ell(\theta_{j-1}), \quad -\ell(\theta_j), \quad \lambda_{jr}\ell(\theta_{j+1}), \quad \underbrace{0,\dots,0}_{N-j-1}\Big)^\top, \\
\boldsymbol{\gamma}_{j1} &:= \Big(\underbrace{0,\dots,0}_{j-2}, \quad -\lambda_{j\ell}\ell(-\theta_{j-1}), \quad \ell(-\theta_j), \quad -\lambda_{jr}\ell(-\theta_{j+1}), \quad \underbrace{0,\dots,0}_{N-j-1}\Big)^\top, \\
\boldsymbol{\eta}_{j} &:= \Big(\underbrace{0,\dots,0}_{j-2}, \quad \lambda_{j\ell}\ell'(\theta_{j-1}), \quad -\ell'(\theta_j), \quad \lambda_{jr}\ell'(\theta_{j+1}), \quad \underbrace{0,\dots,0}_{N-j-1}\Big)^\top.
\end{aligned}$$

Consequently, if $R(\boldsymbol{\theta}) < \infty$, then

$$\begin{aligned}
\log s_{jy}(\boldsymbol{\theta}+\boldsymbol{v}) &= \log\big(s_{jy}(\boldsymbol{\theta}) + \boldsymbol{\gamma}_{jy}^\top \boldsymbol{v} + \frac{1}{2}\,\boldsymbol{v}^\top \operatorname{diag}(\boldsymbol{\eta}_j)\boldsymbol{v} + O(\|\boldsymbol{v}\|^3)\big) \\
&= \log s_{jk}(\boldsymbol{\theta}) + \frac{1}{s_{jy}(\boldsymbol{\theta})}\,\boldsymbol{\gamma}_{jy}^\top \boldsymbol{v} \\
&\quad + \frac{1}{2}\,\boldsymbol{v}^\top\Big(\frac{1}{s_{jy}(\boldsymbol{\theta})}\operatorname{diag}(\boldsymbol{\eta}_j) - \frac{1}{s_{jy}(\boldsymbol{\theta})^2}\,\boldsymbol{\gamma}_{jy}\boldsymbol{\gamma}_{jy}^\top\Big)\boldsymbol{v} + O(\|\boldsymbol{v}\|^3).
\end{aligned}$$

All in all, for $\boldsymbol{\theta}, \boldsymbol{v} \in \mathbb{R}^N$ with $R(\boldsymbol{\theta}) < \infty$,

$$R(\boldsymbol{\theta}+\boldsymbol{v}) = R(\boldsymbol{\theta}) + \nabla R(\boldsymbol{\theta})^\top \boldsymbol{v} + 2^{-1}\boldsymbol{v}^\top D^2 R(\boldsymbol{\theta})\boldsymbol{v} + O(\|\boldsymbol{v}\|^3)$$

as $\boldsymbol{v} \to \mathbf{0}$, where

$$\begin{aligned}
\nabla R(\boldsymbol{\theta}) &= -\sum_{j=2}^{N-1}\Big(\frac{1}{s_{j0}(\boldsymbol{\theta})}\,\boldsymbol{\gamma}_{j0} + \frac{1}{s_{j1}(\boldsymbol{\theta})}\,\boldsymbol{\gamma}_{j1}\Big), \\
D^2 R(\boldsymbol{\theta}) &= \sum_{j=2}^{N-1}\Big(-\Big(\frac{1}{s_{j0}(\boldsymbol{\theta})} + \frac{1}{s_{j1}(\boldsymbol{\theta})}\Big)\operatorname{diag}(\boldsymbol{\eta}_j) + \frac{\boldsymbol{\gamma}_{j0}\boldsymbol{\gamma}_{j0}^\top}{s_{j0}(\boldsymbol{\theta})^2} + \frac{\boldsymbol{\gamma}_{j1}\boldsymbol{\gamma}_{j1}^\top}{s_{j1}(\boldsymbol{\theta})^2}\Big).
\end{aligned}$$

Logarithmic barrier algorithm with the Newton iteration. We implemented (see Algorithm 1) the logarithmic barrier (penalty) algorithm (see, e.g. Boyd and Vandenberghe (2009), p.563 for a general description) with the penalty equal $R(\boldsymbol{\theta})$. It takes the triple $(\mathfrak{X}, \mathfrak{T}, \mathfrak{Y})$ as an input data, where $\mathfrak{X}$ is the sample of the raw data grid points $\{X_1, X_2, \dots, X_{\widetilde{n}}\}$ and $\mathfrak{Y}$ is the corresponding sample of the binary values $(Y_i)_{i=1}^{\widetilde{n}}$. At the very beginning, the preprocessing of the input data is performed in order to determine the unique row data points in the case of ties. The data obtained from the preprocessing are the vectors $\mathbf{X}$ and $\mathbf{Y}$ with unique ordered components. That is, $\boldsymbol{X} = \{x_1 < x_2 < \dots < x_n\}$. Components of $\mathbf{Y} := (y_1, y_2, ..., y_n)$ are the weighted values of $\mathfrak{Y}$ with respect to the frequencies of their occurrence in the original sample $\mathfrak{Y}$. That is, $y_i := \underset{1\le k\le \widetilde{n}}{\text{mean}}(Y_k : X_k = x_i)$ with $1 \le i \le n$. The weights $\boldsymbol{w} := (w_i)_{i=1}^n$ are computed as $w_i := \#\{k \leq \widetilde{n} : X_k = x_i\}$. Notice that the negative log-likelihood which is actually computed, is

$$L(\boldsymbol{\theta}) := -\sum_{i=1}^{n} w_i \big(y_i \log(\ell(\theta_i)) + (1 - y_i)\log(\ell(-\theta_i))\big), \tag{2.4}$$

where $\theta(\cdot) := \operatorname{logit} \mu(\cdot)$.

At the next stage, the starting value β_o for the (logarithmic barrier) penalty parameter β is chosen; it is typically between 0.1 and 10. Then the starting value θ_o of the parameterized solution $\boldsymbol{\theta} := \operatorname{logit}(\mu(\mathfrak{T}))$ to minimization problem

$$\min_{\boldsymbol{\theta} \in \mathbb{R}^N} L_\beta(\boldsymbol{\theta})$$

is obtained by fitting the logistic regression model to raw data $\mathfrak{X}$ and $\mathfrak{Y}$; recall that $L_\beta(\boldsymbol{\theta}) = L(\boldsymbol{\theta}) + \beta R(\boldsymbol{\theta})$.

The algorithm consists of the outer and the inner loops. The purpose of the outer loop is to decrease β. At each iteration of the outer loop at most $k_o > 0$ iterations of the inner loop are executed. We call it a Newton loop because this is where the Newton-Raphson method is actually implemented. At every instance of the Newton loop the gradient

$$\mathbf{g} =: \nabla L_\beta(\boldsymbol{\theta}) = \nabla L(\boldsymbol{\theta}) + \beta \nabla R(\boldsymbol{\theta})$$

and Hessian

$$\mathbf{H} =: D^2 L_\beta(\boldsymbol{\theta}) = D^2 L(\boldsymbol{\theta}) + \beta D^2 R(\boldsymbol{\theta})$$

of the target functional $L_\beta(\boldsymbol{\theta})$ (denoted by **LogLike** in Algorithm 1) are computed. It is done by utilizing the function **PenLogLike** which also computes negative (penalized) log-likelihood (logarithmic barrier function) with penalty parameter β, and checks bi-log-concavity constraints (2.2) and (2.3). This function takes 3 parameters as an input: $\boldsymbol{\theta}$, β and boolean d. The latter one is set to `TRUE` when Hessian and gradient are required; otherwise d is set to `FALSE`.

Having computed the gradient and Hessian, the algorithm solves linear system

$$\mathbf{H} \cdot \Delta\boldsymbol{\theta} = -\mathbf{g}$$

for the Newton step $\Delta\boldsymbol{\theta}$ of the current solution $\boldsymbol{\theta}$. After that, the incremental change is calculated as a scalar product of the solution's step and its gradient $\kappa := \mathbf{g} \cdot \Delta\boldsymbol{\theta}$ which is the directional derivative of L_β.

The next guess for the solution is $\boldsymbol{\theta} + a\Delta\boldsymbol{\theta}$, where a is the step size (initially set to 1) and **PenLogLike** function is applied again to calculate the updated value of the negative (conditional) log-likelihood **LogLike** at a new solution guess. After that everything is ready to correct the step size a. To do this, we use the Armijo condition. This condition can be described as follows (see, e.g., Nocedal and Wright (2006)):

$$L_\beta(\boldsymbol{\theta} + a\Delta\boldsymbol{\theta}) \leq L_\beta(\boldsymbol{\theta}) + c_1 a \kappa,$$

where c_1 is some constant in $(0, 1)$. That is a should first of all give sufficient decrease in the objective function $L_\beta(\cdot)$, as measured by this inequality. In other words, the reduction in L_β should be proportional to both the step length a and directional derivative κ.

In the Armijo (condition) procedure implemented here, c_1 is set to 0.3 and the step size a is divided by 2 at each iteration of procedure's loop. At every iteration of the Armijo procedure the **PenLogLike** function is utilized to calculate the updated value **LogLike** at a new solution guess with corrected step size $a_{cor} := a/2$. Notice that in this procedure bi-log-concavity constraints (2.2) and (2.3) are verified to insure that the next solution guess $\boldsymbol{\theta} + a_{cor}\Delta\boldsymbol{\theta}$ satisfies them. The step size correction loop stops after the sufficient decrease in the negative log-likelihood **LogLike** was achieved and all bi-log-concavity constraints were satisfied.

After getting corrected step size a_{cor}, the solution is updated as

$$\boldsymbol{\theta}_{new} := \boldsymbol{\theta} + a_{cor}\Delta\boldsymbol{\theta}.$$

The algorithm stops after β was decreased sufficiently enough, namely, has dropped below 10^{-8} in the implementation. It returns estimate

$$\widehat{\boldsymbol{\mu}}_n \equiv \widehat{\boldsymbol{\mu}} = (1 + \exp(-\boldsymbol{\theta}))^{-1}$$

of bi-log-concave regression function μ and the log-likelihood value.

There are several strategies to decrease parameter β. The one which proved to be the most (computationally) efficient in our framework, proceeds as follows. During the first 10 iterations the barrier parameter β is reduced by half each time, and then is multiplied by the factor of 0.1 every subsequent iteration. Newton loop is usually restricted to $k_o = 500$ iterations and the safeguard stopping rule is designed in such a way that the loop stops when the incremental change κ becomes less than some predefined value $r \in (0, 1)$. This parameter is typically (and in the algorithm as well) set to $r_o = 10^{-9}$. The algorithm was implemented in the statistical computing environment R. Notice that the actual implementation of the algorithm computes and returns also the isotonic (antitonic in the case when μ is survival function) estimator (see e.g. Groeneboom and Jongbloed (2014) for details), its smoothed version, logistic regression estimator and their corresponding log-likelihoods. In terms of the computation speed, the simulations have shown that the best case performance has the order $O(\log^2 n\sqrt{n})$, where n is the sample size of $\boldsymbol{X}$. The worst case performance has the order $O(n^2)$. The reason is that the number of iterations of the Newton loop is bounded by a constant having the same order as n. The best case performance for this loop is $\log n$. The usual safeguarding rule for Armijo criterion procedure bounds the number of iterations by 20 giving the order of $\log n$ to $\sqrt{n}$, while the number of outer loop iterations is kept quite small (between 12 and 17) having the order of $\sqrt{n}$.

Algorithm 1 Binary bi-log-concave regression: part I

procedure BiLogConcaveRegression($\mathfrak{X}, \mathfrak{T}, \mathfrak{Y}$)
  $n \leftarrow$ numberOfUniqueValues($\mathfrak{X}$)
  $\widetilde{n} \leftarrow$ length($\mathfrak{X}$)
  $\mathbf{X} := (x_i)_{i=1}^n \leftarrow$ uniqueOrderedValues($\mathfrak{X}$)
  $\boldsymbol{w} \leftarrow (w_i)_{i=1}^n : \ \ w_i := \#\{1 \le k \le \widetilde{n} : \ X_k = x_i\}$
  $\boldsymbol{Y} \leftarrow (y_i)_{i=1}^n : \ \ y_i := \underset{1 \le k \le \widetilde{n}}{\text{mean}} (Y_k : \ \ X_k = x_i)$
  $N \leftarrow$ length($\mathfrak{T}$)
  $\beta \leftarrow \beta_o$
  $\mathbf{T} \leftarrow$ sort($\mathfrak{T}$)
  $\mathbf{M} \leftarrow (\mathbf{T}, ..., \mathbf{T}) - (\mathbf{T}', ..., \mathbf{T}')'$
  for $i \leftarrow 2, ..., N-1$ **do**
    $\lambda_{r,i-1} \leftarrow \mathbf{M}_{i,i-1}\mathbf{M}_{i+1,i-1}^{-1}$
    $\lambda_{\ell,i-1} \leftarrow \mathbf{M}_{i+1,i}\mathbf{M}_{i+1,i-1}^{-1}$
  end for
  $\mathbf{W} \leftarrow \text{Matrix}_{i=1,j=1}^{N,2}(0.0)$
  for $j \leftarrow 1, ..., N$ **do**
    for $i \leftarrow 1, ..., \widetilde{n}$ **do**
      $\mathbf{W}_{j2} \leftarrow \mathbf{W}_{j2} + 1_{\{\mathbf{X_i} = \mathbf{T}_j \cap Y_i = 0\}}$
      $\mathbf{W}_{j1} \leftarrow \mathbf{W}_{j1} + 1_{\{\mathbf{X_i} = \mathbf{T}_j \cap Y_i = 1\}}$
    end for
  end for
  function PenLogLike($\boldsymbol{\theta}, \beta, d$)
    $k \leftarrow$length($\boldsymbol{\theta}$)
    $\boldsymbol{\ell}_o \leftarrow \text{vector}_{i=1}^k(10^{-8})$
    $\boldsymbol{\ell}_1 \leftarrow \text{vector}_{i=1}^k(10^{-8})$
    $J_+ \leftarrow (\boldsymbol{\theta} > 0)$
    $\boldsymbol{\ell}_o[J_+] \leftarrow -\boldsymbol{\theta}[J_+] - \log(1 + \exp(-\boldsymbol{\theta}[J_+]))$
    $\boldsymbol{\ell}_1[J_+] \leftarrow -\log(1 + \exp(-\boldsymbol{\theta}[J_+]))$
    $J_- \leftarrow (\boldsymbol{\theta} \le 0)$
    $\boldsymbol{\ell}_o[J_-] \leftarrow -\log(1 + \exp(\boldsymbol{\theta}[J_-]))$
    $\boldsymbol{\ell}_1[J_-] \leftarrow \boldsymbol{\theta}[J_-] - \log(1 + \exp(\boldsymbol{\theta}[J_-]))$
    $\mathbf{s}_o \leftarrow \boldsymbol{\ell}_{o,2:(N-1)} - \lambda_\ell \boldsymbol{\ell}_{o,1:(N-2)} - \lambda_{\mathbf{r}} \boldsymbol{\ell}_{o,3:N}$
    $\mathbf{s}_1 \leftarrow \boldsymbol{\ell}_{1,2:(N-1)} - \lambda_\ell \boldsymbol{\ell}_{1,1:(N-2)} - \lambda_{\mathbf{r}} \boldsymbol{\ell}_{1,3:N}$
    if $\mathbf{s}_o \le 0$ **or** $\mathbf{s}_1 \le 0$ **then**
      $\textbf{SumFunc} \leftarrow \infty$
      $\textbf{SumLog} \leftarrow \infty$
    else
      $\textbf{SumFunc} \leftarrow -\mathbf{W}_{\cdot,1}\boldsymbol{\ell}_1 - \mathbf{W}_{\cdot,2}\boldsymbol{\ell}_o$
      $\textbf{SumLog} \leftarrow -\beta(\sum_{i=1}^{N-2} \log(\mathbf{s}_{oi}) + \sum_{i=1}^{N-2} \log(\mathbf{s}_{1i}))$
    end if
    $\textbf{LogLike} \leftarrow \textbf{SumFunc} + \textbf{SumLog}$
    if $d =$ `TRUE` **then**
      $\mathbf{g}_L \leftarrow \text{vector}_{j=1}^N(0.0)$

Algorithm 1 Binary bi-log-concave regression: part II

$\mathbf{B} \leftarrow \text{vector}_{j=1}^{N}(0.0)$
for $i \leftarrow 1, ..., N$ **do**
  if $\exp(-\boldsymbol{\theta}_i) \neq \infty$ **then**
    $\mathbf{g}_{L,i} \leftarrow -(\mathbf{W}_{i1}e^{-\theta_i} - \mathbf{W}_{i2})(1 + e^{-\theta_i})^{-1}$
    $\mathbf{B}_i \leftarrow (\mathbf{W}_{i1} + \mathbf{W}_{i2})(2 + e^{-\theta_i} + e^{\theta_i})^{-1}$
  end if
end for
$\mathbf{H}_L \leftarrow \text{diag}(\mathbf{B})$
if $\mathrm{s}_o \leq (0, ..., 0)$ **or** $\mathrm{s}_1 \leq (0, ..., 0)$ **then**
  $\mathbf{g} \leftarrow \texttt{NULL}$
  $\mathbf{H} \leftarrow \texttt{NULL}$
  return $\mathbf{LogLike}, \mathrm{s}_o, \mathrm{s}_1, \mathbf{g}, \mathbf{H}, \mathbf{SumFunc}, \mathbf{SumLog}$
else
  $\boldsymbol{\gamma} \leftarrow \text{Matrix}_{i=1,j=1}^{N-2,3}(10^{-8})$
  $\widetilde{\boldsymbol{\gamma}} \leftarrow \text{Matrix}_{i=1,j=1}^{N-2,3}(10^{-8})$
  $\boldsymbol{\eta} \leftarrow \text{Matrix}_{i=1,j=1}^{N-2,3}(10^{-8})$
  $\mathbf{v} \leftarrow (\exp(-\boldsymbol{\theta}) + 1)^{-1}$
  $\mathbf{u} \leftarrow (\exp(\boldsymbol{\theta}) + 1)^{-1}$
  $\boldsymbol{\gamma}_{1:(N-2),1} \leftarrow \lambda_\ell \mathbf{v}_{1:(N-2)}$
  $\boldsymbol{\gamma}_{1:(N-2),2} \leftarrow -\mathbf{v}_{2:(N-1)}$
  $\boldsymbol{\gamma}_{1:(N-2),3} \leftarrow \lambda_\mathbf{r} \mathbf{v}_{3:N}$
  $\widetilde{\boldsymbol{\gamma}}_{1:(N-2),1} \leftarrow -\lambda_\ell \mathbf{u}_{1:(N-2)}$
  $\widetilde{\boldsymbol{\gamma}}_{1:(N-2),2} \leftarrow \mathbf{u}_{2:(N-1)}$
  $\widetilde{\boldsymbol{\gamma}}_{1:(N-2),3} \leftarrow -\lambda_\mathbf{r} \mathbf{u}_{3:N}$
  $\boldsymbol{\eta}_{1:(N-2),1} \leftarrow \lambda_\ell \mathbf{v}_{1:(N-2)} \cdot (\mathbf{1}_{N-2} - \mathbf{v}_{1:(N-2)})$
  $\boldsymbol{\eta}_{1:(N-2),2} \leftarrow -\mathbf{v}_{2:(N-1)} \cdot (\mathbf{1}_{N-2} - \mathbf{v}_{2:(N-1)})$
  $\boldsymbol{\eta}_{1:(N-2),3} \leftarrow \lambda_\mathbf{r} \mathbf{v}_{3:N} \cdot (\mathbf{1}_{N-2} - \mathbf{v}_{3:N})$
  $\mathbf{g}_R \leftarrow -\boldsymbol{\gamma}\mathbf{s}_o^{-1} - \widetilde{\boldsymbol{\gamma}}\mathbf{s}_1^{-1}$
  $\mathbf{H}_R \leftarrow -\text{diag}(\boldsymbol{\eta})(\mathbf{s}_o^{-1} + \mathbf{s}_1^{-1} + \boldsymbol{\gamma}'\boldsymbol{\gamma}\mathbf{s}_o^{-2} + \widetilde{\boldsymbol{\gamma}}'\widetilde{\boldsymbol{\gamma}}\mathbf{s}_1^{-2})$
  $\boldsymbol{H} \leftarrow \boldsymbol{H}_L + \beta \boldsymbol{H}_R$
  $\boldsymbol{g} \leftarrow \boldsymbol{g}_L + \beta \boldsymbol{g}_R$
  return $\mathbf{LogLike}, \mathrm{s}_o, \mathrm{s}_1, \mathbf{g}, \mathbf{H}, \mathbf{SumFunc}, \mathbf{SumLog}$
end if
else
  return $\mathbf{LogLike}, \mathrm{s}_o, \mathrm{s}_1, \mathbf{SumFunc}, \mathbf{SumLog}$
end if
end function
$(\widehat{a}, \widehat{b}) \leftarrow \text{LogisticRegression}(\mathfrak{Y}, \mathfrak{X})$
$\boldsymbol{\theta}_o \leftarrow \widehat{a} + \widehat{b}\mathbf{T}$
$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta}_o$
$\widehat{\boldsymbol{\mu}}_o \leftarrow (\exp(-\boldsymbol{\theta}_{\boldsymbol{o}}) + 1)^{-1}$
$\kappa \leftarrow 1$

Algorithm 1 Binary bi-log-concave regression: part III

$r \leftarrow r_0$
$s \leftarrow 0$
while $\beta > \beta_o$ **do**
  $s \leftarrow s + 1$
  for $k \leftarrow 1, ..., k_o$ **do**
    $a \leftarrow 1$
    $(\boldsymbol{g}, \boldsymbol{H}) \leftarrow$ **PenLogLike**$(\boldsymbol{\theta}, \beta, d = \texttt{TRUE})$
    if $\boldsymbol{g} = \texttt{NULL}$ **then**
      stop
    end if
    $\Delta\boldsymbol{\theta} \leftarrow -\boldsymbol{H}^{-1}\boldsymbol{g}$
    $\kappa \leftarrow \boldsymbol{g}\Delta\boldsymbol{\theta}$
    LogLikeInit $\leftarrow$ **PenLogLike**$(\boldsymbol{\theta}, \beta, d = \texttt{FALSE})$
    LogLike $\leftarrow$ **PenLogLike**$(\boldsymbol{\theta} + a\Delta\boldsymbol{\theta}, \beta, \texttt{FALSE})$
    $i \leftarrow 0$
    while (**LogLike** $-$ **LogLikeInit**) $> 0.3 \cdot a \cdot \kappa$ **do**
      $i \leftarrow i + 1$
      if $i > 20$ **then**
        stop
      end if
      $a \leftarrow a/2$
      LogLike $\leftarrow$ **PenLogLike**$(\boldsymbol{\theta} + a\Delta\boldsymbol{\theta}, \beta, \texttt{FALSE})$
    end while
    $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + a\Delta\boldsymbol{\theta}$
    if $|\kappa/2| < r$ **then**
      stop
    end if
  end for
  if $s < 11$ **then**
    $\beta \leftarrow \beta/2$
  else
    $\beta \leftarrow \beta/10$
  end if
end while
LogLikeBLC $\leftarrow$ **PenLogLike**$(\boldsymbol{\theta}, \beta, \texttt{FALSE})$
$\widehat{\boldsymbol{\mu}}_n \leftarrow (\exp(-\boldsymbol{\theta}) + 1)^{-1}$
return $\widehat{\boldsymbol{\mu}}_n$, **LogLikeBLC**
end procedure

2.5. Numerical Examples

Hereafter and until the end of the chapter we will write mu(x) instead of $\mu(x)$ to provide uniformity of notations.

Example 2.7. There were $n = 81$ data pairs (X_i, Y_i) with $(X_i) =: \mathbf{X} \subset \mathbf{T}$ being a fixed equidistant grid on the interval $[-4, 4]$ with step length $\Delta(\mathbf{X}) = 0.1$. True bi-log-concave regression function mu($\mathbf{T}$) (red curve) was generated as the c.d.f. of the uniform distribution on the interval $[-4, 4]$ with values computed at the knots of the fine grid $\mathbf{T} = (t_1, ..., t_N) = (-4, -3.95, -3.9, ..., 3.95, 4)$ consisting of $N = 161$ points (the step length is $\Delta(\mathbf{T}) = 0.05$). Blue circles in Figure 2.1 represent the values of $\mathbf{X}$ corresponding to (Y_i) which are either 0 or 1. Solid blue curve shows bi-log-concave regression function estimator $\widehat{\text{mu}}$; green curve corresponds to the logistic fit and grey curve shows isotonic least-squares regression estimator (see e.g. Groeneboom and Jongbloed (2014) for details). The starting value for the (logarithmic barrier) penalty parameter was chosen as $\beta = 0.1$ and the algorithm stopped when β had decreased below 10^{-8}, which corresponds to 14 outer loop iterations. During the first 10 iterations the barrier parameter was decreased by a factor of 0.5 each time and then by a factor of 0.1. Inner Newton loop was restricted to 500 iterations and the safeguard stopping rule parameter was set to $d = 10^{-9}$. The increase in the negative conditional log-likelihood from -43.044 to -41.84 was obtained, comparing to the initial value produced by the standard logistic regression estimation procedure. The log-likelihood corresponding to isotonic regression estimator was -36.81.

Example 2.8. There were $n = 101$ pairs (X_i, Y_i) with $(X_i) =: \mathbf{X} \subset \mathbf{T}$ a fixed equidistant grid on the interval $[-10, 10]$ with the step length $\Delta(\mathbf{X}) = 0.2$. True bi-log-concave regression function mu($\mathbf{T}$) was simulated as the c.d.f. of the normal distribution with mean zero and variance 4 with values computed at the knots of the fine grid $\mathbf{T} = (t_1, ..., t_N) = (-10, -9.8, ..., 9.8, 10)$ consisting of $N = 201$ points (i.e. $\Delta(\mathbf{T}) = 0.1$). Blue curve in Figure 2.2 shows bi-log-concave fit $\widehat{\text{mu}}$ together with isotonic regression estimator and the logistic regression fit. The starting value for the barrier parameter was set to $\beta = 0.1$ and the algorithm stopped when it had decreased below 10^{-8}, which corresponds to 14 outer loop iterations. In this case, the increase in the negative (conditional) log-likelihood from -28.731 to -27.169 was obtained, comparing to the initial value produced by the standard logistic regression estimation procedure. The log-likelihood corresponding to isotonic regression estimator is -22.886.

Example 2.9. Here there are $n = 101$ pairs (X_i, Y_i) with $(X_i) =: \mathbf{X} \subset \mathbf{T}$ a fixed equidistant grid on the interval $[-10, 10]$ with the step $\Delta(\mathbf{X}) = 0.2$. True regression function mu($\mathbf{T}$) (red curve) was simulated as the c.d.f. of the logistic distribution with shape parameter $k = 5$ and scale parameter $s = 2$ with values

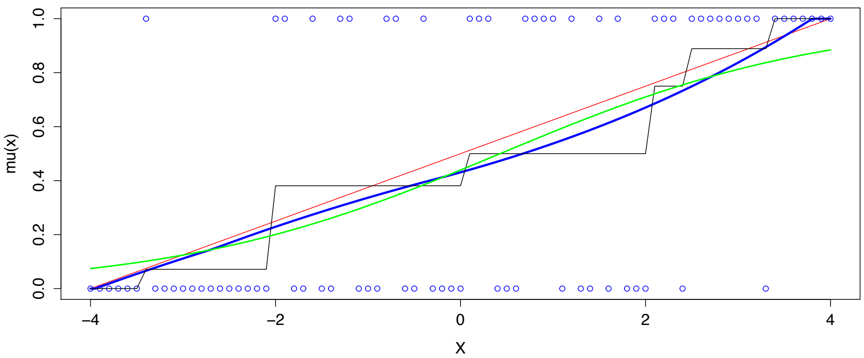


Figure 2.1.: Example of the estimated mu(x) for the data $\mathbf{X}$ simulated from the uniform distribution: blue circles are the values of $\mathbf{X}$ corresponding to (Y_i) - either 0 or 1; solid blue curve is bi-log-concave regression function estimator $\widehat{\text{mu}}$; green curve is the logistic fit; grey curve shows isotonic least-squares regression estimator; red curve is the true bi-log-concave regression function mu(x).

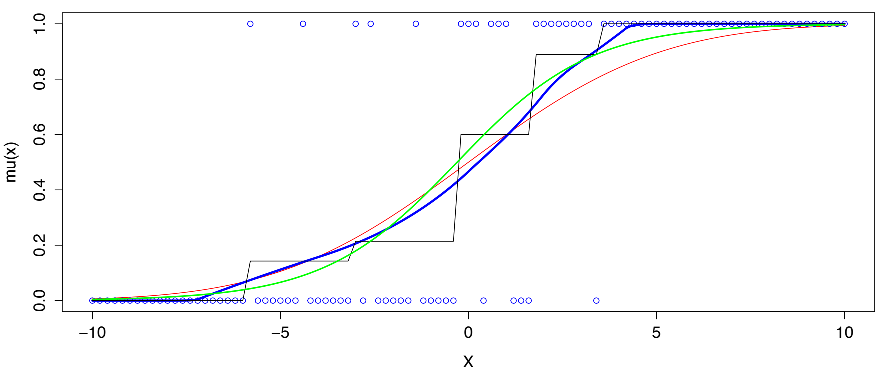


Figure 2.2.: Example of the estimated mu(x) for the data $\mathbf{X}$ simulated from the normal distribution: blue circles are the values of $\mathbf{X}$ corresponding to (Y_i) - either 0 or 1; solid blue curve is bi-log-concave regression function estimator $\widehat{\text{mu}}$; green curve is the logistic fit; grey curve shows isotonic least-squares regression estimator; red curve is the true bi-log-concave regression function mu(x).

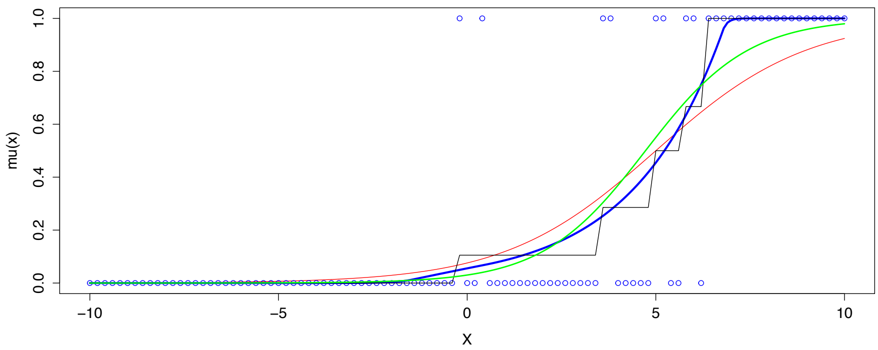


Figure 2.3.: Example of the estimated mu(x) for the data $\mathbf{X}$ simulated from the logistic distribution.

computed at the knots of the fine grid $\mathbf{T} = (-10, -9.9, ..., 9.9, 10)$ consisting of $N = 201$ points (thus $\Delta(\mathbf{T}) = 0.1$). Blue curve in Figure 2.3 shows bi-log-concave regression estimator $\widehat{\text{mu}}$. The starting value for the barrier parameter was set to $\beta = 0.1$ and the algorithm stopped when it had decreased below 10^{-8}, which had corresponded to 14 outer loop iterations. In this case, the increase in the (negative) log-likelihood from -21.639 to -18.897 was obtained, comparing to the initial value produced by the standard logistic regression estimation procedure. The log-likelihood corresponding to isotonic regression estimator is -15.264.

Example 2.10. In this case there are $n = 76$ pairs (X_i, Y_i) with $(X_i) =: \mathbf{X} \subset \mathbf{T}$ a fixed equidistant grid on the interval $[0, 15]$ with the step $\Delta(\mathbf{X}) = 0.2$. True bi-log-concave regression function mu($\mathbf{T}$) (red curve) was simulated as the c.d.f. of the Gamma distribution with shape parameter $k = 2$ and scale parameter $s = 0.5$ with values computed at the knots of the fine grid $\mathbf{T} = (0, 0.1, ..., 14.9, 15)$ consisting of $N = 151$ points (i.e. $\Delta(\mathbf{T}) = 0.1$). Blue curve in Figure 2.4 shows bi-log-concave regression function estimator $\widehat{\text{mu}}$ together with isotonic (grey curve) and logistic regression (green curve) fits. The increase in the log-likelihood from -27.043 to -23.699 was achieved. The log-likelihood corresponding to isotonic regression estimator is -19.245.

Example 2.11. There are $n = 101$ pairs (X_i, Y_i) with $(X_i) =: \mathbf{X} \subset \mathbf{T}$ a fixed equidistant grid on the interval $[-3, 17]$ with the step $\Delta(\mathbf{X}) = 0.2$. True regression function mu($\mathbf{T}$) (red curve) was simulated as the c.d.f. of the Gumbel distribution with shape parameter $k = 0.5$ and scale parameter $s = 2$ with values computed at the knots of the fine grid $\mathbf{T} = (-3, -2.9, ..., 16.9, 17)$ consisting of $N = 201$ points (thus $\Delta(\mathbf{T}) = 0.1$). Blue curve in Figure 2.5 shows bi-log-concave regression

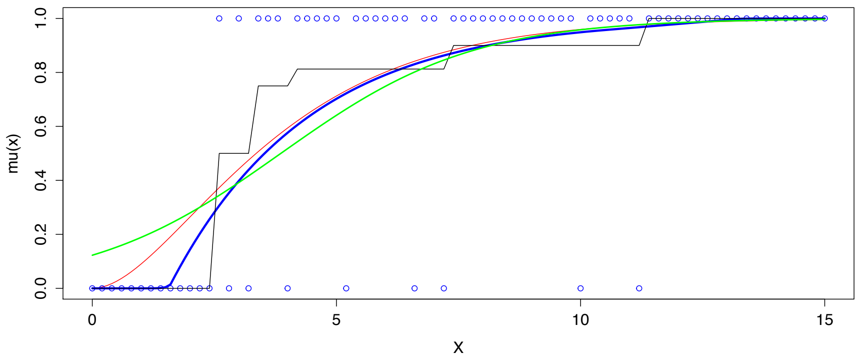


Figure 2.4.: Example of the estimated mu(x) for the data $\mathbf{X}$ simulated from the Gamma distribution.

function estimator $\widehat{\text{mu}}$. The increase in the log-likelihood from -19.874 to -17.249 was achieved. The log-likelihood corresponding to isotonic regression estimator is -13.521.

Example 2.12. Here there are $n = 101$ pairs (X_i, Y_i) with $(X_i) =: \mathbf{X} \subset \mathbf{T}$ a fixed equidistant grid on the interval $[-4, 4]$ with the step length $\Delta(\mathbf{X}) = 0.08$. True regression function mu($\mathbf{T}$) (red curve) was simulated as the c.d.f. of the bimodal distribution in Example 1.15; its values were computed at the knots of the fine grid $\mathbf{T} = (-4, -3.96, ..., 3.96, 4)$ consisting of $N = 201$ points ($\Delta(\mathbf{T}) = 0.04$). Blue curve in Figure 2.6 shows bi-log-concave estimator. In this case, the increase in the log-likelihood from -40.375 to -36.526 was obtained. The log-likelihood corresponding to isotonic regression estimator is -29.603.

Example 2.13. Here $n = 71$ pairs (X_i, Y_i) were simulated with $(X_i) =: \mathbf{X} \subset \mathbf{T}$ a fixed equidistant grid on the interval $[-6, 8]$ with the step $\Delta(\mathbf{X}) = 0.2$. True regression function mu($\mathbf{T}$) (red curve) was simulated as the survival function corresponding to the normal distribution with mean 2 and variance 3 with values computed at the knots of the fine grid $\mathbf{T} = (t_1, ..., t_N) = (-6, -5.9, ..., 7.9, 8)$ consisting of $N = 141$ points (with $\Delta(\mathbf{T}) = 0.1$). Blue curve in Figure 2.7 shows bi-log-concave fit together with isotonic and logistic regression fits. The increase in the log-likelihood from -23.508 to -20.817 was achieved. The log-likelihood corresponding to isotonic regression estimator is -16.668.

Example 2.14. In this example we simulated $n = 51$ observations (X_i, Y_i) with $(X_i) =: \mathbf{X} \subset \mathbf{T}$ a fixed equidistant grid on the interval $[-5, 15]$ with the step $\Delta(\mathbf{X}) = 0.4$. True bi-log-concave regression function mu($\mathbf{T}$) (red curve) was

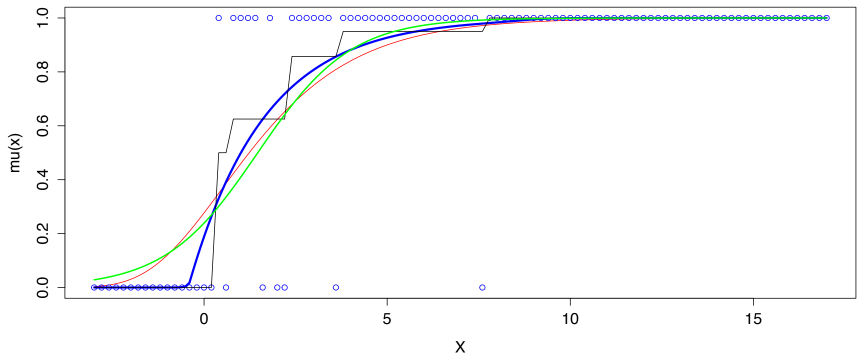


Figure 2.5.: Example of the estimated mu(x) for the data $\mathbf{X}$ simulated from the Gumbel distribution.

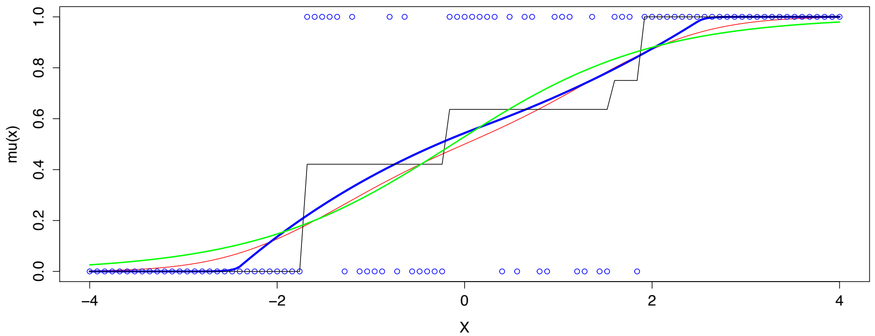


Figure 2.6.: Example of the estimated mu(x) for the data $\mathbf{X}$ simulated from bi-modal distribution in Example 1.15.

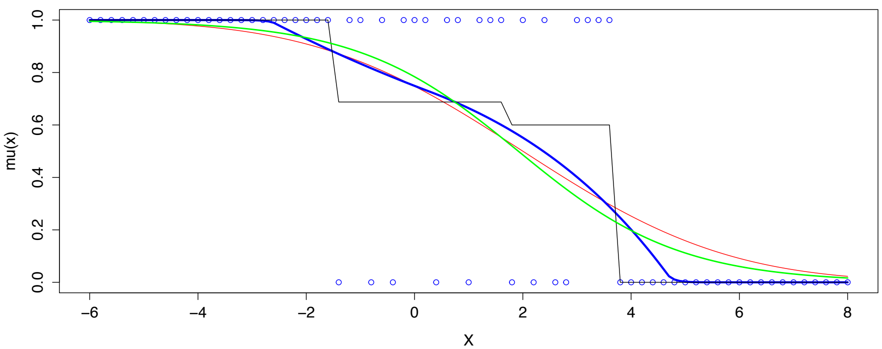


Figure 2.7.: Example of the estimated mu(x) in the case when it is the survival function corresponding to the normal distribution.

simulated as the survival function of the logistic distribution with shape parameter $k = 6$ and scale parameter $s = 2$ with values computed at the knots of the fine grid $\mathbf{T} = (-5, -4.8, ..., 14.8, 15)$ consisting of $N = 101$ points (i.e. $\Delta(\mathbf{T}) = 0.2$). Blue curve in Figure 2.8 shows bi-log-concave regression function estimator $\widehat{\mathrm{mu}}$. The increase (comparing to the starting value produced by the logistic regression estimation) in the log-likelihood from -16.349 to -15.511 was achieved. The log-likelihood corresponding to isotonic regression estimator is -11.402.

Example 2.15. There were $n = 101$ pairs (X_i, Y_i) simulated with $(X_i) =: \mathbf{X} \subset \mathbf{T}$ being a fixed equidistant grid on the interval $[0, 10]$ with the step $\Delta(\mathbf{X}) = 0.1$. True regression function mu($\mathbf{T}$) (red curve) was simulated as the survival function of the Gamma distribution with shape parameter $k = 3$ and scale parameter $s = 1$ with values computed at the knots of the fine grid $\mathbf{T} = (0, 0.05, ..., 9.95, 10)$ consisting of $N = 201$ points ($\Delta(\mathbf{T}) = 0.05$). Blue curve in Figure 2.9 shows bi-log-concave regression function estimator $\widehat{\mathrm{mu}}$. The increase in the log-likelihood is from -25.876 to -23.329.

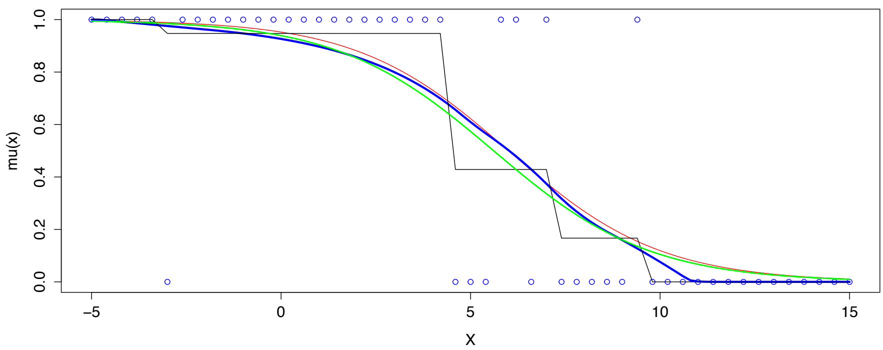


Figure 2.8.: Example of the estimated mu when it is the survival function of the logistic distribution.

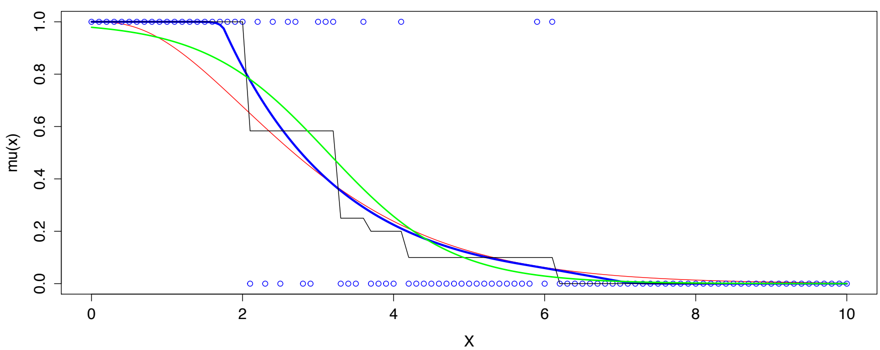


Figure 2.9.: Example of the estimated mu(x) when it is the survival function of the Gamma distribution.

3. Proofs

Proof of Theorem 1.7. When proving Theorem 1.7 we assume that the following facts about concave functions are widely known (cf. Rockafellar (1970), Groeneboom et al. (2001)):

Lemma 3.1. *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 $a \in [h'(x+), h'(x-)]$,

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

The second useful result is the following:

Lemma 3.2. *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 $[a, b]$; this latter means Lipschitz-continuity of h on J.

Proof of Lemma 3.2 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,

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

Equivalence of statements (i-iv) of Theorem 1.7 will be verified in four steps.

Proof of (i) $\Rightarrow$ (ii). Suppose that F is bi-log-concave. First of all, notice that because of non-degeneracy of F, there exist two points $x_1 < x_2$ such that $F(x_1), F(x_2) \in (0,1)$. Since $\log F$ is concave, it follows from Lemma 3.1 that F is continuous on $\{x : \log F(x) > -\infty\} = \{x : F(x) > 0\}$. Furthermore, F is continuous on $\mathbb{R}\setminus\{x_1^o, x_2^o\}$ and $F > 0$ on (x_1^o, x_2^o), where

$$x_1^o := \inf\{x : F(x) > 0\} \leq x_1,$$

$$x_2^o := \sup\{x : F(x) > 0\} \geq x_2.$$

Besides that, $F \equiv 0$ on $(-\infty, x_1^o) \cup (x_2^o, +\infty)$. Analogously, concavity of $\log(1-F)$ implies that F is continuous on $\{y : \log(1-F(y)) > -\infty\} = \{y : F(y) < 1\}$. Furthermore, F is continuous on $\mathbb{R}\setminus\{y_1^o, y_2^o\}$ and $F < 1$ on (y_1^o, y_2^o), where

$$y_1^o := \inf\{y : F(y) < 1\} \leq x_1,$$

$$y_2^o := \sup\{y : F(y) < 1\} \geq x_2.$$

Besides that, $F \equiv 1$ on $(-\infty, y_1^o) \cup (y_2^o, +\infty)$. But this implies that either $F(x_1^o) \equiv 1$ or $F(x_2^o) \equiv 1$ and either $F(y_1^o) \equiv 0$ or $F(y_2^o) \equiv 0$, - a situation which is only possible when points y_1^o, y_2^o coincide with x_1^o, x_2^o and when F is continuous at these points. Therefore F is continuous on $\{x : 0 \leq F(x) \leq 1\}$ and eventually on $\mathbb{R}$. In particular, $J(F) = (a,b)$ for some real a and 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

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

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 (1.1) follow directly from the last part of Lemma 3.1, applied to $h = \log F$ and $h = \log(1 - F)$.

Proof of (ii) ⇒ (iii). Suppose that F is continuous on $\mathbb{R}$, differentiable on $J(F)$ with derivative $f = F'$ and satisfies the inequalities (1.1). 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 (1.1),

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

and

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

whence $h(x) \ge h(y)$ and $\widetilde{h}(x) \le \widetilde{h}(y)$.

Proof of (iii) ⇒ (iv). Suppose that F satisfies the conditions in part (iii). The consequence of these monotonicity properties is the boundedness of f on $J(F)$: if we fix any $x_o \in J(F)$, then for any other point $x \in J(F)$,

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

Finally, local Lipschitz-continuity of f may be verified via Lemma 3.2. Let $c, d \in J(F)$ with $c < d$. For arbitrary different $x, y \in (c, d)$, $x < y$,

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

as $y \to x$ in the case of F non-decreasing on (c, d). Here the first inequality follows from the facts that $F(x) > 0$ and $(h(y) - h(x))(y - x)^{-1} \le 0$. The second inequality is the consequence of inequality (1.1), (1). Hence we obtain

$$\limsup_{y \to x} \frac{f(y) - f(x)}{y - x} \le h(c)^2 F(d) \quad \text{for all } x \in (c, d). \tag{3.1}$$

In the case of F being non-increasing on (c,d), similar derivations will lead to the following inequality:

$$\limsup_{y\to x} \frac{f(y)-f(x)}{y-x} \ \le\ h(c)^2 F(c) \quad \text{for all } x \in (c,d). \tag{3.2}$$

Analogously one can show that

$$\liminf_{y\to x} \frac{f(y)-f(x)}{y-x} \ \ge\ -\widetilde{h}(d)^2(1-F(c)) \quad \text{for all } x \in (c,d) \tag{3.3}$$

in the case of F non-decreasing on (c,d) and

$$\liminf_{y\to x} \frac{f(y)-f(x)}{y-x} \ \ge\ -\widetilde{h}(d)^2(1-F(d)) \quad \text{for all } x \in (c,d) \tag{3.4}$$

in the case of non-increasing F. Therefore, according to Lemma 3.2, f is Lipschitz-continuous on (c,d) with Lipschitz-constant

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

in the case of F non-decreasing on (c,d) and

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

in the case of F non-increasing on (c,d). This proves local Lipschitz-continuity of f on $J(F)$. In particular, f is absolutely continuous with L_1-derivative f'. Then 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 it may be chosen such that

$$f'(x) \ \in\ \Bigl[\liminf_{y\to x} \frac{f(y)-f(x)}{y-x}, \limsup_{y\to x} \frac{f(y)-f(x)}{y-x}\Bigr] \tag{3.5}$$

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

$$\bigl[-\widetilde{h}(d)^2(1-F(c)), h(c)^2F(d)\bigr] \ =\ \Bigl[\frac{-f(d)^2(1-F(c))}{(1-F(d))^2}, \frac{f(c)^2F(d)}{F(c)^2}\Bigr] \tag{3.6}$$

according to (3.1) and (3.3) for non-decreasing F, and in

$$\bigl[-\widetilde{h}(d)^2(1-F(d)), h(c)^2F(c)\bigr] \ =\ \Bigl[\frac{-f(d)^2}{1-F(d)}, \frac{f(c)^2}{F(c)}\Bigr] \tag{3.7}$$

for non-increasing F, according to (3.2) and (3.4). Since F and f are continuous, letting $c, d \to x$ implies (1.2).

Proof of (iv) $\Rightarrow$ (i). One can easily verify that continuous 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 3.2, h is non-increasing on $J(F)$ if, and only if,

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

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 (1.2) and continuity of f that

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

as $y \to x$.

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

Proof of Proposition 1.19. For any fixed $x_o \in J(F)$, monotonicity of $f/F = \log(F)'$ implies that for $x \in J(F)$, $x < x_o$,

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

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

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

Analogously one can show that

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

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

$$\begin{aligned} \int_{\mathbb{R}} e^{tx} F(dx) &= \int_{\mathbb{R}} \int_{\mathbb{R}} 1_{[z \leq x]} t e^{tz} \, dz \, F(dx) \\ &= t \int_{\mathbb{R}} e^{tz} (1 - F(z)) \, dz \\ &= t \int_{\mathbb{R}} \exp\bigl(tz + \log(1 - F(z))\bigr) \, dz. \end{aligned}$$

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 $\int_{\mathbb{R}} e^{tx} F(dx)$ is finite if, and only if,

$$\lim_{z \to \infty} \frac{d}{dz} \bigl(tz + \log(1 - F(z))\bigr) = \lim_{z \to \infty} \Bigl(t - \frac{f(z)}{1 - F(z)}\Bigr) = t - T_2(F)$$

is strictly negative, which is equivalent to $t < T_2(F)$. □

Proof of Lemma 1.23. Assume that F is non-degenerate in the sense of Definition 1.1. Weak convergence of F_n to F means pointwise convergence on the set of all continuity points of F. According to the result from Sengupta and Nanda (1999) (Theorem 2, (d)), distribution function F is log-concave as the pointwise limit of log-concave distribution functions F_n. Then $1 - F_n \to 1 - F$ pointwise on the set of continuity points of $1 - F$ as well. Since each function $1 - F_n$ is log-concave their limit $1 - F$ is log-concave, again by Sengupta and Nanda (1999) (Theorem 2, (d)). Therefore $F \in \mathcal{F}_{\text{blcd}}$.

Bi-log-concave functions F_n, F are continuous on $\mathbb{R}$ and admit strictly positive densities f_n, f on $J(F_n)$, $J(F)$, respectively, by Theorem 1.7, (ii) and (iv)$'$. Therefore, by the standard argument, the sequence of continuous increasing functions F_n converges in probability to increasing and continuous function F. Then, according to the following generalization of Proposition 2.1 from Resnick (2007), we can show the uniform convergence (in other words, convergence in the supremum norm) of F_n to F in probability on $[a, b] \subset \mathbb{R}$ such that $J(F) \cap J(F_n) \cap [a, b] \neq \emptyset$ for each n:

Proposition 3.3. *Let $U_o : K = [a, b] \to \mathbb{R}$ be a fixed continuous and isotonic function, and let (U_n) be a sequence of random isotonic functions $U_n : K \to \mathbb{R}$. If*

$$U_n(x) \to_p U_o(x)$$

for any fixed $x \in K$, then

$$||U_n - U_o||_{\infty, K} \to_p 0.$$

Proof. For fixed $m \in \mathbb{N}$ let

$$x_j := a + \frac{j}{m}(b-a),$$

$0 \le j \le m$. For $x \in [x_{j-1}, x_j]$,

$$\begin{aligned} U_n(x) - U_o(x) &\le (U_n(x_j) - U_o(x_j)) + (U_o(x_j) - U_o(x_{j-1})) \\ &\le M_n + U_o(x_j) - U_o(x_{j-1}), \end{aligned}$$

where

$$M_n := \max_{j=0,1,\ldots,m} |U_n(x_j) - U_o(x_j)| \to_p 0.$$

On the other hand,

$$U_n(x) - U_o(x) \ge -M_n - (U_o(x_j) - U_o(x_{j-1})).$$

Then for $x \in [x_{m-1}, b]$,

$$U_n(x) - U_o(x) \le M_n + U_o(b) - U_o(x_{m-1})$$

and

$$U_n(x) - U_o(x) \ge -M_n - (U_o(b) - U_o(x_{m-1})).$$

For $x \in [x_{m-2}, x_{m-1}]$,

$$U_n(x) - U_o(x) \le M_n + U_o(x_{m-1}) - U_o(x_{m-2})$$

and

$$U_n(x) - U_o(x) \ge -M_n - (U_o(x_{m-1}) - U_o(x_{m-2})).$$

Analogously, for $x \in [x_1, x_2]$,

$$U_n(x) - U_o(x) \le M_n + U_o(x_2) - U_o(x_1),$$

$$U_n(x) - U_o(x) \ge -M_n - (U_o(x_2) - U_o(x_1)).$$

And, finally, for $x \in [a, x_1]$,

$$U_n(x) - U_o(x) \le M_n + U_o(x_1) - U_o(a),$$

$$U_n(x) - U_o(x) \ge -M_n - (U_o(x_1) - U_o(a)).$$

Noticing that

$$U_o(b) - U_o(a) = \max_{j=1,\ldots,m} (U_o(x_j) - U_o(x_{j-1}))$$

and taking into account the definition of M_n we sum up all inequalities with the same sign to obtain

$$n \cdot (U_n(x) - U_o(x)) \leq nM_n + U_o(b) - U_o(a)$$

and

$$n \cdot (U_n(x) - U_o(x)) \geq -nM_n - (U_o(b) - U_o(a))$$

for $x \in [a, b]$. Therefore

$$U_n(x) - U_o(x) \leq M_n + \frac{1}{n}(U_o(b) - U_o(a)),$$

$$U_n(x) - U_o(x) \geq -M_n - \frac{1}{n}(U_o(b) - U_o(a)).$$

Keeping in mind that $M_n \to_p 0$, it is equivalent to

$$U_n(x) - U_o(x) \leq o_p(1)$$

and

$$U_n(x) - U_o(x) \geq -o_p(1),$$

respectively. Hence

$$||U_n - U_o||_{\infty,K} \to_p 0.$$

□

The functions F_n and F are isotonic and therefore $||F_n - F||_{\infty,[a,b]} \to_p 0$ as $n \to \infty$.

Consider now $L_n := \min(F_n, F)$ and $U_n := \max(F_n, F)$. It implies that F, $F_n \in [L_n, U_n]$ and L_n, $U_n \to F$. Then for x, $x' \in [a, b] \cap J(F_n) \cap J(F)$ with $x < x'$ the RHS of the second inequality for hazard rates from Lemma 1.25 implies

$$\begin{aligned} f_n(x) = (1 - F_n)\frac{f_n}{1 - F_n}(x) \quad &\geq \quad (1 - F_n(x))\frac{\log(1 - F_n(x')) - \log(1 - F_n(x))}{x' - x} \\ &\geq \quad (1 - U_n(x))\frac{\log(1 - U_n(x')) - \log(1 - L_n(x))}{x' - x} \\ &\to_p \quad (1 - F(x))\frac{\log(1 - F(x')) - \log(1 - F(x))}{x' - x} \\ &\overset{x' \downarrow x}{\longrightarrow} \quad \frac{f(1 - F)}{1 - F}(x). \end{aligned}$$

Therefore

$$f_n(x) \geq f(x) + o_P(1).$$

Analogously, $f_n(x) \leq f(x) + o_P(1)$ and therefore $f_n \to f$ on $[a, b]$ pointwise. Since f_n is Lipschitz-continuous on $[a, b]$ (Theorem 1.7, (iv)$'$) the sequence (f_n)

is uniformly equicontinuous. Indeed, $\forall \epsilon > 0\ \exists \delta > 0$ such that $\forall x, y \in [a, b]$ and $|x - y| < \delta$:

$$|f_n(x) - f_n(y)| \leq L|x - y| < L\delta = \epsilon$$

if we take $\delta := \epsilon/L$, where $L > 0$ is the corresponding Lipschitz constant. The latter one, as the proof of (iii) $\Rightarrow$ (iv) of Theorem 1.7 shows, does not depend on the choice of x, $y \in [a, b] \subset J(F_n)$. Coupled together, the uniform equicontinuity and pointwise convergence of the sequence (f_n) of (Lipschitz)-continuous functions to (Lipschitz)-continuous function f on the compact $[a, b]$, yields the uniform convergence in probability:

$$||f_n - f||_{\infty,[a,b]} \to_p 0$$

as $n \to \infty$.

Uniform boundedness of f, f_n, f' and f'_n on $[a, b]$ is a consequence of the property (iv)$'$ of bi-log-concave function F (respectively F_n) from Theorem 1.7 and its proof (iii)$\Rightarrow$(iv), (3.5)-(3.7). □

Proof of Corollary 1.24. Recall that (see, e.g. Castillo and Rafeiro (2016), Theorem 3.10, for more general embedding result)

$$L_\infty([a, b]) \subset ... \subset L_2([a, b]) \subset L_1([a, b]),$$

where $L_\infty([a, b])$ is the space of all essentially bounded functions on $[a, b]$. Besides, supremum of a continuous function equals to its essential supremum on Lebesgue-measurable set $E \subseteq \mathbb{R}^m$ whose intersection with any open ball with centre in E has non-zero Lebesgue measure. This is a standard result from the theory of functions and measure theory which obviously holds true for $[a, b] \subset \mathbb{R}$. Thus, taking into account Lemma 1.23, we obtain that for some constants k_q depending on a, b and the exponents $q \in [1, \infty]$ of corresponding Lebesgue spaces, the following inequalities hold true on $[a, b]$:

$$\|f\|_{L_1} \leq k_2\|f\|_{L_2} \leq ... \leq k_\infty\|f\|_{L_\infty} \equiv k_\infty\|f\|_\infty = O_p(k_\infty),$$

and analogously for f_n for all n. By Theorem 1.7, (iv)$'$, f has L_1-derivative f' on $J(F)$ which is integrable on $[a, b]$. Then, using the results of Lemma 1.23, it implies that for some constants k'_q depending on a, b and the exponents $q \in [1, \infty]$ of corresponding Lebesgue spaces, the following inequalities hold true on $[a, b]$:

$$\|f'\|_{L_1} \leq k'_2\|f'\|_{L_2} \leq ... \leq k'_\infty\|f'\|_{L_\infty} \equiv k'_\infty\|f'\|_\infty = O_p(k'_\infty),$$

and analogously for f'_n for all n. Hence f, f_n, f', $f'_n \in L_q([a, b])$ for any $q \in [1, \infty]$. □

Proof of Lemma 1.25. It is well-known (and a consequence of Lemma 3.1) that concavity of $h = \log F$ and $h = \log(1 - F)$ implies that

$$h'(x_1) \geq \frac{h(x_2) - h(x_1)}{x_2 - x_1} \geq h'(x_2).$$

These are the asserted bounds in Lemma 1.25. □

Proof of Lemma 1.31.

We prove this result in several steps. Since $\kappa_{n,\alpha}^{ODW} = O(1)$, it can be replaced with an arbitrary fixed number $\kappa > 0$; so we redefine $\gamma_n(t)$ as

$$\gamma_n(t) = \frac{C(t) + \nu D(t) + \kappa}{n+1},$$

where $\kappa = \kappa_{n,\alpha}^{ODW}$. Let $\check{F}_n = (n+1)^{-1} n \widehat{F}_n$, where $\widehat{F}_n$ is the empirical distribution function of X_1, X_2, ..., X_n.

Step 0. The inequality $T_n^{ODW}(H) \leq \kappa_{n,\alpha}^{ODW}$ has been replaced with $T_n^{ODW}(H) \leq \kappa$. The latter one implies that $J(H) \subset [X_{(1)}, X_{(n)}]$ and

$$K\left(\check{F}_n, H\right) \leq \gamma_n(\check{F}_n) \tag{3.8}$$

on $\{X_{(1)}, X_{(2)}, \ldots, X_{(n)}\}$, where

$$\check{F}_n(X_{(i)}) = \frac{i}{n+1} := t_{ni}.$$

We will show that this inequality holds true for $X \in [X_{(1)}, X_{(n)}]$, i.e. on the set $\{\check{F}_n \in [\delta_n, 1-\delta_n]\}$, where

$$\delta_n = \frac{c_n \log\log n}{n} \to 0,$$

but $0 < c_n \to \infty$.

Indeed, inequality (3.8) means that for $1 \leq i \leq n$

$$K\left(t_{ni}, H(X_{(i)})\right) \leq \gamma_n\left(t_{ni}\right)$$

and

$$K\left(t_{n,i+1}, H(X_{(i+1)})\right) \leq \gamma_n\left(t_{n,i+1}\right) \tag{3.9}$$

on the sets $\{X_{(i)} : \delta_n \leq t_{ni} \leq 1-\delta_n\}$. We distinguish two cases: $H(X) \geq t_{ni}$ and $H(X) \leq t_{ni}$ for $X \in [X_{(i)}, X_{(i+1)}]$.

Case 1. For $X \in [X_{(i)}, X_{(i+1)}]$ such that $H(X) \geq t_{ni}$ it follows from the monotonicity of H and inequalities (3.9) that

$$K\left(t_{ni}, H(X)\right) \leq K\left(t_{ni}, H(X_{(i+1)})\right) =$$

$$K\left(t_{n,i+1}, H(X_{(i+1)})\right) + K\left(t_{ni}, H(X_{(i+1)})\right) - K\left(t_{n,i+1}, H(X_{(i+1)})\right) \leq$$

$$\gamma_n\left(t_{n,i+1}\right) - \frac{1}{n+1} D_1 K\left(t_{n,i+\alpha}, H(X_{(i+1)})\right), \tag{3.10}$$

where $\alpha \in (0,1)$ and

$$t_{n,i+\alpha} =: \frac{i+\alpha}{n+1}.$$

In general, for $s \in [0, n+1]$, one writes

$$t_{ns} =: \frac{s}{n+1}.$$

Since

$$\gamma_n(t) = \frac{C(t) + \nu D(t) + \kappa}{n+1},$$

where $C(t) \geq 0$ and $D(t) \geq 0$ are Lipschitz in $\operatorname{logit}(t)$ (Dümbgen and Wellner (2014)), we can write

$$|\gamma_n(t) - \gamma_n(s)| = \left| \frac{C(t) + \nu D(t) + \kappa - C(s) - \nu D(s) - \kappa}{n+1} \right| \leq$$

$$\frac{|C(t) - C(s)| + \nu |D(t) - D(s)|}{n+1} \leq G_n(\nu) |\operatorname{logit}(t) - \operatorname{logit}(s)|,$$

where

$$G_n(\nu) = \frac{L_C + \nu L_D}{n+1} = O\left(\frac{1}{n+1}\right)$$

with L_C, L_D being the corresponding Lipschitz constants. Thus $\gamma_n(t)$ is Lipschitz in $\operatorname{logit}(t)$ which allows to continue inequality (3.10) in the following way:

$$\gamma_n(t_{n,i+1}) - \frac{1}{n+1} D_1 K\left(t_{n,i+\alpha}, H(X_{(i+1)})\right) \leq$$

$$\gamma_n(t_{ni}) + G_n(\nu)\left(\operatorname{logit}(t_{n,i+1}) - \operatorname{logit}(t_{ni})\right) -$$

$$- \frac{1}{n+1}\left(\operatorname{logit}(t_{n,i+\alpha}) - \operatorname{logit}(H(X_{(i+1)}))\right),$$

where we also utilized the fact (Dümbgen and Wellner (2014), Lemma (K.1), page 16) that

$$\frac{\partial K(s,t)}{\partial s} = \operatorname{logit} s - \operatorname{logit} t$$

for $s, t \in (0,1)$. We will estimate each of three terms in the last inequality.

For the second term we write:

$$G_n(\nu)\left(\operatorname{logit}(t_{n,i+1}) - \operatorname{logit}(t_{ni})\right) \leq \frac{G_n(\nu)}{n+1} \frac{1}{\xi_{ni}(1-\xi_{ni})},$$

where $\xi_{ni} \in [t_{ni}, t_{n,i+1}]$, since

$$(\operatorname{logit} x)' = \frac{1}{x(1-x)}.$$

Notice that the following inequalities hold true

$$\gamma_n(t_{ni}) \geq c\,\frac{\log\log n}{n+1},$$

$$t_{ni}(1-t_{ni}) \geq c'\delta_n \tag{3.11}$$

for some positive constants c and c' uniformly in $t_{ni} \in [\delta_n, 1-\delta_n]$. Therefore we can continue previous inequality for the second term in the following way:

$$\frac{G_n(\nu)}{n+1}\frac{1}{\xi_{ni}(1-\xi_{ni})} \leq \frac{G_n(\nu)}{c'\delta_n(n+1)} \leq O\left(\frac{n}{c'(n+1)^2}\right) = O\left(\frac{1}{n+1}\right)$$

since $\delta_n >> 1/n$.

For the third term one derives analogously:

$$\frac{1}{n+1}\left(\operatorname{logit}(H(X_{(i+1)})) - \operatorname{logit} t_{n,i+\alpha}\right) \leq$$
$$\frac{1}{n+1}(H(X_{(i+1)}) - t_{n,i+\alpha})\frac{1}{\zeta_{ni}(1-\zeta_{ni})} \leq$$
$$\frac{1}{n+1}\frac{1}{n+1}\frac{1}{\zeta_{ni}(1-\zeta_{ni})} \leq O\left(\frac{1}{n+1}\right),$$

where $\zeta_{ni} \in [t_{n,i+\alpha}, H(X_{(i+1)})]$, since $H(X_{(i+1)}) \geq t_{n,i+1} > t_{n,i+\alpha}$ by assumption that for every i and $X \in [X_{(i)}, X_{(i+1)}]$ it holds $H(X) \geq t_{ni}$. Thus we have derived

$$\begin{aligned} K\left(t_{ni}, H(X)\right) \leq \gamma_n\left(t_{ni}\right) + G_n(\nu)\left(\operatorname{logit}\left(t_{n,i+1}\right) - \operatorname{logit}\left(t_{ni}\right)\right) - \\ - \frac{1}{n+1}\left(\operatorname{logit}\left(t_{n,i+\alpha}\right) - \operatorname{logit}(H(X_{(i+1)}))\right) \leq \\ O\left(\frac{\log\log n}{n+1}\right) + O\left(\frac{1}{n+1}\right) + O\left(\frac{1}{n+1}\right), \end{aligned}$$

where $t_{ni}, t_{n,i+1} \in [\delta_n, 1-\delta_n]$ (and therefore $t_{n,i+\alpha} \in [\delta_n, 1-\delta_n]$).

Notice that

$$\frac{1}{n+1}\left(\operatorname{logit}(H(X_{(i+1)})) - \operatorname{logit} t_{n,i+\alpha}\right) \leq$$
$$\frac{1}{(n+1)^2}\frac{1}{\zeta_{ni}(1-\zeta_{ni})} = \frac{1}{(n+1)^2\zeta_{ni}(1-\zeta_{ni})\gamma_n(t_{ni})}\gamma_n(t_{ni})$$

and

$$\frac{1}{(n+1)^2\zeta_{ni}(1-\zeta_{ni})\gamma_n(t_{ni})} \leq$$
$$\frac{n+1}{(n+1)^2 t_{ni}(1-t_{ni})c\log\log n} \leq$$

$$\frac{1}{c' n\delta_n \log\log n} = \frac{1}{c' c_n (\log\log n)^2} \to 0$$

uniformly in $t_{ni} \in [\delta_n, 1-\delta_n]$ since $\delta_n = n^{-1} c_n \log\log n$ and $c_n \to \infty$ as $n \to \infty$.

Analogously,

$$\frac{G_n(\nu)}{n+1} \frac{1}{\xi_{ni}(1-\xi_{ni})} = \frac{G_n(\nu)}{(n+1)\xi_{ni}(1-\xi_{ni})\gamma_n(t_{ni})} \gamma_n(t_{ni})$$

and

$$\frac{G_n(\nu)}{(n+1)\xi_{ni}(1-\xi_{ni})\gamma_n(t_{ni})} \leq$$

$$\frac{n+1}{(n+1)^2 t_{ni}(1-t_{ni}) c \log\log n} \leq$$

$$\frac{1}{c'(n+1)\delta_n \log\log n} = \frac{1}{c' c_n (\log\log n)^2} \to 0$$

uniformly in $t_{ni} \in [\delta_n, 1-\delta_n]$ as $n \to \infty$.

Therefore, more precise upper bound estimate is

$$\begin{aligned} K\left(t_{ni}, H(X)\right) \leq & \gamma_n\left(t_{ni}\right) + G_n(\nu)\left(\text{logit}\left(t_{n,i+1}\right) - \text{logit}\left(t_{ni}\right)\right) - \\ & - \frac{1}{n+1}\left(\text{logit}\left(t_{n,i+\alpha}\right) - \text{logit}(H(X_{(i+1)}))\right) \leq \\ & \gamma_n\left(t_{ni}\right) + \gamma_n\left(t_{ni}\right) o(1) + \gamma_n\left(t_{ni}\right) o(1) = \\ & \gamma_n\left(t_{ni}\right)(1+o(1)). \end{aligned}$$

For $X \in [X_{(i)}, X_{(i+1)}]$ such that $H(X) \geq t_{n,i+1}$ it follows from the monotonicity of H and inequalities (3.9) that

$$K\left(t_{n,i+1}, H(X)\right) \leq K\left(t_{n,i+1}, H(X_{(i+1)})\right) \leq \gamma_n\left(t_{n,i+1}\right).$$

Thus on the set $\{\check{F}_n \in [\delta_n, 1-\delta_n]\}$ we obtained the following upper bound:

$$K(\check{F}_n, H(X)) \leq \gamma_n(\check{F}_n)(1+o(1))$$

(i.e. on $[X_{(i)}, X_{(i+1)}]$ and therefore on $[X_{(1)}, X_{(n)}]$).

Case 2. For $X \in [X_{(i)}, X_{(i+1)}]$ such that $H(X) \leq t_{n,i+1}$ it follows from the monotonicity of H and inequalities (3.9) that

$$K\left(t_{n,i+1}, H(X)\right) \leq K\left(t_{n,i+1}, H(X_{(i)})\right) =$$

$$K\left(t_{ni}, H(X_{(i)})\right) + K\left(t_{n,i+1}, H(X_{(i)})\right) - K\left(t_{ni}, H(X_{(i)})\right) \leq$$

$$\gamma_n\left(t_{ni}\right) - \frac{1}{n+1} D_1 K\left(t_{n,i+\alpha}, H(X_{(i)})\right)) \leq$$

$$\gamma_n(t_{n,i+1}) - G_n(\nu)(\text{logit}(t_{n,i+1}) - \text{logit}(t_{ni})) +$$
$$\frac{1}{n+1}\left(\text{logit}(t_{n,i+\alpha}) - \text{logit}(H(X_{(i)}))\right),$$

where $\alpha \in (0,1)$ as in the previous case. We will estimate each of three terms in the last inequality, analogously to **Case 1**.

For the second term one writes:

$$G_n(\nu)(\text{logit}(t_{n,i+1}) - \text{logit}(t_{ni})) \leq \frac{G_n(\nu)}{n+1}\frac{1}{\xi_{ni}(1-\xi_{ni})},$$

where $\xi_{ni} \in [t_{ni}, t_{n,i+1}]$. Utilizing inequalities (3.11), we continue the last inequality in the following way:

$$\frac{G_n(\nu)}{n+1}\frac{1}{\xi_{ni}(1-\xi_{ni})} \leq \frac{G_n(\nu)}{c'\delta_n(n+1)} \leq O\left(\frac{n}{c'(n+1)^2}\right) = O\left(\frac{1}{n+1}\right)$$

since $\delta_n >> 1/n$.

For the third term one derives analogously:

$$\frac{1}{n+1}\left(\text{logit}\, t_{n,i+\alpha} - \text{logit}(H(X_{(i)}))\right) \leq$$
$$\frac{1}{n+1}(t_{n,i+\alpha} - H(X_{(i+1)}))\frac{1}{\zeta_{ni}(1-\zeta_{ni})} \leq$$
$$\frac{1}{n+1}\frac{1}{n+1}\frac{1}{\zeta_{ni}(1-\zeta_{ni})} \leq O\left(\frac{1}{n+1}\right),$$

where $\zeta_{ni} \in [H(X_{(i)}), t_{n,i+\alpha}]$, since $H(X_{(i)}) \leq t_{ni} < t_{n,i+\alpha}$ by assumption that for every i and $X \in [X_{(i)}, X_{(i+1)}]$ it holds $H(X) \leq t_{ni}$.

Thus we have derived

$$K(t_{n,i+1}, H(X)) \leq \gamma_n(t_{n,i+1}) - G_n(\nu)(\text{logit}(t_{n,i+1}) - \text{logit}(t_{ni})) +$$
$$\frac{1}{n+1}\left(\text{logit}(t_{n,i+\alpha}) - \text{logit}(H(X_{(i)}))\right) \leq O\left(\frac{\log\log n}{n+1}\right) + O\left(\frac{1}{n+1}\right),$$

where $t_{ni}, t_{n,i+1} \in [\delta_n, 1-\delta_n]$. Notice that

$$\frac{1}{n+1}\left(\text{logit}\, t_{n,i+\alpha} - \text{logit}(H(X_{(i)}))\right) \leq$$
$$\frac{1}{(n+1)^2}\frac{1}{\zeta_{ni}(1-\zeta_{ni})} = \frac{1}{(n+1)^2\zeta_{ni}(1-\zeta_{ni})\gamma_n(t_{ni})}\gamma_n(t_{ni})$$

and

$$\frac{1}{(n+1)^2\zeta_{ni}(1-\zeta_{ni})\gamma_n(t_{ni})} \leq$$

$$\frac{n+1}{(n+1)^2 t_{ni}(1-t_{ni})c\log\log n} \leq$$
$$\frac{1}{c' n\delta_n \log\log n} = \frac{1}{c' c_n (\log\log n)^2} \to 0$$

uniformly in $t_{ni} \in [\delta_n, 1-\delta_n]$.

Analogously,

$$\frac{G_n(\nu)}{n+1}\frac{1}{\xi_{ni}(1-\xi_{ni})} = \frac{G_n(\nu)}{(n+1)\xi_{ni}(1-\xi_{ni})\gamma_n(t_{ni})}\gamma_n(t_{ni})$$

and

$$\frac{G_n(\nu)}{(n+1)\xi_{ni}(1-\xi_{ni})\gamma_n(t_{ni})} \leq \frac{1}{c' c_n (\log\log n)^2} \to 0$$

uniformly in $t_{ni} \in [\delta_n, 1-\delta_n]$ as $n \to \infty$.

Therefore, more precise upper bound estimate in this case is

$$K\left(t_{n,i+1}, H(X)\right) \leq \gamma_n\left(t_{n,i+1}\right) - G_n(\nu)\left(\text{logit}\left(t_{n,i+1}\right) - \text{logit}\left(t_{ni}\right)\right) +$$
$$\frac{1}{n+1}\left(\text{logit}\left(t_{n,i+\alpha}\right) - \text{logit}(H(X_{(i)}))\right) \leq$$
$$\gamma_n\left(t_{ni}\right) + \gamma_n\left(t_{ni}\right) o(1) = \gamma_n\left(t_{ni}\right)(1+o(1)).$$

For $X \in [X_{(i)}, X_{(i+1)}]$ such that $H(X) \leq t_{ni}$ it follows from the monotonicity of H and inequalities (3.9) that

$$K\left(t_{ni}, H(X)\right) \leq K\left(t_{ni}, H(X_{(i)})\right) \leq \gamma_n\left(t_{ni}\right).$$

Thus on the set $\{\delta_n \leq \check{F}_n \leq 1-\delta_n\}$ we have the same upper bound as in the previous case:

$$K(\check{F}_n, H(X)) \leq \gamma_n(\check{F}_n)(1+o(1))$$

(i.e. on $[X_{(i)}, X_{(i+1)}]$ and therefore on $[X_{(1)}, X_{(n)}]$).

Step 1. Assume that F and G are two distribution functions such that $H = F, G$ satisfies $[X_{(1)}, X_{(n)}] \subset J(H)$ and $K(\check{F}_n, H) \leq \gamma_n(\check{F}_n)(1+o(1))$ on the set $\{\delta_n \leq \check{F}_n \leq 1-\delta_n\}$ (in particular, on $[X_{(1)}, X_{(n)}]$). Then uniformly on $\{\delta_n \leq \check{F}_n \leq 1-\delta_n\} \cap \{\delta_n \leq H \leq 1-\delta_n\}$

$$|\,\text{logit}(H) - \text{logit}(\check{F}_n)| \to 0$$

as $n \to \infty$.

It follows from inequalities established by Dümbgen and Wellner (2014) (Lemma (K.5), page 18) that $K(\check{F}_n, H) \leq \gamma_n(\check{F}_n)(1+o(1))$ implies

$$|H - \check{F}_n| \leq \sqrt{2\check{F}_n(1-\check{F}_n)\gamma_n(\check{F}_n)(1+o(1))} + \gamma_n(\check{F}_n)(1+o(1))$$

and

$$|H - \check{F}_n| \leq \sqrt{2H(1-H)\gamma_n(\check{F}_n)(1+o(1))} + \gamma_n(\check{F}_n)(1+o(1)).$$

Notice also that

$$|\operatorname{logit}(H) - \operatorname{logit}(\check{F}_n)| = \left|\log\frac{H}{1-H} - \log\frac{\check{F}_n}{1-\check{F}_n}\right| = \left|\log\left(\frac{H}{\check{F}_n}\frac{1-\check{F}_n}{1-H}\right)\right|.$$

The inequality $|H - \check{F}_n| \leq \sqrt{2\check{F}_n(1-\check{F}_n)\gamma_n(\check{F}_n)(1+o(1))} + \gamma_n(\check{F}_n)(1+o(1))$ implies

$$\left|\frac{H}{\check{F}_n} - 1\right|, \left|\frac{1-H}{1-\check{F}_n} - 1\right| \to 0$$

as $n \to \infty$. Indeed, consider the ratio

$$\frac{|H - \check{F}_n|}{\check{F}_n(1-\check{F}_n)}.$$

On the set $\{\delta_n \leq \check{F}_n \leq 1 - \delta_n\}$

$$\frac{|H - \check{F}_n|}{\check{F}_n(1-\check{F}_n)} \leq \sqrt{\frac{2\gamma_n(\check{F}_n)(1+o(1))}{\check{F}_n(1-\check{F}_n)}} + \frac{\gamma_n(\check{F}_n)(1+o(1))}{\check{F}_n(1-\check{F}_n)} \leq$$

$$\sqrt{\frac{2\gamma_n(\delta_n(1-\delta_n))(1+o(1))}{\delta_n(1-\delta_n)}} + \frac{\gamma_n(\delta_n(1-\delta_n))(1+o(1))}{\delta_n(1-\delta_n)} =$$

$$O\left(\sqrt{\frac{\log\log n/n}{c_n \log\log n/n}} + \frac{\log\log n/n}{c_n \log\log n/n}\right) = O\left(\sqrt{\frac{1}{c_n}} + \frac{1}{c_n}\right) \to 0$$

as $n \to \infty$ since $0 < \delta_n \to 0$ and

$$\gamma_n(\delta_n) = O\left(\frac{\log\log(1/\delta_n) + \nu\log\log\log(1/\delta_n) + \kappa}{n}\right) = O\left(\frac{\log\log(1/\delta_n)}{n}\right).$$

In particular,

$$\left|\frac{H}{\check{F}_n} - 1\right| \leq o(1), \left|\frac{1-H}{1-\check{F}_n} - 1\right| \leq o(1).$$

This implies that uniformly on $\{\delta_n \leq \check{F}_n \leq 1 - \delta_n\}$

$$\left|\operatorname{logit}(H) - \operatorname{logit}(\check{F}_n)\right| = \left|\log\left(\frac{H}{\check{F}_n}\frac{1-\check{F}_n}{1-H}\right)\right| \leq$$

$$\left|\log\frac{H}{\check{F}_n} + o(1)\right| + \left|\log\frac{1-\check{F}_n}{1-H} + o(1)\right| \leq \log 1 + o(1) + \log 1 + o(1) = o(1).$$

Consider now the set $\{\delta_n \le H \le 1-\delta_n\}$. Inequality

$$|H - \check{F}_n| \le \sqrt{2H(1-H)\gamma_n(\check{F}_n)(1+o(1))} + \gamma_n(\check{F}_n)(1+o(1))$$

implies

$$\left|1 - \frac{\check{F}_n}{H}\right|, \left|1 - \frac{1-\check{F}_n}{1-H}\right| \to 0$$

as $n \to \infty$. Indeed, consider the ratio

$$\frac{|H - \check{F}_n|}{H(1-H)}.$$

Uniformly on $\{\delta_n \le H \le 1-\delta_n\} \cap \{0 < \check{F}_n < 1\}$

$$\gamma_n(\check{F}_n) \le O\left(\frac{\log\log n}{n}\right)$$

whence

$$\frac{|H - \check{F}_n|}{H(1-H)} \le O\left(\sqrt{\frac{\log\log n}{n\delta_n(1-\delta_n)}} + \frac{\log\log n}{n\delta_n(1-\delta_n)}\right) =$$

$$O\left(\sqrt{\frac{1}{c_n}} + \frac{1}{c_n}\right) \to 0$$

as $n \to \infty$. On the set $\{\delta_n \le H \le 1-\delta_n\} \cap \{\check{F}_n = 0\}$

$$\frac{|H - \check{F}_n|}{H(1-H)} = \frac{1}{1-H} \le \frac{1}{1-H(X_{(1)})} = \frac{H(X_{(1)})}{H(X_{(1)})(1-H(X_{(1)}))} \le$$

$$\frac{|H(X_{(1)}) - \check{F}_n(X_{(1)})| + O(n^{-1})}{H(X_{(1)})(1-H(X_{(1)}))} \le O\left(\sqrt{\frac{1}{c_n}} + \frac{1}{c_n} + \frac{1}{c_n \log\log n}\right) \to 0$$

as $n \to \infty$.

Analogously, on the set $\{\delta_n \le H \le 1-\delta_n\} \cap \{\check{F}_n = 1\}$

$$\frac{|H - \check{F}_n|}{H(1-H)} = \frac{1}{H} \le \frac{1}{H(X_{(n)})} = \frac{1-H(X_{(n)})}{H(X_{(n)})(1-H(X_{(n)}))} \le$$

$$\frac{|H(X_{(n)}) - \check{F}_n(X_{(n)})| + O(n^{-1})}{H(X_{(n)})(1-H(X_{(n)}))} \le O\left(\sqrt{\frac{1}{c_n}} + \frac{1}{c_n} + \frac{1}{c_n \log\log n}\right) \to 0,$$

as $n \to \infty$. In particular, we have

$$\left|1 - \frac{\check{F}_n}{H}\right| \le o(1), \left|1 - \frac{1-\check{F}_n}{1-H}\right| \le o(1).$$

This implies that uniformly on $\{\delta_n \le \check{F}_n \le 1 - \delta_n\} \cap \{\delta_n \le H \le 1 - \delta_n\}$

$$\left|\operatorname{logit}(H) - \operatorname{logit}(\check{F}_n)\right| = \left|\log\left(\frac{H}{\check{F}_n}\frac{1-\check{F}_n}{1-H}\right)\right| \le$$

$$\left|\log\frac{H}{\check{F}_n} + o(1)\right| + \left|\log\frac{1-\check{F}_n}{1-H} + o(1)\right| \le o(1)$$

as $n \to \infty$.

Step 2. There exists a constant $G(\kappa)$ such that for arbitrary $s, t \in (0,1)$ and $n \ge 1$,

$$\left|\frac{\gamma_n(t)}{\gamma_n(s)} - 1\right| \le G(\kappa)|\operatorname{logit}(t) - \operatorname{logit}(s)|.$$

Recall that

$$\gamma_n(t) = \frac{C(t) + \nu D(t) + \kappa}{n+1}$$

is Lipschitz in $\operatorname{logit}(t)$ (see Step 0). Then

$$\left|\frac{\gamma_n(t)}{\gamma_n(s)} - 1\right| = \left|\frac{(n+1)(\gamma_n(t) - \gamma_n(s))}{C(s) + \nu D(s) + \kappa}\right| \le$$

$$\frac{(n+1)G_n(\nu)|\operatorname{logit}(t) - \operatorname{logit}(s)|}{\kappa} = G(\kappa)|\operatorname{logit}(t) - \operatorname{logit}(s)|,$$

where $G_n(\nu) = (L_C + \nu L_D)/(n+1)$ with L_C, L_D being Lipschitz constants corresponding to functions $C(t), D(t)$ and

$$G(\kappa) = \frac{L_C + \nu L_D}{\kappa}.$$

Therefore

$$\left|\frac{\gamma_n(t)}{\gamma_n(s)} - 1\right| \le G(\kappa)|\operatorname{logit}(t) - \operatorname{logit}(s)|.$$

Step 3. Having

$$K(\check{F}_n, H) \le \gamma_n(\check{F}_n)(1 + o(1))$$

on $\{\delta_n \le \check{F}_n \le 1 - \delta_n\} \cap \{\delta_n \le H \le 1 - \delta_n\}$ (i.e. on $[X_{(1)}, X_{(n)}]$), where H is a distribution function such that $J(H) \subset [X_{(1)}, X_{(n)}]$, one can consecutively apply results of Steps 1 and 2. As it was shown in Step 1 (taking $H := F$),

$$|F - \check{F}_n| \le \sqrt{2F(1-F)\gamma_n(\check{F}_n)(1 + o(1))} + \gamma_n(\check{F}_n)(1 + o(1)). \qquad (3.12)$$

Step 1 implies

$$|\operatorname{logit}(F) - \operatorname{logit}(\check{F}_n)| \to 0 \qquad (3.13)$$

uniformly on $\{\delta_n/2 \le \check{F}_n \le 1-\delta_n/2\}$ as $n \to \infty$. Notice also that

$$\{\delta_n \le F \le 1-\delta_n\} \subset \{\delta_n/2 \le \check{F}_n \le 1-\delta_n/2\}$$

eventually as $n \to \infty$. Otherwise there exists x_n such that $\delta_n < F(x_n) < 1-\delta_n$ but $\check{F}_n(x_n) < \delta_n/2$ or $\check{F}_n(x_n) > 1-\delta_n/2$. Then

$$\{\delta_n \le F \le 1-\delta_n\} = \{|\operatorname{logit} F| \le \frac{1-\delta_n}{\delta_n} = \log\frac{1}{\delta_n} + o(1)\}$$

and

$$\{\delta_n/2 \le \check{F}_n \le 1-\delta_n/2\} = \{|\operatorname{logit} \check{F}_n| \le \frac{1-\delta_n/2}{\delta_n/2} = \log 2 + \log\frac{1}{\delta_n} + o(1).\}$$

Taking into account convergence (3.13) and Step 2, we obtain

$$\frac{\gamma_n(F)}{\gamma_n(\check{F}_n)} \to 1$$

as $n \to \infty$. Therefore one can write

$$|F-\check{F}_n| \le \sqrt{(2+o(1))F(1-F)\gamma_n(F)} + (1+o(1))\gamma_n(F) =$$

$$\sqrt{(2+o(1))F(1-F)\gamma_n(F)} + (1+o(1))\sqrt{F(1-F)\gamma_n(F)}\sqrt{\frac{\gamma_n(F)}{F(1-F)}}.$$

Taking into account the properties of the function γ_n established in Step 1, we have that

$$\frac{\gamma_n(F)}{F(1-F)} \to 0$$

as $n \to \infty$. Thus inequality (3.12) is equivalent to

$$|F-\check{F}_n| \le \sqrt{(2+o(1))F(1-F)\gamma_n(F)}. \tag{3.14}$$

Analogously to (3.12) (taking $H := G$),

$$|G-\check{F}_n| \le \sqrt{2G(1-G)\gamma_n(\check{F}_n)(1+o(1))} + \gamma_n(\check{F}_n)(1+o(1)).$$

By means of Step 1,

$$|\operatorname{logit}(G) - \operatorname{logit}(\check{F}_n)| \to 0$$

uniformly on $\{\delta_n \le G \le 1-\delta_n\} \subset \{\delta_n/2 \le \check{F}_n \le 1-\delta_n/2\}$ as $n \to \infty$, analogously to (3.13). But $\{\delta_n \le F \le 1-\delta_n\} \subset \{\delta_n/2 \le \check{F}_n \le 1-\delta_n/2\}$. Thus

$$|\operatorname{logit}(G) - \operatorname{logit}(F)| \to 0$$

uniformly on $\{\delta_n \le F \le 1-\delta_n\}$ as $n \to \infty$. Then (3.14) is valid with G in place of F, too. Combining both inequalities yields

$$|F - G| \le |F - \check{F}_n| + |G - \check{F}_n| \le$$

$$\le \sqrt{(2+o(1))F(1-F)\gamma_n(F)} + \sqrt{(2+o(1))F(1-F)\gamma_n(F)}.$$

Therefore

$$|F - G| \le \sqrt{(8+o(1))F(1-F)\gamma_n(F)}.$$

Step 4. As shown at the beginning of Step 1,

$$|H - \check{F}_n| \le \sqrt{2H(1-H)\gamma_n(\check{F}_n)(1+o(1))} + \gamma_n(\check{F}_n)(1+o(1)), \tag{3.15}$$

and, as $n \to \infty$,

$$\left| \frac{H}{\check{F}_n} - 1 \right| \to 0.$$

It follows from inequalities established by Dümbgen and Wellner (2014) (Lemma (K.5), page 18) that

$$|H - \check{F}_n| \le \sqrt{2\check{F}_n(1-\check{F}_n)\gamma_n(\check{F}_n)(1+o(1))} + \gamma_n(\check{F}_n)(1+o(1)).$$

Consider the set $\{0 < \check{F}_n < 1\}$ and recall that

$$\gamma_n(\check{F}_n) = O\left(\frac{\log\log n}{n}\right)$$

(Dümbgen and Wellner (2014)); then

$$\check{F}_n(1-\check{F}_n)\gamma_n(\check{F}_n) \le$$

$$\frac{\check{F}_n(1-\check{F}_n)C(\check{F}_n) + \nu\check{F}_n(1-\check{F}_n)D(\check{F}_n) + \kappa\check{F}_n(1-\check{F}_n)}{n+1} = O\left(\frac{1}{n+1}\right).$$

Then

$$\sqrt{\check{F}_n(1-\check{F}_n)\gamma_n(\check{F}_n)(1+o(1))} = O\left(\frac{1}{\sqrt{n}}\right).$$

Summing up everything together, one can write

$$|H - \check{F}_n| \le O\left(\frac{1}{\sqrt{n}}\right) + O\left(\frac{\log\log n}{n}\right) = O\left(\frac{1}{\sqrt{n}}\right),$$

that is

$$|F - \check{F}_n| \le O\left(\frac{1}{\sqrt{n}}\right) \text{ and } |G - \check{F}_n| \le O\left(\frac{1}{\sqrt{n}}\right).$$

Therefore

$$|F-G| \leq O\left(\frac{1}{\sqrt{n}}\right).$$

Now consider cases $\{\check{F}_n = 0\}$ and $\{\check{F}_n = 1\}$. For $x < X_{(1)}$, one can write

$$|H-\check{F}_n| \leq H(x) \leq H(X_{(1)}) \leq O\left(\frac{1}{\sqrt{n}}\right) + \frac{1}{n+1} = O\left(\frac{1}{\sqrt{n}}\right).$$

For $x > X_{(n)}$,

$$|H-\check{F}_n| \leq 1-H(x) \leq 1-H(X_{(n)}) \leq O\left(\frac{1}{\sqrt{n}}\right) + \frac{1}{n+1} = O\left(\frac{1}{\sqrt{n}}\right).$$

Hence $|H-\check{F}_n| \leq O\left(n^{-1/2}\right)$ in these cases as well. So finally one obtains

$$||F-G||_\infty = O\left(\frac{1}{n^{1/2}}\right).$$

This accomplishes the proof of the Lemma. □

Proof of Lemma 1.32. Notice that the graph of $\log F_1$ is a straight line connecting the points $(x_1, \log y_1)$ and $(x_2, \log y_2)$, while $\log F$ is a concave function. Thus if $F(x_1) \leq y_1$ and $F(x_2) \geq y_2$, then $\log F \leq \log F_1$ on $(-\infty, x_1]$. Similarly, if $F(x_1) \geq y_1$ and $F(x_2) \geq y_2$, then $\log F \geq \log F_1$ on $[x_1, x_2]$. Finally, if $F(x_1) \geq y_1$ and $F(x_2) \leq y_2$, then $\log F \leq \log F_1$ on $[x_2, \infty)$.

Analogous considerations apply to the concave function $\log(1-F)$ and the function $\log(1-F_2)$ which describes a straight line connecting the points $(x_1, \log(1-y_1))$ and $(x_2, \log(1-y_2))$. □

Proof of Proposition 1.33. By definition of bi-log-concave function $\mu : \mathbb{R} \to [0,1]$ there exist $x_1 < x_2$ such that $\mu(x_1)$, $\mu(x_2) \in (0,1)$. Let $y_i = \mu(x_i)$, where $i = 1$, 2. Assume that $y_1 < y_2 \in (0,1)$. Bi-log-concavity of μ yields concavity of $\log \mu$ and $\log(1-\mu)$ simultaneously. This allows us to utilize Lemma 1.32. Namely, interpolating between the points (x_1, y_1) and (x_2, y_2) by functions F_1 and F_2 from that lemma, and comparing them to μ, we conclude that μ is non-decreasing on $\mathbb{R}$. Besides that, Theorem 1.7 implies that it is continuous on $\mathbb{R}$. Taken together, these facts make μ a cumulative distribution function.

Assume that $y_1 > y_2 \in (0,1)$. As in the previous case, interpolating between the points (x_1, y_1) and (x_2, y_2) by functions F_1 and F_2 and comparing them to μ, we obtain that μ is non-increasing on $\mathbb{R}$. Its continuity on $\mathbb{R}$ implies that μ is a survival function. The case $y_1 = y_2 \in (0,1)$, treated in analogous way, yields $\mu \equiv y_1$, i.e. μ is a constant function. □

Proof of Lemma 1.34. By assumption there exist real numbers $x_L < x_U$ with $L_n(x_L) > 0$ and $U_n(x_U) < 1$. Then

$$\Delta := \frac{\log U_n(x_U) - \log L_n(x_L)}{x_U - x_L} < \infty.$$

The assertion is trivial if $L_n^o \equiv 1$ and $U_n^o \equiv 0$, meaning that no $G \in \mathcal{F}_{\mathrm{blcd}}$ fits in between L_n and U_n. Otherwise let $G \in \mathcal{F}_{\mathrm{blcd}}$ such that $L_n \leq G \leq U_n$. With $g := G'$, it follows from concavity of $\log G$ that for $x \geq x_U$,

$$g(x) \leq \frac{g}{G}(x) \leq \frac{g}{G}(x_U) \leq \frac{\log G(x_U) - \log G(x_L)}{x_U - x_L} \leq \Delta.$$

On the other hand, convexity of $-\log(1-G)$ implies that for $x \leq x_U$,

$$g(x) \leq \frac{g}{1-G}(x) \leq \frac{g}{1-G}(x_U) = \frac{g}{G}(x_U)\frac{G}{1-G}(x_U) \leq \frac{\Delta U_n(x_U)}{1-U_n(x_U)}.$$

Consequently, any $G \in \mathcal{F}_{\mathrm{blcd}}$ with $L_n \leq G \leq U_n$ is Lipschitz-continuous with Lipschitz-constant $\max\big\{\Delta, \Delta U_n(x_U)/(1-U_n(x_U))\big\}$. Hence the pointwise infimum L_n^o and supremum U_n^o have the same property. □

Proof of Lemma 1.35. 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 $G \in \mathcal{F}_{\mathrm{blcd}}$ such that $L_n \leq G \leq U_n$.

For part (i) it suffices to show that for any $x \in J(G)$ the density $g = G'$ satisfies the inequality $g(x) \leq \max\{\gamma_1, \gamma_2\}$. 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 $x \geq b$ it follows from concavity of $\log G$ and $G(a) \geq r$, $G(b) \leq s$ that

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

Similarly convexity of $-\log(1-G)$ and the inequalities $G(a) \geq r$, $G(b) \leq s$ imply that for $x \leq a$,

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

For $a < x < b$ we get the two inequalities

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

and

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

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

$$g(x) \leq \frac{G(x)\log(G(x)/r) + (1 - G(x))\log\bigl((1 - G(x))/(1 - s)\bigr)}{b - a}.$$

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

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

As to part (ii), it suffices to show that $G(x) \leq G(a)\exp(\gamma_1(x-a))$ for $x \leq a$ and $G(x) \geq 1-(1-G(b))\exp(-\gamma_2(x-b))$ for $x \geq b$. We know from Theorem 1.7 (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 $(g/(1 - G))(b) \geq \gamma_2$. □

Proof of Theorem 1.57. Suppose that $F \notin \mathcal{F}_{\mathrm{blcd}}$. This means, either $\log F$ or $\log(1 - F)$ or both are not concave. When $\log F$ is not concave then 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 $F \in \mathcal{F}_{\mathrm{blcd}}$. 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

$$\sup_{G \in \mathcal{F}_{\mathrm{blcd}} : L_n \leq G \leq U_n} \|G - F\|_\infty \leq \|L_n - F\|_\infty + \|U_n - F\|_\infty \to_p 0, \quad (3.16)$$

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 $G \in \mathcal{F}_{\mathrm{blcd}}$. 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 $G \in \mathfrak{F}_{\mathrm{blcd}}$ with $L_n \le G \le U_n$, for any $x \in K$ it follows from monotonicity of h_F and h_G that

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

Analogously,

$$\begin{aligned}
\sup_{x\in K}\big(h_F(x) - h_G(x)\big) &\le \max_{i=1,\ldots,m-1}\big(h_F(a_i) - h_F(a_{i+2})\big) + o_p(1)\\
&\le 2\varepsilon + o_p(1).
\end{aligned}$$

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

$$\sup_{G\in\mathfrak{F}_{\mathrm{blcd}}\,:\,L_n\le G\le U_N} \big\|\log(G)' - \log(F)'\big\|_{K,\infty} = o_p(1). \tag{3.17}$$

Analogously one can show that

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

Moreover, since $G' = \log(G)'\,G$, it follows from (3.16) and (3.17) that

$$\sup_{G\in\mathfrak{F}_{\mathrm{blcd}}\,:\,L_n\le G\le U_N} \|G' - F'\|_{K,\infty} = o_p(1).$$

Finally, let $x_1 < \sup(J(F))$ and $b_1 < f(x_1)/F(x_1)$. As in the proof of Lemma 1.35 (ii) one may argue that for any fixed $x_1' > x_1$, $x_1' \in J(F)$,

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

for all $x \le x' \le x_1$. 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 with asymptotic 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 $[x_2, \infty)$, $x_2 > \inf(J(F))$. □

Proof of Corollary 1.58. 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 $z \in \mathbb{R}$,

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

so by Fubini's theorem,

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

provided that

$$\int |\phi'(x)| \bigl|1_{[x \geq 0]} - G(x)\bigr|\, dx \;<\; \infty. \tag{3.18}$$

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 1.57 (ii) that asymptotically with probability one,

$$U_n^o(x) \;\leq\; U_n(x') \exp(b_1'(x - x')) \quad \text{for } x \leq x' \leq x_1 \tag{3.19}$$

and

$$1 - L_n^o(x) \;\leq\; (1 - L_n(x')) \exp(-b_2'(x - x')) \quad \text{for } x \geq x' \geq x_2. \tag{3.20}$$

If we choose $b_1' > b_1$ and $b_2' > b_2$, the inequalities (3.19) and (3.20) imply (3.18) for arbitrary distribution functions G with $L_n^o \leq G \leq U_n^o$. More precisely, for any fixed $c \geq 0$ and $\delta := \min\{b_1' - b_1, b_2' - b_2\} > 0$,

$$\begin{aligned}
\int_{-\infty}^{x_1 - c} |\phi'(x)| U_n^o(x)\, dx \;&\leq\; U_n(x_1) \int_{-\infty}^{x_1 - c} \exp(a - b_1 x + b_1'(x - x_1))\, dx \\
&\leq\; U_n(x_1) \exp(a - b_1 x_1 - \delta c) \int_{-\infty}^{0} \exp(\delta y)\, dy \\
&=\; \frac{U_n(x_1) \exp(a - b_1 x_1 - \delta c)}{\delta}
\end{aligned}$$

and

$$\int_{x_2+c}^{\infty} |\phi'(x)|(1 - L_n^o(x))\,dx \;\le\; \frac{(1 - L_n(x_1))\exp(a + b_2x_2 - \delta c)}{\delta}.$$

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

$$\sup_{G\,:\,L_n^o\le G\le U_n^o}\Big|\int \phi\,dG - \int \phi\,dF\Big| \;=\; \sup_{G\,:\,L_n^o\le G\le U_n^o}\Big|\int_{-\infty}^{\infty} \phi'(x)(F-G)(x)\,dx\Big| \tag{3.21}$$

is not larger than

$$\begin{aligned}
&\sup_{G\,:\,L_n^o\le G\le U_n^o}\|G-F\|_\infty \int_{x_1-c}^{x_2+c}|\phi'(x)|\,dx \\
&\qquad + \int_{-\infty}^{x_1-c}|\phi'(x)|(U_n^o+F)(x)\,dx + \int_{x_2+c}^{\infty}|\phi'(x)|(2-L_n^o-F)(x)\,dx \\
&\quad\le \frac{2F(x_1)e^{a-b_1x_1-\delta c}}{\delta} + \frac{2(1-F(x_2))e^{a+b_2x_2-\delta c}}{\delta} + o_p(1).
\end{aligned}$$

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

Proof of Theorem 1.59. 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 $1-\varepsilon$,

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

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

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

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

and

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

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

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

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

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

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

$$\frac{|h-F|}{(F(1-F))^\gamma} \;\le\; \frac{|h-\widehat{F}_n|}{(\widehat{F}_n(1-\widehat{F}_n))^\gamma}\Big(\frac{\widehat{F}_n(1-\widehat{F}_n)}{F(1-F)}\Big)^\gamma + \frac{|\widehat{F}_n - F|}{(F(1-F))^\gamma} \;\le\; \nu_\varepsilon n^{-1/2} \tag{3.22}$$

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 (3.22) 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 (3.19) and (3.20) are satisfied, too. Writing $J_n = [x_{n1}, x_{n2}]$, we can deduce from (3.21) and (3.22) that

$$\begin{aligned}\sup_{G\,:\,L_n^o \le G \le U_n^o}\Big|\int \phi\, d(G-F)\Big| &\le \nu_\varepsilon n^{-1/2}\int_{x_{n1}}^{x_{n2}} |\phi'(x)|F(x)^\gamma(1-F(x))^\gamma\,dx \\ &\quad + \int_{-\infty}^{x_{n1}} |\phi'(x)|(F+U_n^o)(x)\,dx \\ &\quad + \int_{x_{n2}}^{\infty} |\phi'(x)|(2-F-L_n^o)(x)\,dx.\end{aligned}$$

Notice that

$$\begin{aligned} F(x) &\le F(x_1)\exp(b_1'(x-x_1)) \quad \text{for } x \le x_1, \\ 1-F(x) &\le (1-F(x_2))\exp(-b_2'(x-x_2)) \quad \text{for } x \ge x_2. \end{aligned}$$

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

$$x_{n1} \;\ge\; O(1) - \frac{\log n}{b_1'(2-2\gamma)} \quad \text{and} \quad x_{n2} \;\le\; O(1) + \frac{\log n}{b_2'(2-2\gamma)}. \tag{3.23}$$

Notice also that by (3.19), (3.20) and (3.22),,

$$\begin{aligned} (F+U_n^o)(x) &\le (F+U_n^o)(x_{n1})\exp(b_1'(x-x_{n1})) \\ &\le \omega_\varepsilon n^{-1/(2-2\gamma)}\exp(b_1'(x-x_{n1})) \quad \text{for } x \le x_{n1}, \\ (2-F-L_n^o)(x) &\le \omega_\varepsilon n^{-1/(2-2\gamma)}\exp(-b_2'(x-x_{n2})) \quad \text{for } x \ge x_{n2}, \end{aligned}$$

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

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

with

$$\begin{aligned}
I_{n0} &:= \nu_\varepsilon n^{-1/2} \int_{x_1}^{x_2} |\phi'(x)|\, dx \;=\; O(n^{-1/2}),\\
I_{n1} &:= \nu_\varepsilon n^{-1/2} \int_{x_{n1}}^{x_1} |\phi'(x)| F(x)^\gamma \, dx \;=\; O\Big(n^{-1/2} \int_{x_{n1}}^{x_1} |\phi'(x)| e^{\gamma b_1' x}\, dx\Big),\\
I_{n1}' &:= \int_{-\infty}^{x_{n1}} |\phi'(x)|(F + U_n^o)(x)\, dx \;=\; O\Big(n^{-1/(2-2\gamma)} \int_{-\infty}^{x_{n1}} |\phi'(x)| e^{b_1'(x - x_{n1})}\, dx\Big),\\
I_{n2} &:= \nu_\varepsilon n^{-1/2} \int_{x_2}^{x_{n2}} |\phi'(x)|(1 - F(x))^\gamma \, dx \;=\; O\Big(n^{-1/2} \int_{x_2}^{x_{n2}} |\phi'(x)| e^{-\gamma b_2' x}\, dx\Big),\\
I_{n2}' &:= \int_{x_{n2}}^{\infty} |\phi'(x)|(2 - F - L_n^o)(x)\, dx \;=\; O\Big(\frac{1}{n^{1/(2-2\gamma)}} \int_{x_{n2}}^{\infty} |\phi'(x)| e^{-b_2'(x - x_{n2})} dx\Big).
\end{aligned}$$

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

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

where $b' := \min\{b_1', b_2'\} > 0$. Moreover, both I_{n1}' and I_{n2}' are of order

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

and

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

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 $b_1 \ne \gamma b_1'$ and $b_2 \ne \gamma b_2'$. Then

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

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,

$$\begin{aligned} I'_{n1} &= O\Big(n^{-1/(2-2\gamma)} \int_{-\infty}^{x_{n1}} \exp(-b_1 x + b'_1(x - x_{n1}))\, dx\Big) \\ &= O\Big(n^{-1/(2-2\gamma)} \exp(-b_1 x_{n1}) \int_0^{\infty} \exp(-(b'_1 - b_1)s)\, ds\Big) \\ &= O\big(n^{-1/(2-2\gamma)} \exp(-b_1 x_{n1})\big) \;=\; O\big(n^{-(1-b_1/b'_1)/(2-2\gamma)}\big) \;=\; O(n^{-\beta_1}) \end{aligned}$$

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 $(1-\widetilde{\gamma})/(2(1-\gamma))$. $\square$

Proof of Corollary 1.60. Notice that for any $\alpha' \in (0,1)$ one has $\|\widehat{F}_n - F\|_\infty \leq \kappa^{\mathrm{KS}}_{n,\alpha'}$ with probability $1-\alpha'$. Thus by the triangle inequality for the sup-norm, the Kolmogorov-Smirnov confidence region of all c.d.f.'s $\widetilde{F}$ such that

$$\|\widetilde{F} - \widehat{F}_n\|_\infty \leq \kappa^{\mathrm{KS}}_{n,\alpha},$$

is contained in the deterministic set

$$\big\{\text{c.d.f.'s } \widetilde{F} : \|\widetilde{F} - F\|_\infty \leq c_n\big\}$$

with $c_n := \kappa^{\mathrm{KS}}_{n,\alpha'} + \kappa^{\mathrm{KS}}_{n,\alpha} = O(n^{-1/2})$ with probability $1-\alpha'$. $\square$

Proof of Remark 2.3. Recall that by Proposition 1.33, any $F \in \mathcal{F}_{\mathrm{blc}}$ is either a constant or a distribution function or a survival function. Hence any vector $\varphi \in \bar{\mathbb{F}}_n$ has to have non-increasing or non-decreasing components.

If $\varphi_1 = \varphi_n$, then $\varphi = (F(X_{ni}))_{i=1}^n$ with $F : \mathbb{R} \to [0,1]$ constant, i.e. $F \in \mathcal{F}_{\mathrm{blc}} \cup \{0\} \cup \{1\}$.

If $\varphi_1 < \varphi_n$, then there exists a sequence $(F_n)_n$ in $\mathcal{F}_{\mathrm{blcd}}$ such that $(F_n(X_{ni}))_{i=1}^n$ converges to φ. Because of $\varphi_1 < \varphi_n$, the sequence $(F_n)_n$ is tight, and we may assume w.l.o.g. that it converges weakly to a some F on $\mathbb{R}$. If F is non-degenerate, then, according to Lemma 1.23, $F \in \mathcal{F}_{\mathrm{blcd}}$ and weak convergence of $(F_n)_n$ to F implies uniform convergence on $\mathbb{R}$, whence $\varphi = (F(X_{ni}))_{i=1}^n \in \mathbb{F}_n$. If F is degenerate, then there exists a real number $X_o \in [X_{n1}, X_{nn}]$ such that

$$\varphi_i = \begin{cases} 0 \text{ if } X_{ni} < X_o, \\ 1 \text{ if } X_{ni} > X_o. \end{cases}$$

But now there exists an interval $[a,b] \subset \mathbb{R}$ such that

$$\{X_{n1}, ..., X_{nn}\} \cap [a,b] \subset \{X_o\} \subset [a,b]$$

and

$$\varphi_i = \frac{X_o - a}{b - a}$$

whenever $X_{ni} = X_o$. Thus $\varphi = (F(X_{ni}))_{i=1}^n$ with

$$F(x) := \begin{cases} 0, \ x \le a, \\ (x-a)/(b-a), \ a \le x \le b, \\ 1, \ x \ge b, \end{cases}$$

the c.d.f. of $U([a,b])$, and the latter belongs to $\mathcal{F}_{\text{blcd}}$.

If $\varphi_1 > \varphi_n$ we may argue analogously for survival functions instead of distribution functions. □

Proof of Theorem 2.2. We will use a modified version of Kolmogorov's maximal inequality:

Lemma 3.4. *Let A_1, A_2, ..., A_n be independent random variables with mean zero and finite variances. Then*

$$\frac{1}{2}\,\mathbb{E}\left(\max_{k=1,..,n}\left|\sum_{i=1}^k A_i\right|\right) \le \sqrt{\sum_{i=1}^n \mathbb{E}(A_i^2)}.$$

Proof. Kolmogorov's maximal inequality states that for any $\eta > 0$,

$$\mathbb{P}\left(\max_{k=1,..,n}\left|\sum_{i=1}^k A_i\right| \ge \eta\right) \le \sigma^2/\eta^2$$

with $\sigma^2 := \mathbb{E}((\sum_{i=1}^n A_i)^2) = \sum_{i=1}^n \mathbb{E}(A_i^2)$. Thus

$$\begin{aligned} \mathbb{E}\left(\max_{k=1,..,n}\left|\sum_{i=1}^k A_i\right|\right) &= \int_0^\infty \mathbb{P}\left(\max_{k=1,..,n}\left|\sum_{i=1}^k A_i\right| \ge \eta\right) d\eta \\ &\le \int_0^\infty \min(\sigma^2/\eta^2, 1) d\eta \\ &= \sigma + \sigma^2 \int_\sigma^\infty \eta^{-2} d\eta = 2\sigma. \end{aligned}$$

Our arguments are inspired by van de Geer (1993). For convenience we write $F_{ni} := F(X_{ni})$ for any function $F : \mathbb{R} \to \mathbb{R}$. Then the normalized negative log-likelihood may be written as

$$L_n(F) = -\frac{1}{n}\sum_{i=1}^n (Y_{ni}\log F_{ni} + (1 - Y_{ni})\log(1 - F_{ni}))$$

for $F \in \mathscr{F}$. By definition of $\widehat{\mu}_n$,

$$
\begin{aligned}
0 &\leq L_n(\widehat{\mu}_n) - L_n(\mu) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(Y_{ni}\log\left(\frac{\widehat{\mu}_{ni}}{\mu_{ni}}\right) + (1-Y_{ni})\log\left(\frac{1-\widehat{\mu}_{ni}}{1-\mu_{ni}}\right)\right) \\
&= \frac{2}{n}\sum_{i=1}^{n}\left(Y_{ni}\log\sqrt{\frac{\widehat{\mu}_{ni}}{\mu_{ni}}} + (1-Y_{ni})\log\sqrt{\frac{1-\widehat{\mu}_{ni}}{1-\mu_{ni}}}\right) \\
&\leq \frac{2}{n}\sum_{i=1}^{n}\left(Y_{ni}\left(\sqrt{\frac{\widehat{\mu}_{ni}}{\mu_{ni}}}-1\right) + (1-Y_{ni})\left(\sqrt{\frac{1-\widehat{\mu}_{ni}}{1-\mu_{ni}}}-1\right)\right) \\
&= \frac{2}{n}\sum_{i=1}^{n}\left(\mu_{ni}\left(\sqrt{\frac{\widehat{\mu}_{ni}}{\mu_{ni}}}-1\right) + (1-\mu_{ni})\left(\sqrt{\frac{1-\widehat{\mu}_{ni}}{1-\mu_{ni}}}-1\right)\right) \\
&+ \frac{2}{n}\sum_{i=1}^{n}\frac{Y_{ni}-\mu_{ni}}{\sqrt{\mu_{ni}}}\sqrt{\widehat{\mu}_{ni}} - \frac{2}{n}\sum_{i=1}^{n}\frac{Y_{ni}-\mu_{ni}}{\sqrt{1-\mu_{ni}}}\sqrt{1-\widehat{\mu}_{ni}} \\
&= -\frac{1}{n}\sum_{i=1}^{n}(2 - 2\sqrt{\mu_{ni}\widehat{\mu}_{ni}} - 2\sqrt{(1-\mu_{ni})(1-\widehat{\mu}_{ni})}) \\
&+ Z_{n1}(\sqrt{\widehat{\mu}_n}) - Z_{n2}(\sqrt{1-\widehat{\mu}_n}),
\end{aligned}
$$

where

$$Z_{n1}(g) := \frac{2}{n}\sum_{i=1}^{n}\frac{Y_{ni}-\mu_{ni}}{\sqrt{\mu_{ni}}}g(X_{ni}),$$

$$Z_{n1}(g) := \frac{2}{n}\sum_{i=1}^{n}\frac{Y_{ni}-\mu_{ni}}{\sqrt{1-\mu_{ni}}}g(X_{ni})$$

for $g : \mathbb{R} \to \mathbb{R}$.

Notice that $\sqrt{\widehat{\mu}_n}$ and $\sqrt{1-\widehat{\mu}_n}$ belong to $\mathscr{F}_{mon}$. Moreover, for arbitrary numbers $a, \widehat{a} \in [0,1]$:

$$
\begin{aligned}
2 - 2\sqrt{a\widehat{a}} - 2\sqrt{(1-a)(1-\widehat{a})} &= (\sqrt{\widehat{a}} - \sqrt{a})^2 + (\sqrt{1-\widehat{a}} - \sqrt{1-a})^2 \\
&= \frac{(\widehat{a}-a)^2}{(\sqrt{\widehat{a}}+\sqrt{a})^2} + \frac{(\widehat{a}-a)^2}{(\sqrt{1-\widehat{a}}+\sqrt{1-a})^2} \\
&\geq \frac{(\widehat{a}-a)^2}{2(\widehat{a}+a)} + \frac{(\widehat{a}-a)^2}{2(2-\widehat{a}-a)} \\
&\geq (\widehat{a}-a)^2,
\end{aligned}
$$

because

$$2(\widehat{b}+b) - (\sqrt{\widehat{b}}+\sqrt{b})^2 = (\sqrt{\widehat{b}}-\sqrt{b})^2 \geq 0$$

for $b, \widehat{b} \in \{a, \widehat{a}, 1-a, 1-\widehat{a}\}$ and

$$\frac{1}{2x} + \frac{1}{2(2-x)} \geq 1$$

for $x = a + \widehat{a} \in [0, 2]$.

Consequently,

$$\frac{1}{n}\sum_{i=1}^{n}(\widehat{\mu}_{ni} - \mu_{ni})^2 \leq \sup_{g \in \mathscr{F}_{mon}} |Z_{n1}(g)| + \sup_{g \in \mathscr{F}_{mon}} |Z_{n2}(g)|.$$

Hence it suffices to show that

$$\mathbb{E}\left(\sup_{g \in \mathscr{F}_{mon}} |Z_{nj}(g)|\right) \leq 4n^{-1/2} \tag{3.24}$$

for $j = 1, 2$.

If $g : \mathbb{R} \to [0, 1]$ is non-increasing, then

$$\begin{aligned}
|Z_{nj}(g)| &= \left|\sum_{i=1}^{n} A_{ni} g(X_{ni})\right| \\
&= \left|\sum_{i=1}^{n} A_{ni} \sum_{k=i}^{n}(g(X_{nk}) - g(X_{n,k+1}))\right| \\
&= \left|\sum_{k=1}^{n}(g(X_{nk}) - g(X_{n,k+1})) \sum_{i=1}^{k} A_{ni}\right| \\
&\leq \sum_{k=1}^{n}(g(X_{nk}) - g(X_{n,k+1})) \left|\sum_{i=1}^{k} A_{ni}\right| \\
&\leq \max_{k=1,\ldots,n} \left|\sum_{i=1}^{k} A_{ni}\right|
\end{aligned}$$

where

$$A_{ni} := \begin{cases} 2\frac{Y_{ni} - \mu_{ni}}{n\sqrt{\mu_{ni}}} & \text{if } j = 1; \\ 2\frac{Y_{ni} - \mu_{ni}}{n\sqrt{1 - \mu_{ni}}} & \text{if } j = 2; \end{cases}$$

with $X_{n,n+1} := +\infty$ and $g(+\infty) := 0$.

Analogously, if $g : \mathbb{R} \to [0, 1]$ is non-decreasing,

$$|Z_{nj}(g)| \leq \max_{k=1,\ldots,n} \left|\sum_{i=k}^{n} A_{ni}\right|.$$

Lemma 3.4 implies that

$$\frac{1}{2}\max\left(\mathbb{E}\left(\max_{k=1,..,n}\left|\sum_{i=1}^{k}A_{ni}\right|\right),\mathbb{E}\left(\max_{k=1,..,n}\left|\sum_{i=k}^{n}A_{ni}\right|\right)\right)\leq\sqrt{\sum_{i=1}^{n}\mathbb{E}(A_{ni}^2)}.$$

But

$$\sqrt{\sum_{i=1}^{n}\mathbb{E}(A_{ni}^2)}\leq\sqrt{\sum_{i=1}^{n}\frac{4\,\mathbb{E}((Y_{ni}-\mu_{ni})^2)}{n^2\mu_{ni}(1-\mu_{ni})}}=2/\sqrt{n},$$

which implies inequality (3.24). □

Proof of Theorem 2.5. Let $\mathscr{P}:=\{P_\mu,\mu\in\mathcal{F}_{\text{blc}}\}$ be a family of probability measures on some measurable space. We assume that $\mathscr{P}$ is dominated by a σ-finite measure ν. Consider i.i.d. observations $(X_i,Y_i)=:Z_i$, $i=1,...,n$ from a probability measure $P_o:=P_{\mu_o}$, where $\mu_o(x)$ is true bi-log-concave regression function. In some cases we will write (X,Y) instead of (X_i,Y_i) so that Y will take the value either 0 or 1. We also define

$$P_n:=\frac{1}{n}\sum_{i=1}^{n}\delta_{Z_i}$$

to be the empirical distribution based on observations Z_i and

$$Q_n:=\frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}$$

to be the empirical distribution based on X_i. Recall that

$$\mu(x)=\mathbb{P}(Y=1|X=x)=1-\mathbb{P}(Y=0|X=x)$$

and $X_i\sim Q$.

Let P_μ be some probability measure from $\mathscr{P}$. Recall that if we take as a dominating measure $\nu=Q\otimes\delta$, where δ is the counting measure on $\{0,1\}$, then the class of densities

$$f_\mu(x,y):=\frac{dP_\mu}{d\nu}$$

in this model is given by

$$\left\{\mu(x)1_{\{y=1\}}(y)+(1-\mu(x))1_{\{y=0\}}(y)=\mu(x)^y(1-\mu(x))^{1-y}\right\},$$

where $x\in[a,b]$ and $y\in\{0,1\}$. Then

$$\frac{dP_\mu}{d\nu}(\cdot,y)=\begin{cases}\mu\ \text{ if }\ y=1,\\ 1-\mu\ \text{ if } y=0.\end{cases}$$

Consider now the set

$$\mathscr{G} = \left\{ \left(\sqrt{\frac{\mu^Y (1-\mu)^{1-Y}}{\mu_o^Y (1-\mu_o)^{1-Y}}} - 1 \right) 1_{\{\mu_o^Y (1-\mu_o)^{1-Y} > 0\}} \ : \ \mu \in \mathcal{F}_{\text{blc}}, \ Y \in \{0,1\} \right\}.$$

This set is bounded. Indeed, according to Theorem 1.7, (ii), bi-log-concave functions μ and μ_o are continuous. Besides that, μ_o and $1-\mu_o$ are strictly positive since $0 < \mu_o < 1$ on $J(\mu_o)$ by definition. Therefore μ/μ_o and $(1-\mu)/(1-\mu_o)$ are continuous and, according to the Weierstrass theorem, are bounded on a fixed compact interval $[a,b] \subseteq J(\mu) \cap J(\mu_o)$. Then the set $\mathscr{G}$ is uniformly bounded on $[a,b]$ in the sense of van de Geer (1993) (see Appendix B), namely:

$$\left\| \sup_{g \in \mathscr{G}} |g| \right\|_\infty < \infty.$$

Consider further two subsets that comprise the set $\mathscr{G}$ of continuous functions restricted to $[a,b]$:

$$\mathscr{G}_{j,n} = \left\{ \sqrt{\frac{\mu}{\mu_o}} - 1 \ : \ h(\mu, \mu_o) \le 2^j \delta_n, \ \mu \in \mathcal{F}_{\text{blc}} \right\} \tag{3.25}$$

and

$$\bar{\mathscr{G}}_{j,n} = \left\{ \sqrt{\frac{1-\mu}{1-\mu_o}} - 1 \ : \ h(\mu, \mu_o) \le 2^j \delta_n, \ \mu \in \mathcal{F}_{\text{blc}} \right\}, \tag{3.26}$$

where δ_n is some sequence converging to zero, $j = 1, 2, \ldots$ and $h(\mu, \mu_o)$ is the Hellinger distance

$$h(\widehat{\mu}_n, \mu) = \sqrt{\frac{1}{2} \int_{\mathbb{R}} ((\sqrt{1-\mu} - \sqrt{1-\widehat{\mu}_n})^2 + (\sqrt{\mu} - \sqrt{\widehat{\mu}_n})^2) dQ}.$$

We proceed by considering the set $\mathscr{G}_{j,n}$; all derivations for the set $\bar{\mathscr{G}}_{j,n}$ follow in a similar way.

Consider the δ_n-entropy (see Kolmogorov and Tikhomirov (1959) and Appendix B for definitions and details) and metric entropy with bracketing (see, e.g., van de Geer (1993) and Appendix B) of the set $\mathscr{G}_{j,n}$, namely

$$\mathscr{H}(\delta_n, \mathscr{G}_{j,n}, || \cdot ||_{P_n})$$

and

$$\mathscr{H}^B(\delta_n, \mathscr{G}_{j,n}, || \cdot ||_{P_o}),$$

respectively. Here $|| \cdot ||_{P_o}$ and $|| \cdot ||_{P_n}$ are the restrictions of the norms $|| \cdot ||_{2,P_o}$ and $|| \cdot ||_{2,P_n}$ to the space $L_2([a,b]) \supset \mathscr{G}_{j,n} \mid_{[a,b]}$ (i.e. with respect to the first component Q of $P_\mu = Q \otimes \delta$), that is

$$||g||_{\mathcal{M}} = \left(\int_a^b g^2 d\mathcal{M} \right)^{1/2},$$

for $g \in L_2([a,b])$ and a given probability measure $\mathcal{M}$. Notice that the embedding $\mathscr{G}_{j,n}\,|_{[a,b]} \subset L_2([a,b])$ holds true indeed because

$$\begin{aligned}\int_a^b \left(\sqrt{\frac{\mu}{\mu_o}(t)}-1\right)^2 dt &\le \frac{1}{\mu_o(a)}\int_a^b \left(\sqrt{\mu(t)}-\sqrt{\mu_o(t)}\right)^2 dt \\ &\le \frac{2h^2(\mu,\mu_o)}{\mu_o(a)} \le \frac{2\cdot(2^j\delta_n)^2}{\mu_o(a)} \\ &= \frac{2^{2j+1}}{\mu_o(a)n^{2\alpha}} \to 0\end{aligned}$$

as $n\to\infty$, where the last inequality follows from (3.25) and the last equality holds true with $\delta_n := 1/n^\alpha$ and $\alpha \in (0,1/2]$.

Consider now a sequence $Q_n \xrightarrow{w} Q$ (almost surely) and recall that $a,b \in \mathrm{supp}(Q) \subset [a,b]$ and, furthermore, $Q([a,a+\epsilon)) > 0$ and $Q((b-\epsilon,b]) > 0$ for some small $\epsilon > 0$. Notice that the previous result, i.e.

$$\int_a^b \left(\sqrt{\frac{\mu}{\mu_o}(t)}-1\right)^2 dt \to 0$$

implies

$$\int_a^b |\widehat{\mu}_n - \mu| dQ_n \to 0,$$

as $n\to\infty$. Let now $I \subset [a,b]$ and $Q_n(I) \ge \delta > 0$. Then there are two sequences $\widehat{a}_n$ and $\widehat{b}_n$ such that $\widehat{a}_n \to_p a$ and $\widehat{b}_n \to_p b$. Furthermore, it holds true that

$$\int_a^b |\widehat{\mu}_n - \mu| dQ_n \ge Q_n(I)\min_{x\in\widehat{I}_n} |\widehat{\mu}_n(x) - \mu(x)|,$$

where $\widehat{I}_n \subset [\widehat{a}_n, \widehat{b}_n]$. Then

$$(\widehat{\mu}_n - \mu)(\widehat{a}_n) \to_p 0$$

and

$$(\widehat{\mu}_n - \mu)(\widehat{b}_n) \to_p 0.$$

Notice that

$$Q_n([a,a+\epsilon)) = Q_n((-\infty,a+\epsilon)) \ge Q([a,a+\epsilon)) + o_p(1).$$

Consider $\mu(a)$, $\mu(b) \in [\delta, 1-\delta]$, where $\delta \in (0,1/2)$. Then

$$\begin{aligned}\widehat{\mu}_n(a) &= \exp(\log(\widehat{\mu}_n(a))) \\ &= \exp(\log(\widehat{\mu}_n(a)) - \log(\widehat{\mu}_n(\widehat{a}_n)))\widehat{\mu}_n(\widehat{a}_n) \\ &= \exp\left((a-\widehat{a}_n)\frac{\log(\widehat{\mu}_n(a)) - \log(\widehat{\mu}_n(\widehat{a}_n))}{a-\widehat{a}_n}\right)\widehat{\mu}_n(\widehat{a}_n).\end{aligned}$$

Furthermore, according to the first inequality from Lemma 1.25,

$$\begin{aligned}\frac{\log(\widehat{\mu}_n(a)) - \log(\widehat{\mu}_n(\widehat{a}_n))}{a - \widehat{a}_n} &\leq \frac{\widehat{\mu}_n'(\widehat{a}_n)}{\widehat{\mu}_n(\widehat{a}_n)} \\ &\leq \frac{\log(\widehat{\mu}_n(\widehat{a}_n)) - \log(\widehat{\mu}_n(\widehat{b}_n))}{\widehat{a}_n - \widehat{b}_n} \\ &\to_p \frac{\log(\mu(a)) - \log(\mu(b))}{a - b}.\end{aligned}$$

Since $\widehat{a}_n - a \to_p 0$ and $\widehat{\mu}_n(\widehat{a}_n) \to_p \mu(a)$ we finally obtain

$$\widehat{\mu}_n(a) \to_p \mu(a).$$

Analogously one can proceed in the case of $\widehat{\mu}_n(b)$ deriving

$$\widehat{\mu}_n(b) \to_p \mu(b).$$

These results imply

$$\sup_{t\in[a,b]} \left| \sqrt{\frac{\widehat{\mu}_n}{\mu_o}(t)} - 1 \right| = O_P(1). \tag{3.27}$$

It follows from the result of van de Geer (1993) (Theorem 3.2 and Appendix B), that, having a uniform boundedness of $\mathscr{G}$ (which we have already shown above) and $\delta_n\sqrt{n} \geq 1$,

$$h(\widehat{\mu}_n, \mu_o) = O_p(\delta_n),$$

provided that the following conditions hold true:

$$\lim_{j\to\infty} \limsup_{n\to\infty} \frac{\sqrt{\mathscr{H}^B(\delta_n, \mathscr{G}_{j,n}, ||\cdot||_{P_o})}}{2^j\delta_n\sqrt{n}} = 0, \tag{3.28}$$

and

$$\limsup_{n\to\infty} \mathbb{P}\left(\sum_{i=1}^{\infty} \frac{\sqrt{\mathscr{H}(2^{-i}\delta_n, \mathscr{G}_{j,n}, ||\cdot||_{P_n})}}{2^i 2^j \delta_n\sqrt{n}} > \beta_j \text{ for some } j\right) = 0 \tag{3.29}$$

for some sequence $\beta_j \to 0$.

We will construct such sequence δ_n which fulfils these conditions and thus represents the rate of convergence in the Hellinger distance for the estimator of μ_o. In order to do so, we will first need some result from the theory of empirical processes. This result makes it possible to estimate δ_n-entropy and, subsequently, metric entropy with bracketing. Let us recall some preliminaries.

For $0 < \alpha < \infty$, we consider a class of all functions on a bounded set Ω in $\mathbb{R}^d$ that possess uniformly bounded partial derivatives up to order $\underline{\alpha}$ (the greatest integer smaller than α) and whose highest partial derivatives are Lipschitz of order $\alpha - \underline{\alpha}$.

Define for any vector $k = (k_1, k_2, ..., k_d)$ of d integers the following differential operator

$$D^k = \frac{\delta^{k.}}{\delta x_1^{k_1} \cdots \delta x_d^{k_d}},$$

where $k. = \sum_{i=1}^{d} k_i$. Then for a function $g : \Omega \to \mathbb{R}$ let

$$||g||_\alpha = \max_{k. \leq \underline{\alpha}} \sup_x |D^k g(x)| + \max_{k. = \underline{\alpha}} \sup_{x,y} \frac{|D^k g(x) - D^k g(y)|}{||x-y||^{\alpha - \underline{\alpha}}}, \tag{3.30}$$

where the suprema are taken over all x, y in the interior of Ω with $x \neq y$. Let $C_M^\alpha(\Omega)$ be the set of all continuous functions $g : \Omega \to \mathbb{R}$ with $\|g\|_\alpha \leq M$. Then the following result (van der Vaart and Wellner (1996), Theorem 2.7.1) holds true:

Theorem 3.5. *Let Ω be a bounded, convex subset of $\mathbb{R}^d$ with non-empty interior. There exists a constant K depending only on α and d such that*

$$\mathscr{H}(\delta_n, C_1^\alpha(\Omega), \|\cdot\|_\infty) \leq K\lambda(\Omega^1)\delta_n^{-d/\alpha},$$

where $\lambda(\Omega^1)$ is the Lebesgue measure of the set $\{x : \|x - \Omega\| < 1\}$.

First we notice that, by the scaling arguments, this theorem is also valid in the case of $C_M^\alpha(\Omega)$ with $M > 1$. Namely, for the functions f and g such that $\|f\|_\alpha \leq M$ and $\|g\|_\alpha \leq M$ we have

$$\|f/M\|_\alpha \leq 1$$

and

$$\|g/M\|_\alpha \leq 1$$

and

$$\left\| \frac{f}{M} - \frac{g}{M} \right\|_\infty = \frac{\|f-g\|_\infty}{M}.$$

In our case $C_M^\alpha(\Omega) \equiv \mathscr{G}_{j,n}$ (and, subsequently $C_M^\alpha(\Omega) \equiv \bar{\mathscr{G}}_{j,n}$) with $\alpha = 2$ and $\underline{\alpha} = 1$. We have to show that there exists such $M < \infty$ that $\|g\|_\alpha \leq M$ for any $g \in \mathscr{G}_{j,n}$ (and $g \in \bar{\mathscr{G}}_{j,n}$). That is, that the following norm is finite:

$$\begin{aligned} \left\| \frac{\widehat{\mu}_n}{\mu_o} \right\|_2 &= \max \left\{ \sup_{t \in [a,b]} \left| \sqrt{\frac{\widehat{\mu}_n(t)}{\mu_o(t)}} - 1 \right|, \sup_{t \in [a,b]} \left| \left(\sqrt{\frac{\widehat{\mu}_n(t)}{\mu_o(t)}} - 1 \right)' \right| \right\} \\ &+ \sup_{x,y \in [a,b]} \frac{\left| \left(\sqrt{\frac{\widehat{\mu}_n(x)}{\mu_o(x)}} - 1 \right)' - \left(\sqrt{\frac{\widehat{\mu}_n(y)}{\mu_o(y)}} - 1 \right)' \right|}{|x-y|}. \end{aligned}$$

First of all we will verify that

$$\sup_{t \in [a,b]} \left| \left(\sqrt{\frac{\widehat{\mu}_n(t)}{\mu_o(t)}} - 1 \right)' \right|$$

is finite. Assume $y > x$ and let $\widehat{\mu}_n$ and μ_o be bi-log-concave cumulative distribution functions (the case of constant bi-log-concave regression functions is trivial with respect to the statement of the theorem we are proving). In what follows we write μ instead of $\widehat{\mu}_n$. Consider

$$\begin{aligned}
\frac{\sqrt{\frac{\mu(y)}{\mu_o(y)}} - \sqrt{\frac{\mu(x)}{\mu_o(x)}}}{y-x} &= \frac{\sqrt{\mu_o(x)\mu(y)} - \sqrt{\mu(x)\mu_o(y)}}{(y-x)\sqrt{\mu_o(x)\mu_o(y)}} \\
&\leq \frac{\sqrt{\mu_o(y)\mu(y)} - \sqrt{\mu(x)\mu_o(y)}}{(y-x)\sqrt{\mu_o(a)}\sqrt{\mu_o(y)}} \\
&= \frac{\sqrt{\mu(y)} - \sqrt{\mu(x)}}{(y-x)\sqrt{\mu_o(a)}} \underset{y\to x}{\longrightarrow} \frac{(\sqrt{\mu(x)})'}{\mu_o(a)} \\
&= \frac{\mu'(x)}{2\sqrt{\mu_o(a)}\sqrt{\mu(x)}} = \frac{\mu'(x)}{\mu(x)} \cdot \frac{\sqrt{\mu(x)}}{2\sqrt{\mu_o(a)}} \\
&= h(x) \cdot \frac{\sqrt{\mu(x)}}{2\sqrt{\mu_o(a)}} \leq h(a) \cdot \frac{\sqrt{\mu(b)}}{2\sqrt{\mu_o(a)}},
\end{aligned}$$

where $h(x) = \mu'(x)/\mu(x)$. On the other hand,

$$\begin{aligned}
\frac{\sqrt{\frac{\mu(x)}{\mu_o(x)}} - \sqrt{\frac{\mu(y)}{\mu_o(y)}}}{x-y} &= \frac{\sqrt{\mu(x)\mu_o(y)} - \sqrt{\mu_o(x)\mu(y)}}{(x-y)\sqrt{\mu_o(x)\mu_o(y)}} \\
&\leq \frac{\sqrt{\mu_o(x)\mu(x)} - \sqrt{\mu_o(x)\mu(y)}}{(x-y)\sqrt{\mu_o(b)}\sqrt{\mu_o(x)}} \\
&= \frac{\sqrt{\mu(x)} - \sqrt{\mu(y)}}{(x-y)\sqrt{\mu_o(b)}} \underset{y\to x}{\longrightarrow} \frac{(\sqrt{\mu(x)})'}{\mu_o(b)} \\
&= \frac{\mu'(x)}{2\sqrt{\mu_o(b)}\sqrt{\mu(x)}} = \frac{\mu'(x)}{\mu(x)} \cdot \frac{\sqrt{\mu(x)}}{2\sqrt{\mu_o(b)}} \\
&= h(x) \cdot \frac{\sqrt{\mu(x)}}{2\sqrt{\mu_o(b)}} \geq h(b) \cdot \frac{\sqrt{\mu(a)}}{2\sqrt{\mu_o(b)}}.
\end{aligned}$$

Therefore

$$\limsup_{y\to x} \frac{\sqrt{\frac{\mu}{\mu_o}(y)} - \sqrt{\frac{\mu}{\mu_o}(x)}}{y-x} \leq \frac{1}{2} h(a) \sqrt{\frac{\mu(b)}{\mu_o(a)}}.$$

On the other hand,

$$\liminf_{y\to x} \frac{\sqrt{\frac{\mu}{\mu_o}(y)} - \sqrt{\frac{\mu}{\mu_o}(x)}}{y-x} \geq \frac{1}{2} h(b) \sqrt{\frac{\mu(a)}{\mu_o(b)}}.$$

Thus, by Lemma 3.2, $\sqrt{\frac{\mu}{\mu_o}(t)}$ is Lipschitz-continuous on $[a,b]$ with Lipschitz-constant

$$L := L(a,b) \in \left[\frac{1}{2}h(b)\sqrt{\frac{\mu(a)}{\mu_o(b)}}, \frac{1}{2}h(a)\sqrt{\frac{\mu(b)}{\mu_o(a)}}\right].$$

Therefore, as it is well-known from the theory of functions,

$$\sup_{t\in[a,b]} |(\sqrt{\widehat{\mu}_n(t)/\mu_o(t)} - 1)'| < \infty$$

and $\sqrt{\frac{\mu}{\mu_o}(t)} - 1$ is continuous on $[a,b]$. Hence, by the Weierstrass theorem, the latter function is bounded on $[a,b]$. Analogously one proceeds in the case when μ and μ_o are bi-log-concave survival functions.

Now we turn our attention to the second summand of $\|\widehat{\mu}_n/\mu_o\|_2$, namely:

$$\mathscr{S}(x,y) := \frac{\left|\left(\sqrt{\frac{\widehat{\mu}_n(x)}{\mu_o(x)}} - 1\right)' - \left(\sqrt{\frac{\widehat{\mu}_n(y)}{\mu_o(y)}} - 1\right)'\right|}{|x-y|}.$$

Let $x < y$ and write:

$$\begin{aligned}\mathscr{S}(x,y) &= \frac{\frac{\mu'(y)}{\mu_o(y)} \cdot \frac{\sqrt{\mu_o(y)}}{\sqrt{\mu(y)}} - \frac{\mu_o'(y)}{\mu_o(y)} \cdot \frac{\sqrt{\mu(y)}}{\sqrt{\mu_o(y)}} - \frac{\mu'(x)}{\mu_o(x)} \cdot \frac{\sqrt{\mu_o(x)}}{\sqrt{\mu(x)}} + \frac{\mu_o'(x)}{\mu_o(x)} \cdot \frac{\sqrt{\mu(x)}}{\sqrt{\mu_o(x)}}}{2(y-x)} \\ &= \frac{\frac{\mu'(y)}{\mu_o(y)} \cdot O_P(1) - \frac{\mu_o'(y)}{\mu_o(y)} \cdot O_P(1) - \frac{\mu'(x)}{\mu_o(x)} \cdot O_P(1) + \frac{\mu_o'(x)}{\mu_o(x)} \cdot O_P(1)}{2(y-x)},\end{aligned}$$

where the last equality is due to equation (3.27). Therefore, we can consider the following expression equivalent to the last one:

$$\frac{\frac{\mu'(y)}{\mu_o(y)} - \frac{\mu'(x)}{\mu_o(x)} + \left(\frac{\mu_o'(x)}{\mu_o(x)} - \frac{\mu_o'(y)}{\mu_o(y)}\right)}{2(y-x)} = \frac{\frac{1-\mu(y)}{\mu_o(y)}\widetilde{h}(y) - \frac{1-\mu(x)}{\mu_o(x)}\widetilde{h}(x)}{2(y-x)} + \frac{h_o(x) - h_o(y)}{2(y-x)}, \tag{3.31}$$

where $h_o = \mu_o'/\mu_o$ and $\widetilde{h} = \mu'/(1-\mu)$. Notice that

$$\frac{\frac{1-\mu(y)}{\mu_o(y)}\widetilde{h}(y) - \frac{1-\mu(x)}{\mu_o(x)}\widetilde{h}(x)}{2(y-x)} + \frac{h_o(x) - h_o(y)}{2(y-x)} \geq \frac{\frac{1-\mu(y)}{\mu_o(b)}\widetilde{h}(y) - \frac{1-\mu(x)}{\mu_o(a)}\widetilde{h}(x)}{2(y-x)}. \tag{3.32}$$

Here the inequality holds true since h_o is non-increasing (due to Theorem 1.7, (iii)) and therefore the summand $(h_o(x) - h_o(y))/(2(y-x))$ is non-negative. Consider now the case when μ and μ_o are cumulative distribution functions. Notice that in this case μ', μ_0' and $\widetilde{h}$ are strictly positive (see Theorem 1.7, (iv)$'$). We can continue the last inequality in the following way:

$$
\begin{aligned}
\frac{\frac{1-\mu(y)}{\mu_o(b)}\widetilde{h}(y) - \frac{1-\mu(x)}{\mu_o(a)}\widetilde{h}(x)}{2(y-x)} &\geq \frac{(1-\mu(y))\widetilde{h}(y)}{2\mu_o(b)(y-x)} - \frac{(1-\mu(x))\widetilde{h}(x)}{2\mu_o(a)\mu_o(b)(y-x)} \\
&= \frac{\mu_o(a)(1-\mu(y))-(1-\mu(x))}{2(y-x)} \cdot \frac{\widetilde{h}(y)}{\mu_o(a)\mu_o(b)} + \frac{\widetilde{h}(y)-\widetilde{h}(x)}{2(y-x)} \cdot \frac{1-\mu(x)}{\mu_o(a)\mu_o(b)} \\
&\geq \frac{\widetilde{h}(a)}{2\mu_o(a)\mu_o(b)} \cdot \frac{\mu_o(a)(1-\mu(x))\exp(\mu'(x)(y-x)/(1-\mu(x)))-(1-\mu(x))}{y-x} \\
&+ \frac{\widetilde{h}(y)-\widetilde{h}(x)}{2(y-x)} \cdot \frac{1-\mu(b)}{\mu_o(a)\mu_o(b)} \\
&\geq \frac{\widetilde{h}(a)}{2\mu_o(a)\mu_o(b)} \cdot \frac{(1-\mu(x))(\exp(\mu'(x)(y-x)/(1-\mu(x)))-1)}{y-x} \\
&+ \frac{\widetilde{h}(y)-\widetilde{h}(x)}{(y-x)} \cdot \frac{1-\mu(b)}{2\mu_o(a)\mu_o(b)} \\
&\underset{y\to x}{\longrightarrow} -\frac{\widetilde{h}(a)\widetilde{h}(x)(1-\mu(x))}{2\mu_o(a)\mu_o(b)} + \frac{(1-\mu(b))\widetilde{h}'(x)}{2\mu_o(a)\mu_o(b)} \\
&\geq -\frac{\widetilde{h}(a)\widetilde{h}(b)(1-\mu(a))}{2\mu_o(a)\mu_o(b)} + \min_{x\in[a,b]}\widetilde{h}'(x)\frac{1-\mu(b)}{2\mu_o(a)\mu_o(b)}.
\end{aligned}
$$

Consider now

$$\widetilde{h}'(x) = \left(\frac{\mu'(x)}{1-\mu(x)}\right)' = \frac{\mu''(x)}{1-\mu(x)} + \left(\frac{\mu'(x)}{1-\mu(x)}\right)^2.$$

Since $\widetilde{h}(x)$ is non-decreasing (see Theorem 1.7, (iii)) we have $\widetilde{h}'(x) \geq 0$, and therefore

$$\frac{\mu''(x)}{1-\mu(x)} \geq -\left(\frac{\mu'(x)}{1-\mu(x)}\right)^2.$$

But $\mu'(x)/(1-\mu(x))$ is bounded. Indeed, since $\mu'(x)$ is continuous (as locally Lipschitz-continuous), and $1-\mu(x)$ is continuous and non-zero (see Theorem 1.7) the ratio $\mu'(x)/(1-\mu(x))$ is continuous and, according to the Weierstrass theorem, is bounded on the compact $[a,b]$. Thus

$$0 < \frac{\mu'(x)}{1-\mu(x)} \leq M$$

for some $\infty > M > 0$ and therefore

$$-\left(\frac{\mu'(x)}{1-\mu(x)}\right)^2 \geq -M^2.$$

Hence $\min_{x\in[a,b]} \widetilde{h}'(x)$ is finite and thus the lower bound for $\mathscr{S}(x,y)$ is finite as well. Let us now find the upper bound. Consider

$$\begin{aligned}
\frac{\frac{\mu'(y)}{\mu_o(y)} - \frac{\mu'(x)}{\mu_o(x)} + \left(\frac{\mu_o'(x)}{\mu_o(x)} - \frac{\mu_o'(y)}{\mu_o(y)}\right)}{2(y-x)} &= \frac{\mu'(y)}{2\mu_o(y)(y-x)} + \frac{\mu'(x)}{2\mu_o(x)(x-y)} \\
&- \frac{h_o(y)-h_o(x)}{2(y-x)} \\
&\leq \frac{\mu'(y)}{2\mu_o(x)(y-x)} + \frac{\mu'(x)}{2\mu_o(x)(x-y)} \\
&- \frac{h_o(y)-h_o(x)}{2(y-x)} \\
&= \frac{1}{2\mu_o(x)} \cdot \frac{\mu'(x)-\mu'(y)}{x-y} - \frac{h_o(y)-h_o(x)}{2(y-x)}.
\end{aligned}$$

Recall that $h = \mu'/\mu$ is non-increasing function; then

$$\begin{aligned}
\frac{\mu'(x)-\mu'(y)}{x-y} &= \frac{\mu(x)h(x)-\mu(y)h(y)}{x-y} \\
&= h(x)\frac{\mu(x)-\mu(y)}{x-y} + \mu(y)\frac{h(x)-h(y)}{x-y} \\
&\leq h(x)\frac{\mu(x)-\mu(y)}{x-y},
\end{aligned}$$

because $x < y$ and $h(x) - h(y) \geq 0$. Continue the last inequality:

$$\begin{aligned}
h(x)\frac{\mu(y)-\mu(x)}{y-x} &\leq h(a)\frac{\mu(y)-\mu(x)}{y-x} \\
&\leq h(a)\mu(x)\frac{\exp(h(x)(y-x))-1}{y-x} \\
&\underset{y\to x}{\longrightarrow} h(a)h(x)\mu(x) \\
&\leq h^2(a)\mu(b).
\end{aligned}$$

Taking everything together we finally obtain the upper bound:

$$\begin{aligned}
\frac{1}{2\mu_o(x)} \cdot \frac{\mu'(x)-\mu'(y)}{x-y} - \frac{h_o(y)-h_o(x)}{2(y-x)} &\underset{y\to x}{\longrightarrow} \frac{h^2(a)\mu(b)}{2\mu_o(x)} - \frac{h_o'(x)}{2} \\
&\leq \frac{h^2(a)\mu(b)}{2\mu_o(a)} - \frac{1}{2}\min_{x\in[a,b]} h_o'(x),
\end{aligned}$$

where $\min_{x\in[a,b]} h_o'(x)$ is finite. Indeed, consider

$$h_o'(x) = \left(\frac{\mu_o'(x)}{\mu_o(x)}\right)' = \frac{\mu_o''(x)}{\mu_o(x)} - \left(\frac{\mu_o'(x)}{\mu_o(x)}\right)^2.$$

Since $h_o(x)$ is non-increasing (see Theorem 1.7, (iii)) we have $h_o'(x) \leq 0$, and therefore

$$\frac{\mu_o''(x)}{\mu_o(x)} \leq \left(\frac{\mu_o'(x)}{\mu_o(x)}\right)^2 .$$

But $\mu_o'(x)/\mu_o(x)$ is bounded. Indeed, since $\mu_o'(x)$ is continuous (as locally Lipschitz-continuous), and $\mu_o(x)$ is continuous and non-zero (see Theorem 1.7), the ratio $\mu_o'(x)/\mu_o(x)$ is continuous and, according to the Weierstrass theorem, is bounded on the compact $[a,b]$. Thus

$$0 < \frac{\mu_o'(x)}{\mu_o(x)} \leq M_o$$

for some $0 < M_o < \infty$, and therefore

$$\left(\frac{\mu_o'(x)}{\mu_o(x)}\right)^2 \leq M_o^2.$$

Hence $\min\limits_{x\in[a,b]} h_o'(x)$ is finite and thus the upper bound for $\mathscr{S}(x,y)$ is finite as well. Thus we have shown that $\left\|\frac{\widehat{\mu}_n}{\mu_o}\right\|_2$ is finite in the case when μ and μ_o are bi-log-concave distribution functions.

Consider now the case when μ and μ_o are bi-log-concave survival functions. Notice that in this case μ', μ_o' and $\widetilde{h}$, $\widetilde{h}_o$, h and h_o are strictly negative. Indeed, let us consider μ' for simplicity. Condition (iii) of Theorem 1.7 implies that $\mu' < 0$ on $J(\mu)$. For if $\mu'(x_o) = 0$ for some $x_o \in J(\mu)$, then isotonicity of $\widetilde{h} = \mu'/(1-\mu)$ would imply that $\mu'(x) = 0$ for $x \leq x_o$, and antitonicity of $h = \mu'/\mu$ would yield $\mu'(x) = 0$ for $x \geq x_o$. Hence μ would be constant on $J(\mu)$, a violation of μ being a survival function on $\mathbb{R}$. Having that in mind, we start from (3.31) and (3.32) in order to derive the lower bound. We continue the latter inequality in the following way:

$$\begin{aligned}
&\frac{\frac{1-\mu(y)}{\mu_o(b)}\widetilde{h}(y) - \frac{1-\mu(x)}{\mu_o(a)}\widetilde{h}(x)}{2(y-x)} \geq \frac{\frac{1-\mu(y)}{\mu_o(a)\mu_o(b)}\widetilde{h}(y) - \frac{1-\mu(x)}{\mu_o(a)}\widetilde{h}(x)}{2(y-x)} \\
= \;& \frac{\mu_o(b)(1-\mu(x)) - (1-\mu(y))}{2(x-y)} \cdot \frac{\widetilde{h}(x)}{\mu_o(a)\mu_o(b)} + \frac{\widetilde{h}(y) - \widetilde{h}(x)}{2(y-x)} \cdot \frac{1-\mu(y)}{\mu_o(a)\mu_o(b)} \\
= \;& \frac{(1-\mu(x)) - (1-\mu(y))}{2(x-y)} \cdot \frac{\widetilde{h}(a)}{\mu_o(a)\mu_o(b)} + \frac{\widetilde{h}(y) - \widetilde{h}(x)}{2(y-x)} \cdot \frac{1-\mu(a)}{\mu_o(a)\mu_o(b)},
\end{aligned}$$

where the last inequality is due to the fact that the first summand in the numerator is negative as $\widetilde{h} < 0$ (and so by multiplying it with $\mu_o(a) < 1$ will only make the whole fraction smaller). We continue the last inequality:

$$\begin{aligned}
& \frac{(1-\mu(x))-(1-\mu(y))}{2(x-y)} \cdot \frac{\widetilde{h}(a)}{\mu_o(a)\mu_o(b)} + \frac{\widetilde{h}(y)-\widetilde{h}(x)}{2(y-x)} \cdot \frac{1-\mu(a)}{\mu_o(a)\mu_o(b)} \\
= \quad & \frac{\widetilde{h}(a)}{2\mu_o(a)\mu_o(b)} \cdot \frac{(1-\mu(y))\exp(\mu'(y)(x-y)/(1-\mu(y)))-(1-\mu(y))}{x-y} \\
+ \quad & \frac{\widetilde{h}(y)-\widetilde{h}(x)}{2(y-x)} \cdot \frac{1-\mu(a)}{\mu_o(a)\mu_o(b)} \\
\underset{x\to y}{\longrightarrow} \quad & -\frac{\widetilde{h}(a)\widetilde{h}(y)(1-\mu(y))}{2\mu_o(a)\mu_o(b)} + \frac{(1-\mu(a))\widetilde{h}'(y)}{2\mu_o(a)\mu_o(b)} \\
\geq \quad & -\frac{\widetilde{h}(a)\widetilde{h}(b)(1-\mu(b))}{2\mu_o(a)\mu_o(b)} + \min_{y\in[a,b]} \widetilde{h}'(y)\frac{1-\mu(a)}{2\mu_o(a)\mu_o(b)}.
\end{aligned}$$

Consider now

$$\widetilde{h}'(y) = \left(\frac{\mu'(y)}{1-\mu(y)}\right)' = \frac{\mu''(y)}{1-\mu(y)} + \left(\frac{\mu'(y)}{1-\mu(y)}\right)^2.$$

Since $\widetilde{h}(y)$ is non-decreasing $\widetilde{h}'(y) \geq 0$, and therefore

$$\frac{\mu''(y)}{1-\mu(y)} \geq -\left(\frac{\mu'(x)}{1-\mu(x)}\right)^2.$$

But $\mu'(x)/(1-\mu(x))$ is bounded. Indeed, since $\mu'(x)$ is continuous (as locally Lipschitz-continuous), and $1-\mu(x)$ is continuous and non-zero (see Theorem 1.7) the ratio $\mu'(x)/(1-\mu(x))$ is continuous and, according to the Weierstrass theorem, is bounded on the compact $[a,b]$. Therefore, $\min_{x\in[a,b]} \widetilde{h}'(x)$ is finite and thus the lower bound for $\mathscr{S}(x,y)$ is finite, too.

Let us now find the upper bound. Consider

$$\begin{aligned}
\frac{\frac{\mu'(y)}{\mu_o(y)} - \frac{\mu'(x)}{\mu_o(x)} + \left(\frac{\mu_o'(x)}{\mu_o(x)} - \frac{\mu_o'(y)}{\mu_o(y)}\right)}{2(y-x)} \quad = \quad & \frac{\mu'(y)}{2\mu_o(y)(y-x)} + \frac{\mu'(x)}{2\mu_o(x)(x-y)} \\
- \quad & \frac{h_o(y)-h_o(x)}{2(y-x)} \\
\leq \quad & \frac{\mu'(y)}{2\mu_o(x)(y-x)} + \frac{\mu'(x)}{2\mu_o(y)(x-y)} \\
- \quad & \frac{h_o(y)-h_o(x)}{2(y-x)} \\
= \quad & \frac{1}{2\mu_o(y)} \cdot \frac{\mu'(x)-\mu'(y)}{x-y} - \frac{h_o(y)-h_o(x)}{2(y-x)}.
\end{aligned}$$

Recall that $h = \mu'/\mu$ is non-increasing function; then

$$\begin{aligned}\frac{\mu'(x)-\mu'(y)}{x-y} &= \frac{\mu(x)h(x)-\mu(y)h(y)}{x-y}\\ &= h(x)\frac{\mu(x)-\mu(y)}{x-y} + \mu(y)\frac{h(x)-h(y)}{x-y}\\ &\leq h(x)\frac{\mu(x)-\mu(y)}{x-y},\end{aligned}$$

because $x < y$ and $h(x) - h(y) \geq 0$. Continue the last inequality:

$$\begin{aligned}h(x)\frac{\mu(y)-\mu(x)}{y-x} &\leq h(a)\frac{\mu(y)-\mu(x)}{y-x}\\ &\leq h(a)\mu(x)\frac{\exp(h(x)(y-x))-1}{y-x}\\ &\underset{y\to x}{\longrightarrow} h(a)h(x)\mu(x)\\ &\leq h^2(a)\mu(a).\end{aligned}$$

Taking everything together we finally obtain the upper bound:

$$\begin{aligned}\frac{1}{2\mu_o(y)}\cdot\frac{\mu'(x)-\mu'(y)}{x-y} - \frac{h_o(y)-h_o(x)}{2(y-x)} &\underset{y\to x}{\longrightarrow} \frac{h^2(a)\mu(a)}{2\mu_o(y)} - \frac{h_0'(x)}{2}\\ &\leq \frac{h^2(a)\mu(a)}{2\mu_o(b)} - \frac{1}{2}\min_{x\in[a,b]} h_0'(x),\end{aligned}$$

where $\min_{x\in[a,b]} h_0'(x)$ is finite due to the similar arguments as in the case of the lower bound.

Thus we have proved that $\mathscr{S}$ is bounded and therefore the norm $\left\|\frac{\widehat{\mu}_n}{\mu_o}\right\|_2$ is finite also in the case when μ and μ_o are bi-log-concave survival functions.

Now all conditions of Theorem 3.5 are verified, and together with Corollary 1.14 and equation (3.27) this theorem yields the upper bound for δ_n-entropy with respect to the supremum norm:

$$\mathscr{H}(\delta_n, \mathscr{G}_{j,n}, ||\cdot||_\infty) \leq \frac{c}{\sqrt{\delta_n}}, \tag{3.33}$$

where $c > 0$ and $g \in \mathscr{G}_{j,n}$. It is well known (see, e.g. van der Vaart and Wellner (1996), p.84), that

$$\mathscr{H}(\delta_n, \mathscr{G}_{j,n}, ||\cdot||_{P_n}) \leq \mathscr{H}^B(2\delta_n, \mathscr{G}_{j,n}, ||\cdot||_{P_n}).$$

Further, it follows from van de Geer (1993) (see also Appendix B) and the continuity of the functions from $\mathscr{G}_{j,n}$ (on $[a,b]$) that

$$\mathscr{H}^B(2\delta_n, \mathscr{G}_{j,n}, ||\cdot||_{P_o}) \leq \mathscr{H}(\delta_n, \mathscr{G}_{j,n}, ||\cdot||_{P_o,\infty}) = \mathscr{H}(\delta_n, \mathscr{G}_{j,n}, ||\cdot||_\infty),$$

where

$$||g||_{P_o,\infty} = \underset{x\in[a,b]}{\text{esssup}}|g(x)|$$

is the essential supremum of $g \in \mathscr{G}_{j,n}$ on $[a,b]$. This result, applied with regard to the measure P_n, together with the previous inequality, yields:

$$\mathscr{H}(\delta_n, \mathscr{G}_{j,n}, ||\cdot||_{P_n}) \leq \mathscr{H}(\delta_n, \mathscr{G}_{j,n}, ||\cdot||_{P_n,\infty}) = \mathscr{H}(\delta_n, \mathscr{G}_{j,n}, ||\cdot||_{\infty}).$$

The equalities in the preceding two expressions follow from equality of supremum and essential supremum of a continuous function on Lebesgue-measurable set $E \subseteq \mathbb{R}^m$ whose intersection with any open ball with centre in E has non-zero Lebesgue measure. This is a standard result from the theory of functions and measure theory which obviously holds true for $[a,b] \subset \mathbb{R}$.

First we consider condition (3.28):

$$\begin{aligned} \frac{\sqrt{\mathscr{H}^B(\delta_n, \mathscr{G}_{j,n}, ||\cdot||_{P_o})}}{2^j\delta_n\sqrt{n}} &\leq \frac{\sqrt{\mathscr{H}(\delta_n/2, \mathscr{G}_{j,n}, ||\cdot||_{\infty})}}{2^j\delta_n\sqrt{n}} \\ &= \frac{\sqrt{c}}{2^j}\frac{\sqrt{\sqrt{2}/\sqrt{\delta_n}}}{\delta_n\sqrt{n}}. \end{aligned}$$

Hence, in order to satisfy condition (3.28), it is sufficient if

$$\limsup_{n\to\infty} \frac{1}{\delta_n^{5/4} n^{1/2}} < \infty$$

since

$$\lim_{j\to\infty} \frac{1}{2^j} = 0.$$

Let $\delta_n = 1/n^{\alpha}$, where $0 < \alpha \leq 1/2$. Then condition $\delta_n\sqrt{n} \geq 1$ is satisfied. Notice that we don't consider sequences like $\delta_n = 1/\log^{\kappa} n$ with $\kappa > 0$ to ensure faster rate of convergence. Further, it should be the case that

$$\limsup_{n\to\infty} \frac{1}{n^{1/2-5\alpha/4}} < \infty.$$

Therefore $\alpha \leq 2/5$ and we choose $\alpha = 2/5$ so that δ_n will have the fastest possible rate of convergence.

Otherwise, if $\alpha > 2/5$, one obtains

$$\limsup_{n\to\infty} n^{5\alpha/4-1/2} = \infty.$$

Now we consider condition (3.29):

$$\begin{aligned}\sum_{i=1}^{\infty}\frac{\sqrt{\mathscr{H}(2^{-i}\delta_n,\mathscr{G}_{j,n},||\cdot||_{P_n})}}{2^i2^j\delta_n\sqrt{n}} &\leq \sum_{i=1}^{\infty}\frac{\sqrt{\mathscr{H}(2^{-i}\delta_n,\mathscr{G}_{j,n},||\cdot||_{\infty})}}{2^i2^j\delta_n\sqrt{n}}\\ &= \sum_{i=1}^{\infty}\frac{\sqrt{c}}{2^j}\frac{\sqrt{2^{i/2}/\sqrt{\delta_n}}}{2^i\delta_n\sqrt{n}}\\ &= \sum_{i=1}^{\infty}\frac{\sqrt{c}}{2^j}\frac{\sqrt{2^{i/2}n^{\alpha/2}}}{2^in^{1/2-\alpha}}.\end{aligned}$$

Hence, in order to satisfy that condition, it would be sufficient if

$$\limsup_{n\to\infty}\sum_{i=1}^{\infty}\frac{1}{2^{3i/4}}\frac{n^{\alpha/4}}{n^{1/2-\alpha}}=0.$$

Since the series

$$\sum_{i=1}^{\infty}\frac{1}{2^{3i/4}}$$

converges it must hold true that

$$\limsup_{n\to\infty}\frac{1}{n^{1/2-5\alpha/4}}=0.$$

Applying the arguments similar to the ones above we obtain $\alpha\leq 2/5$ and again the choice is $\alpha=2/5$.

Otherwise, if $\alpha>2/5$, it follows that

$$\limsup_{n\to\infty}n^{5\alpha/4-1/2}=\infty.$$

Having established that $\delta_n=n^{-2/5}$, we can finally write

$$h(\widehat{\mu}_n,\mu)=O_p(\delta_n)=O_p\left(\frac{1}{n^{2/5}}\right).$$

□

Remark 3.6. *Notice that the boundedness of $\mathscr{S}$ and Lemma 3.2, (i) implies straightforwardly that the first derivatives of the functions from $\mathscr{G}_{j,n}$ (and subsequently from $\bar{\mathscr{G}}_{j,n}$) are Lipschitz-continuous of order 1 and therefore bounded on $[a,b]\subset\mathbb{R}$. Thus, by verifying conditions of Theorem 3.5 we also verified conditions of Theorem XV from Kolmogorov and Tikhomirov (1959) which yields the same upper bound for the entropy as in equation (3.33):*

Theorem 3.7 (Theorem XV, Kolmogorov and Tikhomirov (1959)). *Let $\mathscr{A}$ be a set of functions g on $[a,b]$ such that for every $x \in [a,b]$:*

(i) $|g^{(k)}(x)| \leq C_k$, $k = 0, ..., p$;

(ii) $|g^{(p)}(x+h) - g^{(p)}(x)| \leq C|h|^{\alpha}$, $0 < \alpha \leq 1$.

If the constants $C_0, C_1, ..., C_p$, C are positive then

$$\mathscr{H}(\epsilon, \mathscr{A}, ||\cdot||_{\infty}) \sim \left(\frac{1}{\epsilon}\right)^{\frac{1}{p+\alpha}}.$$

In our setting $(\mathscr{G}_{j,n} \cup \bar{\mathscr{G}}_{j,n}) \subset \mathscr{A}$ and $p = \alpha = 1$ whereas $\epsilon = \delta_n$ which immediately yields the result as in (3.33).

Proof of Corollary 2.6. Notice that bi-log-concave function μ belongs to the class of functions for which the optimality of convergence rate of estimators $\widehat{\mu}$ on the compact $[a,b]$ with respect to L_1-norm (see Stone (1982), Eggermont and LaRiccia (2009), p. 10 and pp. $17-19$, and also Appendix A) holds true. Indeed, consider the class $\mathfrak{G}_C([a,b])$ consisting of the functions g on $[a,b]$ satisfying the following conditions:

(i) $g \in C^{m-1}([a,b])$;

(ii) $g^{(m-1)}$ is absolutely continuous;

(iii) $||g^{(m)}||_{\infty} \leq C$,

where $m \geq 1$ and $0 < C < \infty$. Then the rate $O(n^{-m/(2m+1)})$ is the optimal rate of convergence of nonparametric estimators $\widehat{g}_n$ of $g \in \mathfrak{G}_C([a,b])$ with respect to L_q-norm, $q \in [1,\infty)$. Moreover, the corresponding estimators $\widehat{g}_n$ are asymptotically optimal.

When applied to our setting, condition (iii) is satisfied indeed because Lemma 1.23 guarantees that μ'' is uniformly bounded on $[a,b]$. Besides, conditions (i) and (ii) are satisfied as well due to Theorem 1.7, (iv). Thus for bi-log-concave μ we have $m = 2$.

Notice that the classes $\mathscr{G}_{j,n}$ and $\bar{\mathscr{G}}_{j,n}$ also satisfy conditions for the optimality of convergence rates on the compact $[a,b]$ with $m = 2$. Indeed, Lipschitz-continuity (and therefore the boundedness) on $[a,b]$ of the first derivatives of the functions from $\mathscr{G}_{j,n}$ and $\bar{\mathscr{G}}_{j,n}$ was pointed out in Remark 3.6 and implies also the boundedness of the second derivatives of these functions.

Finally, since $||P_n - P||_1 \leq 2H(P_n, P)$ for probability measures P_n, P (see e.g. Pollard (2002)), we obtain

$$||\widehat{\mu}_n - \mu||_1 \leq 2H(P_n, P) \equiv 2h(\widehat{\mu}_n, \mu),$$

that is, $q = 1$, and the result of the corollary follows. □

Appendix

A. Nonparametric Rates of Convergence

The following definitions can be found in Stone (1982).

Let (X, Y) be a pair of random variables and let θ denote the regression function of the response Y on X, that is $\mathbb{E}(Y \mid X) = \theta(X)$. Let $\widehat{\theta}_n$ with $n \geq 1$ denote estimators of θ, so that $\widehat{\theta}_n$ is based on a random sample (X_1, Y_1), ..., (X_n, Y_n) of size n from the distribution of (X, Y).

The unknown regression function θ is assumed to belong to a collection Θ of suitably smooth functions on $\mathbb{R}^d$, and, besides, $\theta(x) \in J$, where J is an open interval in $\mathbb{R}$ for $\theta \in \Theta$ and $x \in \mathbb{R}^d$. Let $\mathbb{P}_\theta$ denote the dependence of various probabilities on θ.

Let further $T(\theta) = T(\cdot;\theta)$ denote an arbitrary finite linear combination (with constant coefficients) of the derivatives of θ. One of the examples is $T(x;\theta) = \theta(x)$. Let $\widehat{T}_n$ be an arbitrary (measurable) estimator of $T(\theta)$ based on the training sample $X_1, Y_1, ..., X_n, Y_n$.

Let C be a compact subset of $\mathbb{R}^d$ having non-empty interior and $q \in (0, +\infty]$. Let $\{b_n\}$ be a sequence of (eventually) positive constants.

Definition A.1. *The sequence $\{b_n\}$ is called a lower rate of convergence if there is a $c > 0$ such that*

$$\lim_n \inf_{\widehat{T}_n} \sup_{\Theta} \mathbb{P}_\theta(||\widehat{T}_n - T(\theta)||_{q,C} \geq cb_n) = 1.$$

Definition A.2. *The sequence $\{b_n\}$ is called an achievable rate of convergence if there is a sequence of estimators $\{\widehat{T}_n\}$ and a $c > 0$ such that*

$$\lim_n \sup_{\Theta} \mathbb{P}_\theta(||\widehat{T}_n - T(\theta)||_{q,C} \geq cb_n) = 0. \tag{A.1}$$

Definition A.3. *The sequence $\{b_n\}$ is called an optimal rate of convergence if it is both a lower and an achievable rate of convergence.*

Notice that if $\{b_n\}$ is a lower rate of convergence and $\{b'_n\}$ is an achievable rate of convergence, then there are positive constants c and n_o such that $b'_n \geq cb_n$ for $n \geq n_o$. If $\{b_n\}$ and $\{b'_n\}$ are both optimal rates of convergence, then there are positive constants c and n_o such that

$$cb_n \leq b'_n \leq c^{-1}b_n$$

for $n \geq n_o$. Thus it is reasonable to refer to any optimal rate of convergence as *the* optimal rate of convergence.

Definition A.4. *If* $\{b_n\}$ *is the optimal rate of convergence and* $\{\widehat{T}_n\}$ *satisfies (A.1), estimators* $\widehat{T}_n$*,* $n \geq 1$ *are said to be asymptotically optimal.*

B. Some Notions from the Theory of Empirical Processes

The following definitions and results can be found in van de Geer (1993).

Let (W, d) be a space with a semi-metric d, and let Λ be a subset of W.

Definition B.1. *A collection T of subsets $U \subset W$ is called a δ-covering of Λ, if diam$(U) \leq 2\delta$ for every $U \in T$ and $\Lambda = \underset{U \in T}{\cup} U$.*

Definition B.2. *T is called a δ-covering set.*

Notice that one can always take T to be a collection of balls

$$U = \{w \in W \mid d(u, w) \leq \delta\},$$

where $u \in W$. The collection of centres of these balls will also be referred to as a δ-covering set.

Definition B.3. *The δ-covering number of Λ for the metric d is the number of elements of the smallest δ-covering set of Λ.*

Notation B.4. *$N(\delta, \Lambda, d)$ is the δ-covering number of Λ.*

Definition B.5. *The δ-entropy of Λ is $\mathscr{H}(\delta, \Lambda, d) := \log N(\delta, \Lambda, d)$.*

Definition B.6. *If $\mathscr{H}(\delta, \Lambda, d) < \infty$ for all $\delta > 0$, then Λ is called totally bounded for d.*

Let $(\mathscr{X}, \mathscr{A})$ be some measurable space and $g \in \mathscr{G} \subseteq L_q(P)$, where P is some probability measure on $(\mathscr{X}, \mathscr{A})$ and $1 \leq q \leq \infty$. Define

$$\|g\|_{P,q} := \left(\int |g|^q dP \right)^{1/q} \text{ for } 1 \leq q < \infty,$$

$$\|g\|_{P,\infty} := \underset{x}{\text{esssup}} |g(x)|$$

and $\|g\|_\infty = \sup_x |g(x)|$ as usually.

Further define $N^B(\delta, \mathscr{G}, \|\cdot\|_{P,q})$ as the minimal k for which there exist g_1^L, g_1^U, ..., g_k^L, g_k^U such that for each $g \in \mathscr{G}$

$$g_i^L \leq g \leq g_i^U$$

for some i and

$$\|g_i^U - g_i^L\|_{P,q} \leq \delta.$$

Definition B.7. *The metric entropy with bracketing is defined as*

$$\mathscr{H}(\delta, \mathscr{G}, \| \cdot \|_{P,q}) := \log N^B(\delta, \mathscr{G}, \| \cdot \|_{P,q}).$$

Remark B.8.

$$\mathscr{H}^B(2\delta, \mathscr{G}, || \cdot ||_{P,q}) \leq \mathscr{H}(\delta, \mathscr{G}, || \cdot ||_{P,\infty}).$$

Definition B.9. *Set $\mathscr{G}$ is called uniformly bounded if* $\left\| \sup_{g \in \mathscr{G}} |g| \right\|_\infty < \infty$.

Consider a class of probability measures $\{P_\theta, \theta \in \Omega\}$ on $(\mathscr{X}, \mathscr{A})$ dominated by a σ-finite measure ν. Denote by $f_\theta := dP_\theta/d\nu$ the density of P_θ, $\theta \in \Omega$. Let $\widehat{\theta}_n$ be a maximum likelihood estimator of θ_o based on a sequence of i.i.d. observations X_1, X_2, ... from $P_o := P_{\theta_o}$, $\theta_o \in \Omega$. Define

$$P_n := \frac{1}{n} \sum_{k=1}^{n} \delta_{X_k}$$

to be the empirical distribution based on the first n observations. Write $f_o := f_{\theta_o}$ and $\widehat{f}_n := f_{\widehat{\theta}_n}$. Consider further a class of functions

$$g(f) := \{(\sqrt{f/f_o} - 1) 1_{\{f_o > 0\}}\}.$$

Denote by $H(P_\theta, P_o)$ the Hellinger distance between probability measures P_θ and P_o or, equivalently, $h(f_\theta, f_o)$ between the corresponding densities f_θ and f_o:

$$h(f_\theta, f_o) \ = H(P_\theta, P_o) \equiv \sqrt{\frac{1}{2} \int (\sqrt{f_\theta} - \sqrt{f_o})^2 d\nu}.$$

Consider now a subclass of $g(f)$:

$$\mathscr{G}_{n,j} := \{g(f_\theta),\ \theta \in \Omega:\ h(f_o, f_\theta) \leq 2^j \delta_n\},$$

where $j = 1$, 2, ..., and δ_n is some sequence of positive numbers.

Theorem B.10 (van de Geer (1993), Theorem 3.2)**.** *Suppose $\mathscr{G}$ is uniformly bounded. Let $\{\delta_n\}$ be a sequence for which $\delta_n \sqrt{n} \geq 1$ and*

$$\lim_{j \to \infty} \limsup_{n \to \infty} \frac{\sqrt{\mathscr{H}^B(\delta_n, \mathscr{G}_{j,n}, || \cdot ||_{P_o})}}{2^j \delta_n \sqrt{n}} = 0$$

and

$$\limsup_{n \to \infty} \mathbb{P} \left(\sum_{i=1}^{\infty} \frac{\sqrt{\mathscr{H}(2^{-i}\delta_n, \mathscr{G}_{j,n}, || \cdot ||_{P_n})}}{2^i 2^j \delta_n \sqrt{n}} > \beta_j \textit{ for some } j \right) = 0$$

for some sequence $\beta_j \to 0$. Then

$$h(\widehat{f}_n, f_o) = O_p(\delta_n).$$

References

An, M. Y. (1996). Log-concave probability distributions: Theory and statistical testing. Game Theory and Information 9611002, EconWPA.

An, M. Y. (1998). Logconcavity versus logconvexity: a complete characterization. *J. Econom. Theory*, 80(2):350–369.

Bagnoli, M. and Bergstrom, T. (2005). Log-concave probability and its applications. *Econom. Theory*, 26(2):445–469.

Bahadur, R. R. and Savage, L. J. (1956). The nonexistence of certain statistical procedures in nonparametric problems. *Annals of Mathematical Statistics*, 27:1115–1122.

Barlow, R. E. and Proschan, F. (1965). *Mathematical theory of reliability*. With contributions by Larry C. Hunter. The SIAM Series in Applied Mathematics. John Wiley & Sons, Inc., New York-London-Sydney.

Barlow, R. E. and Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing. Probability Models.* Holt, Rinehart and Winston, New York.

Berk, R. H. and Jones, D. H. (1979). Goodness-of-fit statistics that dominate the kolmogorov statistics. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 47:47–59.

Boyd, S. and Vandenberghe, L. (2009). *Convex optimization*. Cambridge University Press, Cambridge, UK.

Castillo, R. and Rafeiro, H. (2016). *An Introductory Course in Lebesgue Spaces*. CMS Books in Mathematics. Springer.

Cule, M., Gramacy, R., and Samworth, R. (2010). Maximum likelihood estimation of a multidimensional log-concave density (with discussion). *J. Roy. Statist. Soc., Ser. B*, 72:545–600.

Cule, M. and Samworth, R. (2010). Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density. *Electron. J. Statist.*, 4:254–270.

Dümbgen, L. (2010). *Empirische Prozesse*. Lecture Notes. Universität Bern, Bern.

Dümbgen, L., Freitag-Wolf, S., and Jongbloed, G. (2004). Consistency of concave regression with an application to current-status data. *Mathematical Methods of Statistics*, 13(1):69–81.

Dümbgen, L., Freitag-Wolf, S., and Jongbloed, G. (2006). Estimating a unimodal distribution from interval-censored data. *J. Amer. Statist. Assoc.*, 101(475):1094–1106.

Dümbgen, L., Kolesnyk, P., and Wilke, R. A. (2016). Bi-log-concave distribution functions. *Journal of Statistical Planning and Inference, to appear*, arXiv:1508.07825v5.

Dümbgen, L. and Rufibach, K. (2009). Maximum likelihood estimation of a log-concave density: Basic properties and uniform consistency. *Bernoulli*, 15(1):40–68.

Dümbgen, L. and Rufibach, K. (2011). logcondens: Computations related to univariate log-concave density estimation. *Journal of Statistical Software*, 39(6):1–28.

Dümbgen, L. and Wellner, J. A. (2014). Confidence bands for distribution functions: A new look at the law of the iterated logarithm. *arXiv:1402.2918v2*.

Durrett, R. (2010). *Probability: theory and examples. 4-th edition*. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, Cambridge, UK.

Eggermont, P. P. and LaRiccia, V. N. (2009). *Maximum Penalized Likelihood Estimation: Regression.*, volume II. Wiley-Blackwell.

Groeneboom, P. and Jongbloed, G. (2014). *Nonparametric Estimation under Shape Constraints*. Cambridge Books. Cambridge University Press, Cambridge, UK.

Groeneboom, P., Jongbloed, G., and Wellner, J. A. (2001). Estimation of a convex function: Characterizations and asymptotic theory. *The Annals of Statistics*, 29:1653–1698.

Ibragimov, I. (1956). On the composition of unimodal distributions. *Theory Probab. Appl.*, 1(2):255–260.

Jager, L. and Wellner, J. A. (2007). Goodness-of-fit tests via phi-divergences. *Ann. Statist.*, 35(5):2018–2053.

Johnson, N. L. and Kotz, S. (1970). *Distributions in statistics. Continuous univariate distributions. 2.* Houghton Mifflin Co., Boston, Mass.

Kleiber, C. and Kotz, S. (2003). *Statistical size distributions in economics and actuarial sciences*. Hoboken, NJ: Wiley.

Kolmogorov, A. N. and Tikhomirov, V. M. (1959). ε-entropy and ε-capacity of sets in function spaces. *Uspehi Mat. Nauk*, 14(2 (86)):3–86.

Kotz, S., Read, C. B., Balakrishnan, N., and Vidakovic, B., editors (2006). *Encyclopedia of statistical sciences*. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York.

Le Cam, L. and Lo Yang, G. (2000). *Asymptotics in Statistics: Some Basic Concepts*. Springer Series in Statistics. Springer-Verlag New York.

Li, K.-C. (1989). Honest confidence regions for nonparametric regression. *Ann. Statist.*, 17(3):1001–1008.

Massart, P. (1990). The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *The Annals of Probability*, 18(3).

Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*. Springer, New York.

Owen, A. B. (1995). Nonparametric likelihood confidence bands for a distribution function. *Journal of the American Statistical Association*, 90:516–521.

Patel, J., Kapadia, C., and Owen, D. (1976). *Handbook of Statistical Distributions*. Statistics, textbooks and monographs. M. Dekker.

Pollard, D. (2002). *A User's Guide to Measure Theoretic Probability*. Cambridge University Press, Cambridge, UK.

Prekopa, A. (1973). On logarithmically concave measures and functions. *Acta Scientarium Mathematicarum*, 34:335–343.

R Development Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Resnick, S. I. (2007). *Heavy-tail phenomena: probabilistic and statistical modeling*. Springer series in operations research and financial engineering. Springer, New York.

Robertson, T., Wright, F., and Dykstra, R. (1988). *Order restricted statistical inference*. Wiley & Sons.

Rockafellar, R. T. (1970). *Convex analysis*. Princeton Mathematical Series. Princeton University Press.

Schuhmacher, D., Hüsler, A., and Dümbgen, L. (2011). Multivariate log-concave distributions as a nearly parametric model. *Statistics and Risk Modeling*, 28(3):277–295.

Sengupta, D. and Nanda, A. (1999). Log-concave and concave distributions in reliability. *Naval Research Logistics (NRL)*, 46:419 – 433.

Seregin, A. and Wellner, J. A. (2010). Nonparametric estimation of multivariate convex-transformed densities. *Ann. Statist.*, 38(6):3751–3781.

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

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

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.*, 10(4):1040–1053.

van de Geer, S. (1993). Hellinger-consistency of certain nonparametric maximum likelihood estimators. *Ann. Statist.*, 21(1):14–44.

van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, Cambridge, UK.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes : with applications to statistics*. Springer series in statistics. Springer, New York.

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

Woolridge, J. (2000). *CEOSAL2 (Instructional Stata data sets for Econometrics)*. Boston College Department of Economics.

Index

List of Symbols

Symbols used in Chapters $1 - 3$. This list is not exhaustive.

Symbol	Description
F	cumulative distribution function
$\widehat{F}_n$	empirical distribution function
f	probability density function
f_μ	density of probability measure P_μ w.r.t. the dominating measure ν, p. 90
$\mathcal{F}_{\text{blc}}$	class of all bi-log-concave functions on $\mathbb{R}$, p. 4
$\mathcal{F}_{\text{blcd}}$	class of distributions with bi-log-concave F, p. 9
$\Gamma(\cdot)$	Gamma function, p. 13
$h(\cdot, \cdot)$	Hellinger distance, p. 44
$H(P, P_o)$	Hellinger distance between probability measures P and P_o, p. 91
$H(f, f_o)$	Hellinger distance between probability densities, p. 91
$\mathscr{H}(\delta, \cdot, \cdot)$	δ-entropy, $\delta > 0$, p. 92
$\mathscr{H}^B(\delta, \cdot, \cdot)$	δ-entropy with bracketing, $\delta > 0$, p. 92
$J(F)$	$\{x \in \mathbb{R} : 0 < F(x) < 1\}$, where $F : \mathbb{R} \to [0, 1]$, p. 3
$L_p(\Omega)$	space of functions on $\Omega \subset \mathbb{R}^d$ for which the p-th power of the absolute value is Lebesgue integrable, $d \geq 1$ and $p \geq 1$, p. 13
$\mathcal{N}(m, v)$	Gaussian distribution with mean m and variance v, p. 6
$U([a, b])$	uniform distribution on $[a, b] \in \mathbb{R}$, p. 10

$\kappa_{n,\alpha}^{KS}$	$(1-\alpha)$-quantile of the Kolmogorov-Smirnov test statistic, p. 14
$\kappa_{n,\alpha}^{WKS}$	$(1-\alpha)$-quantile of weighted Kolmogorov-Smirnov statistic, p. 14
$\mathrm{logit}(u)$	inverse logistic function $\log(\frac{u}{1-u})$ for $0 < u < 1$, p. 40
(L_n, U_n)	unconstrained confidence band, p. 13
(L_n^o, U_n^o)	shape-constrained confidence band, p. 17
μ	bi-log-concave function, p. 39
$\widehat{\mu} \equiv \widehat{\mu}_n$	maximum likelihood estimator of bi-log-concave regression function, p. 41
$\mathscr{P}$	family $\{P_\mu, \mu \in \mathfrak{F}_{\mathrm{blc}}\}$ of probability measures indexed by μ, p. 90
$P_n,\ Q_n$	empirical distributions, p. 90
ν	σ-finite measure dominating P_μ, p. 90
$T_1(F)$	$\sup_{x \in J(F)} \frac{f(x)}{F(x)}$, p. 9
$T_2(F)$	$\sup_{x \in J(F)} \frac{f(x)}{1-F(x)}$, p. 9
θ	$\mathrm{logit}(\mu)$, p. 40
$\mathscr{F}_{mon}$	family of functions that are either non-decreasing or non-increasing, p. 43

Erklärung

gemäss Art. 28 Abs. 2 RSL 05

Name/Vorname:	Kolesnyk Petro
Matrikelnummer:	09–991–373
Studiengang:	Statistik, Dissertation
Titel der Arbeit:	Bi-log-concave Distribution and Regression Functions
Leiter der Arbeit:	Prof. Dr. L. Dümbgen

Ich erkläre hiermit, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen benutzt habe. Alle Stellen, die wörtlich oder sinngemäss aus Quellen entnommen wurden, habe ich als solche gekenn-zeichnet. Mir ist bekannt, dass andernfalls der Senat gemäss Artikel 36 Absatz 1 Buchstabe r des Gesetzes vom 5. September 1996 über die Universität zum Entzug des auf Grund dieser Arbeit verliehenen Titels berechtigt ist.

Bern, 10. 11. 2016

Petro Kolesnyk