Confidence Bands for Distribution Functions: The Law of the Iterated Logarithm and Shape Constraints

Lutz Duembgen (Bern)

Jon A. Wellner (Seattle)
Petro Kolesnyk (Bern)
Ralf Wilke (Copenhagen)

November 2014

I. The LIL for Brownian Motion and Bridge

Standard Brownian motion $\mathbb{W} = (\mathbb{W}(t))_{t \geq 0}$

LIL for BM:

$$\limsup_{t \downarrow 0} \frac{\pm \mathbb{W}(t)}{\sqrt{2t \log \log(t^{-1})}} = 1 \quad \text{a.s.}$$

$$\limsup_{t \uparrow \infty} \frac{\pm \mathbb{W}(t)}{\sqrt{2t \log \log(t)}} = 1 \quad \text{a.s.}$$

Refined half of LIL for BM:

For any constant $\nu > 3/2$,

$$\lim_{t \to \{0,\infty\}} \Big(\frac{\mathbb{W}(t)^2}{2t} - \log\log(t + t^{-1}) - \nu \log\log\log(t + t^{-1}) \Big) = -\infty \quad \text{a.s.}$$

Reformulation for standard Brownian bridge $\mathbb{U} = (\mathbb{U}(t))_{t\in(0,1)}$:

$$(0,1) \ni t \mapsto \operatorname{logit}(t) := \log\Big(\frac{t}{1-t}\Big) \in \mathbb{R},$$
$$\mathbb{R} \ni x \mapsto \ell(x) := \frac{e^x}{1+e^x} \in (0,1).$$

Refined half of LIL for BB:

For arbitrary constants $\nu > 3/2$,

$$\sup_{t\in(0,1)} \Big(\frac{\mathbb{U}(t)^2}{2t(1-t)} - C(t) - \nu D(t) \Big) \ < \ \infty \quad \text{a.s.}$$

where

$$C(t) \ := \ \log\sqrt{1+\text{logit}(t)^2/2} \qquad \approx \ \log\log\Big(\frac{1}{t(1-t)}\Big)$$

$$D(t) \ := \ \log\sqrt{1+C(t)^2/2} \qquad \approx \ \log\log\log\Big(\frac{1}{t(1-t)}\Big)$$

as $t \to \{0,1\}$.

II. A General LIL for Sub-Exponential Processes

Nonnegative stochastic process

$$X = (X(t))_{t \in \mathcal{T}} \quad \text{on} \quad \mathcal{T} \subset (0,1).$$

Locally uniform sub-exponentiality:

LUSE$_0$: For arbitrary $a \in \mathbb{R}$, $c \geq 0$ and $\eta \geq 0$,

$$\mathbb{P}\Big(\sup_{t \in [\ell(a), \ell(a+c)]} X(t) \geq \eta \Big) \leq M \exp(-L(c)\,\eta),$$

where $M \geq 1$ and $L : [0, \infty) \to [0,1]$ satisfies

$$L(c) = 1 - O(c) \quad \text{as } c \downarrow 0$$

Refinement for $\zeta \in [0, 1]$:

LUSE$_\zeta$: For arbitrary $a \in \mathbb{R}$, $c \geq 0$ and $\eta \geq 0$,

$$\mathbb{P}\Big(\sup_{t \in [\ell(a), \ell(a+c)]} X(t) \geq \eta \Big) \ \leq \ \frac{M \exp(-L(c)\,\eta)}{\max(1, L(c)\,\eta)^{\zeta}},$$

with M and $L(\cdot)$ as in LUSE$_0$.

Example:

$$X(t) \ := \ \frac{\mathbb{U}(t)^2}{2t(1-t)}$$

satisfies LUSE$_{1/2}$ with

$$M \ = \ 2 \quad \text{and} \quad L(c) \ = \ e^{-c}.$$

Proposition. Suppose that X satisfies LUSE_ζ.

For any $L_o \in (0,1)$ and $\nu > 2-\zeta$ there exists a constant $M_o = M_o(M, L(\cdot), \zeta, L_o, \nu) \geq 1$ such that

$$\mathbb{P}\Big(\sup_{\mathcal{T}} (X - C - \nu D) \geq \eta\Big) \;\leq\; M_o \exp(-L_o\, \eta)$$

for arbitary $\eta \geq 0$.

III. Implications for the Uniform Empirical Process

Let $U_1, U_2, \ldots, U_n$ be i.i.d. $\sim \mathrm{Unif}[0,1]$.

Auxiliary function $K : [0,1] \times (0,1) \to [0,\infty]$,

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

i.e. Kullback-Leibler divergence between $\mathrm{Bin}(1,x)$ and $\mathrm{Bin}(1,p)$.

Two key properties:

$$K(x,p) \; = \; \frac{(x-p)^2}{2p(1-p)}\bigl(1+o(1)\bigr) \quad \text{as } x \to p.$$

$$K(x,p) \le c \quad \Longrightarrow \quad |x-p| \le \begin{cases} \sqrt{2c\,p(1-p)} + c \\ \sqrt{2c\,x(1-x)} + c \end{cases}$$

Implication 1: Uniform empirical distribution function

$$\widehat{G}_n(t) \;:=\; \frac{1}{n}\sum_{i=1}^{n} 1_{[U_i \le t]}$$

Lemma 1. The process $X_n = (X_n(t))_{t\in(0,1)}$ with

$$X_n(t) \;:=\; n\,K(\widehat{G}_n(t), t)$$

satisfies LUSE_0 with $M = 2$ and $L(c) = e^{-c}$.

Theorem 1. For any fixed $\nu > 2$,

$$\sup_{(0,1)}\bigl(X_n - C - \nu D\bigr) \;\to_{\mathcal{L}}\; \sup_{t\in(0,1)}\Bigl(\frac{\mathbb{U}(t)^2}{2t(1-t)} - C(t) - \nu D(t)\Bigr).$$

Main ingredients for proofs:

- $\left(\widehat{G}_n(t)/t\right)_{t\in(0,1]}$ is a reverse martingale.
- Exponential transform and Doob's inequality for submartingales.
- Analytical properties of $K(\cdot,\cdot)$.
- Donsker's invariance for uniform empirical process.

Implication 2: Uniform order statistics

$$0 \; < \; U_{n:1} \; < \; U_{n:2} \; < \; \cdots \; < \; U_{n:n} \; < \; 1.$$

$$\mathcal{T}_n \; := \; \{t_{n1}, t_{n2}, \ldots, t_{nn}\} \quad \text{with} \quad t_{ni} \; := \; \mathbb{E}(U_{n:i}) = \frac{i}{n+1}.$$

Lemma 2. The process $\tilde{X}_n = (\tilde{X}_n(t))_{t \in \mathcal{T}_n}$ with

$$\tilde{X}_n(t_{ni}) \; := \; (n+1) K(t_{ni}, U_{n:i})$$

satisfies LUSE_0 with $M = 2$ and $L(c) = e^{-c}$.

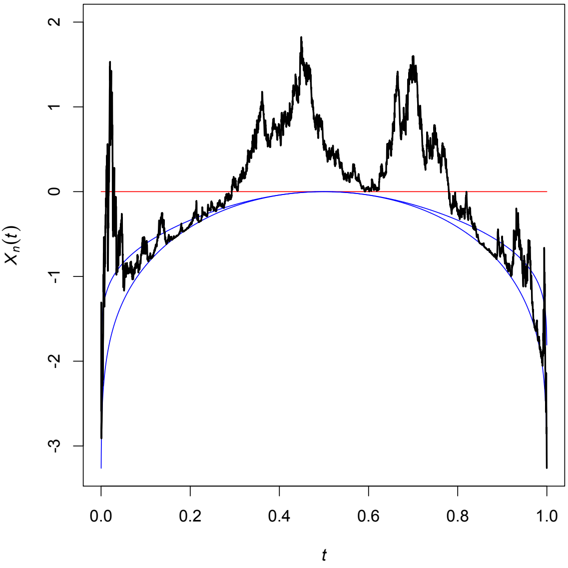
Theorem 2. For any fixed $\nu > 2$,

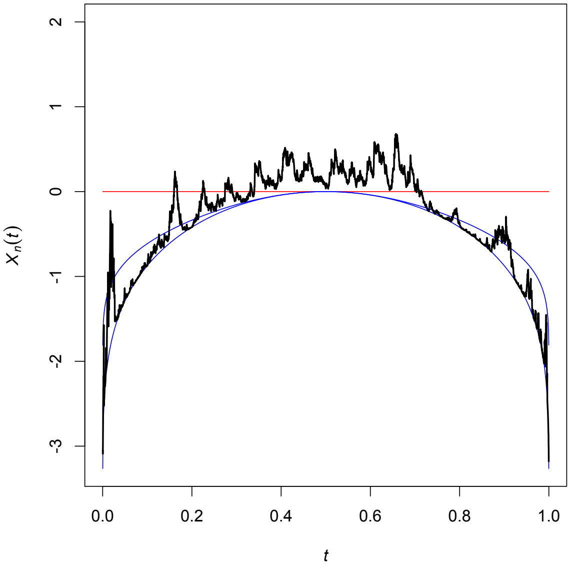
$$\sup_{\mathcal{T}_n} \bigl(\tilde{X}_n - C - \nu D \bigr) \; \to_{\mathcal{L}} \; \sup_{t \in (0,1)} \Bigl(\frac{\mathbb{U}(t)^2}{2t(1-t)} - C(t) - \nu D(t) \Bigr).$$

Main ingredients for proofs:

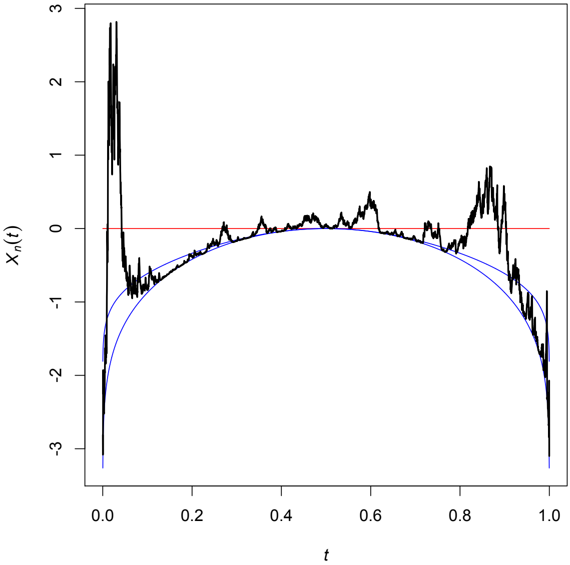
- $\left(U_{n:i}/t_{ni}\right)_{i=1}^{n}$ is a reverse martingale.
- Exponential transform and Doob's inequality for submartingales.
- Connection between Beta and Gamma distributions.
- Analytical properties of $K(\cdot,\cdot)$.
- Donsker's invariance principle for uniform quantile process.

Some realizations of $\tilde{X}_n$ for $n = 5000$ and $\nu = 3$:

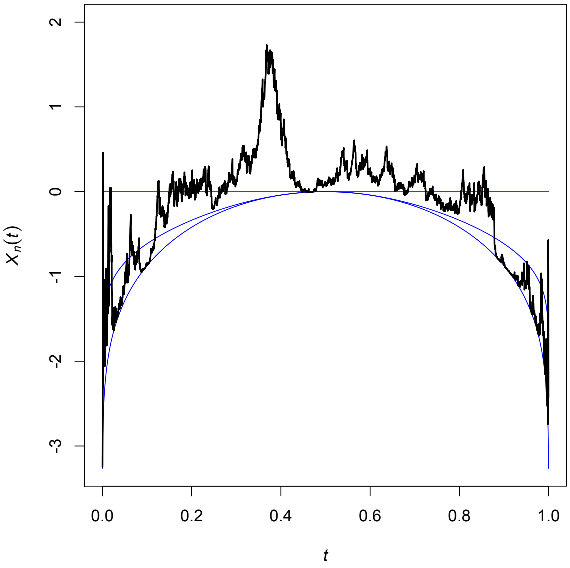




2
1
0
-1
-2
-3
$X_n(t)$
0.0
0.2
0.4
0.6
0.8
1.0
t

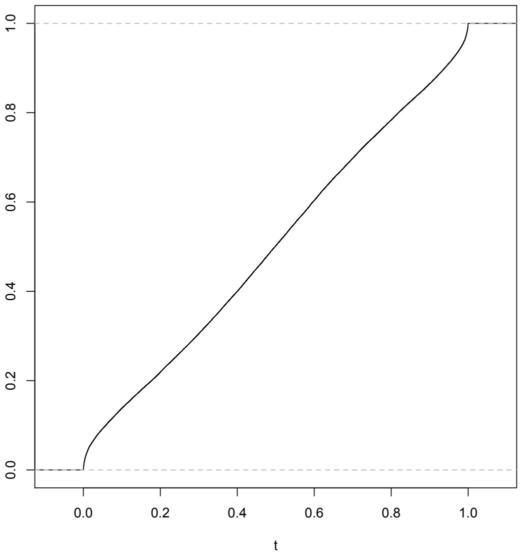


3
2
1
0
-1
-2
-3
$X_n(t)$
0.0
0.2
0.4
0.6
0.8
1.0
t



2
1
0
-1
-2
-3
$X_n(t)$
0.0
0.2
0.4
0.6
0.8
1.0
t

Distribution function of $\arg\max_t \tilde{X}_n(t)$:



III.1 Goodness-of-Fit Tests

Let $X_1, X_2, \ldots, X_n$ be i.i.d. with unknown c.d.f. F on $\mathbb{R}$.

Empirical c.d.f.

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

Testing problem:

$$H_o \ : \ F \equiv F_o \quad \text{versus} \quad H_A \ : \ F \not\equiv F_o.$$

Berk–Jones (1979) proposed the test statistic

$$T_n(F_o) \ := \ \sup_{\mathbb{R}} \, n\, K(\widehat{F}_n, F_o)$$

with critical value

$$\kappa_{n,\alpha}^{\mathrm{BJ}} \ := (1-\alpha) - \text{quantile of } \sup_{t \in (0,1)} \, n\, K_n(\widehat{G}_n(t), t)$$

$$= \ \log\log(n) + O\big(\log\log\log(n)\big).$$

New proposal:

$$T_n(F_o) \;:=\; \sup_{\mathbb{R}}\big(n\,K(\widehat{F}_n, F_o) - C(F_o) - \nu D(F_o)\big)$$

with critical value

$$\begin{aligned}
\kappa_{n,\alpha}^{\mathrm{new}} \;&:= (1-\alpha) - \text{quantile of} \\
&\qquad\qquad \sup_{t\in(0,1)} \big(n\,K(\widehat{G}_n(t), t) - C(t) - \nu D(t)\big) \\
&\to (1-\alpha) - \text{quantile of} \\
&\qquad\qquad \sup_{t\in(0,1)} \Big(\frac{\mathbb{U}(t)^2}{2t(1-t)} - C(t) - \nu D(t)\Big).
\end{aligned}$$

Power. For any fixed $\kappa > 0$,

$$\mathbb{P}_{F_n}\big(T_n(F_o) > \kappa\big) \;\to\; 1$$

as

$$\sup_{\mathbb{R}} \frac{\sqrt{n}|F_n - F_o|}{\sqrt{(1 + C(F_o))F_o(1 - F_o)} + C(F_o)/\sqrt{n}} \;\to\; \infty.$$

Special case: Detecting heterogeneous Gaussian mixtures
(Ingster 1997, 1998; Donoho–Jin 2004)

Setting 1:

$$\begin{aligned} F_o &:= \Phi, \\ F_n &:= (1 - \varepsilon_n)\,\Phi + \varepsilon_n\,\Phi(\cdot - \mu_n) \end{aligned}$$

with

$$\begin{aligned} \varepsilon_n &= n^{-\beta + o(1)}, \quad \beta \in (1/2, 1), \\ \mu_n &\to \infty. \end{aligned}$$

Theorem. For any fixed $\kappa > 0$,

$$\mathbb{P}_{F_n}\big(T_n(F_o) > \kappa\big) \ \to \ 1$$

provided that

$$\mu_n = \sqrt{2r\log(n)} \quad \text{with} \quad r > \begin{cases} \beta - 1/2 & \text{if } \beta \leq 3/4, \\ \big(1 - \sqrt{1-\beta}\big)^2 & \text{if } \beta \geq 3/4. \end{cases}$$

Setting 2 (Contiguous alternatives):

$$\begin{aligned} F_o &:= \Phi, \\ F_n &:= \Big(1 - \frac{\pi}{\sqrt{n}}\Big)\,\Phi + \frac{\pi}{\sqrt{n}}\,\Phi(\cdot - \mu), \qquad \pi, \mu > 0. \end{aligned}$$

Optimal level-α test of F_o versus F_n has asymptotic power

$$\Phi\Big(\Phi^{-1}(\alpha) + \frac{\pi^2(\exp(\mu^2) - 1)}{4}\Big).$$

Theorem. Let $\mu = \sqrt{2s \log(1/\pi)}$ for fixed $s > 0$. As $\pi \downarrow 0$,

$$\Phi\Big(\Phi^{-1}(\alpha) + \frac{\pi^2(\exp(\mu^2) - 1)}{4}\Big) \;\to\; \begin{cases} \alpha & \text{if } s < 1, \\ 1 & \text{if } s > 1, \end{cases}$$

while for any fixed $\kappa > 0$,

$$\mathbb{P}_{F_n}\big(T_n(F_o) > \kappa\big) \;\to\; \quad 1 \quad \text{if } s > 1.$$

III.2 Confidence Bands

Owen (1995) proposed $(1-\alpha)$-confidence band

$$\Big\{F : \sup_{\mathbb{R}} n\,K(\widehat{F}_n, F) \le \kappa_{n,\alpha}^{\mathrm{BJ}}\Big\}.$$

New proposal: With order statistics $X_{n:1} \le X_{n:2} \le \cdots \le X_{n:n}$,

$$\Big\{F : \max_{1\le i\le n}\big((n+1)K(t_{ni}, F(X_{n:i})) - C(t_{ni}) - \nu D(t_{ni})\big) \le \tilde{\kappa}_{n,\alpha}\Big\}$$

Resulting bounds for $F(x)$:

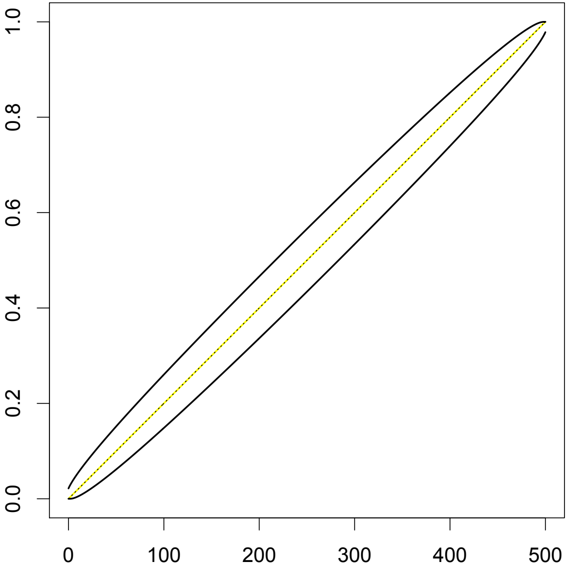
With confidence $1-\alpha$, on $[X_{n:i}, X_{n:i+1})$, $0 \le i \le n$,

$$F \in \begin{cases} [a_{ni}^{\mathrm{BJO}}, b_{ni}^{\mathrm{BJO}}] & \text{with Owen's (1995) proposal,} \\ [a_{ni}^{\mathrm{new}}, b_{ni}^{\mathrm{new}}] & \text{with new proposal,} \end{cases}$$

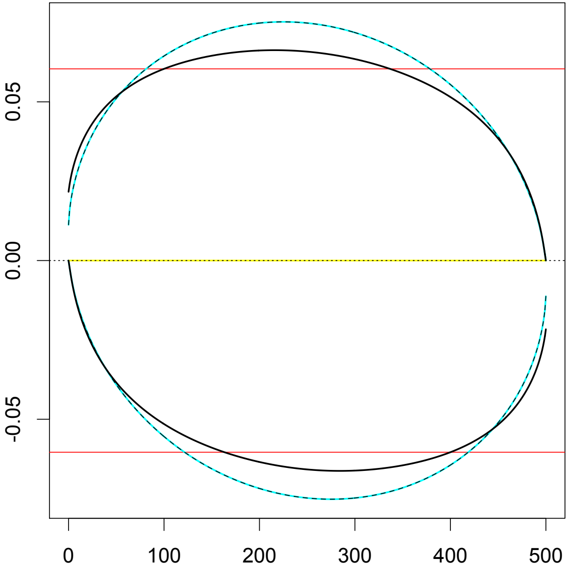
while

$$\widehat{F}_n = s_{ni} := \frac{i}{n}.$$

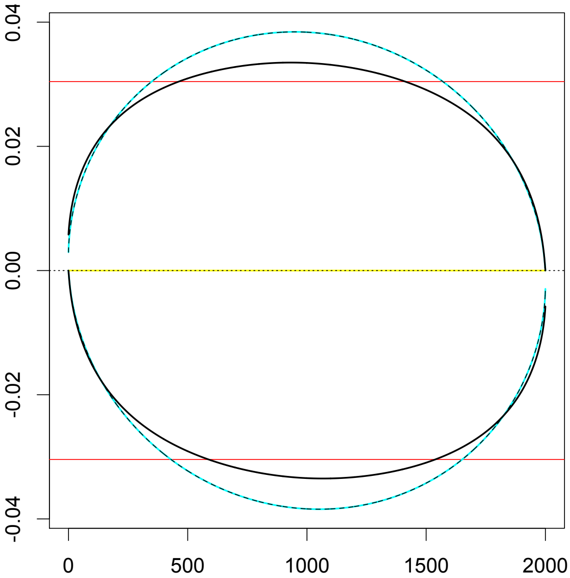
$n = 500$: $\quad i \mapsto a_{ni}^{\mathrm{new}}, s_{ni}, b_{ni}^{\mathrm{new}}$



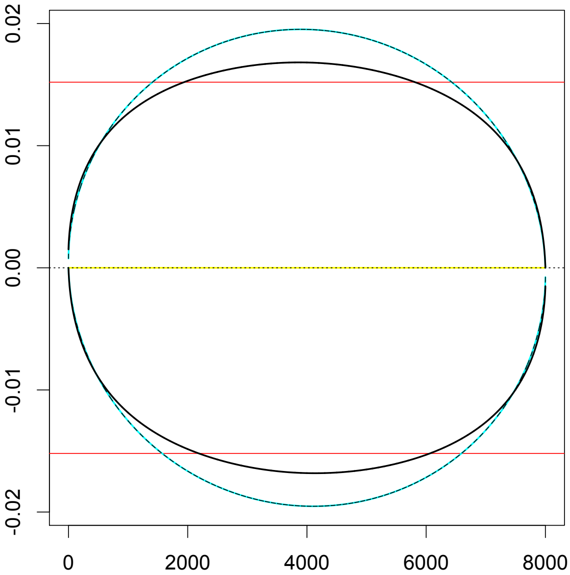
$n = 500$: $i \mapsto a^*_{ni} - s_{ni}, b^*_{ni} - s_{ni}$



$n = 2000$: $\quad i \mapsto a^*_{ni} - s_{ni}, b^*_{ni} - s_{ni}$



$n = 8000$: $\quad i \mapsto a^*_{ni} - s_{ni}, b^*_{ni} - s_{ni}$



Theorem. For any fixed $\alpha \in (0, 1)$,

$$\max_{0 \le i \le n} \frac{b_{ni}^{\mathrm{new}} - a_{ni}^{\mathrm{new}}}{b_{ni}^{\mathrm{BJO}} - a_{ni}^{\mathrm{BJO}}} \;\to\; 1,$$

while

$$\begin{aligned} \max_{0 \le i \le n} (b_{ni}^{\mathrm{BJO}} - a_{ni}^{\mathrm{BJO}}) \;&=\; (1 + o(1))\sqrt{\frac{2 \log \log n}{n}}, \\ \max_{0 \le i \le n} (b_{ni}^{\mathrm{new}} - a_{ni}^{\mathrm{new}}) \;&=\; O(n^{-1/2}). \end{aligned}$$

IV. Bi-Log-Concave Distribution Functions

Shape constraint 1: Log-concave density.

F has density $f = e^{\phi}$ with $\phi : \mathbb{R} \to [-\infty, \infty)$ concave.

Shape constraint 2: Bi-log-concave distribution function.

Both $\log(F)$ and $\log(1 - F)$ are concave.

- Log-concave density $\Longrightarrow$ bi-log-concave c.d.f.
- A bi-log-concave c.d.f. may have arbitrarily many modes!

Theorem. Let $J(F) := \{x \in \mathbb{R} : 0 < F(x) < 1\} \neq \emptyset$.
Four equivalent statements:

- F bi-log-concave.
- F has a density f. On $J(F)$,

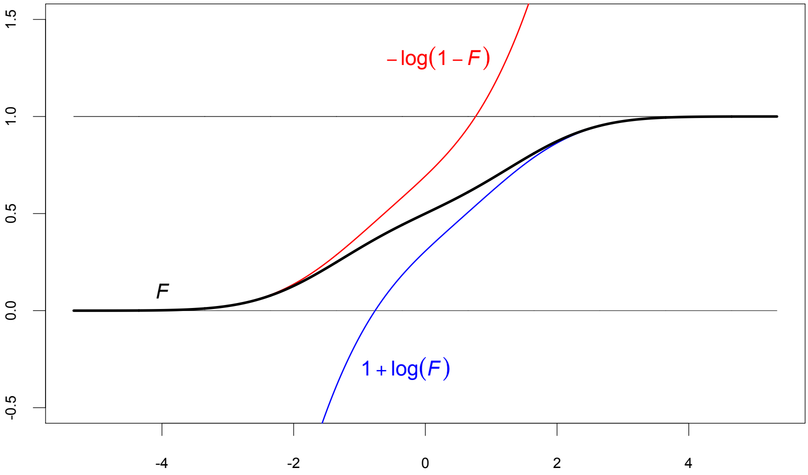
$$f = F' > 0, \quad \frac{f}{F} \searrow \quad \text{and} \quad \frac{f}{1-F} \nearrow.$$

- F has a bounded density f. On $J(F)$,

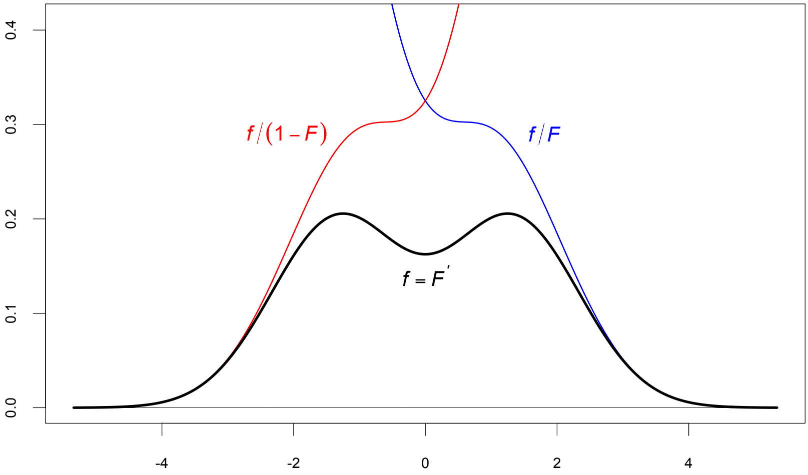
$$f = F' > 0 \quad \text{and} \quad \frac{-f^2}{1-F} \leq f' \leq \frac{f^2}{F}.$$

- F has a density f s.t. for arbitrary $x \in J(F)$ and $t \in \mathbb{R}$,

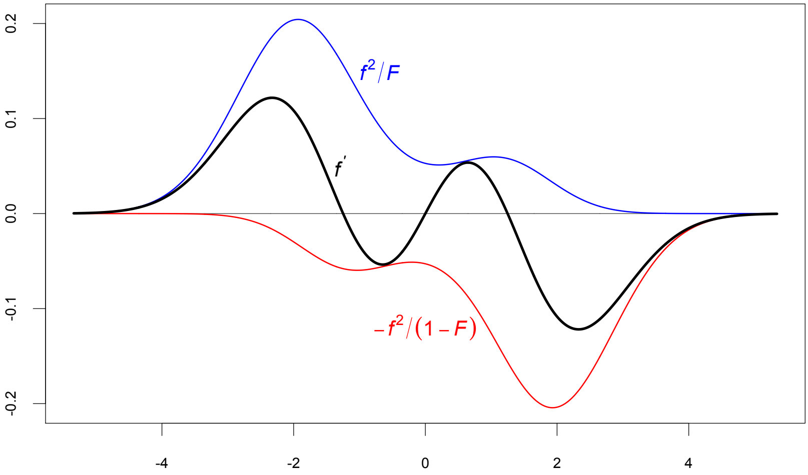
$$F(x+t) \begin{cases} \leq F(x) \exp\Big(\dfrac{f}{F}(x) \cdot t\Big), \\ \geq 1 - (1 - F(x)) \exp\Big(-\dfrac{f}{1-F}(x) \cdot t\Big). \end{cases}$$



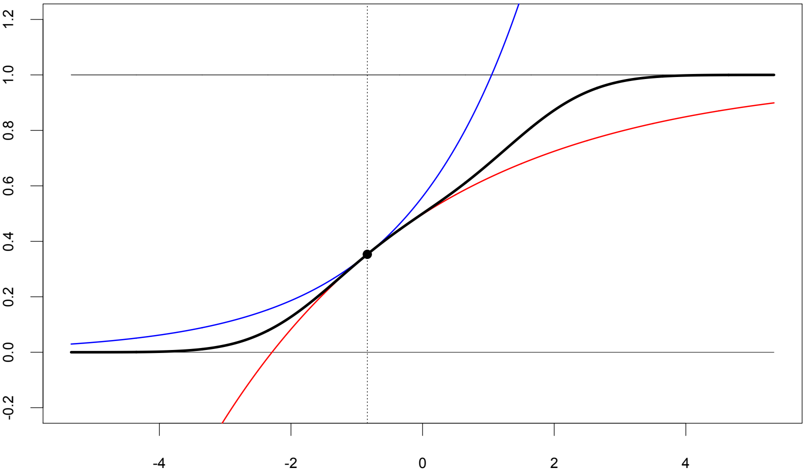
−log(1 − F)
F
1 + log(F)
1.5
1.0
0.5
0.0
-0.5
-4
-2
0
2
4



0.4
0.3
0.2
0.1
0.0
-4
-2
0
2
4
f/(1 − F)
f/F
f = F′



0.2
0.1
0.0
-0.1
-0.2
-4
-2
0
2
4
f^2/F
f'
$-f^2/(1-F)$



-0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
-4
-2
0
2
4

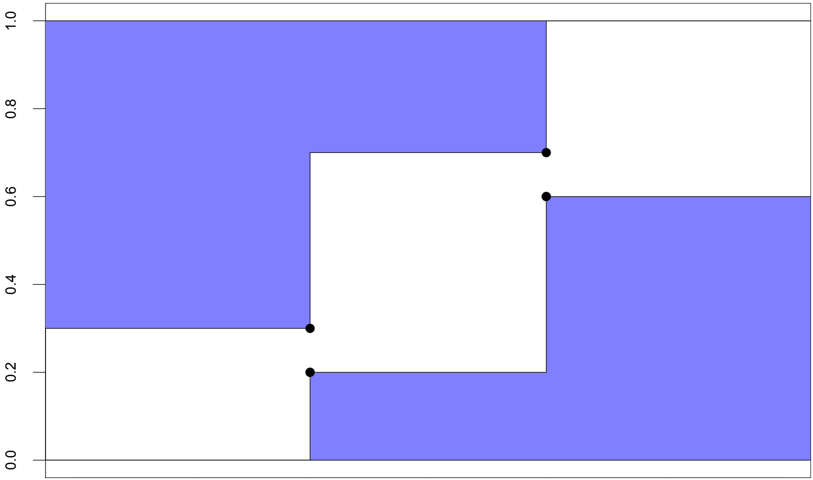
Estimation. Presumably no NPMLE of a bi-log-concave F :-(

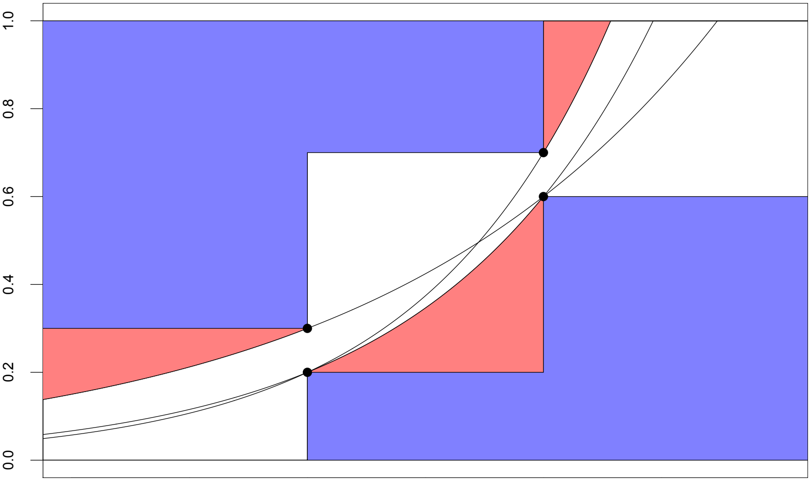
Confidence bands. Starting from a standard $(1-\alpha)$-confidence band (L_n, U_n) for F,

$$\mathbb{P}\big(L_n \le F \le U_n \text{ on } \mathbb{R}\big) \;=\; 1-\alpha,$$

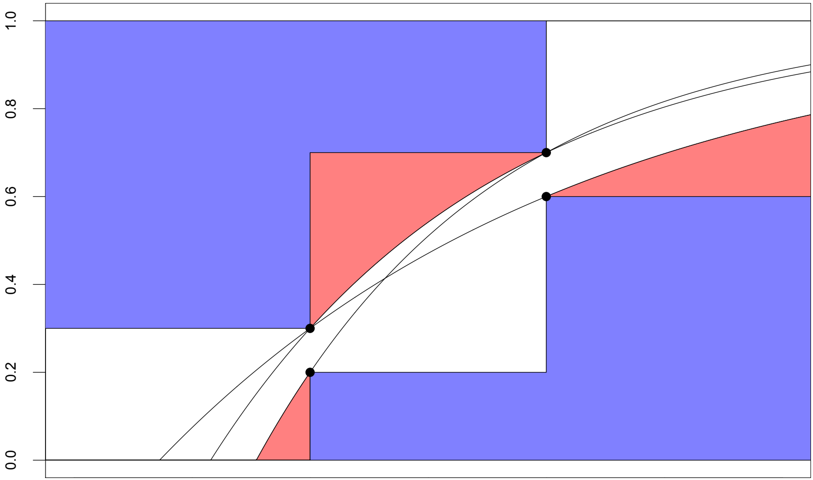
define

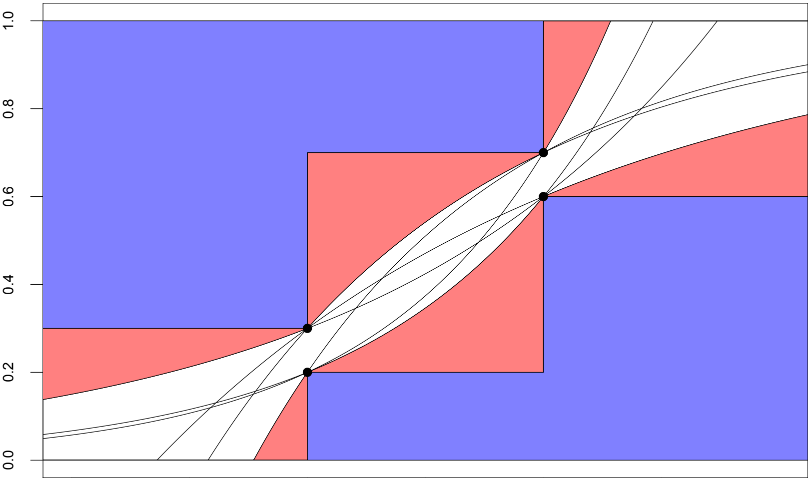
$$\begin{aligned} L_n^o(x) \;&:=\; \inf\big\{G(x) : G \text{ bi-log-concave}, L_n \le G \le U_n \text{ on } \mathbb{R}\big\}, \\ U_n^o(x) \;&:=\; \sup\big\{G(x) : G \text{ bi-log-concave}, L_n \le G \le U_n \text{ on } \mathbb{R}\big\}. \end{aligned}$$


0.0
0.2
0.4
0.6
0.8
1.0

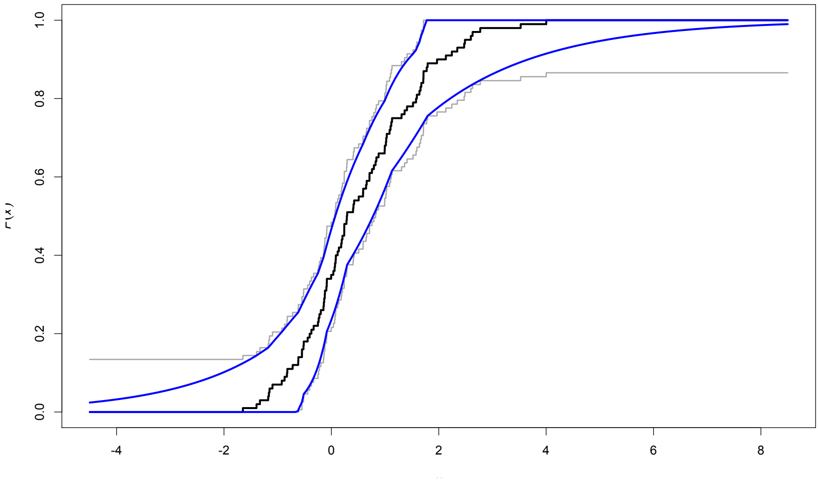


1.0
0.8
0.6
0.4
0.2
0.0

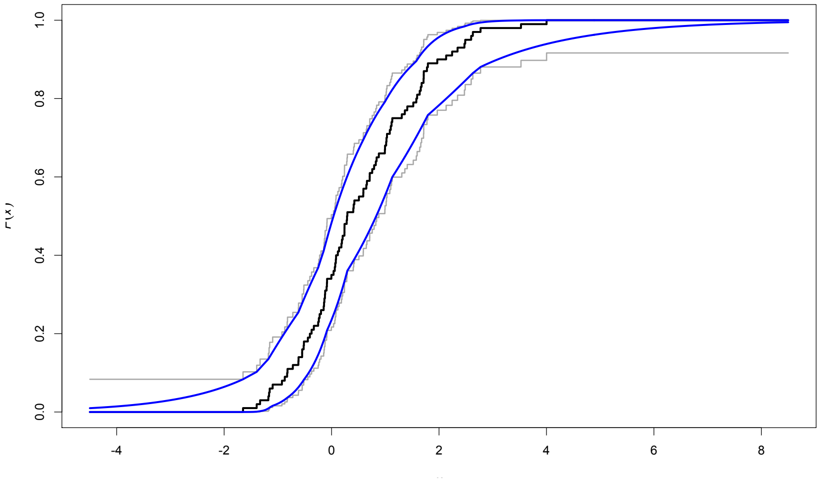

0.0
0.2
0.4
0.6
0.8
1.0



1.0
0.8
0.6
0.4
0.2
0.0



0.0
0.2
0.4
0.6
0.8
1.0
-4
-2
0
2
4
6
8



1.0
0.8
0.6
0.4
0.2
0.0
F(x)
-4
-2
0
2
4
6
8

Theorem. For any integer $k > 0$,

$$\sup_{G\,:\,L_n^o \le G \le U_n^o} \Big| \int x^k \, G(dx) - \int x^k \, F(dx) \Big| = \begin{cases} O_p\big((\log n)^k n^{-1/2}\big) & \text{with KS band,} \\ O_p\big(n^{-1/2}\big) & \text{with new band.} \end{cases}$$

Whenever $\int e^{\lambda x} \, F(dx) < \infty$,

$$\sup_{G\,:\,L_n^o \le G \le U_n^o} \Big| \int e^{\lambda x} \, G(dx) - \int e^{\lambda x} \, F(dx) \Big| \; = \; o_p(1).$$

V. Bi-Log-Concave Binary Regression

Generic observation: $(X, Y) \in \mathbb{R} \times \{0, 1\}$ (or $\mathbb{R} \times [0, 1]$).

Shape constraint:

$$\mathbb{E}(Y \mid X = x) \;=\; \mu(x)$$

with $\mu : \mathbb{R} \to [0, 1]$ bi-log-concave:

$$\log(\mu) \text{ and } \log(1 - \mu) \text{ both concave.}$$

Nonparametric extension of logistic regression, because

$$x \;\mapsto\; \ell(a + bx) \;\text{ strictly bi-log-concave}$$

for arbitrary $a, b \in \mathbb{R}$.