Log–Concavity and Density Estimation

Lutz Dümbgen, Kaspar Rufibach, André Hüsler
(Universität Bern)

Geurt Jongbloed
(Amsterdam)

Nicolai Bissantz
(Göttingen)

I. A Consulting Case

II. Log–Concave Densities

III. ML Estimation

IV. Algorithmic Aspects

V. Censored Data

VI. Outlook

I. A Consulting Case

Critical quantity M_A in a production process

- depending on five other quantities μ_{Ga}, μ_{Ka}, μ_{Ki}, μ_R, μ_S
- should not exceed a certain threshold.

$$
\begin{aligned}
\boldsymbol{M_A} \;=\; & \left[\left(\frac{\boldsymbol{\mu_R} M_{Ga}}{r(1+\boldsymbol{\mu_{Ka}} M_{Ga})}\left(K_1 + \frac{\boldsymbol{\mu_{Ka}} r_{Ka}}{\boldsymbol{\mu_{Ki}} r_{Ki}}\right) - \frac{M_H \boldsymbol{\mu_{Ka}} r_{Ka}}{\boldsymbol{\mu_{Ki}^2} r_{Ki}^2}\right) K_2 + F_z\right] \\
& \cdot \left[0.16P + \left(\frac{D_k}{2} + 0.58 d_2\right)\boldsymbol{\mu_S}\right]
\end{aligned}
$$

with

$$
\begin{aligned}
M_{Ga} &= \tan\left(\beta - \arctan\left(\frac{\boldsymbol{\mu_{Ga}}}{\cos\alpha_r}\right)\right), \\
M_H &= \left(\frac{262}{d_2 \tan(\rho' + \phi) + D_k \boldsymbol{\mu_S}} - F_z\right) \cdot d_2 \tan(\rho' - \phi), \\
\rho' &= \arctan(1.155\,\boldsymbol{\mu_S}) \\
&\vdots
\end{aligned}
$$

Five stochastically independent random quantities

quantity	sample size	sample mean	sample st.d.
μ_{Ga}	10	0.1167	0.0081
μ_{Ka}	10	0.1463	0.0072
μ_{Ki}	10	0.2030	0.0182
μ_{R}	787	1681.03	76.25
μ_{S}	2000	0.0977	0.0091

Estimate distribution of

$$M_A\left(\mu_{Ga}, \mu_{Ka}, \mu_{Ki}, \mu_R, \mu_S\right)$$

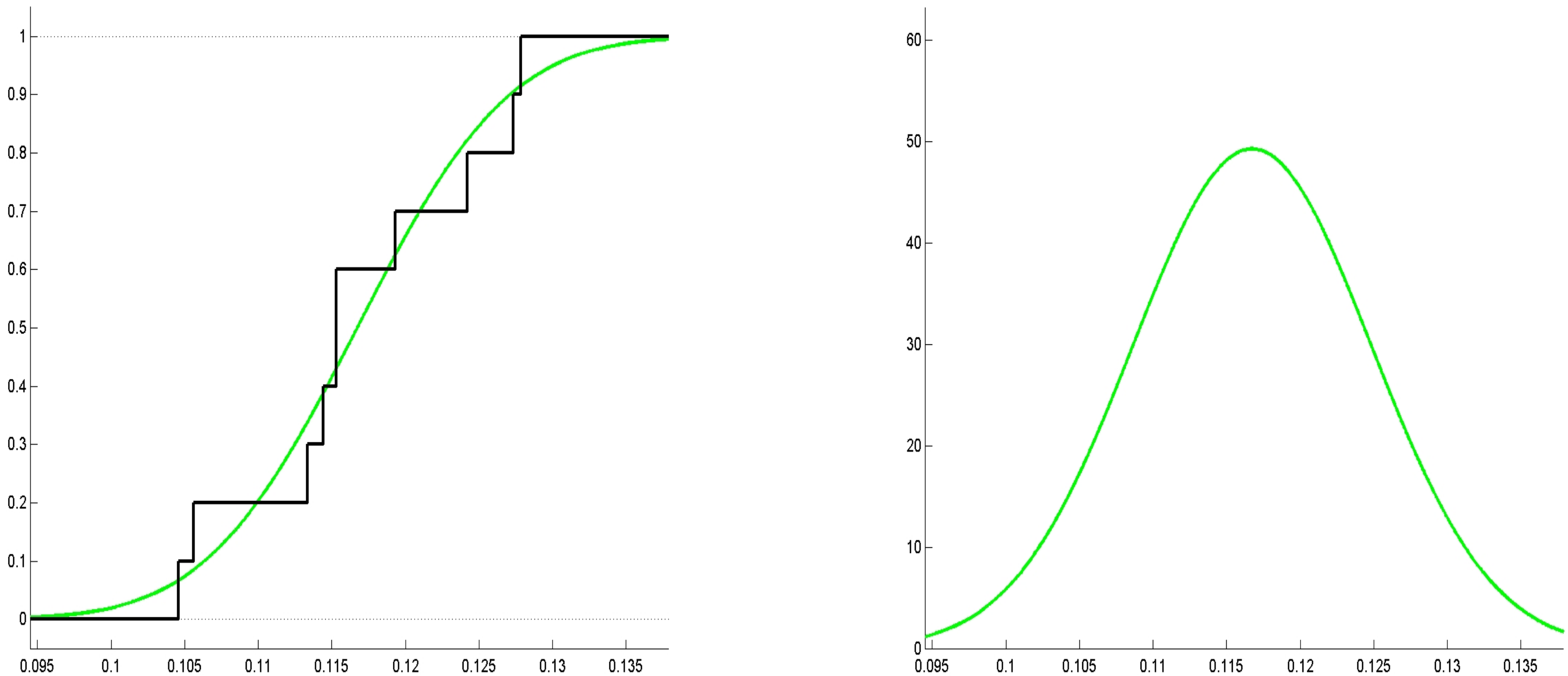
via Monte–Carlo simulations.

Variant 1: Resampling

Variant 2: Normally distributed quantities with estimated moments

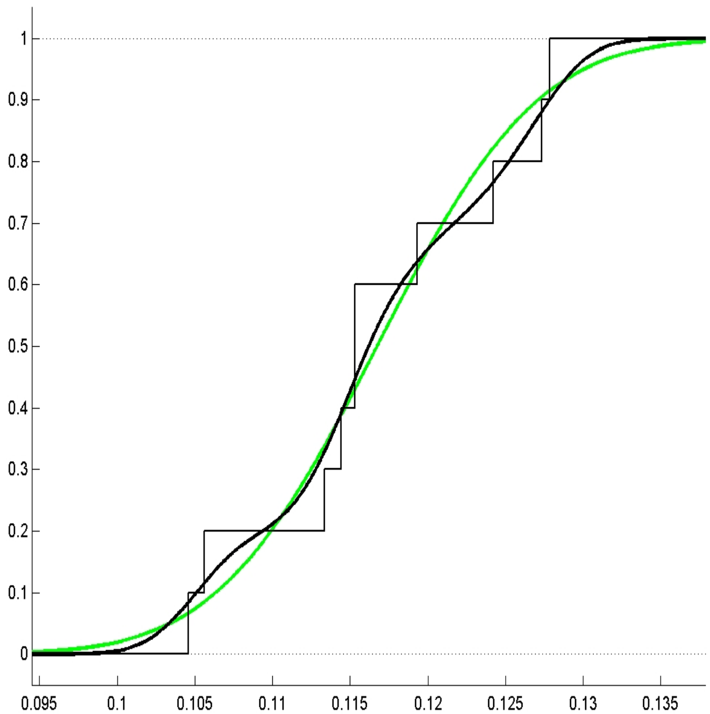
Variant 3: Resampling plus centered gaussian noise

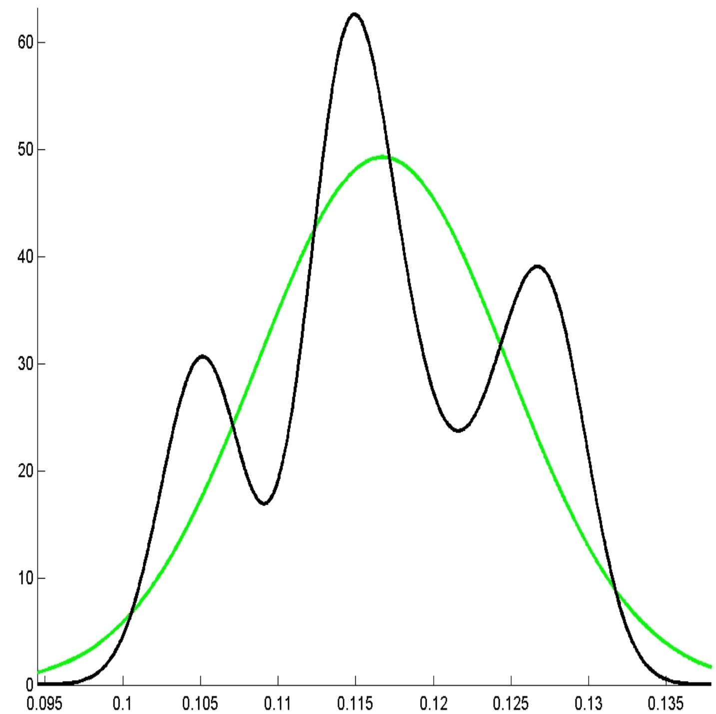
μ_{Ga} $(n = 10)$



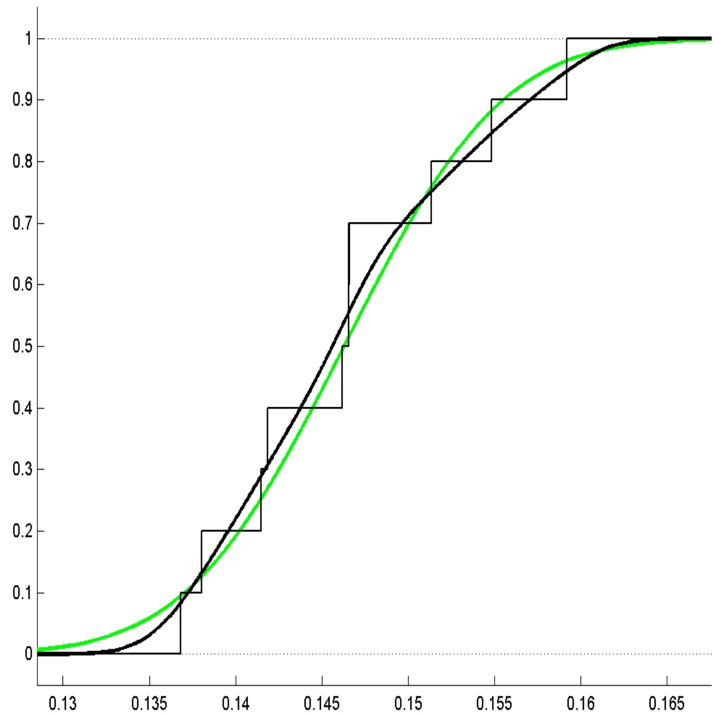
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.095
0.1
0.105
0.11
0.115
0.12
0.125
0.13
0.135
60
50
40
30
20
10
0
0.095
0.1
0.105
0.11
0.115
0.12
0.125
0.13
0.135

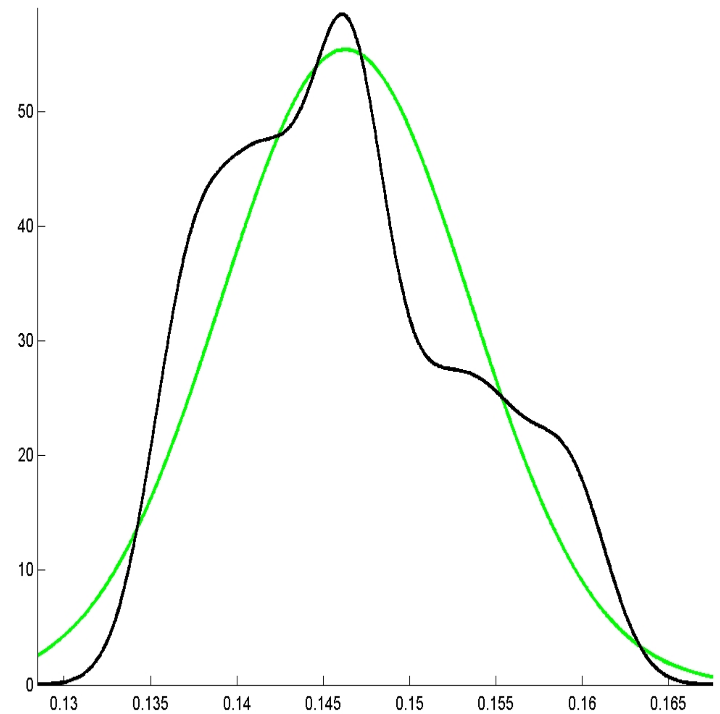
μ_{Ga} $(n = 10)$


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.095
0.1
0.105
0.11
0.115
0.12
0.125
0.13
0.135

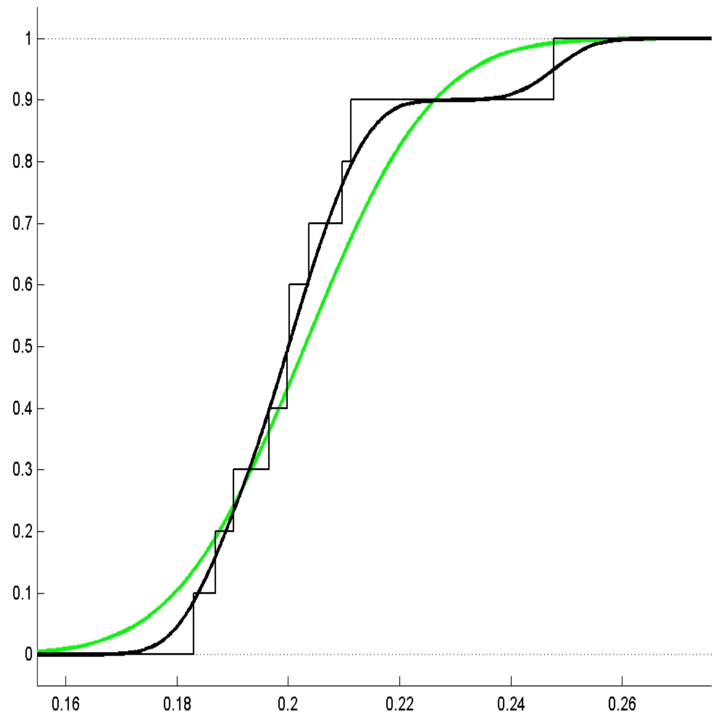

60
50
40
30
20
10
0
0.095
0.1
0.105
0.11
0.115
0.12
0.125
0.13
0.135

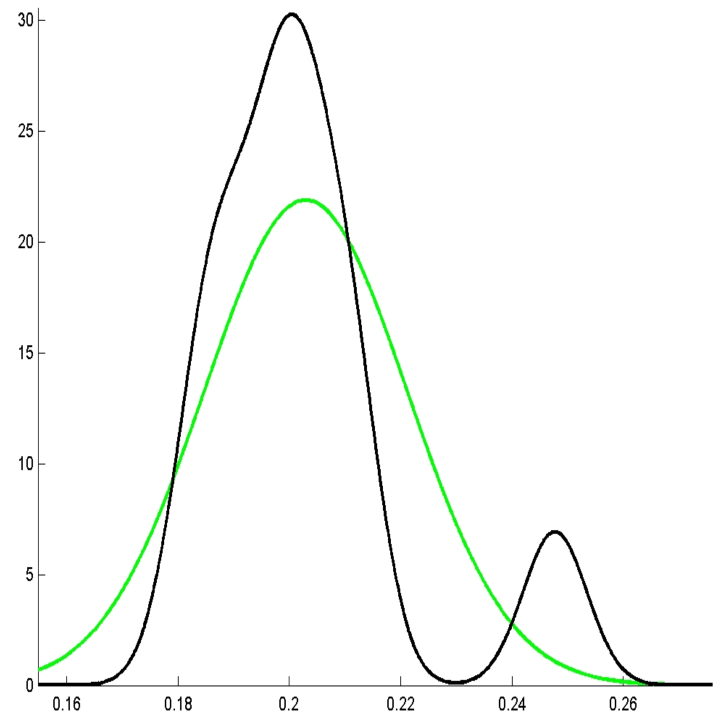
$\boldsymbol{\mu_{Ka}}$ $(n = 10)$



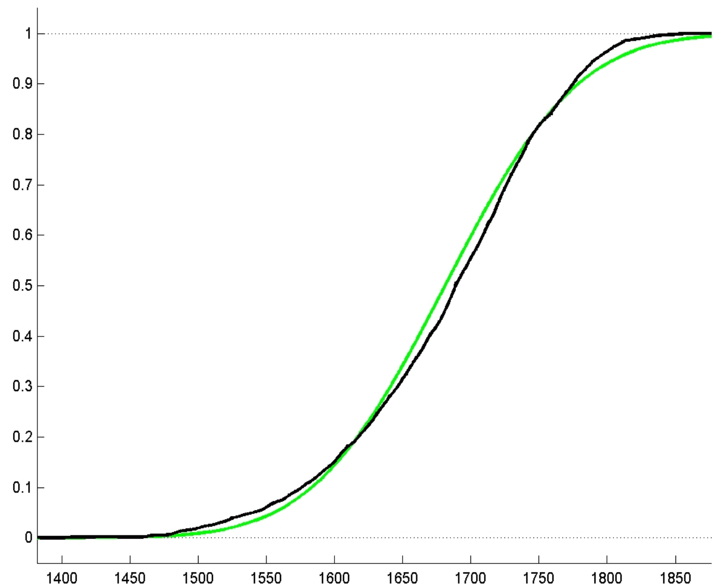


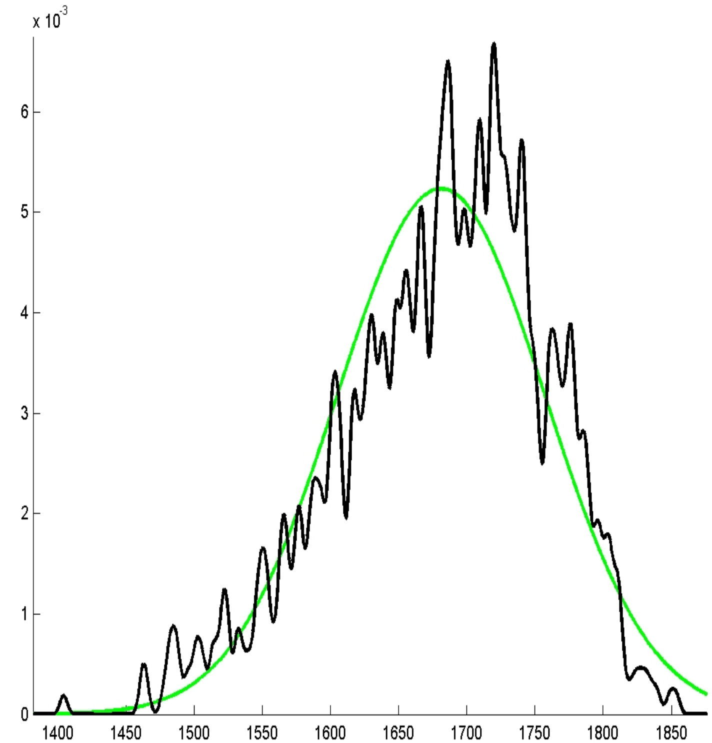
$\boldsymbol{\mu_{Ki}}$ $(n = 10)$



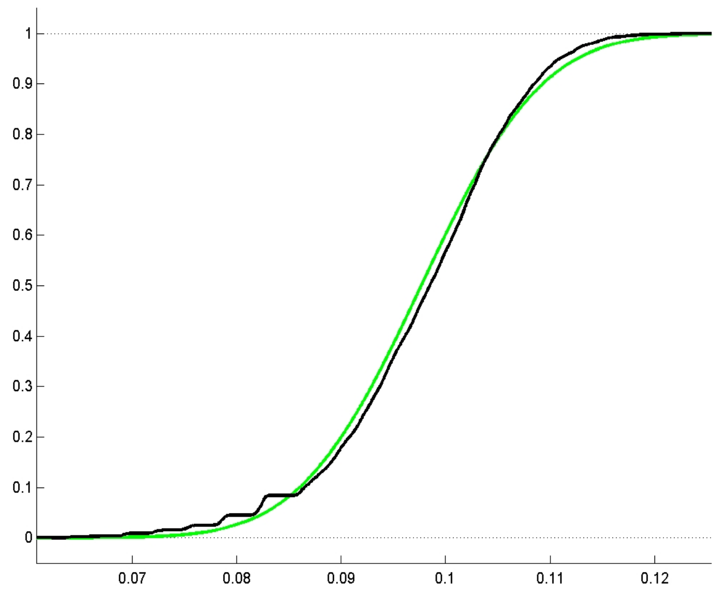


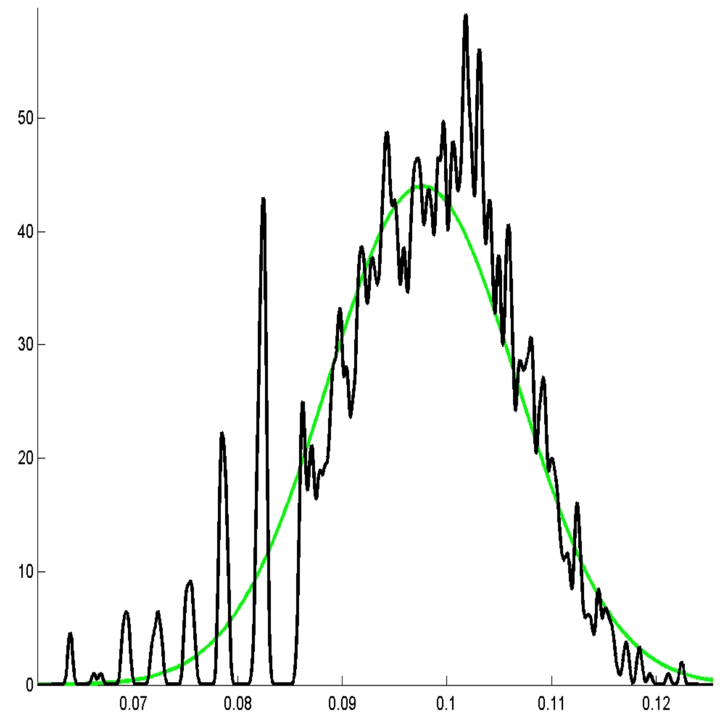
$\boldsymbol{\mu_R}$ $(n = 787)$





$\mu_S \quad (n = 2000)$



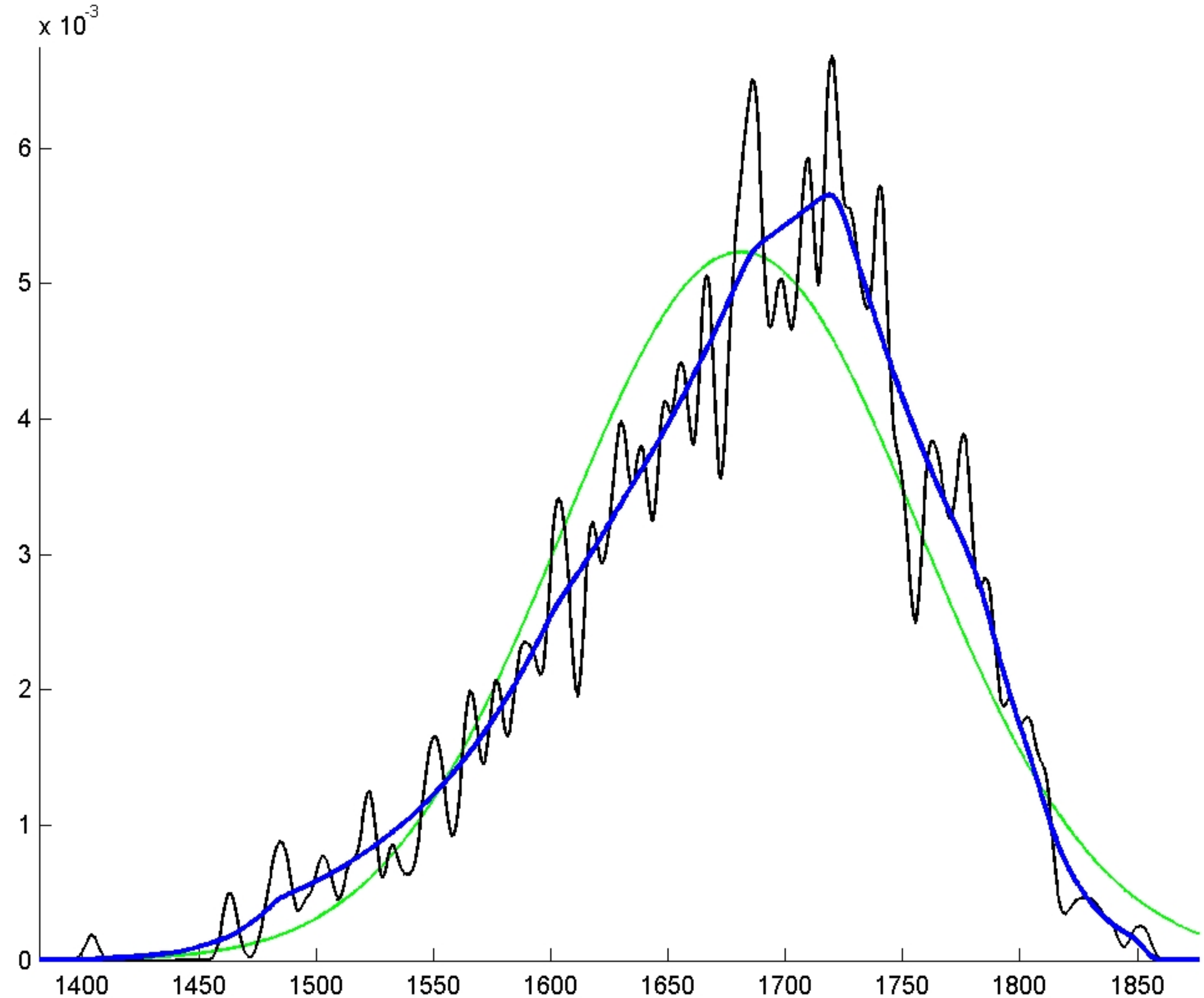


Goal

Fit a unimodal distribution to given data points which is

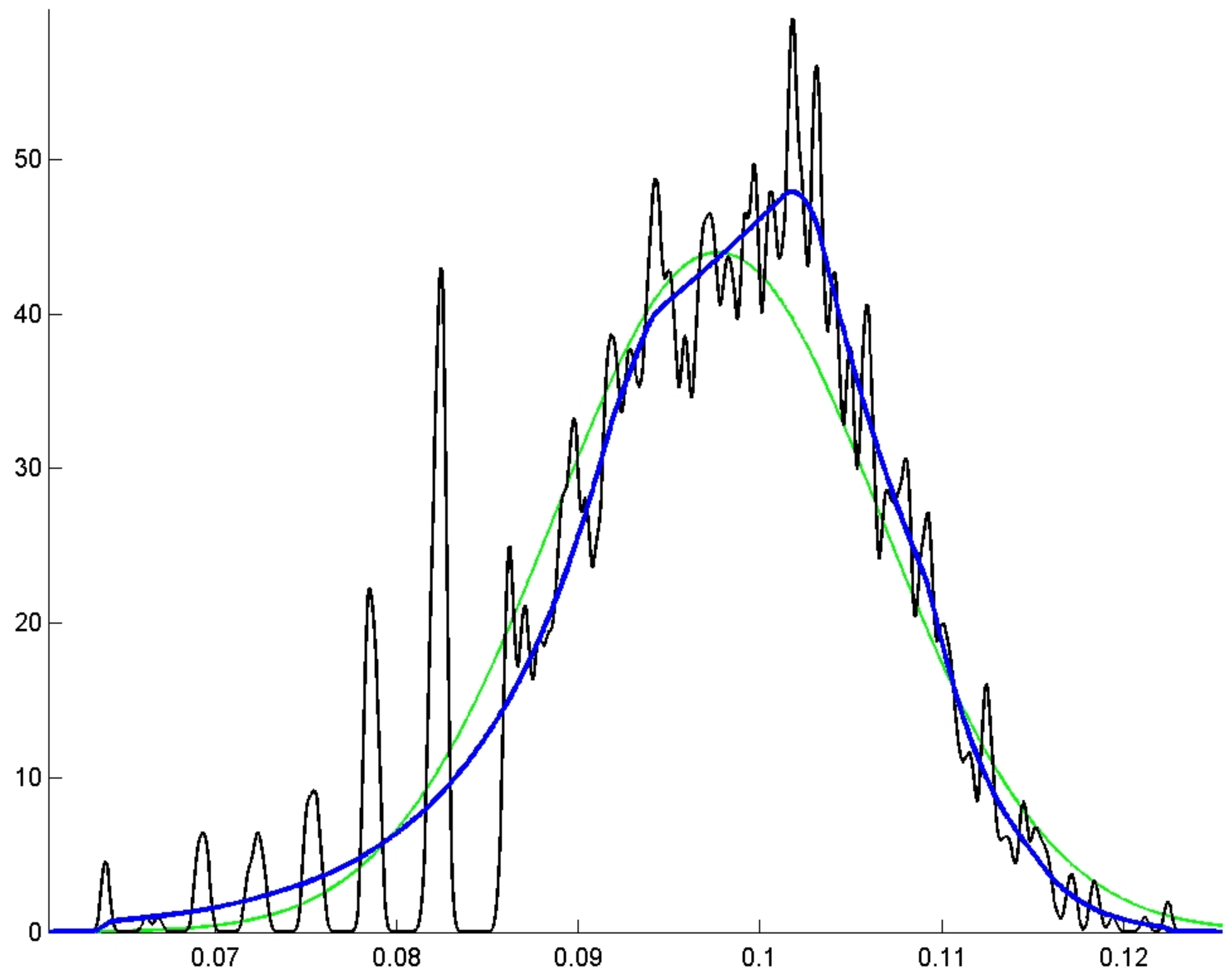
"as close as possible" to their empirical distribution.

μ_R $(n = 787)$



x 10^-3
6
5
4
3
2
1
0
1400
1450
1500
1550
1600
1650
1700
1750
1800
1850

$\boldsymbol{\mu_S}$ $(n = 2000)$



II. Log–Concave Densities

Probability density f on $\mathbb{R}^d$ is log–concave if

$$\boldsymbol{f} \;=\; \mathbf{exp}(\boldsymbol{\psi}) \quad \text{with} \quad \boldsymbol{\psi} : \mathbb{R}^{\boldsymbol{d}} \to [-\infty, \infty) \text{ concave} .$$

- Many standard distributions fulfill this constraint:

$$\mathcal{N}_d(\mu, \Sigma)$$

$$\begin{array}{rl} \mathsf{Gamma}(a,b) & (a \geq \mathbf{1}, b > 0) \\ \mathsf{Weibull}(a,b) & (a \geq \mathbf{1}, b > 0) \\ \mathsf{Beta}(a,b) & (a \geq \mathbf{1}, b \geq \mathbf{1}) \\ \text{Gumbel distr.} & \\ \text{logistic distr.} & \\ \ldots & \end{array}$$

- Log–concave densities are unimodal, i.e.

$$\left\{x \in \mathbb{R}^d : f(x) \geq r\right\} \text{ convex for all } r \geq 0 .$$

- NPMLE is well–defined (s. later)

Further facts about log–concavity and unimodality

- **Prékopa (1971, 1973)**

Let $P(dx) = f(x)dx$. For convex sets $A, B \subset \mathbb{R}^d$ and $\lambda \in [0, 1]$,

$$P((1-\lambda)A + \lambda B) \;\geq\; P(A)^{1-\lambda} P(B)^{\lambda}\,.$$

$$f, g \;\; \text{log–concave} \quad \Longrightarrow \quad f * g \;\; \text{log–concave}\,.$$

- **Ibragimov (1956, $d = 1$)**

$$\text{i.g.} \quad f, g \;\; \text{unimodal} \;\; \not\Longrightarrow \;\; f * g \;\; \text{unimodal}$$

$$f \;\; \text{log–concave},\; g \;\; \text{unimodal} \;\; \Longrightarrow \;\; f * g \;\; \text{unimodal}$$

III. ML Estimation $(d = 1)$

Random sample $X_1 < X_2 < \cdots < X_n$ from unknown log–concave density $f = \exp(\psi)$ with c.d.f. F.

$$\widehat{\psi} \;:=\; \operatorname*{arg\,max}_{\psi \text{ concave}, \int \exp(\psi)=1} \sum_{i=1}^{n} \psi(X_i)$$

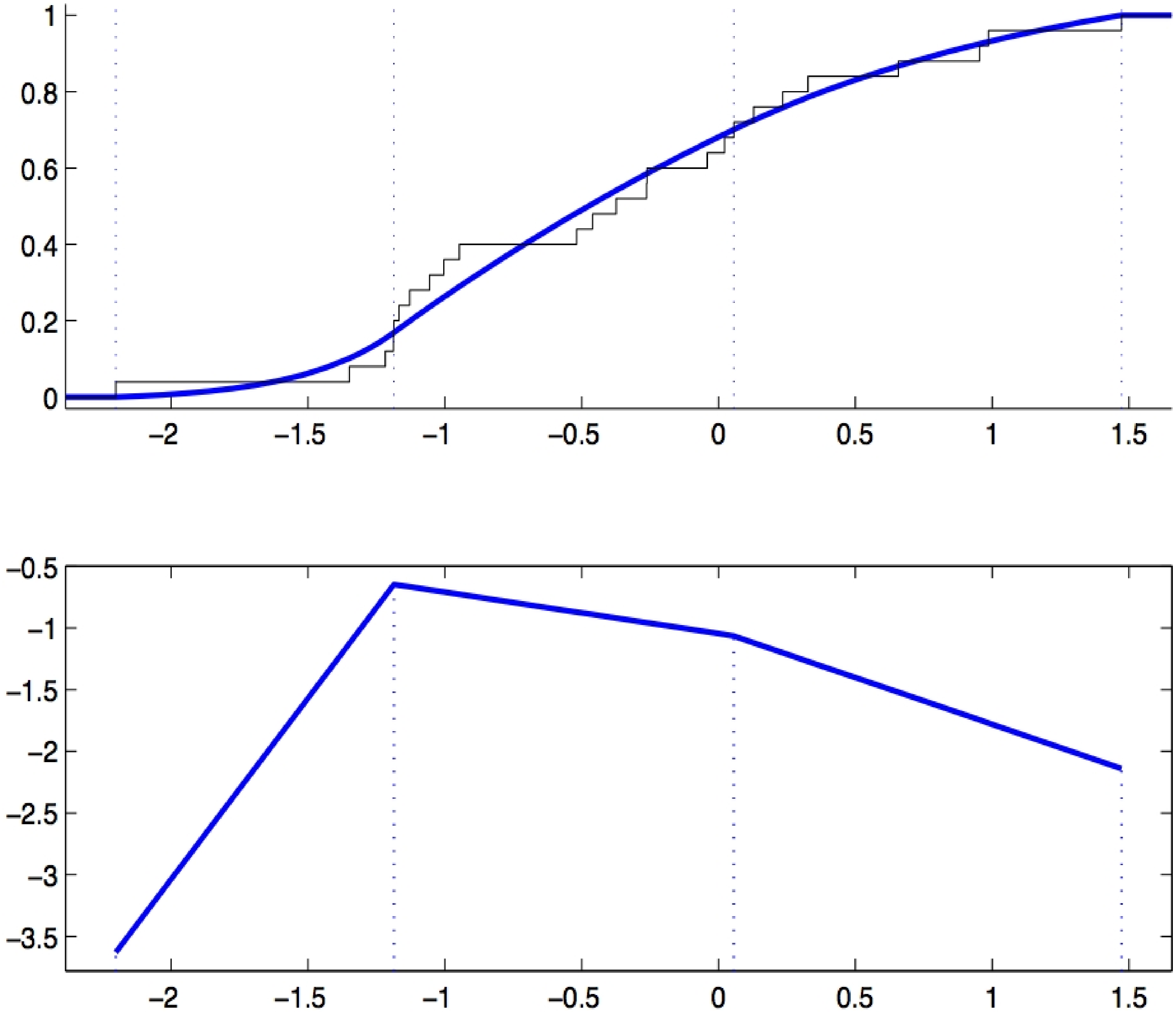
$$= \operatorname*{arg\,max}_{\psi \text{ concave}} \left(\underbrace{\int \psi \, d\widehat{F}_{\text{emp}}}_{\text{Log-Likelihood}} - \underbrace{\int \exp(\psi(x)) \, dx}_{\text{Lagrange-Term}} \right)$$

with empirical c.d.f.

$$\widehat{F}_{\text{emp}}(r) \;:=\; \frac{1}{n} \sum_{i=1}^{n} 1\{X_i \le r\}.$$

Theorem 1 (Existence, uniqueness, type)

- $\widehat{\psi}$ exists and is unique,
- $\widehat{\psi}$ is continuous and piecewise linear on $[X_1, X_n]$ with knots in $\{X_1, X_2, \ldots, X_n\}$,
- $\widehat{f} = \exp(\widehat{\psi}) \equiv 0$ on $\mathbb{R} \setminus [X_1, X_n]$.



1
0.8
0.6
0.4
0.2
0
−2
−1.5
−1
−0.5
0
0.5
1
1.5
−0.5
−1
−1.5
−2
−2.5
−3
−3.5
−2
−1.5
−1
−0.5
0
0.5
1
1.5

Characterizations and further properties of the estimators

$$L(\psi) \ := \ \int \psi \, d\widehat{F}_{\mathrm{emp}} - \int \exp(\psi(x)) \, dx$$

$$\frac{d}{dt}\Big|_{t=0} L\left(\widehat{\psi} + t\Delta\right) \ \leq \ 0 \quad \text{if } \widehat{\psi} + t\Delta \text{ concave for some } t > 0$$

Theorem 2 (Characterisation of $\widehat{f}$)

$$\int \boldsymbol{\Delta}\, d\widehat{\boldsymbol{F}}_{\mathrm{emp}} \;\le\; \int \boldsymbol{\Delta}(\boldsymbol{x})\,\widehat{\boldsymbol{f}}(\boldsymbol{x})\, d\boldsymbol{x}$$

whenever $\widehat{\psi} + \boldsymbol{t\Delta}$ concave for some $\boldsymbol{t} > \mathbf{0}$.

Corollary 1 The c.d.f. $\widehat{F}$ of $\widehat{f}$ satisfies

$$\mathbf{Mean}(\widehat{\boldsymbol{F}}) \;=\; \mathbf{Mean}(\widehat{\boldsymbol{F}}_{\mathbf{emp}}) \qquad (\Delta(x) = \pm x)$$

$$\mathbf{Var}(\widehat{\boldsymbol{F}}) \;\le\; \mathbf{Var}(\widehat{\boldsymbol{F}}_{\mathbf{emp}})\,. \qquad (\Delta(x) = -x^2)$$

Set of knots of $\widehat{\psi}$:

$$\widehat{\mathcal{S}} := \left\{ x : \widehat{\psi}'(x-) > \widehat{\psi}'(x+) \right\} \quad \begin{cases} \supset \{X_1, X_n\} \\ \subset \{X_1, X_2, \ldots, X_n\} \end{cases}$$

Corollary 2

$$\int \Delta \, d\widehat{F}_{\mathrm{emp}} = \int \Delta(x) \widehat{f}(x) \, dx$$

if Δ is continuous and piecewise linear with knots in $\widehat{\mathcal{S}}$.

Theorem 3 (Characterisation of $\widehat{F}$)

For $a < t < b$ with $a, b \in \widehat{\mathcal{S}}$,

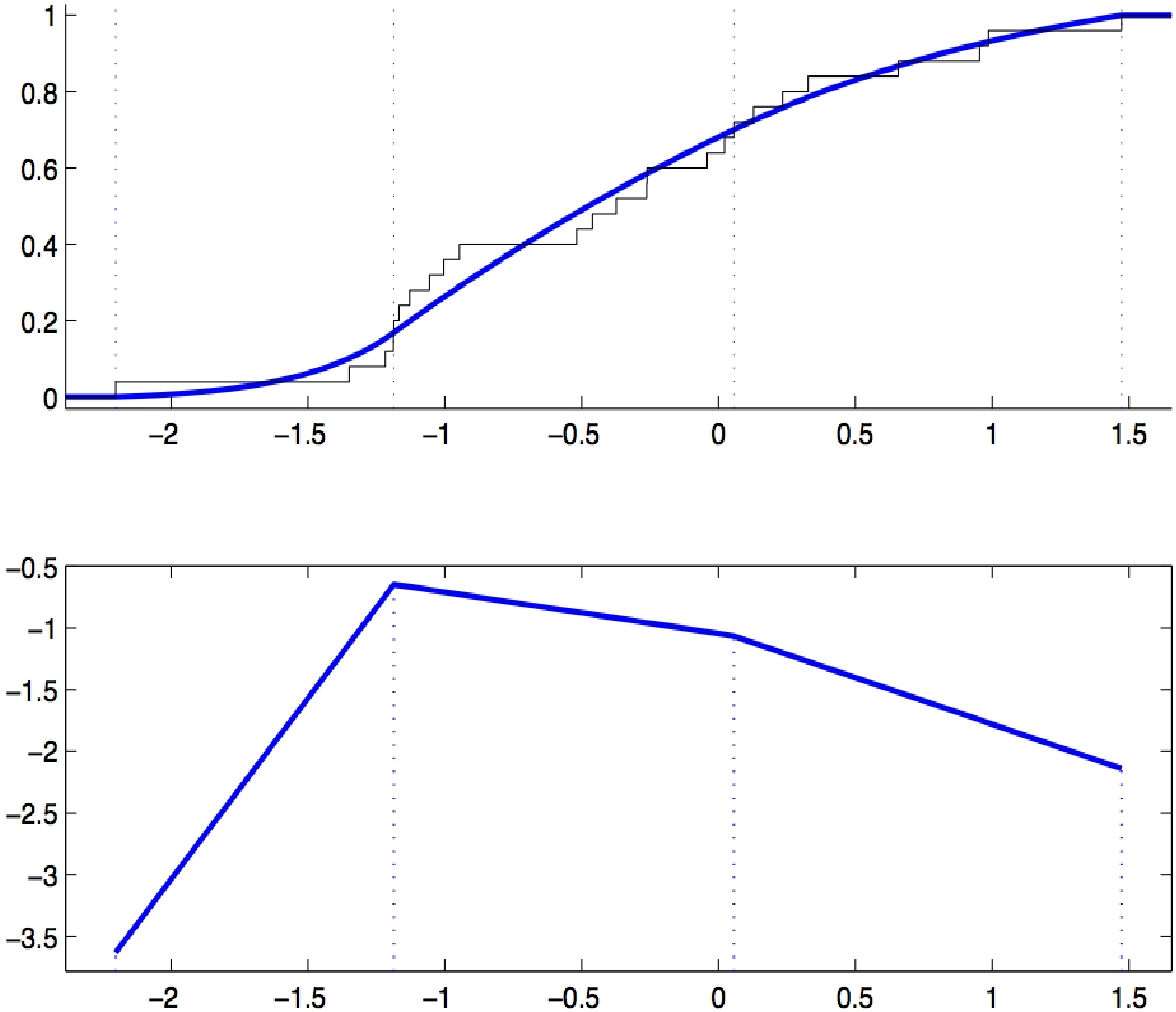
$$\int_a^t \widehat{F}_{\text{emp}}(x)\,dx \;\geq\; \int_a^t \widehat{F}(x)\,dx\,,$$

$$\int_t^b \widehat{F}_{\text{emp}}(x)\,dx \;\leq\; \int_t^b \widehat{F}(x)\,dx\,,$$

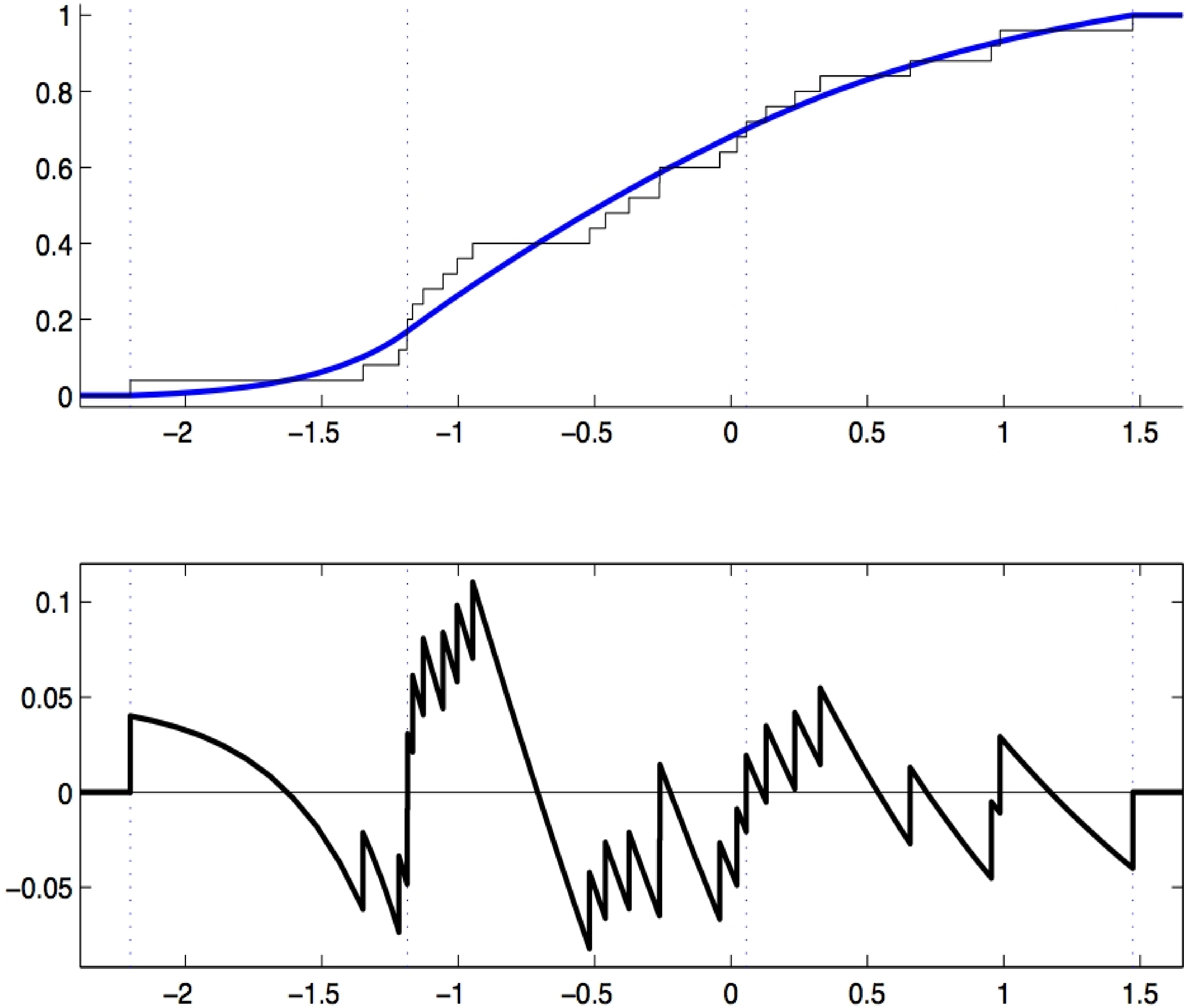
$$\int_a^b \widehat{F}_{\text{emp}}(x)\,dx \;=\; \int_a^b \widehat{F}(x)\,dx\,.$$

Corollary 2 $\widehat{F}(X_1) = 0$, $\widehat{F}(X_n) = 1$ and

$$\widehat{F}_{\text{emp}} - \frac{1}{n} \leq \widehat{F} \leq \widehat{F}_{\text{emp}} \quad \text{on } \widehat{\mathcal{S}}\,.$$



1
0.8
0.6
0.4
0.2
0
−2
−1.5
−1
−0.5
0
0.5
1
1.5
−0.5
−1
−1.5
−2
−2.5
−3
−3.5
−2
−1.5
−1
−0.5
0
0.5
1
1.5



1
0.8
0.6
0.4
0.2
0
−2
−1.5
−1
−0.5
0
0.5
1
1.5
0.1
0.05
0
−0.05
−2
−1.5
−1
−0.5
0
0.5
1
1.5

Consistency of the estimators

Theorem 4 (Consistency of $\widehat{F}$)

$$\left\|\widehat{F} - F\right\|_\infty = O_p\left(\left(\frac{\log n}{n}\right)^{1/2}\right)$$

Conjecture:

$$\mathbb{P}\left(\left\|\widehat{F} - F\right\|_\infty \leq \left\|\widehat{F}_{\mathrm{emp}} - F\right\|_\infty\right) \rightarrow 1$$

But

$$\text{in gen.} \quad \left\|\widehat{F} - F\right\|_\infty \not\leq C \left\|\widehat{F}_{\mathrm{emp}} - F\right\|_\infty !$$

Theorem 5 (Consistency of $\widehat{\psi}$)

Let ψ be Hölder–continuous on $[a, b] \subset \{f > 0\}$ with exponent $\beta \in [1, 2]$, i.e. for some constant L,

$$\left|\psi'(x) - \psi'(y)\right| \leq L|x - y|^{\beta-1} \quad \text{for all } x, y \in [a, b] .$$

Then

$$\sup_{[a+\delta_n, b-\delta_n]} \left|\widehat{\psi} - \psi\right| = O_p\left(\left(\frac{\log n}{n}\right)^{\beta/(2\beta+1)}\right)$$

with $\delta_n = (\log(n)/n)^{1/(2\beta+1)} \to 0$.

Theorem 6 (Asympt. equivalence of $\widehat{F}$ and $\widehat{F}_{\text{emp}}$)

Let ψ be twice continuously differentiable on $[a, b] \subset \{f > 0\}$ with $\psi'' < 0$. Then

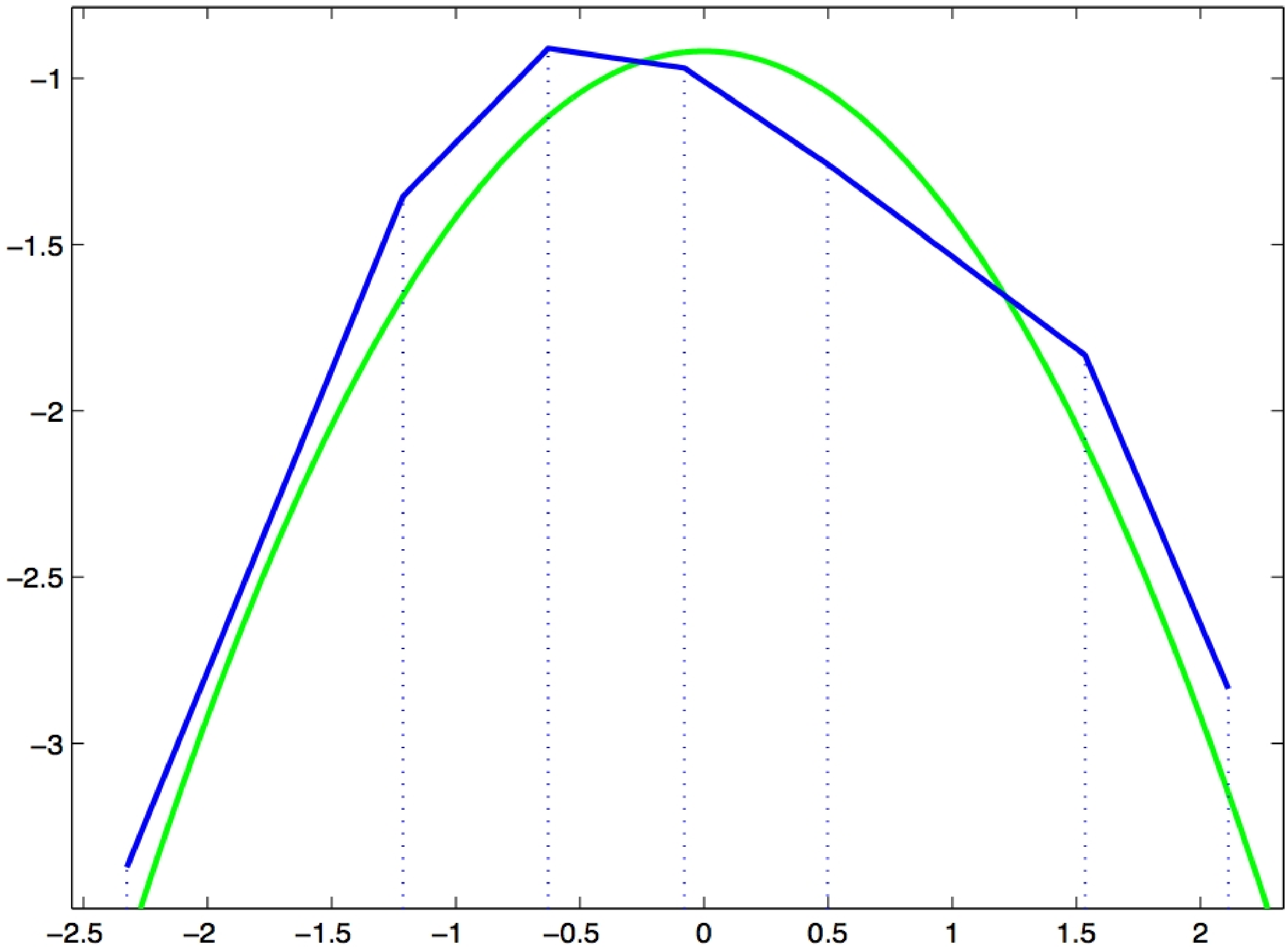
$$\sup_{[a+\delta_n, b-\delta_n]} \left| \widehat{F} - \widehat{F}_{\text{emp}} \right| = o_p\left(\frac{1}{\sqrt{n}}\right).$$

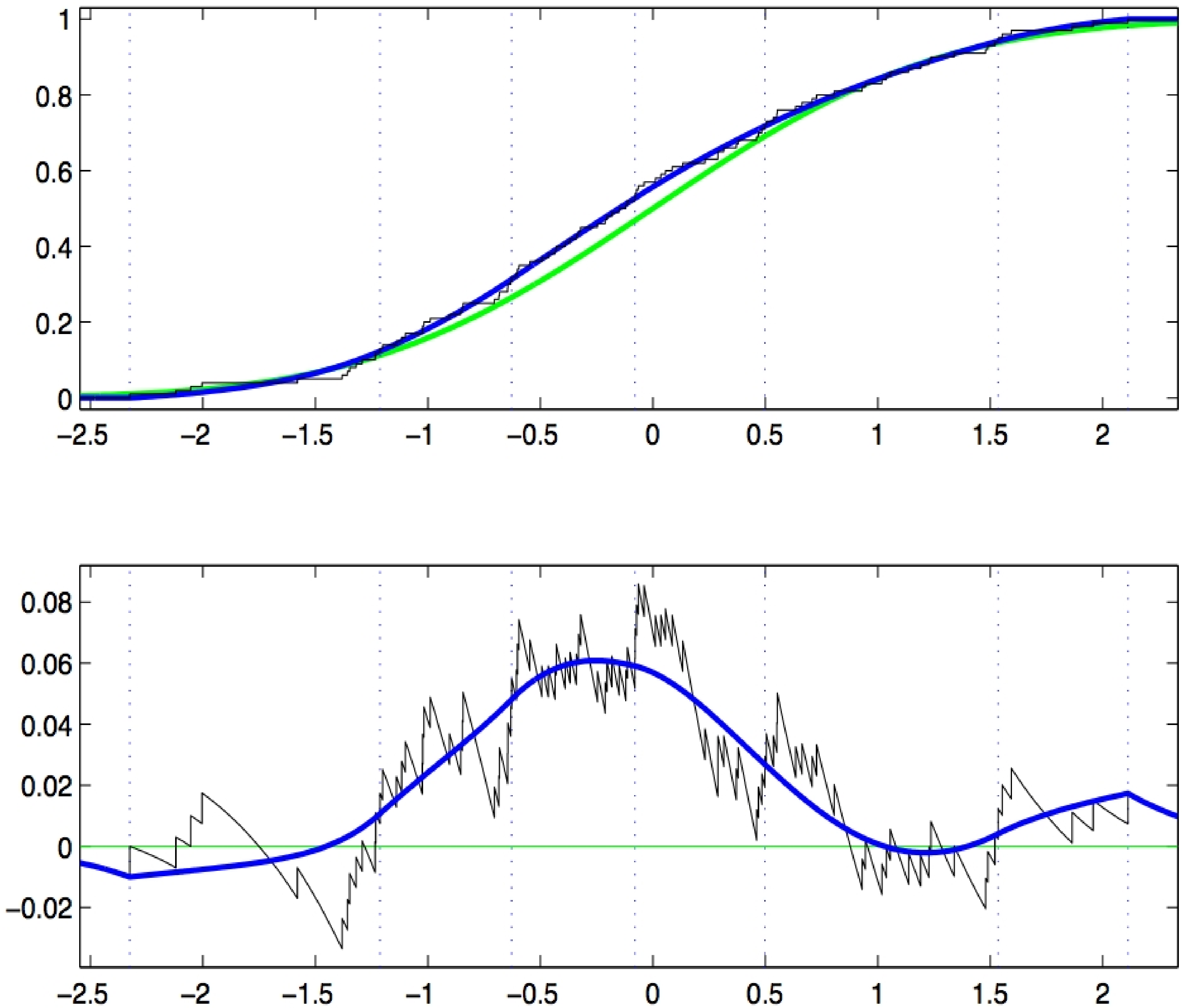
Numerical examples I

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

$$\phi(x) = -\frac{x^2}{2} + c$$

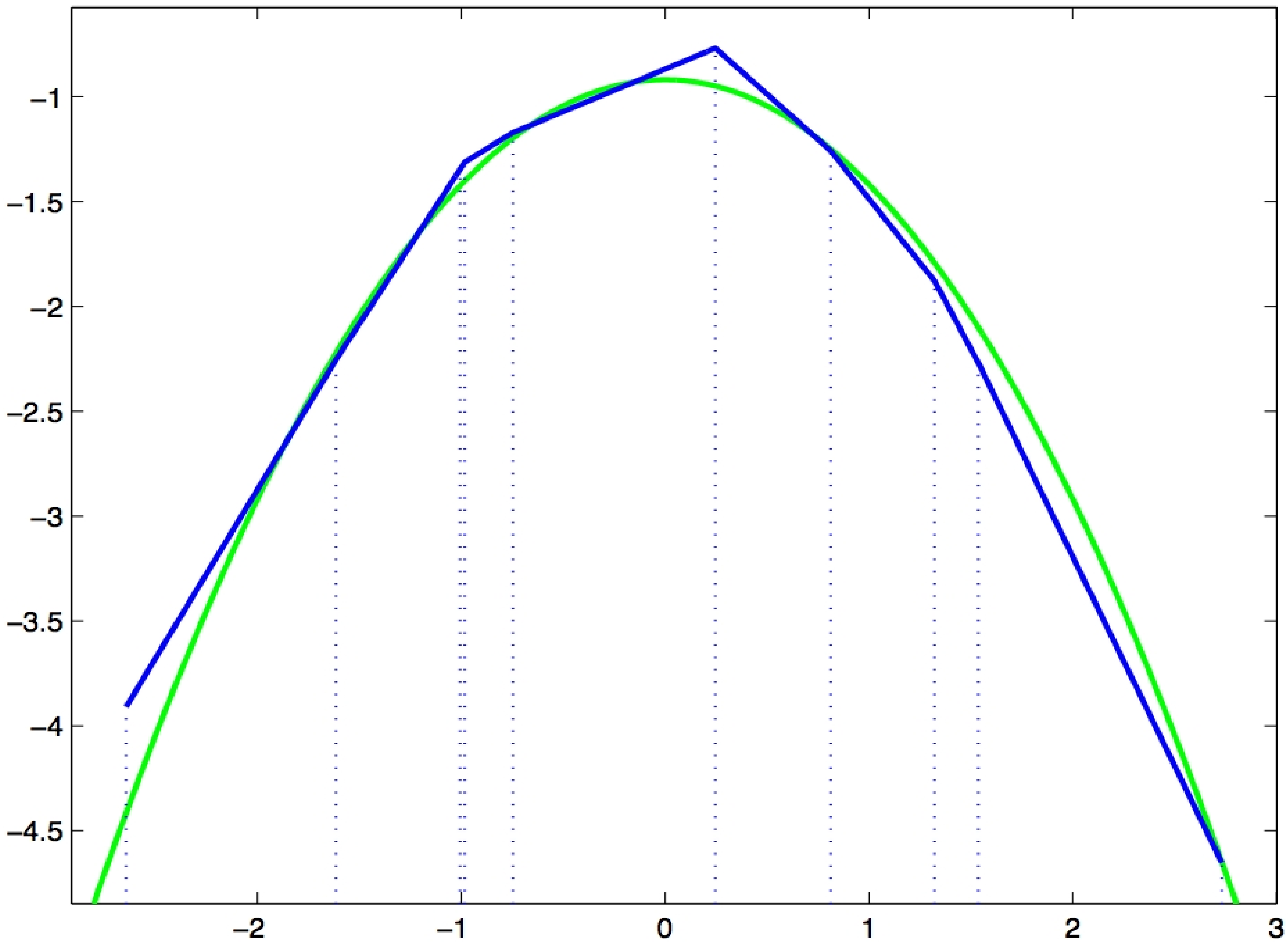
$n = 100$:

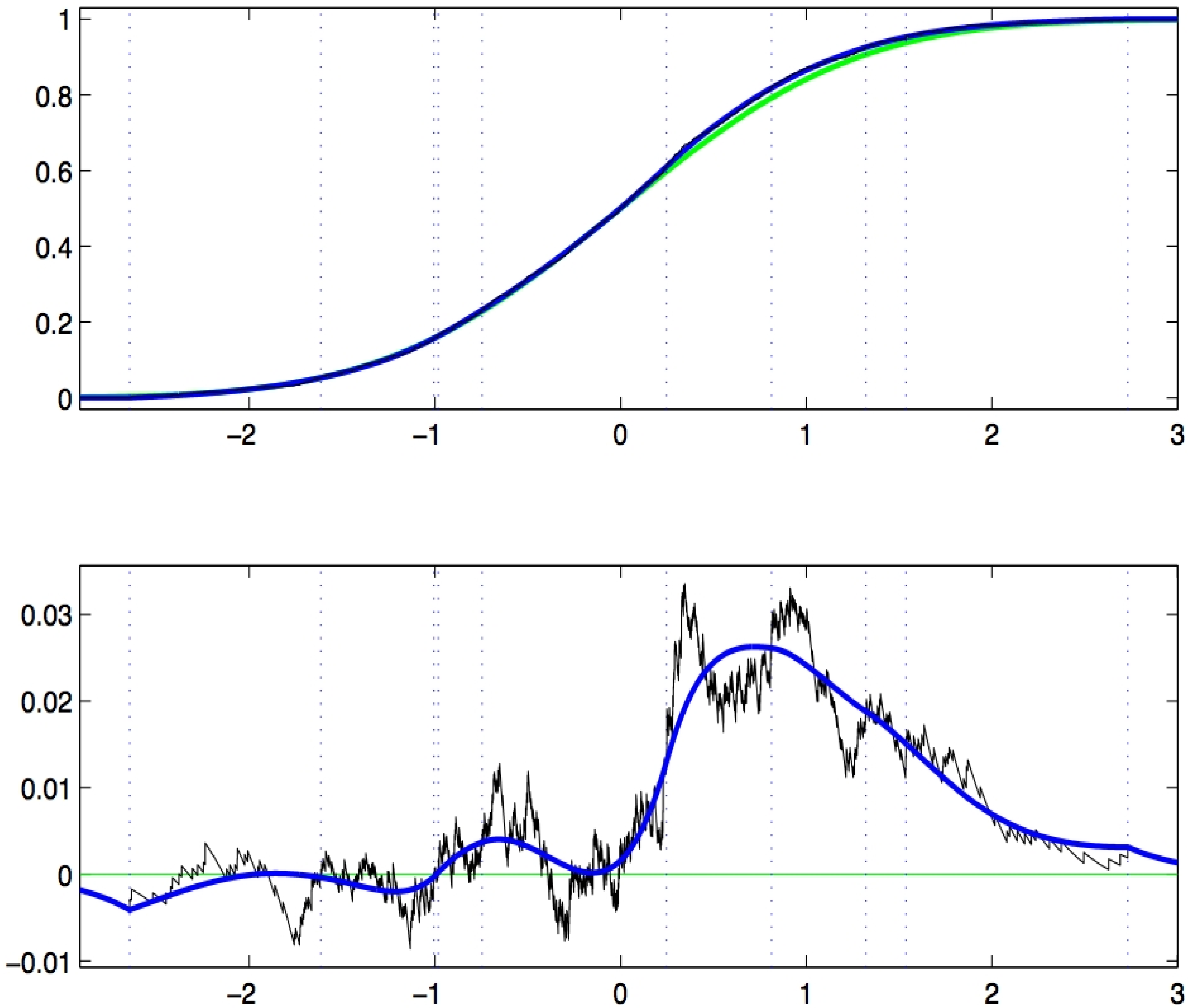




1
0.8
0.6
0.4
0.2
0
−2.5
−2
−1.5
−1
−0.5
0
0.5
1
1.5
2
0.08
0.06
0.04
0.02
0
−0.02
−2.5
−2
−1.5
−1
−0.5
0
0.5
1
1.5
2

$n = 800$:





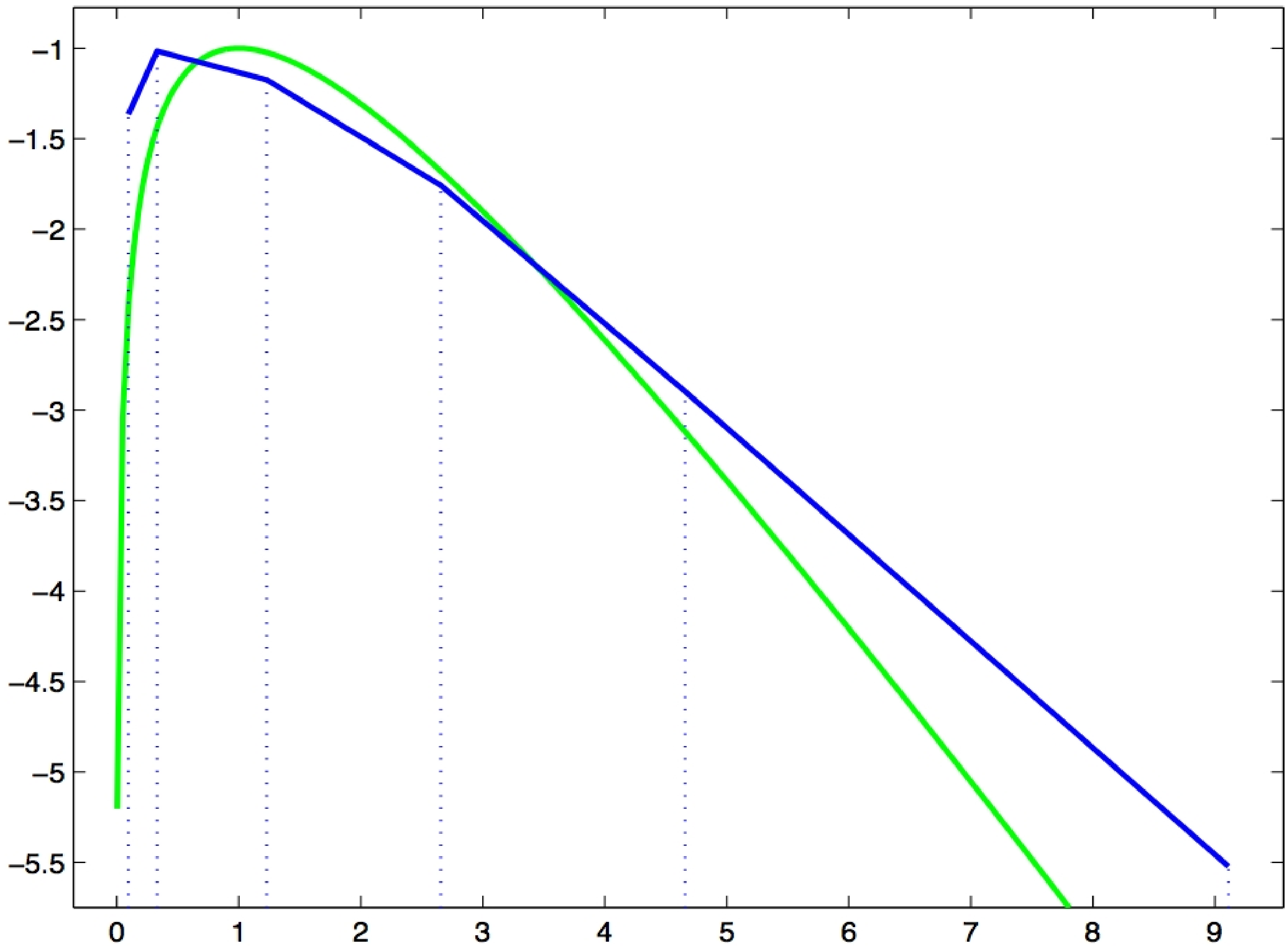
1
0.8
0.6
0.4
0.2
0
−2
−1
0
1
2
3
0.03
0.02
0.01
0
−0.01
−2
−1
0
1
2
3

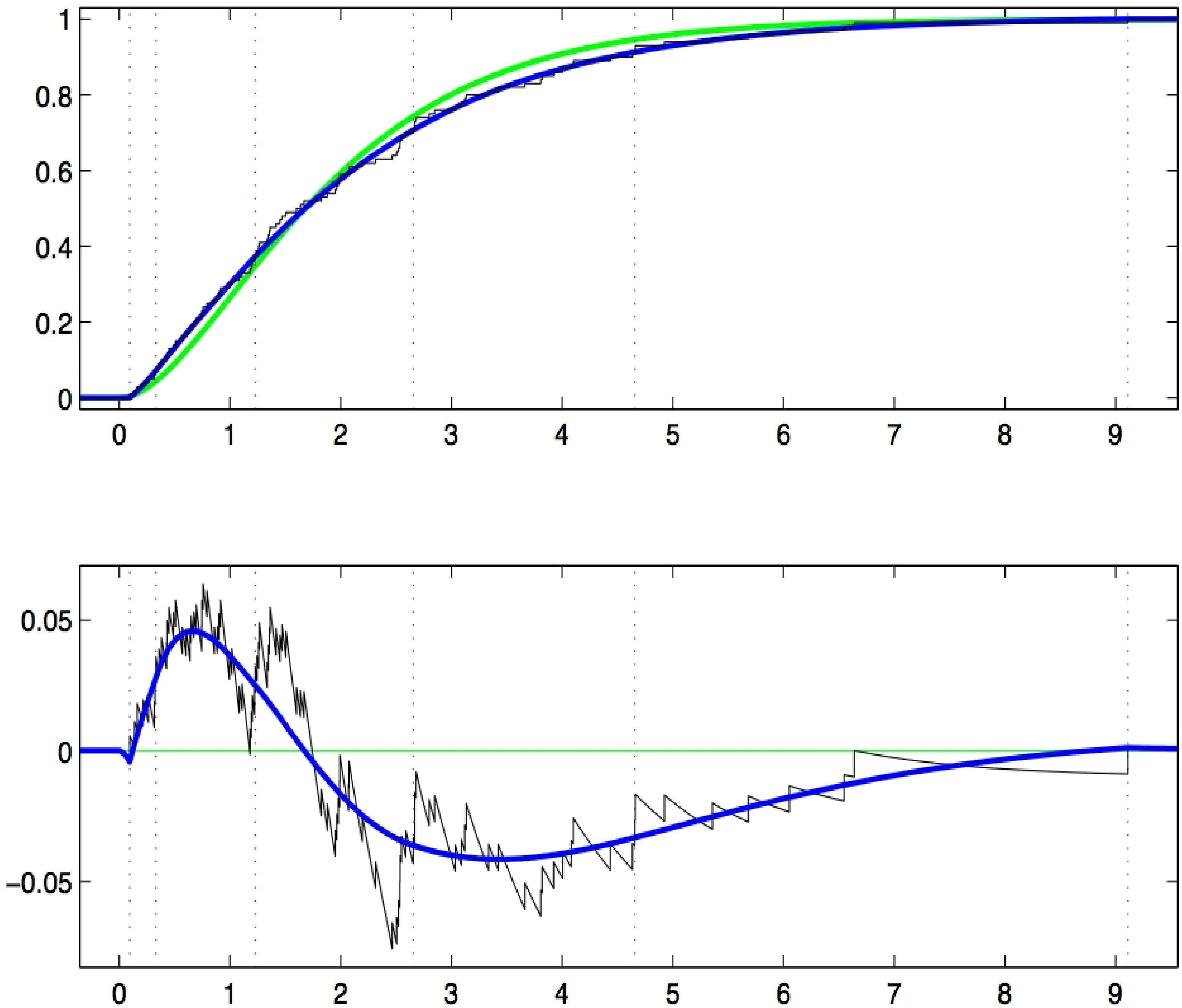
Numerical examples II

$$\boldsymbol{f(x) \ = \ \frac{x \exp(-x)}{\Gamma(2)}}$$

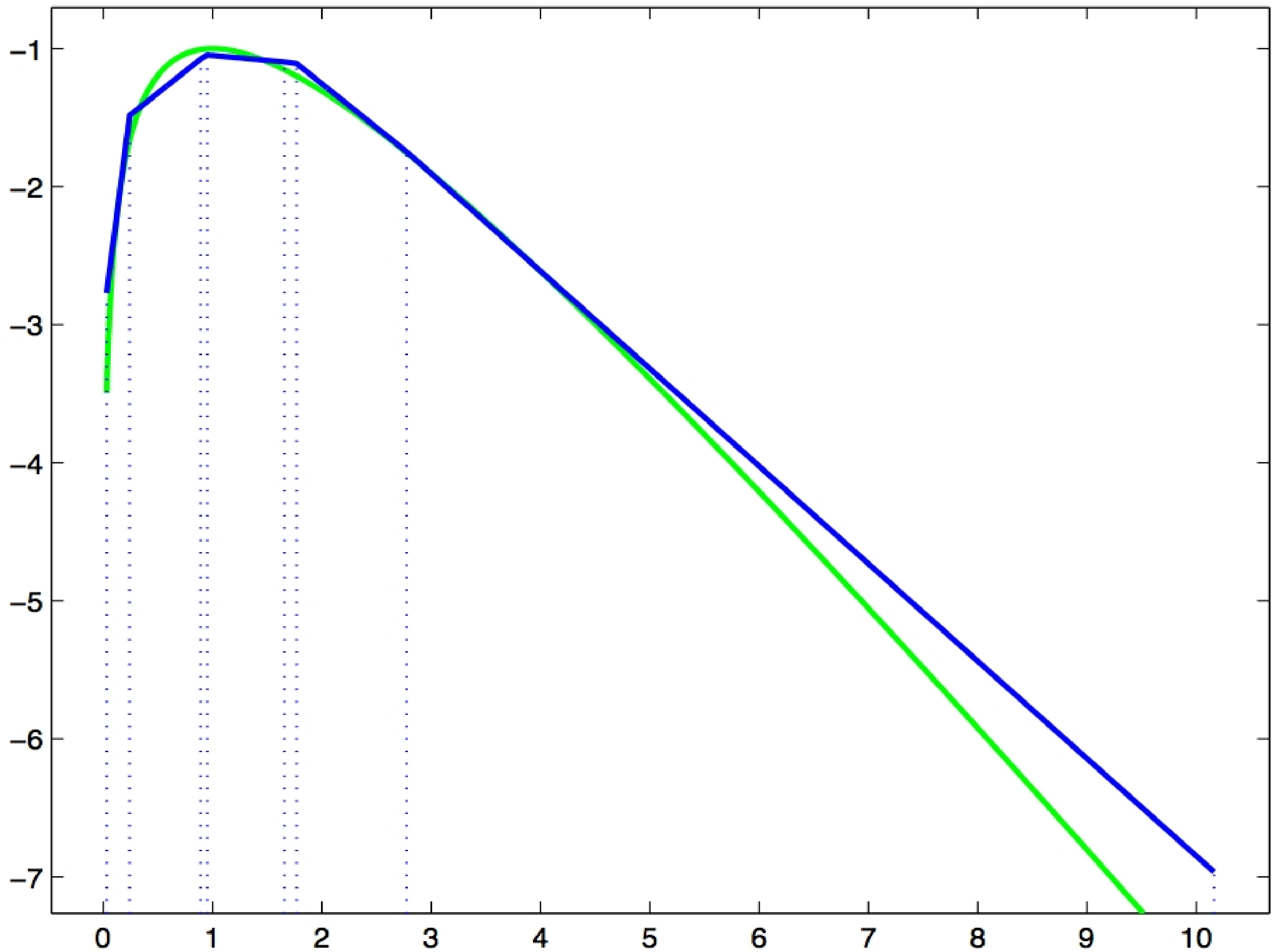
$$\boldsymbol{\phi(x) \ = \ \log(x) - x \ + c}$$

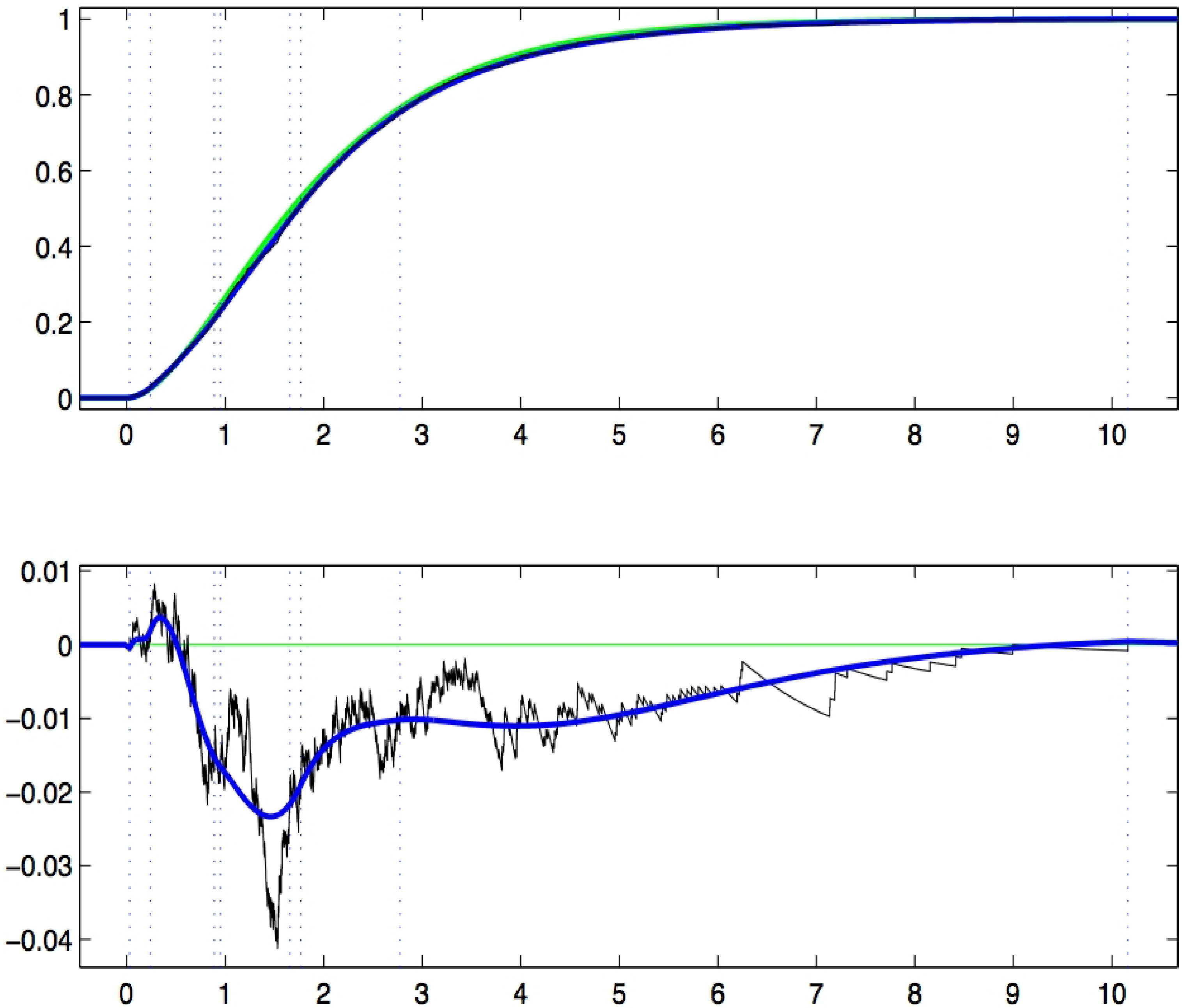
$n = 100$:




1
0.8
0.6
0.4
0.2
0
0 1 2 3 4 5 6 7 8 9
0.05
0
−0.05
0 1 2 3 4 5 6 7 8 9

$n = 800$:





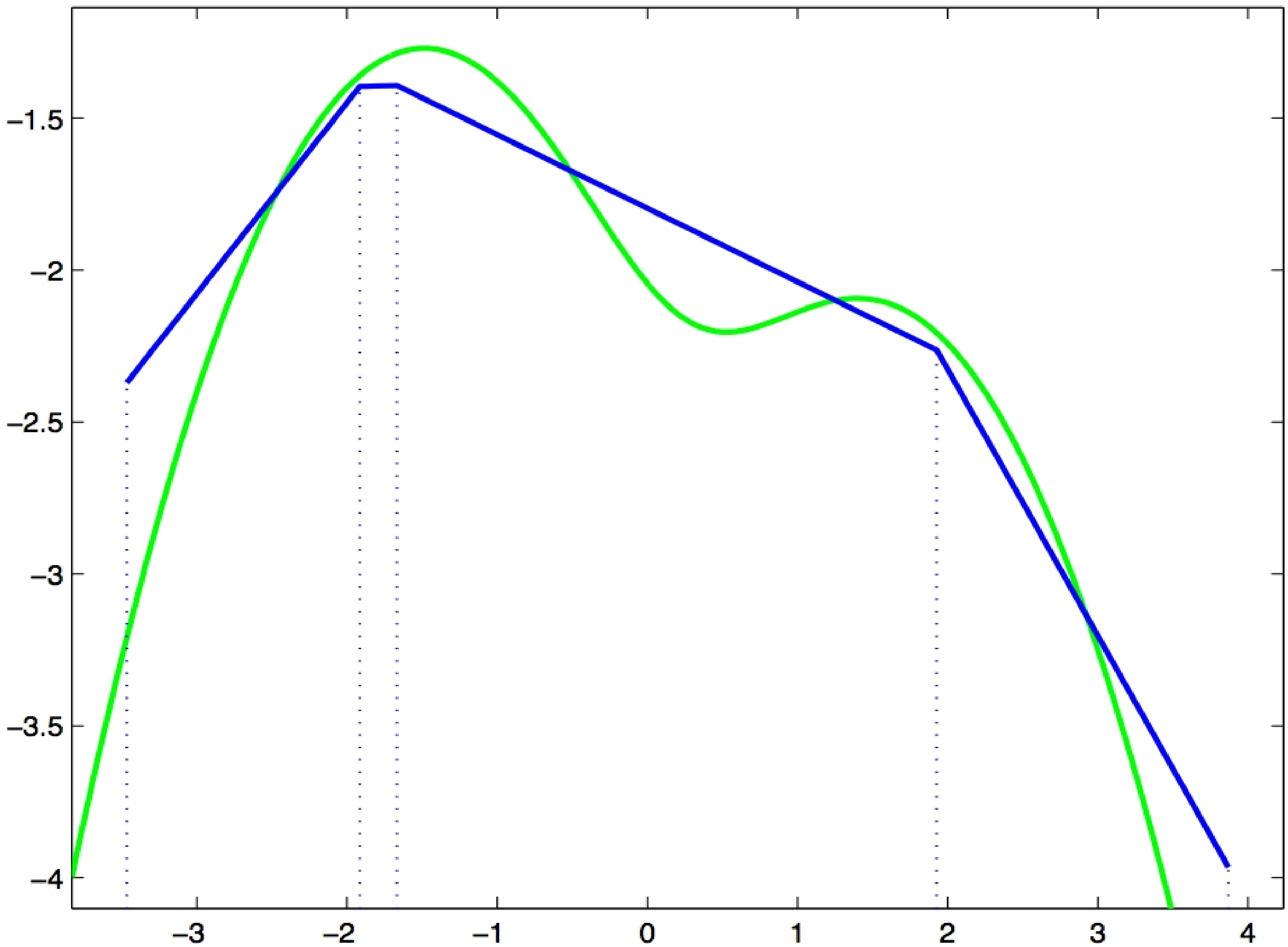
1
0.8
0.6
0.4
0.2
0
0
1
2
3
4
5
6
7
8
9
10
0.01
0
−0.01
−0.02
−0.03
−0.04
0
1
2
3
4
5
6
7
8
9
10

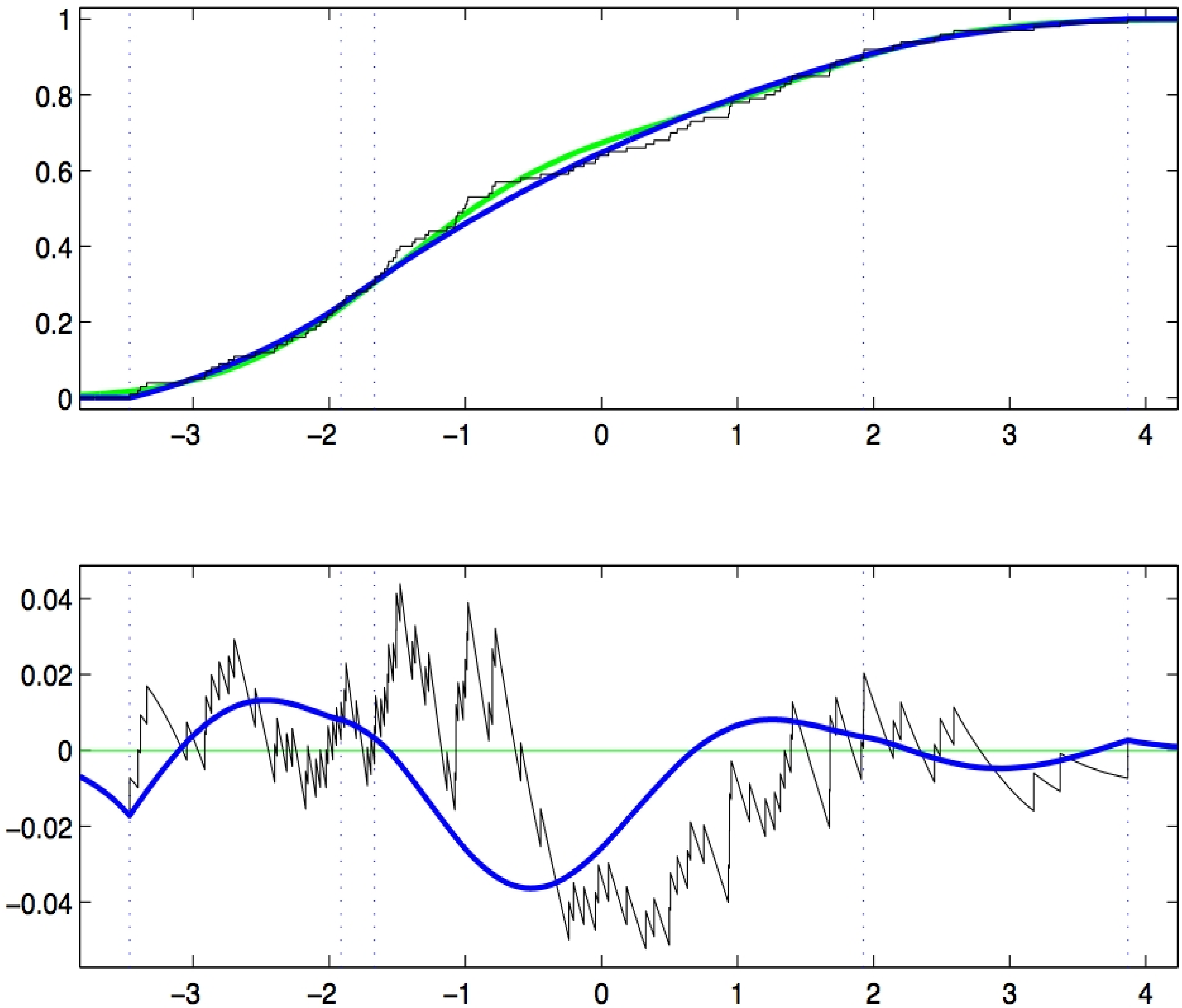
Numerical examples III

$$\boldsymbol{F} \;=\; \mathbf{0.7} \cdot \boldsymbol{\mathcal{N}}(\mathbf{-1.5, 1}) + \mathbf{0.3} \cdot \boldsymbol{\mathcal{N}}(\mathbf{1.5, 1})$$

(bimodal distribution)

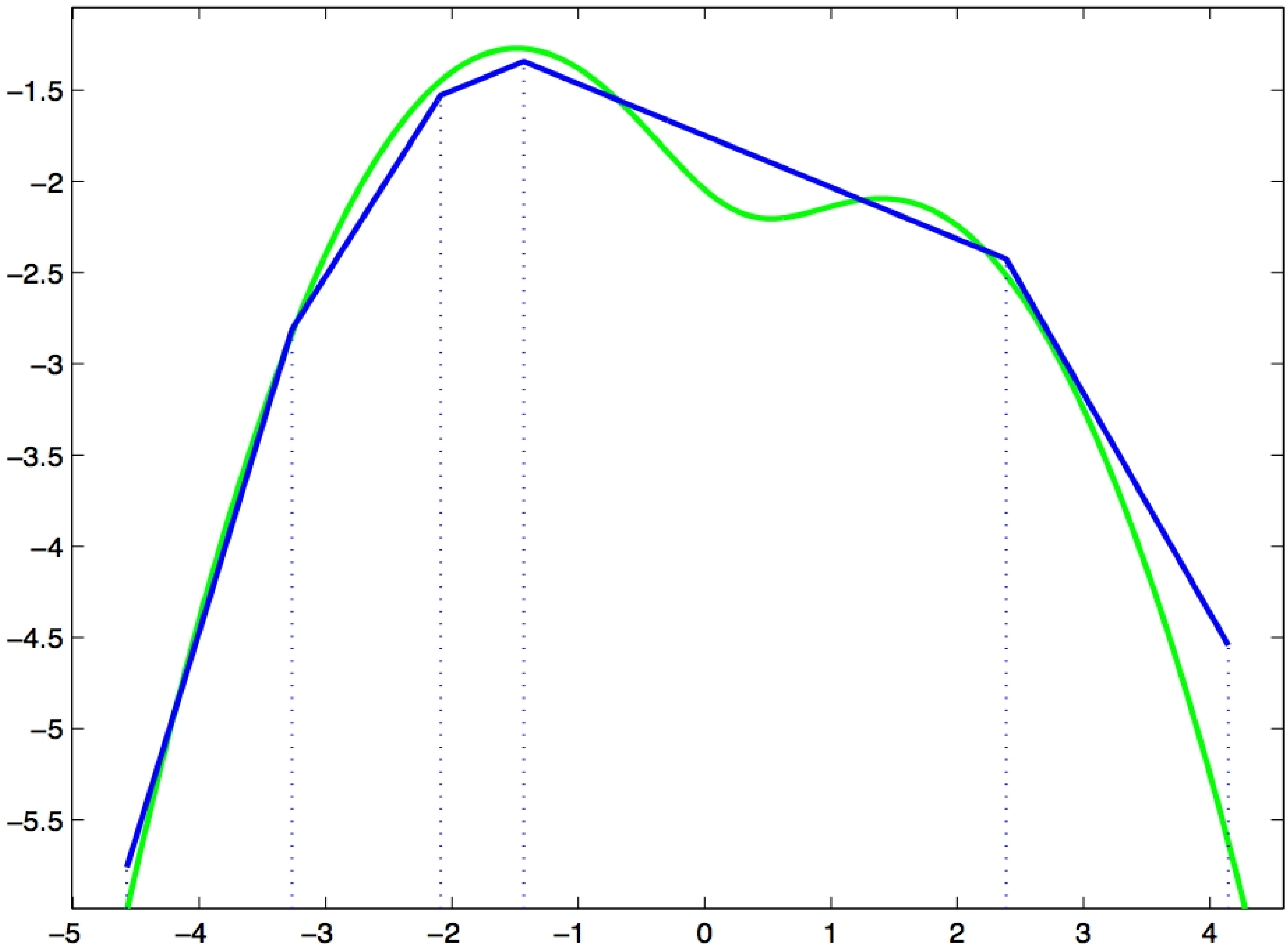
$n = 100$:

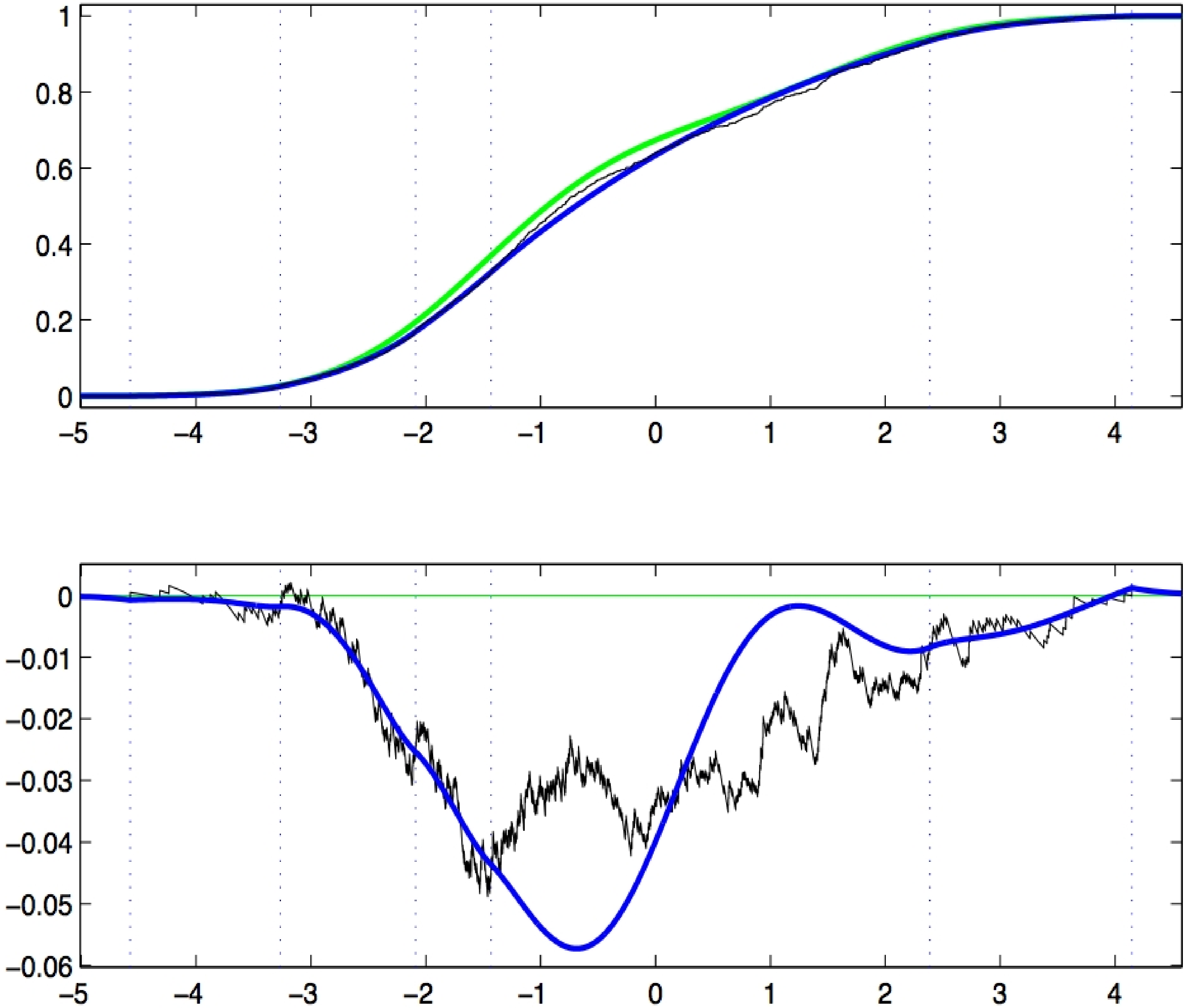




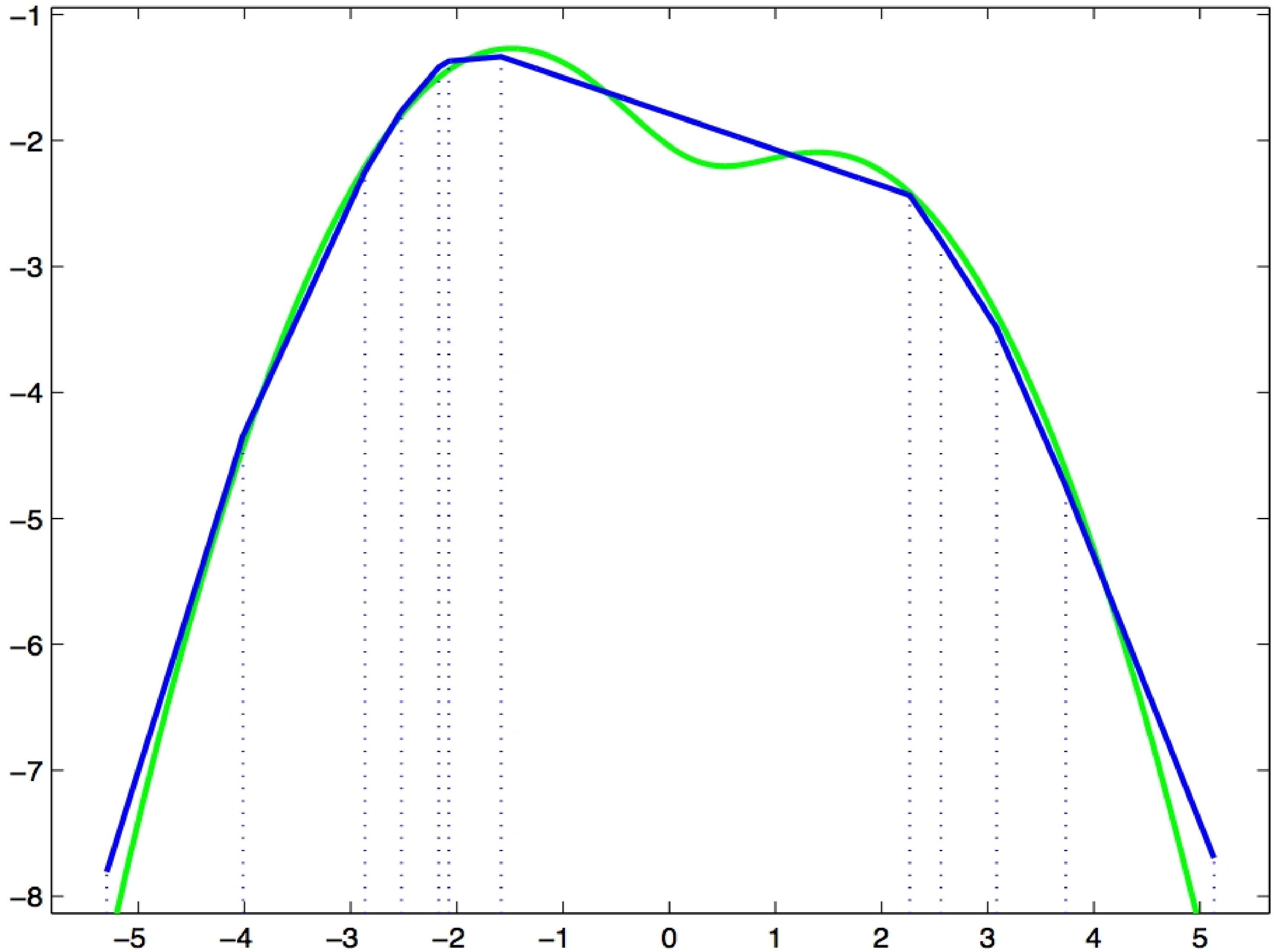
1
0.8
0.6
0.4
0.2
0
−3
−2
−1
0
1
2
3
4
0.04
0.02
0
−0.02
−0.04
−3
−2
−1
0
1
2
3
4

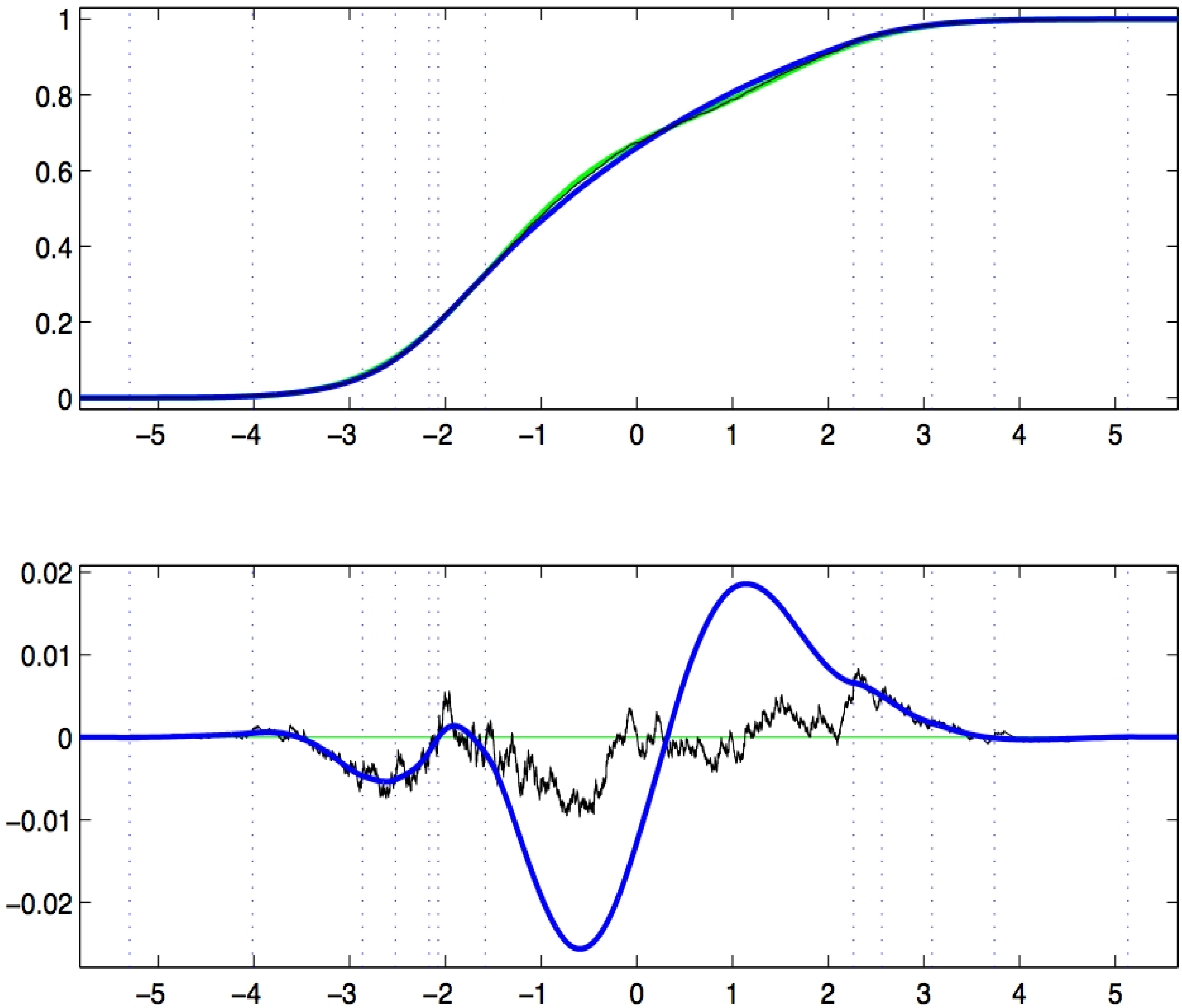
$n = 800$:




1
0.8
0.6
0.4
0.2
0
−5
−4
−3
−2
−1
0
1
2
3
4
0
−0.01
−0.02
−0.03
−0.04
−0.05
−0.06
−5
−4
−3
−2
−1
0
1
2
3
4

$n = 3200$:




1
0.8
0.6
0.4
0.2
0
−5 −4 −3 −2 −1 0 1 2 3 4 5
0.02
0.01
0
−0.01
−0.02
−5 −4 −3 −2 −1 0 1 2 3 4 5

Remark 1 The rate

$$O_p\left(\left(\frac{\log n}{n}\right)^{\beta/(2\beta+1)}\right)$$

for $\widehat{f}$ is optimal under the given conditions.

Remark 2 Integrating the density estimator

$$\widehat{f} = f + O_p\left(\left(\frac{\log n}{n}\right)^{\beta/(2\beta+1)}\right)$$

yields automatically an estimator

$$\widehat{F} = F + O_p\left(\frac{1}{\sqrt{n}}\right).$$

For kernel density estimators with non-negative kernel this is impossible!

IV. Algorithmic Aspects

Abstract setting:

Potential knots $x_1 < x_2 < \cdots < x_m$

Probability weights $p_1, p_2, \ldots, p_m$ with $p_1, p_m > 0$

$$\mathcal{G} := \Big\{\psi \in \mathcal{C}[x_1, x_m] \ : \ \text{piecewise linear with knots in } \{x_1, \ldots, x_m\}\Big\}$$

$$L(\psi) := \sum_{i=1}^{m} p_i \psi(x_i) - \int \exp(\psi(x))\, dx$$

Remark:

$$\widehat{\psi} := \underset{\psi \in \mathcal{G} \text{ concave}}{\arg\max} \; L(\psi)$$

has often $<< m$ knots.

$\Longrightarrow$ Restrict set of possible knots:

$$\{1, m\} \subset J \subset \{1, 2, \ldots, m\}$$

$$\widehat{\psi}_J := \underset{\psi \in \mathcal{G}:\, \text{knots } \in \{x_i : i \in J\}}{\arg\max} \; L(\psi)$$

Fact:

$$\widehat{\psi} = \widehat{\psi}_{\widehat{J}} \quad \text{mit} \quad \widehat{J} := \{i : x_i \text{ knot of } \widehat{\psi}\}$$

Algorithm: $\boldsymbol{J \leftarrow \widehat{J}}$

- Start with $J \leftarrow \{1, m\}$.
- Add knots by means of directional derivatives

$$H(\psi, j) \;:=\; \frac{\partial}{\partial t}\Big|_{t=0} L(\psi + t\Delta_j) \qquad \text{with} \qquad \Delta_j(x) \;:=\; -|x - x_j|$$

- Cautious removing of knots if $\widehat{\psi}_j$ is not concave.

Algorithm $\psi \leftarrow \widehat{\psi}$ **(active sets, vertex exchange)**

```
J ← {1, m}
ψ ← ψ̂_J
while max_j H(ψ, j) > 0 do
    J ← J ∪ K(ψ)
    ψ_new ← ψ̂_J
    while ψ_new is not concave do
        ψ ← (1 − λ(ψ, ψ_new))ψ + λ(ψ, ψ_new)ψ_new
        J ← J \ K(ψ, ψ_new)
        ψ_new ← ψ̂_J
    end while
    ψ ← ψ_new
end while
```

V. Censored Data

Independent event times $\boldsymbol{X_1, X_2, \ldots, X_n} \in \boldsymbol{(0, \infty]}$

($X = \infty$: event does not happen.)

Censoring yields intervals $B_1, B_2, \ldots, B_n$ like

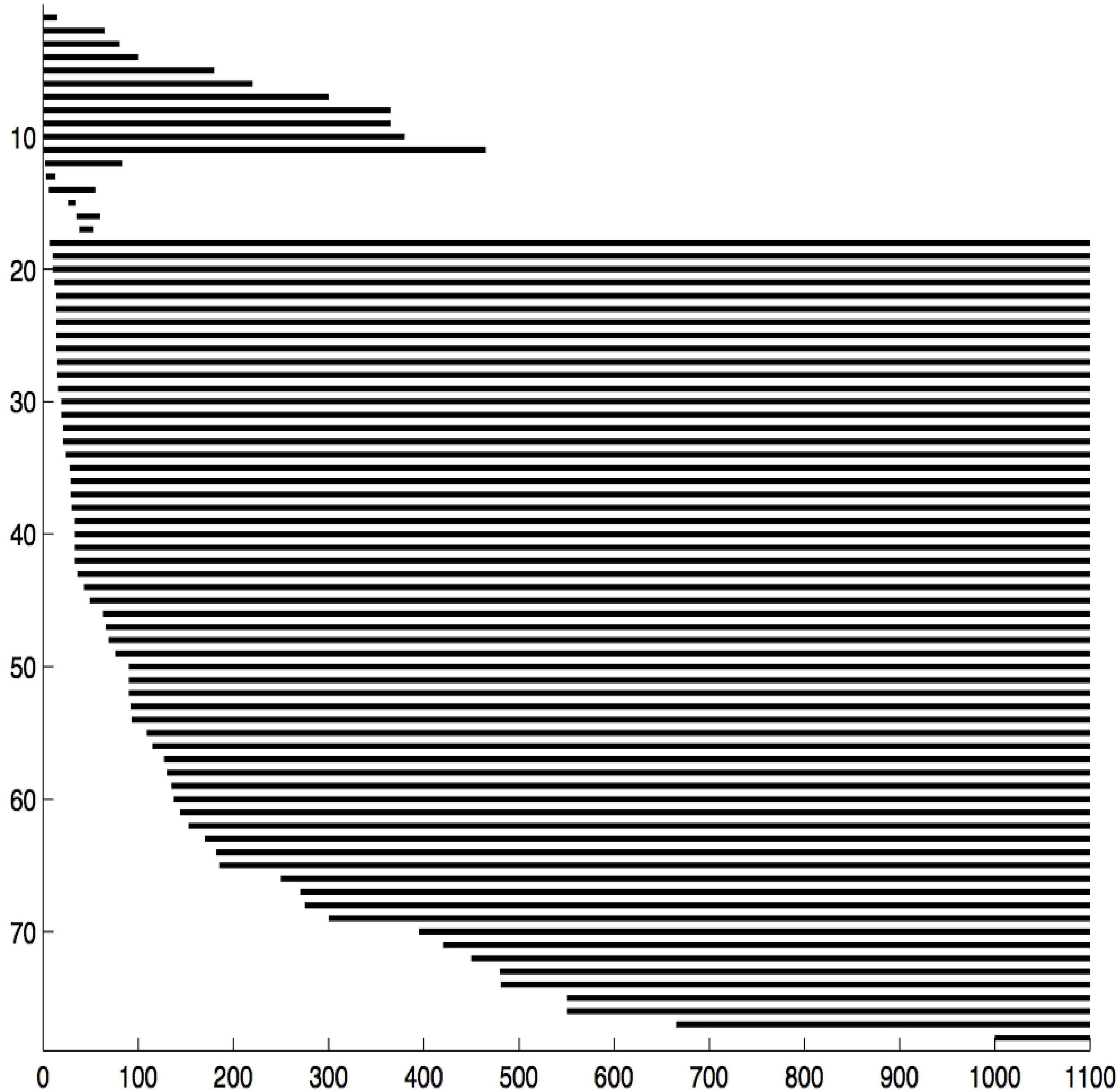
$$\boldsymbol{B_i} = \boldsymbol{\{X_i\}} \subset \boldsymbol{(0, \infty)} \quad \text{or} \quad \boldsymbol{B_i} = \boldsymbol{(L_i, R_i]} \ni \boldsymbol{X_i}$$

Data example:

Rh^- patients that are administered Rh^+ blood transfusion during emergency surgery.

Event: immuno reaction against donated blood; time after surgery.

Interval-censored data:
$n = 79$ patienten have been examined on one or several days after surgery.
Each time blood testing whether immuno reaction occured yet.


10
20
30
40
50
60
70
0
100
200
300
400
500
600
700
800
900
1000
1100

Model:

$$\mathbb{P}(\boldsymbol{X} \in \boldsymbol{B} \,|\, \boldsymbol{X} < \infty) \;=\; \boldsymbol{P}_{\boldsymbol{\psi}}(\boldsymbol{B}) := \int_{\boldsymbol{B}} \exp(\boldsymbol{\psi}(\boldsymbol{x}))\, \boldsymbol{dx}\,,$$
$$\mathbb{P}(\boldsymbol{X} < \infty) \;=\; \boldsymbol{p} \;\in\; (\boldsymbol{0}, \boldsymbol{1}]\,.$$

Parameters of interest:

- Mean, median or other quantiles of P_ψ,
- $\boldsymbol{p}$ (esp. for data example above)

Log–likelihood–functions (without Lagrange term):

$$\begin{aligned} L_o(\psi, p) &= \frac{1}{n} \sum_{i:X_i<\infty}^{n} \big(\log p + \psi(X_i)\big) \\ &+ \frac{\#\{i : X_i = \infty\}}{n} \log(1-p) \end{aligned}$$

$$\begin{aligned} L(\psi, p) &= \frac{1}{n} \sum_{i\,:\,B_i=\{X_i\}} \big(\log p + \psi(X_i)\big) \\ &+ \frac{1}{n} \sum_{i\,:\,|B_i|>0} \log\Big(p P_\psi(B_i) + 1\{R_i = \infty\}(1-p)\Big) \end{aligned}$$

EM algorithm

Fine grid of knot points $0 = x_1 < x_2 < \ldots, < x_m < x_{m+1} = \infty$,

$$[0, x_m] \supset \left\{L_i, R_i : i = 1, \ldots, n\right\} \cap [0, \infty) \, .$$

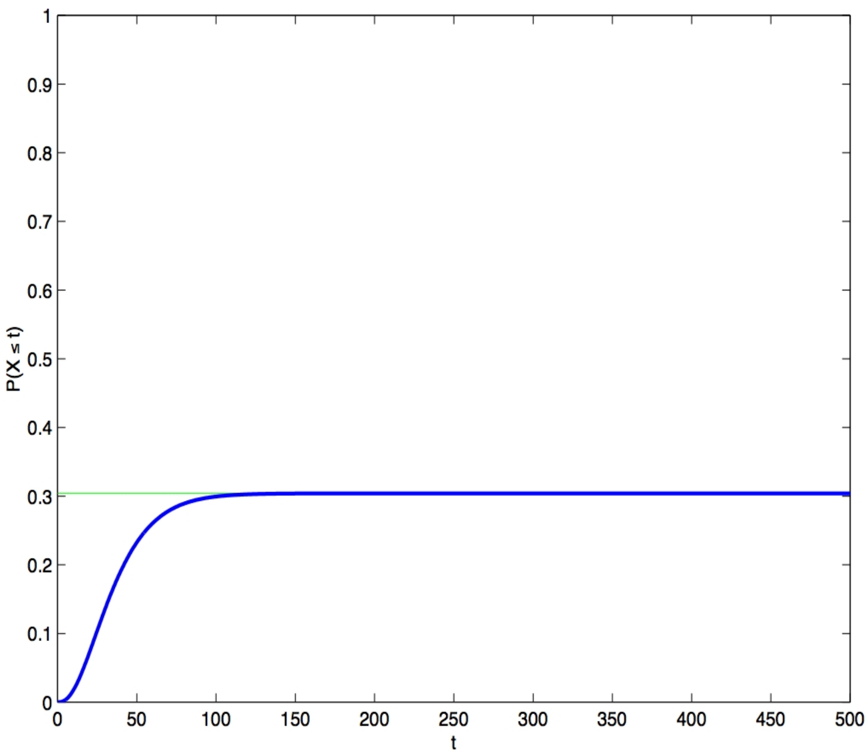
$\boldsymbol{\psi^{(k)}(p)} \rightsquigarrow \psi^{(k+1)}(p)$: Maximize

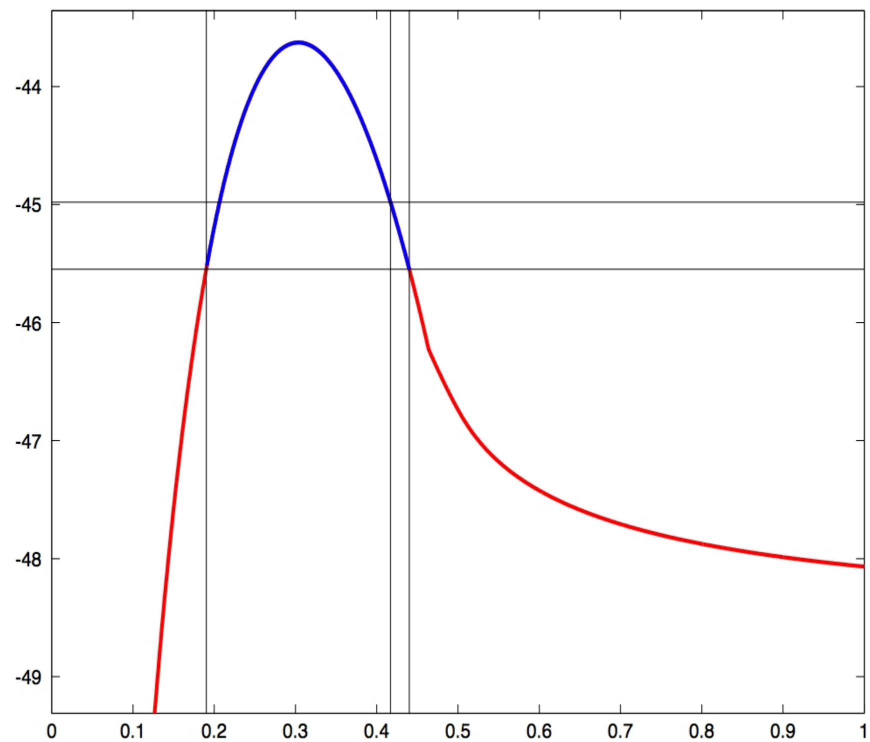
$$\boldsymbol{L^{(k)}(\psi)} \;:=\; \mathbb{E}_{f=\exp(\psi^{(k)})}\left(\boldsymbol{L_o(\psi, p)} \,\middle|\, \boldsymbol{B_1, B_2, \ldots, B_n}\right)$$

$$\sim \;\; \boldsymbol{\sum_{i=1}^{m} p_i^{(k)} \psi(x_i) - \int \exp(\psi(x))\, dx}$$

Data example (cont.)

MLE $\qquad\qquad L(p) := \max_\psi L(\psi, p)$





$\widehat{p} = 0.304$, $\widehat{\text{Med}} = 33.522$ $\qquad$ $p \leq 0.417$ (95% conf.)

VI. Outlook

- Multivariate data ($d > 1$) (Bissantz)

- c –log–concavity (Walther, Bissantz)

$$f(x) = \exp\left(\psi(x) + h(x)\right)$$

$$h \in \mathcal{C}^2(\mathbb{R}^d) \text{ konvex mit } \begin{cases} \sup_x \lambda_{\max}(D^2 h(x)) \leq c \\ \int \operatorname{trace}(D^2 h(x))\, dx \leq c \end{cases}$$

- Semiparametric regression models (Hüsler)

Data pairs $(X, Y) \in \mathbb{R}^p \times \mathbb{R}$:

$$\mathbb{P}(\boldsymbol{Y} \in [\boldsymbol{y}, \boldsymbol{y} + \boldsymbol{dy}) \,|\, \boldsymbol{X} = \boldsymbol{x}) \;=\; \exp(\boldsymbol{\psi}(\boldsymbol{y}) + \boldsymbol{\beta}^\top \boldsymbol{x}\, \boldsymbol{y} - \boldsymbol{C}(\boldsymbol{\beta}^\top \boldsymbol{x}))$$

$$\boldsymbol{\beta} \;\in\; \mathbb{R}^{\boldsymbol{p}} \quad \text{und} \quad \boldsymbol{\psi} : \mathbb{R} \to [-\infty, \infty)$$

- Quantil curves in nonparametric regression (Jongbloed)

$$\mathbb{P}(\boldsymbol{Y} \in [\boldsymbol{y}, \boldsymbol{y} + \boldsymbol{dy}) \,|\, \boldsymbol{X} = \boldsymbol{x}) \;=\; \exp(\boldsymbol{\psi}(\boldsymbol{y} \,|\, \boldsymbol{x}))$$