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 $\mu_{Ga}$, $\mu_{Ka}$, $\mu_{Ki}$, $\mu_{R}$, $\mu_{S}$
- should not exceed a certain threshold.
\[ M_A = \left[ \left( \frac{\mu_R M_{Ga}}{r(1 + \mu_K a M_{Ga})} \left( K_1 + \frac{\mu_K a r K_a}{\mu_K i r K_i} \right) - \frac{M_H \mu_K r K_a}{\mu_K i r K_i} \right) K_2 + F_z \right] \]
\cdot \left[ 0.16 P + \left( \frac{D_k}{2} + 0.58 d_2 \right) \mu_S \right]

with

\[ M_{Ga} = \tan(\beta - \arctan\left( \frac{\mu_{Ga}}{\cos \alpha_r} \right)) \],
\[ M_H = \left( \frac{262}{d_2 \tan(\rho' + \phi) + D_k \mu_S} - F_z \right) \cdot d_2 \tan(\rho' - \phi) \],
\[ \rho' = \arctan(1.155 \mu_S) \]
Five stochastically independent random quantities

<table>
<thead>
<tr>
<th>quantity</th>
<th>sample size</th>
<th>sample mean</th>
<th>sample st.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_{Ga}$</td>
<td>10</td>
<td>0.1167</td>
<td>0.0081</td>
</tr>
<tr>
<td>$\mu_{Ka}$</td>
<td>10</td>
<td>0.1463</td>
<td>0.0072</td>
</tr>
<tr>
<td>$\mu_{Ki}$</td>
<td>10</td>
<td>0.2030</td>
<td>0.0182</td>
</tr>
<tr>
<td>$\mu_{R}$</td>
<td>787</td>
<td>1681.03</td>
<td>76.25</td>
</tr>
<tr>
<td>$\mu_{S}$</td>
<td>2000</td>
<td>0.0977</td>
<td>0.0091</td>
</tr>
</tbody>
</table>
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
\( \mu_{Ga} \quad (n = 10) \)
$\mu_{Ga}$ \hspace{0.5cm} (n = 10)
$\mu_{K\alpha} \quad (n = 10)$
$\mu_{K_i} \quad (n = 10)$
\( \mu_R \quad (n = 787) \)
$\mu_S \quad (n = 2000)$
Goal

Fit a unimodal distribution to given data points which is “as close as possible” to their empirical distribution.
$\mu_R \quad (n = 787)$
$\mu_S \quad (n = 2000)$
II. Log–Concave Densities

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

$$f = \exp(\psi) \quad \text{with} \quad \psi : \mathbb{R}^d \rightarrow [-\infty, \infty) \text{ concave}.$$
Many standard distributions fulfill this constraint:

\[ \mathcal{N}_d(\mu, \Sigma) \]

- Gamma\((a, b)\) \( (a \geq 1, b > 0) \)
- Weibull\((a, b)\) \( (a \geq 1, b > 0) \)
- Beta\((a, b)\) \( (a \geq 1, b \geq 1) \)
- Gumbel distr.
- Logistic distr.

Log–concave densities are unimodal, i.e.

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

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 $\lambda \in [0, 1]$,  
  \[
  P((1 - \lambda)A + \lambda B) \geq P(A)^{1-\lambda}P(B)^{\lambda}.
  \]

  $f, g$ log–concave $\implies f \ast g$ log–concave.

- Ibragimov (1956, $d = 1$)
  \[
  \text{i.e. } f, g \text{ unimodal } \not\implies f \ast g \text{ unimodal}
  \]
  \[
  f \text{ log–concave, } g \text{ unimodal } \implies f \ast g \text{ unimodal}
  \]
III. ML Estimation \hspace{1cm} (d = 1)

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

\[
\hat{\psi} := \arg \max_{\psi \text{ concave}, \int \exp(\psi) = 1} \sum_{i=1}^{n} \psi(X_i)
\]

\[
= \arg \max_{\psi \text{ concave}} \left( \int \psi \, d\hat{F}_{\text{emp}} \right) - \int \exp(\psi(x)) \, dx
\]

with empirical c.d.f.

\[
\hat{F}_{\text{emp}}(r) := \frac{1}{n} \sum_{i=1}^{n} 1\{X_i \leq r\}.
\]
Theorem 1 (Existence, uniqueness, type)

- \( \hat{\psi} \) exists and is unique,
- \( \hat{\psi} \) is continuous and piecewise linear on \([X_1, X_n]\) with knots in \(\{X_1, X_2, \ldots, X_n\}\),
- \( \hat{f} = \exp(\hat{\psi}) \equiv 0 \) on \(\mathbb{R} \setminus [X_1, X_n]\).
Characterizations and further properties of the estimators

\[ L(\psi) := \int \psi \, d\hat{F}_{\text{emp}} - \int \exp(\psi(x)) \, dx \]

\[ \left. \frac{d}{dt} L \left( \hat{\psi} + t\Delta \right) \right|_{t=0} \leq 0 \quad \text{if} \quad \hat{\psi} + t\Delta \text{ concave for some } t > 0 \]
Theorem 2 (Characterisation of $\hat{f}$)

$$\int \Delta \, d\hat{F}_{\text{emp}} \leq \int \Delta(x) \hat{f}(x) \, dx$$

whenever $\hat{\psi} + t\Delta$ concave for some $t > 0$.

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

$$\text{Mean}(\hat{F}) = \text{Mean}(\hat{F}_{\text{emp}}) \quad (\Delta(x) = \pm x)$$

$$\text{Var}(\hat{F}) \leq \text{Var}(\hat{F}_{\text{emp}}). \quad (\Delta(x) = -x^2)$$
Set of knots of $\hat{\psi}$:

$$\hat{S} := \left\{ x : \hat{\psi}'(x-) > \hat{\psi}'(x+) \right\} \subset \{X_1, X_2, \ldots, X_n\} \subset \{X_1, X_n\}$$

Corollary 2

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

if $\Delta$ is continuous and piecewise linear with knots in $\hat{S}$. 

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

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

\[
\int_{a}^{t} \hat{F}_{\text{emp}}(x) \, dx \geq \int_{a}^{t} \hat{F}(x) \, dx,
\]
\[
\int_{t}^{b} \hat{F}_{\text{emp}}(x) \, dx \leq \int_{t}^{b} \hat{F}(x) \, dx,
\]
\[
\int_{a}^{b} \hat{F}_{\text{emp}}(x) \, dx = \int_{a}^{b} \hat{F}(x) \, dx.
\]

Corollary 2 \quad \hat{F}(X_1) = 0, \hat{F}(X_n) = 1 \quad \text{and} \quad \hat{F}_{\text{emp}} - \frac{1}{n} \leq \hat{F} \leq \hat{F}_{\text{emp}} \quad \text{on} \ \hat{S}.
Consistency of the estimators

**Theorem 4** (Consistency of $\hat{F}$)

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

**Conjecture:**

\[ \mathbb{P}\left( \left\| \hat{F} - F \right\|_\infty \leq \left\| \hat{F}_{\text{emp}} - F \right\|_\infty \right) \rightarrow 1 \]

But

in gen. \[ \left\| \hat{F} - F \right\|_\infty \not\leq C \left\| \hat{F}_{\text{emp}} - F \right\|_\infty \]

**Theorem 5** (Consistency of \( \hat{\psi} \))

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

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

Then

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

with \( \delta_n = (\log(n)/n)^{1/(2\beta+1)} \to 0. \)
Theorem 6 (Asympt. equivalence of $\hat{F}$ and $\hat{F}_{\text{emp}}$)

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

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

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

\[ \phi(x) = -\frac{x^2}{2} + c \]
\( n = 100: \)
\( n = 800: \)
Numerical examples II

\[ f(x) = \frac{x \exp(-x)}{\Gamma(2)} \]

\[ \phi(x) = \log(x) - x + c \]
$n = 100$: 
$n = 800$: 
Numerical examples III

\[ F = 0.7 \cdot \mathcal{N}(-1.5, 1) + 0.3 \cdot \mathcal{N}(1.5, 1) \]

(bimodal distribution)
$n = 100$: 
$n = 800$: 
$n = 3200$: 
Remark 1  The rate

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

for \( \hat{f} \) is optimal under the given conditions.

Remark 2  Integrating the density estimator

\[ \hat{f} = f + O_p \left( \left( \frac{\log n}{n} \right)^{\beta/(2\beta+1)} \right) \]

yields automatically an estimator

\[ \hat{F} = F + O_p \left( \frac{1}{\sqrt{n}} \right). \]

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 $p_1, p_m > 0$

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

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

\[ \hat{\psi} := \arg \max_{\psi \in G} \text{concave} L(\psi) \]

has often \( \ll m \) knots.

\[ \implies \text{Restrict set of possible knots:} \]

\[ \{1, m\} \subset J \subset \{1, 2, \ldots, m\} \]

\[ \hat{\psi}_J := \arg \max_{\psi \in G: \text{knots } \in \{x_i : i \in J\}} L(\psi) \]

Fact:

\[ \hat{\psi} = \hat{\psi}_\hat{J} \text{ mit } \hat{J} := \{i : x_i \text{ knot of } \hat{\psi}\} \]
Algorithm: \[ J \leftarrow \hat{J} \]

- Start with \[ J \leftarrow \{1, m\}. \]
- Add knots by means of directional derivatives
  \[ H(\psi, j) := \left. \frac{\partial}{\partial t} \right|_{t=0} L(\psi + t\triangle_j) \quad \text{with} \quad \triangle_j(x) := -|x - x_j| \]
- Cautious removing of knots if \( \hat{\psi}_j \) is not concave.
Algorithm  \( \psi \leftarrow \hat{\psi} \)  (active sets, vertex exchange)

\[
J \leftarrow \{1, m\}
\]

\[
\psi \leftarrow \hat{\psi}_J
\]

while \( \max_j H(\psi, j) > 0 \) do

\[
J \leftarrow J \cup K(\psi)
\]

\[
\psi_{\text{new}} \leftarrow \hat{\psi}_J
\]

while \( \psi_{\text{new}} \) is not concave do

\[
\psi \leftarrow (1 - \lambda(\psi, \psi_{\text{new}}))\psi + \lambda(\psi, \psi_{\text{new}})\psi_{\text{new}}
\]

\[
J \leftarrow J \setminus K(\psi, \psi_{\text{new}})
\]

\[
\psi_{\text{new}} \leftarrow \hat{\psi}_J
\]

end while

\[
\psi \leftarrow \psi_{\text{new}}
\]

end while
V. Censored Data

Independent event times \( X_1, X_2, \ldots, X_n \in (0, \infty] \)

\( (X = \infty : \text{event does not happen.}) \)

Censoring yields intervals \( B_1, B_2, \ldots, B_n \) like

\[
B_i = \{X_i\} \subset (0, \infty) \quad \text{or} \quad B_i = (L_i, R_i] \ni X_i
\]
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.
Model:

\[ P( X \in B \mid X < \infty ) = P_{\psi}(B) := \int_B \exp(\psi(x)) \, dx , \]

\[ P(X < \infty ) = p \in (0, 1]. \]

Parameters of interest:

- Mean, median or other quantiles of \( P_{\psi} \),
- \( p \) (esp. for data example above)
Log–likelihood–functions (without Lagrange term):

\[ L_0(\psi, p) = \frac{1}{n} \sum_{i: X_i < \infty} (\log p + \psi(X_i)) \]
\[ + \frac{\# \{i : X_i = \infty\}}{n} \log(1 - p) \]

\[ L(\psi, p) = \frac{1}{n} \sum_{i: B_i = \{X_i\}} (\log p + \psi(X_i)) \]
\[ + \frac{1}{n} \sum_{i: |B_i| > 0} \log(pP_\psi(B_i) + 1\{R_i = \infty\})(1 - p) \]
EM algorithm

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

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

\( \psi^{(k)}(p) \sim \psi^{(k+1)}(p) \): Maximize

\[
L^{(k)}(\psi) := \mathbb{E}_{f=\exp(\psi^{(k)})}\left( L_o(\psi, p) \mid B_1, B_2, \ldots, B_n \right)
\]

\[
\sim \sum_{i=1}^{m} p_i^{(k)} \psi(x_i) - \int \exp(\psi(x)) \, dx
\]
Data example (cont.)

\[ L(p) := \max_{\psi} L(\psi, p) \]

\[ \hat{p} = 0.304, \ \text{Med} = 33.522 \]

\[ p \leq 0.417 \quad (95\% \text{ conf.}) \]
VI. Outlook

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

- \(c\) –log–concavity (Walther, Bissantz)

\[
f(x) = \exp(\psi(x) + h(x))
\]

\[h \in C^2(\mathbb{R}^d) \text{ konvex mit } \left\{ \begin{array}{l}
\sup_x \lambda_{\max}(D^2 h(x)) \leq c \\
\int \text{trace}(D^2 h(x)) \, dx \leq c
\end{array} \right\}
\]
• Semiparametric regression models (Hüsler)

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

\[
P(Y \in [y, y + dy] \mid X = x) = \exp(\psi(y) + \beta^\top x y - C(\beta^\top x))
\]
\[\beta \in \mathbb{R}^p \text{ und } \psi : \mathbb{R} \to [-\infty, \infty)\]

• Quantil curves in nonparametric regression (Jongbloed)

\[
P(Y \in [y, y + dy] \mid X = x) = \exp(\psi(y \mid x))
\]