Nonparametric estimation under shape constraints, part 1

Piet Groeneboom, Delft University

August 6, 2013

What to expect?

- History of the subject
- Distribution of order restricted monotone estimates.
- Connection with Brownian motion functionals and Airy functions.
- Application to inverse problems. Smoothed maximum likelihood estimator (SMLE) and maximum smoothed likelihood estimator (MSLE) for models for incomplete data.
- Does bootstrapping work for these models?
- (Lecture 2:)
  Limit behavior of the order restricted convex and concave estimates. "Invelope" of integrated Brownian motion $+t^4$.

Lake Mendota

First example in:
'Statistical Inference under Order Restrictions:
the Theory and Application of Isotonic Regression
Barlow, Bartholomew, Bremner, and Brunk (1972):

Number of days until freezing in the years 1854 $+i$, $i = 1, \ldots, 111$,
in Lake Mendota, Wisconsin ("most studied lake in the USA").

Question: Can a warming trend be deduced from the data?

Lake Mendota

Lake Mendota

Barlow, Bartholomew, Bremner, and Brunk (1972):

"According to a simple, useful (if not completely realistic) model, the days till freezing X_i are observations on a normal distribution with unknown means μ_i, $i = 1, 2, \ldots, 111$, and a common variance σ^2."

The maximum likelihood estimates of μ_i under the restriction $\mu_1 \leq \cdots \leq \mu_{111}$ minimize (as a function of the μ_i):

$$\sum_{i=1}^{111} \{X_i - \mu_i\}^2 \,,$$

subject to $\mu_1 \leq \cdots \leq \mu_{111}$.

Monotone least squares

We prefer to use the data on duration of ice:

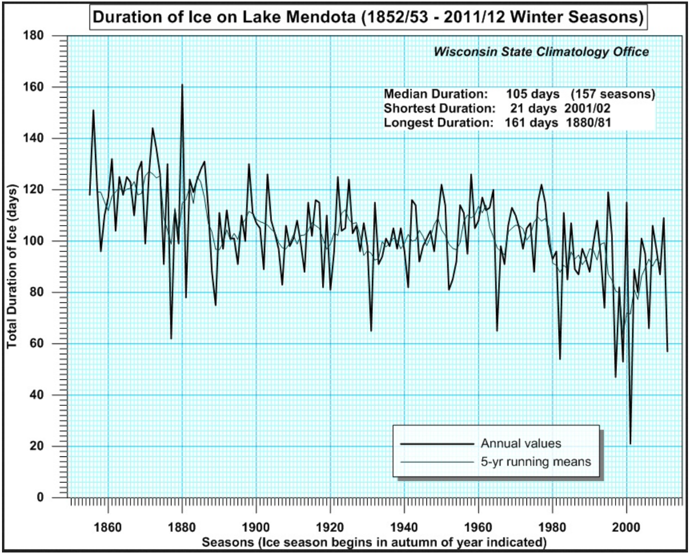
Y_i : number of days the lake was frozen in year i.

Monotone (or "isotonic") least squares estimate is:

$$\hat{\nu} = \text{argmin}_{\nu} \left\{ \sum_{i=1}^{157} \{ Y_i - \nu_i \}^2 \right\}.$$

subject to $\nu_1 \geq \cdots \geq \nu_{157}$.

Lake Mendota ice data

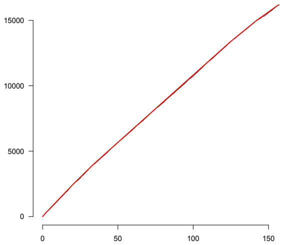


Computation of monotonic least squares estimate

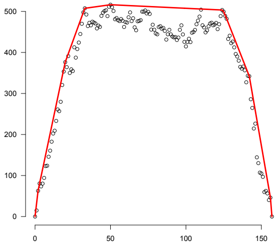
We consider the cusum diagram, consisting of the points

$$(0,0), (1, Y_1), (2, Y_1 + Y_2), \ldots, \left(i, \sum_{j=1}^{i} Y_j\right), \ldots, \left(157, \sum_{j=1}^{157} Y_j\right).$$

For this set of points we compute the least concave majorant.
Solution is left continuous slope of least concave majorant.

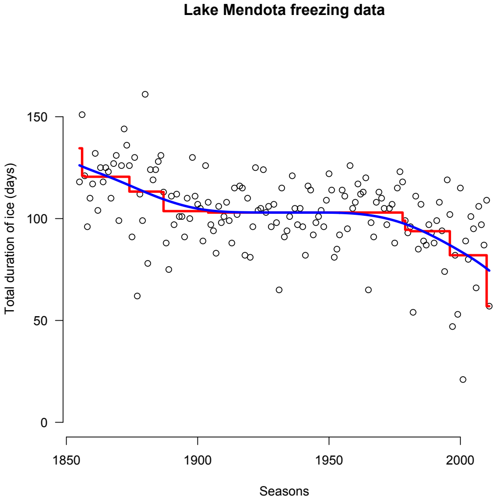


(a) Cusum diagram (+ least concave majorant)

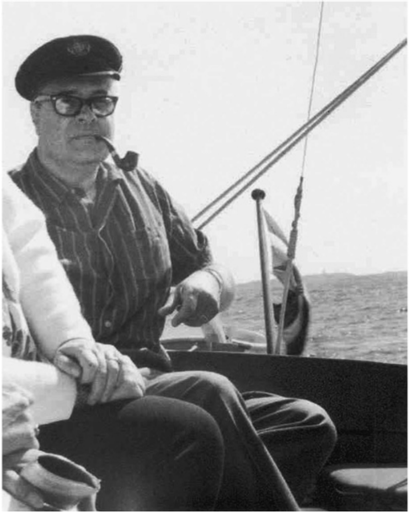


(b) Cusum diagram minus line between endpoints

Isotonic regression function



Ulf Grenander



Monotone density estimation and isotonic regression

The log likelihood of a specific density f is given by

$$\ell(f) = \frac{1}{n}\sum_{i=1}^{n}\log f(X_i) = \int \log f(x)\, d\mathbb{F}_n(x).$$

The Grenander maximum likelihood estimator maximizes this function over all decreasing densities on $[0,\infty)$.

Theorem (Grenander (1956))

The maximum likelihood estimator is the left derivative of the least concave majorant $\hat{F}_n$ of the empirical distribution function $\mathbb{F}_n$.

The Grenander estimator

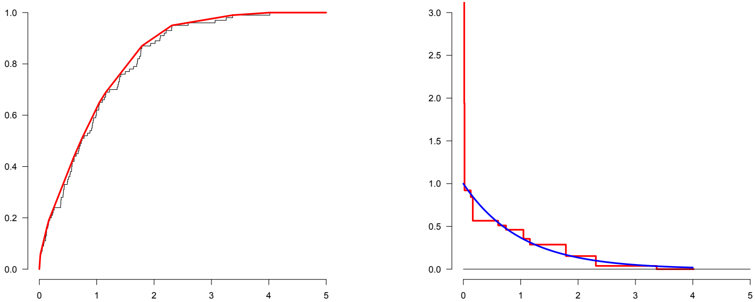


Figure: The least concave majorant $\hat{F}_n$ and its derivative $\hat{f}_n$ (the Grenander estimator) for a sample of size $n = 100$ from a standard exponential.

The Grenander estimator is also least squares estimator

The Grenander estimator $\hat{f}_n$ maximizes

$$\ell(f) = \frac{1}{n} \sum_{i=1}^{n} \log f(X_i) = \int_0^\infty \log f(x)\, d\mathbb{F}_n(x).$$

but minimizes

$$\int_0^\infty f(t)^2\, dt - 2 \int_0^\infty f(t)\, d\mathbb{F}_n(t)$$

over all decreasing densities f on $[0, \infty)$.

Cator (2011): $\hat{f}_n$ is locally asymptotically minimax

Limit distribution of Grenander estimator

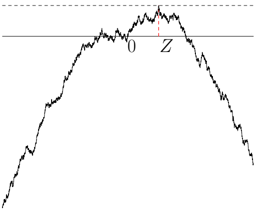
Theorem (Prakasa Rao (1969))

Let $\hat{f}_n$ be the Grenander estimate of f under the monotonicity restriction. Then, if f has a strictly negative derivative f' at the interior point t:

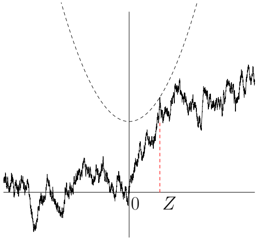
$$n^{1/3}\left\{\hat{f}_n(x) - f(x)\right\} / \left|4f(x)f'(x)\right|^{1/3} \xrightarrow{\mathcal{D}} Z, \; n \to \infty,$$

where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution, and $Z = \mathrm{argmax}_t\{W(t) - t^2\}$, that is: Z is the (almost surely unique) location of the maximum of two-sided Brownian motion minus the parabola $y(t) = t^2$.

$Z = \text{argmax}_t\{W(t) - t^2\}$



(a) $W(t) - t^2$ and Z



(b) $W(t)$ and Z

Bhagavatula Lakshmi Surya Prakasa Rao

The switch relation

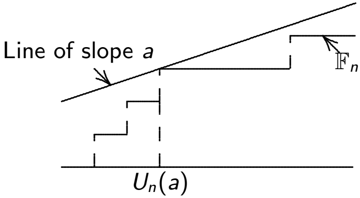


Figure: The switch relation.

Different proof of Prakasa Rao's result: Groeneboom (1985).
Key observation: we have the switch relation

$$\hat{f}_n(t) \geq a \iff t \leq U_n(a) = \text{argmax}\,\{x \geq 0 : \mathbb{F}_n(x) - ax\}$$

and consider the process $\{U_n(a),\ a \in (0, \infty)\}$.

Slopes become the time variable of the process U_n!

The switch relation

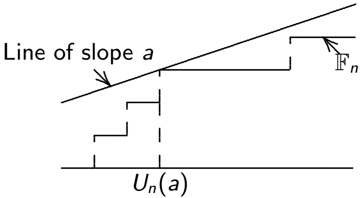


Figure: The switch relation.

So we have, if $a = f_0(t)$ and $U_n(a) = \text{argmax}\{x \geq 0 : \mathbb{F}_n(x) - ax\}$,

$$\mathbb{P}\left\{n^{1/3}\{\hat{f}_n(t) - f_0(t)\} \geq x\right\} = \mathbb{P}\left\{\hat{f}_n(t) \geq a + n^{-1/3}x\right\}$$
$$\stackrel{\text{switch relation}}{=} \mathbb{P}\left\{U_n(a + n^{-1/3}x) \geq t\right\}.$$

This + "Hungarian embedding" leads to new proof of Prakasa Rao's result (see Groeneboom (1985)).

Analytical characterization of distribution of $Z = \operatorname{argmax}\{W(t) - t^2 : t \in \mathbb{R}\}$?

Figure: Herman Chernoff

Chernoff's heat equation (Chernoff (1964))

- Define $u(s,x)$ by

$$u(s,x) = \mathbb{P}\left\{W(t) > t^2 \text{ for some } t > s | W(s) = x\right\}.$$

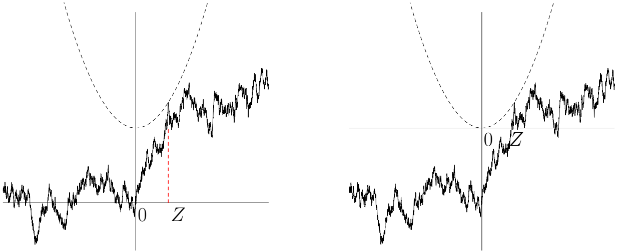
Then $u(s,x) = 1,\ x \geq s^2, \qquad u(s,x) \to 0,\ x \to -\infty$.
If $x < s^2$:

$$\begin{aligned} u(s,x) &= \mathbb{E}\left\{u\left(s+\epsilon, W(s+\epsilon)\right)\right\} \\ &= u(s,x) + \frac{\partial}{\partial s}u(s,x)\epsilon + \frac{1}{2}\frac{\partial^2}{\partial x^2}u(s,x)\epsilon + o(\epsilon). \end{aligned}$$

Hence $u(s,x)$ satisfies the heat equation:

$$\frac{\partial}{\partial s}u(s,x) = -\frac{1}{2}\frac{\partial^2}{\partial x^2}u(s,x).$$

Chernoff's heat equation (Chernoff (1964))



- Define: $M_h = max_{t\in[s-h,s]} W(t)$.
  Then (using space homogeneity of Brownian motion):

$$\mathbb{P}\left\{\max_{t\geq s}\left\{W(t)-t^2\right\}\geq M_h - s^2 \;\middle|\; W(s), M_h\right\}$$
$$= u(s,s^2)\,(=1) - \left\{M_h - W(s)\right\}\partial_2 u(s,s^2) + O_p(h).$$

Chernoff's heat equation (Chernoff (1964))

- Similarly:

$$\mathbb{P}\left\{\max_{t\le s-h}\left\{W(t)-t^2\right\}\ge M_h-(s-h)^2 \;\Big|\; W(s-h),M_h\right\}$$
$$= u(-s,s^2)\,(=1) - \left\{M_h - W(s-h)\right\}\partial_2 u(-s,s^2) + O_p(h).$$

- Conclusion

$$\mathbb{P}\{Z\in[s-h,s]\}$$
$$\sim \mathbb{E}\left\{M_h - W(s)\right\}\left\{M_h - W(s-h)\right\}\partial_2 u(s,s^2)\partial_2 u(-s,s^2)$$
$$\sim h\mathbb{E}\left(\max_{x\in[0,1]} B(x)\right)^2 \partial_2 u(s,s^2)\partial_2 u(-s,s^2)\;(B = \text{Brownian Bridge})$$
$$= \tfrac{1}{2}h\partial_2 u(s,s^2)\partial_2 u(-s,s^2),\; h\downarrow 0.$$

Chernoff's theorem

Theorem (Chernoff (1964))

The density f_Z of $Z = argmax\{W(x) - x^2\}$ is given by:

$$f_Z(s) = \tfrac{1}{2}\partial_2 u(-s, s^2)\partial_2 u(s, s^2).$$

where $u(s,x)$ solves the heat equation:

$$\frac{\partial}{\partial s}u(s,x) = -\frac{1}{2}\frac{\partial^2}{\partial x^2}u(s,x).$$

subject to:

$$u(s,x) = 1,\ x \geq s^2, \qquad u(s,x) \to 0,\ x \to -\infty.$$

Computation of density

Original computations of this density were based on numerically solving Chernoff's heat equation.

But (Groeneboom (1984)):

$$\partial_2 u(-s, s^2) \sim c_1 \exp\left\{-\tfrac{2}{3}s^3 - cs\right\}, \; s \to \infty,$$

where $c \approx 2.9458$ and $c_1 \approx 2.2638$. This entails that a numerical solution of this partial differential equation on a grid will not give a really accurate solution!

Theorem (Groeneboom (1984),Daniels and Skyrme (1985))

The probability density f of the location of the maximum of the process $t \mapsto W(t) - t^2$, $t \in \mathbb{R}$, is given by

$$f(s) = \tfrac{1}{2} g(s) g(-s),$$

where

$$g(s) = \frac{1}{2^{2/3}\pi} \int_{-\infty}^{\infty} \frac{e^{-ius}}{\mathrm{Ai}(i2^{-1/3}u)}\, du.$$

where Ai is the Airy function Ai.

Density also given in: Janson (2013), Groeneboom, Lalley, and Temme (2013).

Distribution of the maximum itself:
Janson, Louchard, and Martin-Löf (2010), Groeneboom (2010)
Groeneboom and Temme (2011).

Henry Daniels

Steven Lalley and Nico Temme

Density of $Z = \text{argmax}\{W(t) - t^2,\, t \in \mathbb{R}\}$

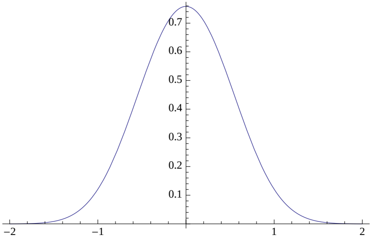


Figure: The density f_Z of the location of the maximum Z of $W(t) - t^2,\, t \in \mathbb{R}$.

Also:

$$\text{var}(Z) = \tfrac{1}{3}\mathbb{E}\max_t \left\{W(t) - t^2\right\},$$

as proved in Groeneboom (2011) and Janson (2013), and (not using the relation with Airy functions) in Pimentel (2013).

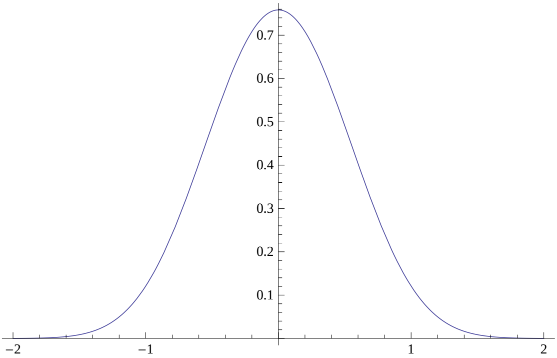
The density can be computed by two lines in Mathematica:

```
In[1]:= f[x_] := (1 / (2 * Pi)) * 2^(1 / 3) *
          Re[NIntegrate[Exp[-I * u * x] / AiryAi[I * 2^(-1 / 3) * u], {u, -10, 10}]]

In[2]:= g[x_] := (1 / 2) * f[x] * f[-x]

In[3]:= Plot[g[x], {x, -2, 2}]
```

Out[3]=



How do the Airy functions enter?

Theorem (Groeneboom (1989))

$Q^{(s,x)}$ is probability distribution of $\{X(t) : t \geq s\}$, where $X(t) = W(t) - t^2$ and $X(s) = x < 0$.
$\tau = \inf\{t \geq s : X(t) = 0\}$.

1. *(Cameron-Martin-Girsanov)*

$$Q^{(s,x)}\{\tau \in dt\} = \phi(s,t)E^0\left\{e^{-2\int_0^{t-s} B(u)\,du} \,\middle|\, B(t-s) = -x\right\} dt,$$

for a (specified) function $\phi(s,t)$, where B is Bessel(3).

2. $$Q^{(s,x)}\{\tau \in dt\} = e^{-\frac{2}{3}\left(t^3-s^3\right)+2sx} h_x(t-s)\,dt,$$

where (Feynman-Kac or Ito's formula):

$$h_x(t) = \frac{1}{2\pi}\int_{v=-\infty}^{\infty} e^{itv}\frac{\mathrm{Ai}(i2^{-1/3}v - 4^{1/3}x)}{\mathrm{Ai}(i2^{-1/3}v)}\,dv,\ t > 0.$$

Bootstrapping the Grenander estimator

Theorem (Kosorok (2008))

The nonparametric bootstrap is inconsistent for the Grenander estimator, i.e.,

$$n^{1/3}\left\{\hat{f}_n^*(t)-\hat{f}_n(t)\right\}\xrightarrow{\mathcal{D}}\left|4f'(t)f(t)\right|^{1/3}Z$$

does not hold (in probability), conditionally on the data, where

$$Z = argmax\{W(t)-t^2 : t\in\mathbb{R}\}.$$

Proof by contradiction

1. Suppose $n^{1/3}\{\hat{f}_n^*(t) - \hat{f}_n(t)\} \xrightarrow{\mathcal{D}} |4f'(t)f(t)|^{1/3}Z$ (in probability), conditionally on the data. Then:

$$n^{1/3}\left\{\hat{f}_n^*(t) - f(t)\right\} = n^{1/3}\left\{\hat{f}_n^*(t) - \hat{f}_n(t)\right\} + n^{1/3}\left\{\hat{f}_n(t) - f(t)\right\}$$
$$\xrightarrow{\mathcal{D}} \left|4f'(t)f(t)\right|^{1/3}\{Z_1 + Z_2\},$$

   where the Z_i are independent copies of $\text{argmax}\{W(x) - x^2\}$.

2. On the other hand,

$$n^{1/3}\left\{\hat{f}_n^*(t) - f(t)\right\}$$
$$\xrightarrow{\mathcal{D}} \left|4f'(t)f(t)\right|^{1/3}\text{argmax}_x\left\{W_1(x) + W_2(x) - x^2\right\},$$

   Right side (with smaller variance than limit in (1)) comes from

$$\mathbb{F}_n^*(t + n^{-1/3}x) - \mathbb{F}_n^*(t) - \left(\mathbb{F}_n(t + n^{-1/3}x) - \mathbb{F}_n(t)\right)$$
$$+ \mathbb{F}_n(t + n^{-1/3}x) - \mathbb{F}_n(t) - f(t)n^{-1/3}x. \qquad \square$$

Does generating bootstrap samples for the Grenander estimator itself work?

Sen, Banerjee, and Woodroofe (2010): No!

In this case we have:

$$n^{1/3}\left\{\hat{f}_n^*(t) - f(t)\right\} \to \left|4f'(t)f(t)\right|^{1/3} \operatorname{argmax}_x \left\{W_1(x) + V_2(x)\right\},$$

where W_1 is two-sided Brownian motion and V_2 is the least concave majorant of the drifting two-sided Brownian motion $W_2(x) - x^2$, independent of W_1.

Current status model

$X_1, X_2, \ldots, X_n \sim F_0$.
Instead of observing the X_i's, one only observes for each i whether or not $X_i \leq T_i$ for some random T_i, T_i independent of X_i.

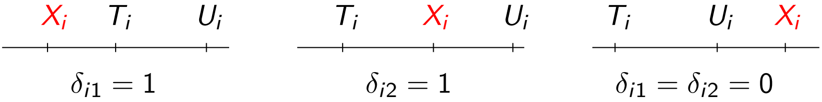
So, instead of observing X_i's, one observes

$$(T_i, \delta_i) = (T_i, 1_{\{X_i \leq T_i\}}).$$

The i-th observation (T_i, δ_i) represents the current status of the hidden variable X_i at time T_i.

Interval censoring, case 2

$X_1, X_2, \ldots, X_n \sim F_0$.
Instead of observing the X_i's, one only observes for each i whether $X_i \le T_i$ or $X_i \in (T_i, U_i]$ or $X_i > U_i$, for some random pair (T_i, U_i), where $T_i < U_i$, where (T_i, U_i) is independent of X_i.



So, instead of observing X_i's, one observes

$$(T_i, U_i, \delta_{i1}, \delta_{i2}) = (T_i, U_i, 1_{\{X_i \le T_i\}}, 1_{\{X_i \in (T_i, U_i]\}}).$$

Tomorrow! Much harder!

Current status model

We want to estimate the unknown distribution function F_0 of X_i, based on the data $(T_i, \delta_i) = (T_i, 1_{\{X_i \le T_i\}})$.

If the X_i are independent of the T_i, the log likelihood function in F (conditional on the T_i's) is given by

$$\ell(F) = \sum_{i=1}^{n} \{\delta_i \log F(T_i) + (1 - \delta_i) \log(1 - F(T_i))\}.$$

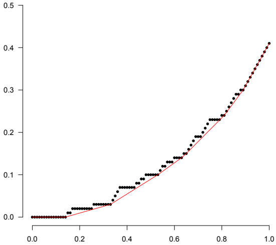
The (nonparametric) maximum likelihood estimator (MLE) $\hat{F}_n$ maximizes $\ell(F)$ over the class of *all* distribution functions F.

MLE is least squares estimator

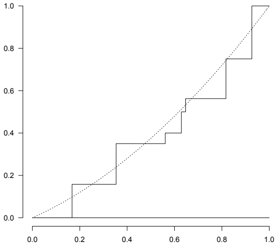
The MLE $\hat{F}_n$ minimizes over $y_1 = F(T_{(1)}) \leq \cdots \leq y_n = F(T_{(n)})$ the sum:

$$\sum_{i=1}^{n} \left\{ \delta_{(i)} - y_i \right\}^2 .$$

where $\delta_{(i)}$ corresponds to $T_{(i)}$. Isotonic regression on the $\delta_{(i)}$'s!



(a) $\left(\frac{i}{n}, \frac{1}{n}\sum_{j=1}^{i} \delta_{(j)}\right)$



(b) $F_0(t) = \frac{1}{2}t(1+t)$, $\hat{F}_{100}$

Local asymptotic distribution

Theorem (Groeneboom (1987))

Let F_0 and G be differentiable at t with strictly positive derivatives $f_0(t)$ and $g(t)$. Let $\hat{F}_n$ be the MLE of F_0. Then, as $n \to \infty$,

$$\frac{n^{1/3}\{\hat{F}_n(t) - F_0(t)\}}{\{4F_0(t)(1-F_0(t))f_0(t)/g(t)\}^{1/3}} \xrightarrow{\mathcal{D}} Z,$$

where $Z = argmax_t\{W(t) - t^2\}$.

Note the analogy with the behavior of the Grenander estimator, but this time we have a distribution function instead of a density.

Niels Keiding

Hepatitis A in Bulgaria

388 KEIDING [Part 3,

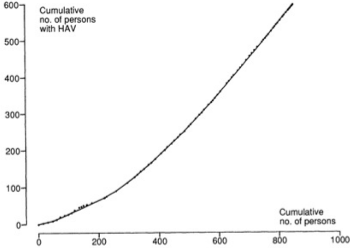


Fig. 5. Hepatitis A in Bulgaria: function *H* and its largest convex minorant

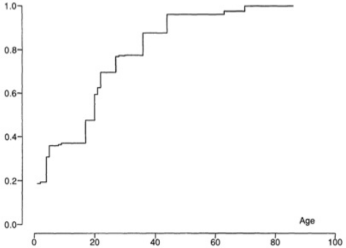


Fig. 6. Hepatitis A in Bulgaria: estimated distribution of age of occurrence of hepatitis A

Hazard

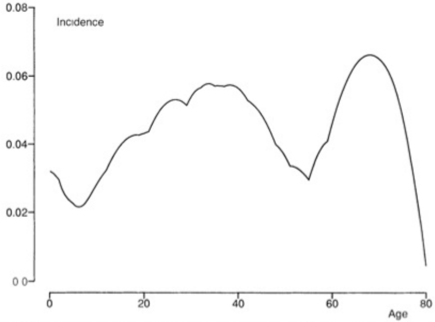


Fig. 7. Estimated hepatitis A incidence in Bulgaria, based on the Epanechnikov kernel $0.75(1-x^2)$ with bandwidth 15 years (chosen by visual inspection)

Density and hazard estimate in discussion

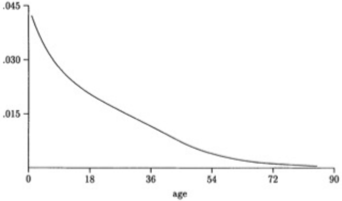


Fig. 8. Density estimate $\hat{g}_{n,h}$

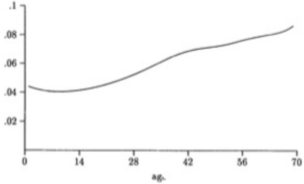


Fig. 9. Incidence estimate

I took $n_1 = 100$ and the kernel

$$K(u) = \frac{35}{32}(1-u^2)^3 \mathbf{1}_{[-1,1]}(u), \qquad u \in \mathbf{R},$$

Smoothed maximum likelihood estimator (SMLE)

The smoothed maximum likelihood estimator (SMLE) $\hat{F}_{n,h}^{(SML)}$ is (modulo a boundary correction), defined by

$$\hat{F}_{n,h}^{(SML)}(t) = \int \mathbb{K}_h(t-x)\,d\hat{F}_n(x), \qquad \mathbb{K}_h(y) = \int_{-\infty}^{y/h} K(u)\,du,$$

where K is (for example) the triweight kernel

$$K(u) = \tfrac{35}{32}\left\{1-u^2\right)^3 1_{[-1,1]}(u).$$

We take $h \asymp n^{-1/5}$ (the usual bandwidth in density estimation).

Theorem (Groeneboom, Jongbloed, and Witte (2010))

Let F_0 be differentiable at t with second derivative $f_0'(t) \neq 0$ and let $g(t) > 0$ and let g have a bounded derivative at t. If $h_n = cn^{-1/5}$, for some $c > 0$, then

$$n^{2/5}\big(\hat{F}_{n,h_n}^{(SML)}(t) - F_0(t)\big) \xrightarrow{\mathcal{D}} N(\mu, \sigma^2), \text{ where}$$

$$\mu = \tfrac{1}{2}c^2 f_0'(t) \int u^2 K(u)\, du, \qquad \sigma^2 = \frac{F_0(t)\{1 - F_0(t)\}}{cg(t)} \int K(u)^2\, du.$$

Figure: Geurt Jongbloed and Birgit Witte

Estimation of distribution functions and densities

Theorem (Groeneboom, Jongbloed, and Witte (2010))

If $h_n = cn^{-1/5}$, *for* $\hat{F}^{(SML)}_{n,h_n}(t) = \int \mathbb{K}_h(t-x)\, d\hat{F}_n(x)$

$$n^{2/5}\big(\hat{F}^{(SML)}_{n,h_n}(t) - F_0(t)\big) \xrightarrow{\mathcal{D}} N(\mu, \sigma^2),\ \textit{where}$$

$$\mu = \tfrac{1}{2}c^2 f_0'(t) \int u^2 K(u)\, du, \qquad \sigma^2 = \frac{F_0(t)\{1 - F_0(t)\}}{cg(t)} \int K(u)^2\, du.$$

Theorem (Groeneboom, Jongbloed, and Witte (2010))

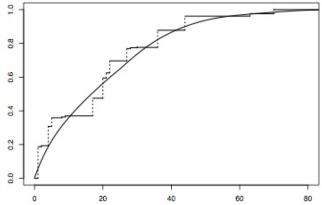
If $h_n = cn^{-1/7}$, *then, for* $\hat{f}^{(SML)}_{n,h_n}(t) = \int K_h(t-x)\, d\hat{F}_n(x)$,

$$n^{2/7}\big(\hat{f}^{(SML)}_{n,h_n}(t) - f_0(t)\big) \xrightarrow{\mathcal{D}} N(\mu, \sigma^2),\ \textit{where}$$

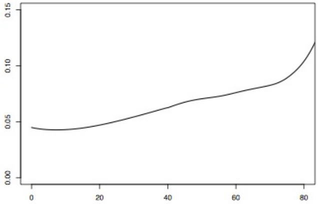
$$\mu = \tfrac{1}{2}c^2 f_0''(t) \int u^2 K(u)\, du, \qquad \sigma^2 = \frac{F_0(t)\{1 - F_0(t)\}}{c^3 g(t)} \int K'(u)^2\, du.$$

SMLE and hazard for Hepatitis A data

Smooth bootstrap estimates of optimal bandwidths in Groeneboom, Jongbloed, and Witte (2010) produces for the Keiding data:



(a) MLE + SMLE



(b) Hazard

The results in Groeneboom, Jongbloed, and Witte (2010) are proved by using local smooth functional theory.

Smooth functionals

Theorem (Groeneboom (1991))

Let F_0 be differentiable on $[0,B]$ and let the observation density g stay away from zero on $[0,B]$. Then:

$$\sqrt{n}\left\{\int x\,d\hat{F}_n(x)-\int x\,dF_0(x)\right\}\xrightarrow{\mathcal{D}}N(0,\sigma^2),$$

where

$$\sigma^2=\int\frac{F_0(x)\{1-F_0(x)\}}{g(x)}\,dx.$$

The mapping $F\mapsto K(F)=\int x\,dF(x)$ is a smooth functional for the current status model.
$F\mapsto F(t)$ is not a smooth functional in the current status model: van der Vaart (1991)

Smoothed maximum likelihood (SMLE)

The mapping $F \mapsto K(F) = \int \mathbb{K}_h(x-y)\,dF(x)$ is a local smooth functional for the current status model, if $h \gg n^{-1/3}$.
We try to find a representation of the form:

$$\int \mathbb{K}_h(x-y)\,d(\hat{F}_n - F_0)(x) = \int \theta_{h_n,\hat{F}_n}(u,\delta)\,d\left(\mathbb{P}_n - P\right)(u,\delta) + o_p((nh)^{-1/2}),$$

where $\theta_{h_n,\hat{F}_n}(u,\delta)$ is (generally) the solution of an integral equation.

Smooth functionals for current status

- The nonlinear aspect of the functional is negligible. This means:

$$\sqrt{n}\left\{K(\hat{F}_n) - K(F_0)\right\} = \sqrt{n}\int \kappa_{F_0}\, d\big(\hat{F}_n - F_0\big) + o_p(1).$$

- Transformation to the observation space measure.

$$\int \kappa_{F_0}\, d\big(\hat{F}_n - F_0\big) = -\int \theta_{\hat{F}_n}(t,\delta)\, dQ_0(t,\delta),$$

where

$$\theta_{\hat{F}_n}(t,\delta) = \frac{\delta - \hat{F}_n(t)}{g(t)}\,.$$

More generally $\theta_{\hat{F}_n}(t,\delta)$ is the solution of an integral equation (Groeneboom (2013b)).

- Use that $\hat{F}_n$ is the MLE.
  Replace $\theta_{\hat{F}_n}$ by $\bar{\theta}_{\hat{F}_n}(t,\delta) = (\delta - \hat{F}_n(t))/\bar{g}(t)$, where $\bar{g}$ is constant on the same intervals as $\hat{F}_n$. Then:

$$\int \bar{\theta}_{\hat{F}_n}(t,\delta)\, dQ_n = 0, \tag{1}$$

  and

$$\int \kappa_{F_0}\, d\big(\hat{F}_n - F_0\big) \overset{\text{preceding step}}{=} -\int \theta_{\hat{F}_n}\, dQ_0$$
$$\overset{(1)}{=} \int \bar{\theta}_{\hat{F}_n}\, d\big(Q_n - Q_0\big) - \int \left\{\theta_{\hat{F}_n} - \bar{\theta}_{\hat{F}_n}\right\} dQ_0.$$

- Asymptotic variance equals information lower bound
  Show:

$$\int \bar{\theta}_{\hat{F}_n}\, d\big(Q_n - Q\big) = \int \theta_{F_0}\, d\big(Q_n - Q_0\big) + o_p\left(n^{-1/2}\right),$$

  where

$$\theta_{F_0}(t,\delta) = \frac{\delta - F_0(t)}{g(t)}.$$

Maximum smoothed likelihood estimator (MSLE)

Instead of smoothing the MLE, we can first smooth the likelihood and then compute the maximum likelihood estimator: maximum smoothed likelihood estimators (MSLE's).
The MSLE (maximum smoothed likelihood estimator) minimizes the Kullback-Leibler distance:

$$\mathcal{K}(\tilde{Q}_n, P) = \int \log \frac{\tilde{Q}_n}{dP}\, d\tilde{Q}_n$$

over the measures P is the allowed class, for a $\tilde{Q}_n$ which is a smoothed empirical $\mathbb{Q}_n$.

MLE does the same for

$$\mathcal{K}(\mathbb{Q}_n, P) = \int \log \frac{d\mathbb{Q}_n}{dP}\, d\mathbb{Q}_n$$

Maximum smoothed likelihood estimator (MSLE)

MLE is slope of greatest convex minorant of

$$\left(\int_{[0,t]} d\mathbb{G}_n(u), \int_{[0,t]} \delta \, d\mathbb{P}_n(u,\delta)\right), \; t \geq 0.$$

The MSLE is slope of greatest convex minorant of

$$\left(\int \mathbb{K}_h(t-x)\, d\mathbb{G}_n(x), \int \delta \, \mathbb{K}_h(t-x)\, d\mathbb{P}_n(x,\delta)\right), \; t \geq 0.$$

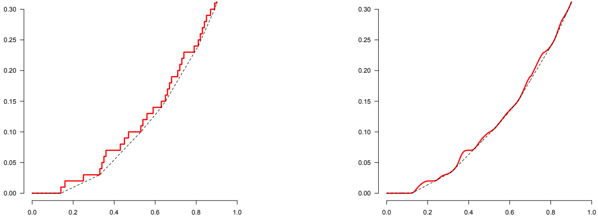


Figure: Unsmoothed and smoothed cusum diagram

MSLE for current status

The MSLE $\hat{F}_{nh}^{(MSLE)}$ maximizes the object function

$$\int_0^M \log F(u)\, dG_{nh}^{\delta}(u) + \int_0^M \log\{1 - F(u)\}\, d\big(G_{nh} - G_{nh}^{\delta}\big)(u),$$

where:

$$G_{nh}(t) = \int \mathbb{K}_h(t - x)\, d\mathbb{G}_n(x),\ t \in [0, M],$$

and

$$G_{nh}^{\delta}(t) = \int \delta\, \mathbb{K}_h(t - x)\, d\mathbb{P}_n(x, \delta),\ t \in [0, M],$$

and where $\mathbb{K}_h(u) = \int_{-\infty}^{u/h} K(y)\, dy$, for a symmetric kernel K.

MLE, SMLE and MSLE

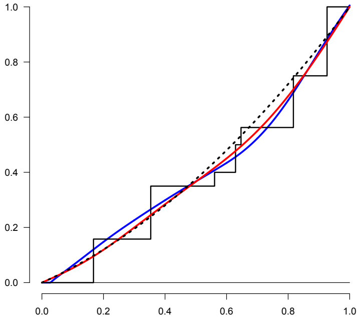


Figure: MLE (black) and SMLE (red) and MSLE (blue) for a current status sample of size 100 from $F_0(t) = \frac{1}{2}t(1+t)$ (dotted). Observation distribution is uniform.

Asymptotic distribution of MSLE

Theorem (Groeneboom, Jongbloed, and Witte (2010))

Fix $t \in (0, M)$ so that f_0'' and g'' exist and are continuous at t. Let $h = h_n \sim cn^{-1/5}$ $(c > 0)$ be the bandwidth used in the definition of g_{nh} and g_{nh}^{δ}. Then

$$n^{2/5}\left\{\hat{F}_{nh}^{MSLE} - F_0(t)\right\} \xrightarrow{\mathcal{D}} N(\mu, \sigma^2),$$

where

$$\mu = \frac{1}{2}c^2\left\{f_0'(t) + \frac{2f_0(t)g'(t)}{g(t)}\right\}\int u^2 K(u)\,du,$$

and

$$\sigma^2 = \frac{F_0(t)\bigl(1 - F_0(t)\bigr)}{cg(t)}\int K(u)^2\,du.$$

Key to proof: Monotonicity constraint is asymptotically not active.

General remarks on the MLE, SMLE and MSLE

- The MSLE can be used in situations where the MLE and SMLE fail.
  Example: The continuous mark model with current status data. The MLE and SMLE are inconsistent, in contrast with the MSLE: Groeneboom, Jongbloed, and Witte (2012).
- The MLE gives asymptotically efficient estimates of smooth functionals for the interval censoring model (so in particular for the current status model). The proof for the general interval censoring model is pretty hard! (Lecture 2).
- The difference between the MSLE and SMLE mainly turns up in the bias; in general one cannot say that one of them is uniformly better than the other estimator.
- There exists an LR-type two-sample test for current status data, based on MSLE's, which is independent of the observation distributions in the two samples (Groeneboom (2012)).

Concluding remarks

Discussed:

- The computation of isotonic estimates via greatest convex minorants and least concave majorants
- The limiting distribution of Grenander's estimator
- Chacterization of the limit density via Chernoff's heat equation
- Characterization of the limit density via Airy functions
- Does the bootstrap work?
- The MLE, SMLE and MSLE for the current status model

Lecture 2:

- Theory and open problems for interval censoring, case 2.
- Same for the bivariate current status model.
- Convex regression.
- The convex envelope of one-sided Brownian motion without drift.
- Does bootstrapping from the Grenander estimate work for global statistics?

Bibliography I

R. E. Barlow, D. J. Bartholomew, J. M. Bremner, and H. D. Brunk. *Statistical inference under order restrictions. The theory and application of isotonic regression*. John Wiley & Sons, London-New York-Sydney, 1972.

Eric Cator. Adaptivity and optimality of the monotone least-squares estimator. *Bernoulli*, 17(2):714–735, 2011.

Herman Chernoff. Estimation of the mode. *Ann. Inst. Statist. Math.*, 16:31–41, 1964.

H. E. Daniels and T. H. R. Skyrme. The maximum of a random walk whose mean path has a maximum. *Adv. in Appl. Probab.*, 17(1):85–99, 1985.

Ulf Grenander. On the theory of mortality measurement. II. *Skand. Aktuarietidskr.*, 39:125–153 (1957), 1956.

P. Groeneboom. Brownian Motion With A Parabolic Drift And Airy Functions. CWI Technical Report Department of Mathematical Statistics-R 8413, CWI, 1984.

Bibliography II

P. Groeneboom. Estimating a monotone density. In *Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer, Vol. II (Berkeley, Calif., 1983)*, Wadsworth Statist./Probab. Ser., pages 539–555, Belmont, CA, 1985. Wadsworth.

P. Groeneboom. Asymptotics for incomplete censored observations. Report 87-18, Mathematical Institute, University of Amsterdam., 1987.

P. Groeneboom. Brownian motion with a parabolic drift and airy functions. *Probab. Theory Related Fields*, 81(1):79–109, 1989.

P. Groeneboom. Nonparametric maximum likelihood estimators for interval censoring and deconvolution. Technical Report 378, Department of Statistics, Stanford University, 1991. URL `http://statistics.stanford.edu/~ckirby/reports/1990_1999/reports1991.html`.

Bibliography III

P. Groeneboom. Nonparametric (smoothed) likelihood and integral equations. Discussion paper. To appear in the Journal of Statistical Planning and Inference, 2013b.

P. Groeneboom, G. Jongbloed, and B. I. Witte. Maximum smoothed likelihood estimation and smoothed maximum likelihood estimation in the current status model. *Ann. Statist.*, 38:352–387, 2010.

Piet Groeneboom. The maximum of Brownian motion minus a parabola. *Electron. J. Probab.*, 15:no. 62, 1930–1937, 2010.

Piet Groeneboom. Vertices of the least concave majorant of Brownian motion with parabolic drift. *Electron. J. Probab.*, 16: no. 84, 2234–2258, 2011.

Piet Groeneboom. Likelihood ratio type two-sample tests for current status data. *Scand. J. Stat.*, 39(4):645–662, 2012.

Bibliography IV

Piet Groeneboom, Geurt Jongbloed, and Birgit I. Witte. A maximum smoothed likelihood estimator in the current status continuous mark model. *J. Nonparametr. Stat.*, 24(1):85–101, 2012.

Piet Groeneboom, Steven Lalley, and Nico Temme. Chernoffs distribution and differential equations of parabolic and airy type. Submitted, 2013.

Piet Groeneboom and Nico M. Temme. The tail of the maximum of Brownian motion minus a parabola. *Electron. Commun. Probab.*, 16:458–466, 2011.

Svante Janson. Moments of the location of the maximum of Brownian motion with parabolic drift. *Electron. Commun. Probab.*, 18:no. 15, 1–8, 2013.

Svante Janson, Guy Louchard, and Anders Martin-Löf. The maximum of Brownian motion with parabolic drift. *Electron. J. Probab.*, 15:no. 61, 1893–1929, 2010.

Bibliography V

Michael R. Kosorok. Bootstrapping the Grenander estimator. In *Beyond parametrics in interdisciplinary research: Festschrift in honor of Professor Pranab K. Sen*, volume 1 of *Inst. Math. Stat. Collect.*, pages 282–292. Inst. Math. Statist., Beachwood, OH, 2008.

Leandro P.R. Pimentel. On the location of the maximum of a continuous stochastic process. Submitted, 2013.

B. L. S. Prakasa Rao. Estimation of a unimodal density. *Sankhyā Ser. A*, 31:23–36, 1969.

Bodhisattva Sen, Moulinath Banerjee, and Michael Woodroofe. Inconsistency of bootstrap: the Grenander estimator. *Ann. Statist.*, 38(4):1953–1977, 2010.

Aad van der Vaart. On differentiable functionals. *Ann. Statist.*, 19 (1):178–204, 1991. ISSN 0090-5364. URL `http://dx.doi.org/10.1214/aos/1176347976`.