Smooth M-estimation in the Current Status Model

Statistik unter einem Dach

Bielefeld

March 27, 2007

Birgit Witte

Joint work with Piet Groeneboom and Geurt Jongbloed

vrije Universiteit Amsterdam, The Netherlands

Illustrating Example: Rubella

- Some facts about Rubella (also called three-days measels or German measels)
  - Once infected, immunity is life-long.
  - Disease can pass unnoticed.
- Interested in stochastic behaviour of variable

$$X \quad = \quad \text{time someone is infected by rubella virus.}$$

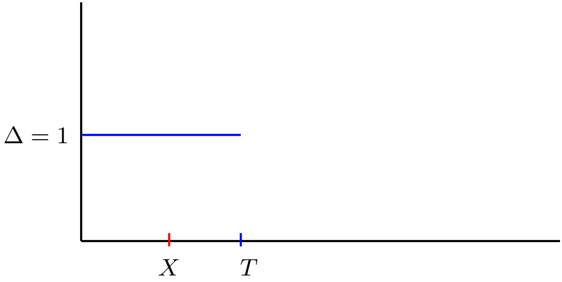
- When people are tested for antibody presence in blood, the vector $Z = (T, \Delta)$ is observed, where

$$T \quad = \quad \text{current age}$$
$$\Delta \quad = \quad \text{current immunization status} \quad = \quad 1_{\{X \le T\}}.$$

General Setup Current Status Model

- **Settings:**

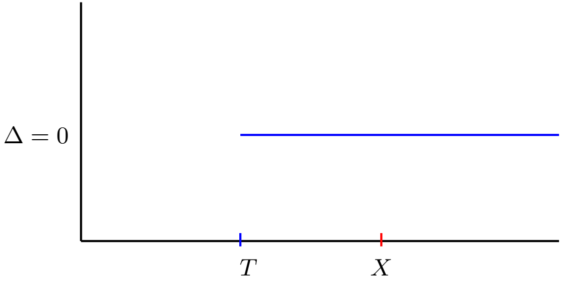
X	variable of interest with **unknown** cdf F_0.
T	censoring variable with (possibly unknown) density g.
$\Delta = 1_{\{X \leq T\}}$	variable indicating whether X hapened before time T or not.
$Z = (T, \Delta)$	observed vector



General Setup Current Status Model

- **Settings:**

X	variable of interest with **unknown** cdf F_0.
T	censoring variable with (possibly unknown) density g.
$\Delta = 1_{\{X \le T\}}$	variable indicating whether X hapened before time T or not.
$Z = (T, \Delta)$	observed vector



General Setup Current Status Model

- **Settings:**

X	variable of interest with **unknown** cdf F_0.
T	censoring variable with (possibly unknown) density g.
$\Delta = 1_{\{X \leq T\}}$	variable indicating whether X hapened before time T or not.
$Z = \big(T, \Delta\big)$	observed vector

- **Known:**

Density of observed vector Z is given by

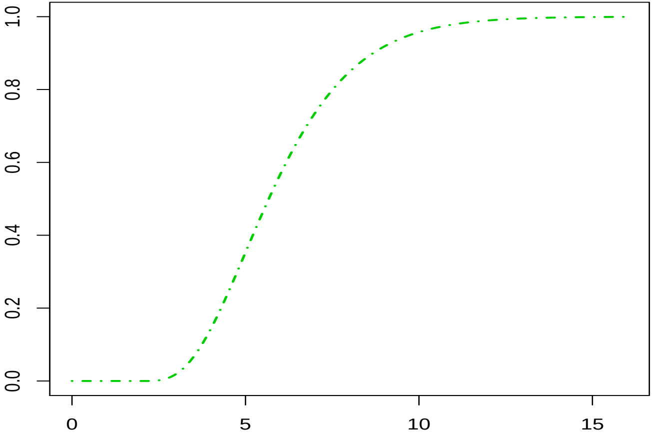
$$f_Z(t, \delta) \quad = \quad g(t)\big\{F(t)\big\}^{\delta}\big\{\big(1 - F(t)\big)\big\}^{(1-\delta)}, \; t > 0$$

- **Goal:**

Estimate the distribution function F_0 and its density f_0, using n iid copies of Z collected during epidemiological studies.

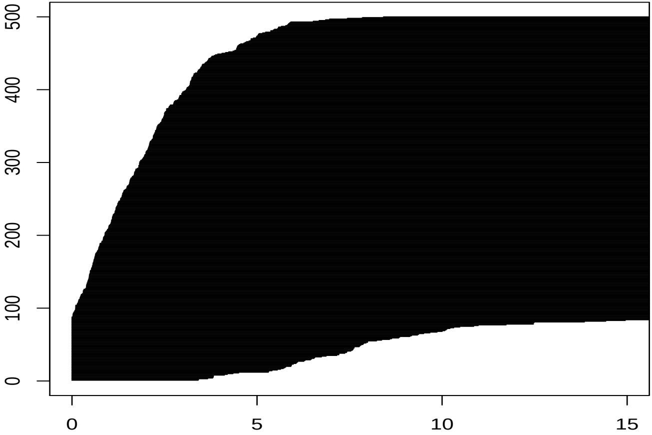
An example

$F_0(x) = 1 - \exp(-x),\ g(t) = \exp(-t),\ n = 250.$



An example

$F_0(x) = 1 - \exp(-x),\ g(t) = \exp(-t),\ n = 250.$



Overview

- Definition of the Estimators
  - Nonparametric Maximum Likelihood Estimator (NPMLE)
  - Maximum Smoothed Likelihood Estimator (MSLE)
  - Smoothed Maximum Likelihood Estimator (SMLE)
- Characterization and Illustration of the Estimators
- Asymptotic Results
  - Distribution Function
  - Density
- Discussion

Some Notation

- K is a kernel density with compact support, i.e. [-1,1], and distribution function $\bar{K}(u) = \int_{-1}^{u} K(x)\,dx$
- h_n is a smoothing parameter (bandwidth) such that $K_{h_n}(u) = 1/h_n K(u/h_n)$, $\bar{K}_{h_n} = \bar{K}(u/h_n)$.
- $\mathbb{G}_n$ is the empirical distribution function (edf) of the T_i's, and $\mathbb{P}_n$ is the edf of the Z_i's.
- g is unknown and estimated by

$$\hat{g}_n(t) = \int K_{h_n}(x-t)\,d\mathbb{G}_n(t),$$

such that $\hat{g}_n(t) = \hat{g}_{n,0}(t) + \hat{g}_{n,1}(t)$

$$\hat{g}_{n,0}(t) = \int (1-\delta) K_{h_n}(x-t)\,d\mathbb{P}_n(t,\delta),$$

$$\hat{g}_{n,1}(t) = \int \delta K_{h_n}(x-t)\,d\mathbb{P}_n(t,\delta).$$

- $\hat{P}_n(t,\delta) = (1-\delta)\hat{G}_{n,0}(t) + \delta\hat{G}_{n,1}(t)$ is a smoothed version of $\mathbb{P}_n$, where

$$\hat{G}_{n,i}(t) = \int_0^t \hat{g}_{n,i}(u)\,du.$$

Estimators for F_0

- Log-likelihood for n iid copies of Z:

$$l(F) = \int \left\{\delta \log F(t) + \left(1-\delta\right) \log \left(1 - F(t)\right)\right\} d\mathbb{P}_n(t,\delta)$$

Nonparametric Maximum Likelihood Estimator (NPMLE):

$$\hat{F}_n = \arg\max_{F \in \mathcal{F}} l(F)$$

Studied by P. Groeneboom & J.A. Wellner (1992), *Information Bounds an Nonparametric Maximum Likelihood Estimation*, New York: Cambridge University Press.

Estimators for F_0

- Log-likelihood for n iid copies of Z:

$$l(F) = \int \left\{\delta \log F(t) + \big(1-\delta\big) \log\big(1 - F(t)\big)\right\} d\mathbb{P}_n(t,\delta)$$

Nonparametric Maximum Likelihood Estimator (NPMLE):

$$\hat{F}_n = \arg\max_{F \in \mathcal{F}} l(F)$$

- Smoothed log-likelihood for n iid copies of Z:

$$l^S(F) = \int \left\{\delta \log F(t) + \big(1-\delta\big) \log\big(1 - F(t)\big)\right\} d\hat{P}_n(t,\delta)$$

Maximum Smoothed Likelihood Estimator (MSLE):

$$\hat{F}_n^{MS} = \arg\max_{F \in \mathcal{F}} \; l^S(F)$$

Introduced by P.P.B. Eggermont & V.N. LaRiccia (2001), *Maximum Penalized Likelihood Estimation*, New York: Springer-Verlag.

Estimators for F_0

- Log-likelihood for n iid copies of Z:

$$l(F) = \int \left\{\delta \log F(t) + \left(1-\delta\right) \log\left(1 - F(t)\right)\right\} d\mathbb{P}_n(t,\delta)$$

Nonparametric Maximum Likelihood Estimator (NPMLE):

$$\hat{F}_n = \arg\max_{F \in \mathcal{F}} l(F)$$

- Smoothed log-likelihood for n iid copies of Z:

$$l^S(F) = \int \left\{\delta \log F(t) + \left(1-\delta\right) \log\left(1 - F(t)\right)\right\} d\hat{P}_n(t,\delta)$$

Maximum Smoothed Likelihood Estimator (MSLE):

$$\hat{F}_n^{MS} = \arg\max_{F \in \mathcal{F}} l^S(F)$$

- Smoothed Maximum Likelihood Estimator (SMLE):

$$\hat{F}_n^{SM}(x) = \int \bar{K}_{h_n}(x-y)\, d\hat{F}_n(y)$$

Estimators for f_0

Based on the relation

$$f_0(x) = \frac{d}{dx} F_0(x)$$

we define the MSLE for $f_0(x)$ as

$$\hat{f}_n^{MS}(x) = \frac{d}{dx} \hat{F}_n^{MS}(x).$$

The SMLE for $f_0(x)$ is defined as

$$\hat{f}_n^{SM}(x) = \frac{d}{dx} \hat{F}_n^{SM}(x) = \int K_{h_n}(x-y)\, d\hat{F}_n(y)$$

Note the relation between $\hat{f}_n^{MS}(x)$ and the estimator $\hat{g}_n(x)$, defined by

$$\hat{g}_n(x) = \int K_{h_n}\big(x-y\big)\, d\mathbb{G}_n(y).$$

Overview

- Definition of the Estimators
  - Nonparametric Maximum Likelihood Estimator (NPMLE)
  - Maximum Smoothed Likelihood Estimator (MSLE)
  - Smoothed Maximum Likelihood Estimator (SMLE)
- Characterization and Illustration of the Estimators
- Asymptotic Results
  - Distribution Function
  - Density
- Discussion

Proposition (Groeneboom & Wellner (1992))

Let H^* be the convex minorant of the cusum diagram (CSD) consisting of the points

$$t \mapsto \left(\mathbb{G}_n(t), \int_{u\in[0,t]} \delta \, d\mathbb{P}_n(u,\delta)\right), \; t \in [0, T_{(n)}]$$

Let $\hat{F}_n(t)$ be the right-continuous slope of H^*. Then $\hat{F}_n$ maximizes $l(F)$ over the class of all distribution functions.

Illustration of $\hat{F}_n$

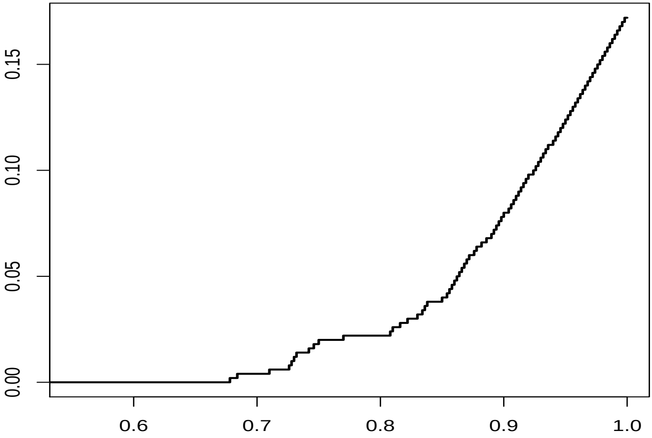


Illustration of $\hat{F}_n$

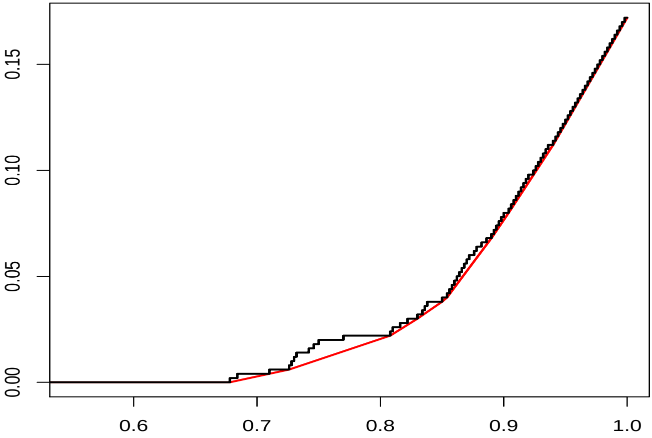


Illustration of $\hat{F}_n$

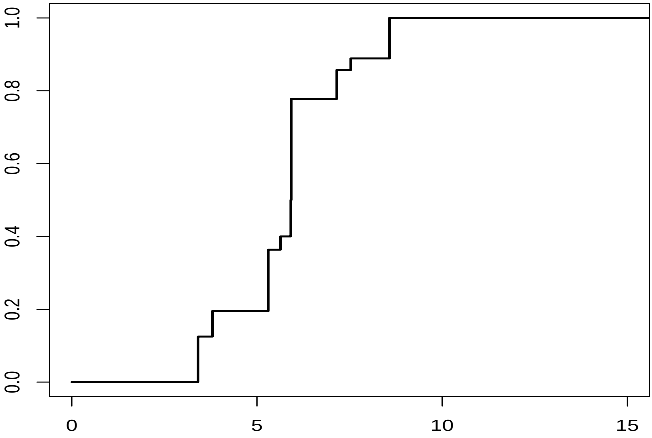
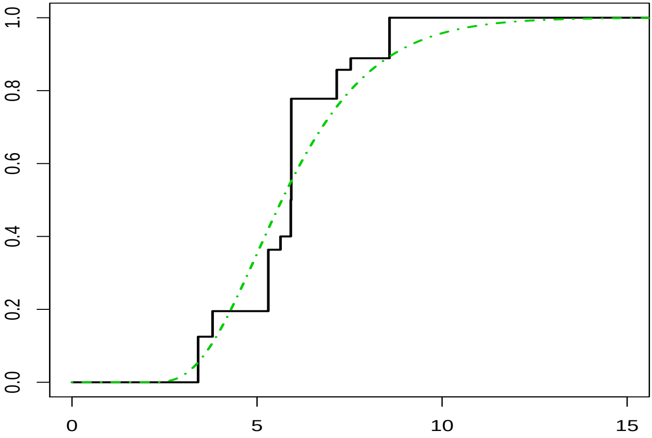


Illustration of $\hat{F}_n$



Naive MSLE

$$l^S(F) = \int \{\delta \log F(t) + (1-\delta)\log(1-F(t))\} d\hat{P}_n(t,\delta)$$

Pointwise maximization of the integrand, by differentiating w.r.t. $F(t)$, leads to

$$\hat{F}_n^{\text{naive}}(t) = \frac{\hat{g}_{n,1}(t)}{\hat{g}_n(t)}, \quad \hat{f}_n^{\text{naive}}(t) = \frac{d}{dt}\frac{\hat{g}_{n,1}(t)}{\hat{g}_n(t)}.$$

Done by M.H. Maathuis (2003), Nonparametric Maximum Likelihood Estimator for Bivariate Censored Data in Continuous Time, Master's Thesis, Delft University of Technology.

Naive MSLE

$$l^S(F) \quad = \quad \int \{\delta \log F(t) + (1-\delta) \log (1 - F(t))\} d\hat{P}_n(t,\delta)$$

Pointwise maximization of the integrand, by differentiating w.r.t. $F(t)$, leads to

$$\hat{F}_n^{\text{naive}}(t) \quad = \quad \frac{\hat{g}_{n,1}(t)}{\hat{g}_n(t)}, \quad \hat{f}_n^{\text{naive}}(t) \quad = \quad \frac{d}{dt} \frac{\hat{g}_{n,1}(t)}{\hat{g}_n(t)}.$$

Advantage:

- Easy to compute numerically.
- Easy to derive asymptotics.

Disadvantage:

- Might decrease locally.

Theorem

Consider the following parameterized curve in $\mathbb{R}_+^2$, which we call the continuous cusum diagram (CCSD):

$$t \mapsto \left(\hat{G}_n(t), \hat{G}_{n,1}(t)\right), \; t \in [0, \tau].$$

Let $\hat{F}_n^{MS}(t)$ be the right-continuous slope of the lower convex hull of the CCSD, evaluated at the point with x-coordinate $\hat{G}_n(t)$. Then

(i) *$\hat{F}_n^{MS}$ maximizes $l^S(F)$ over the class of all sub-distribution functions.*

(ii) *$\hat{F}_n^{MS}$ is the unique maximizer of $l^S(F)$.*

Illustration of $\hat{F}_n^{MS}$

$h = 0.4,\ K(u) = \frac{35}{32}(1-u^2)^3 1_{[-1,1]}(u).$

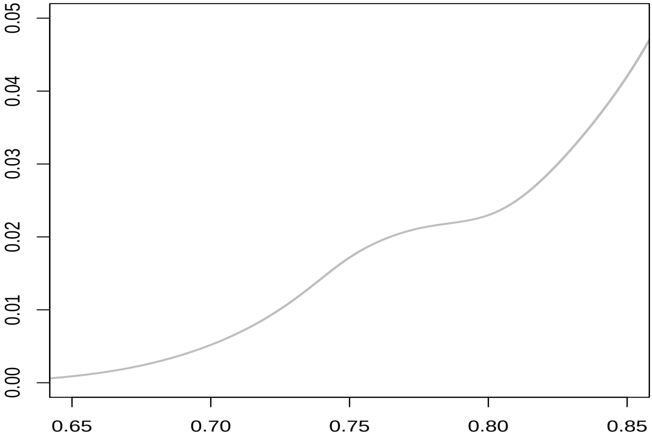


Illustration of $\hat{F}_n^{MS}$

$$h = 0.4,\ K(u) = \frac{35}{32}(1 - u^2)^3 1_{[-1,1]}(u).$$

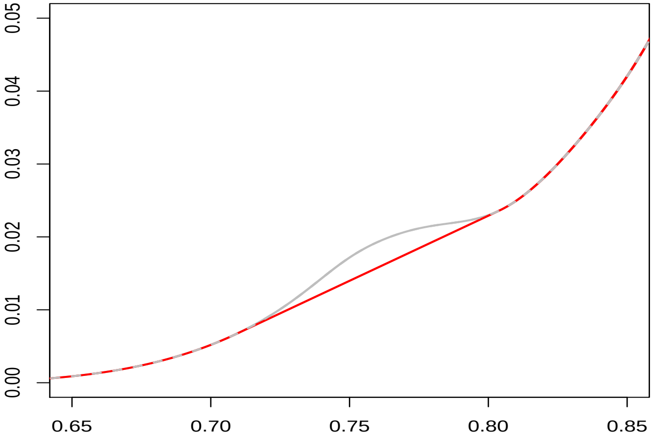


Illustration of $\hat{F}_n^{MS}$

$h = 0.4,\ K(u) = \frac{35}{32}(1-u^2)^3 1_{[-1,1]}(u).$

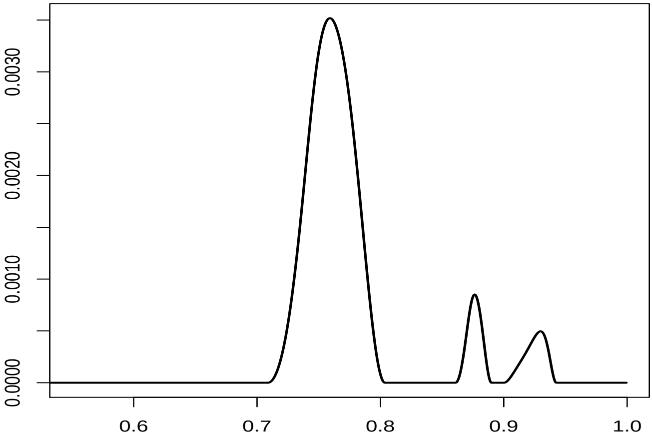


Illustration of $\hat{F}_n^{MS}$

$$h = 0.4,\ K(u) = \tfrac{35}{32}(1 - u^2)^3 1_{[-1,1]}(u).$$

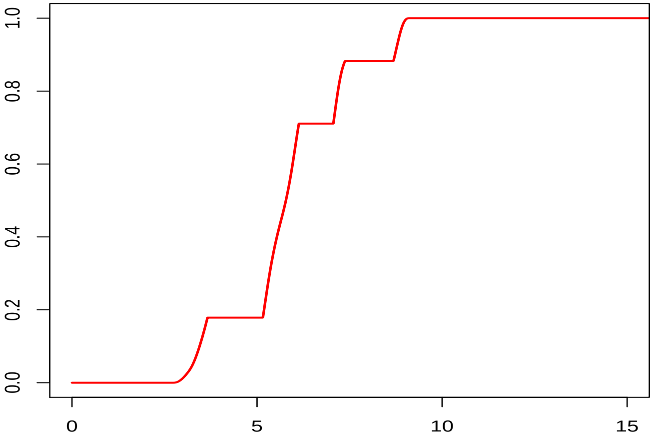


Illustration of $\hat{F}_n^{MS}$

$h = 0.4,\ K(u) = \frac{35}{32}(1-u^2)^3 1_{[-1,1]}(u).$

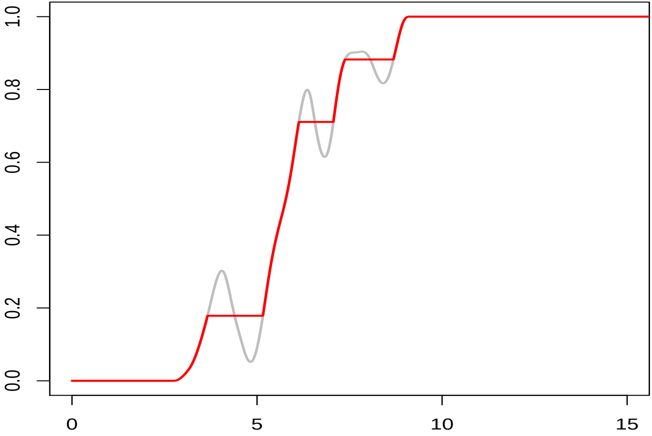


Illustration of $\hat{F}_n^{MS}$

$h = 0.4, \; K(u) = \frac{35}{32}(1-u^2)^3 1_{[-1,1]}(u).$

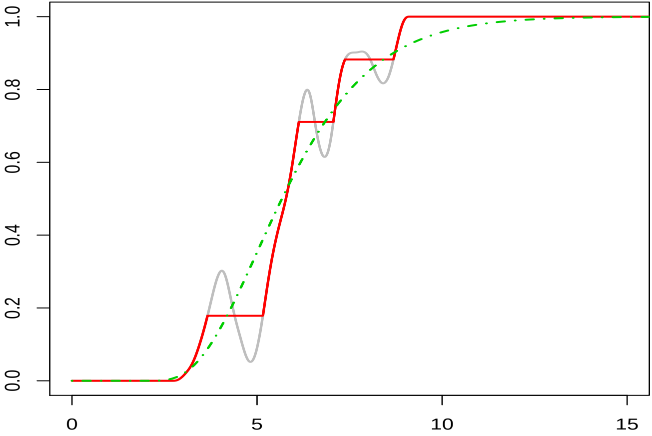


Illustration of $\hat{F}_n^{SM}$

$$h = 0.4,\ K(u) = \frac{35}{32}(1-u^2)^3 1_{[-1,1]}(u)$$

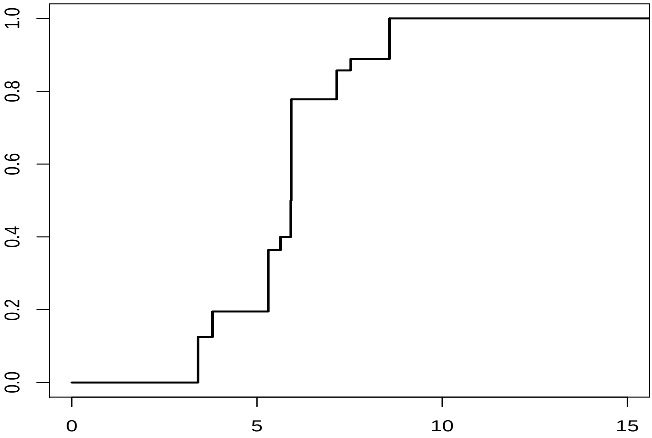


Illustration of $\hat{F}_n^{SM}$

$$h = 0.4,\ K(u) = \frac{35}{32}(1-u^2)^3 1_{[-1,1]}(u)$$

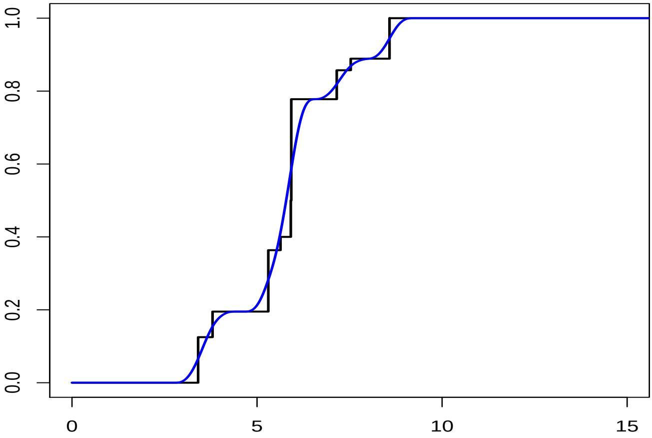


Illustration of $\hat{F}_n^{SM}$

$h = 0.4,\ K(u) = \frac{35}{32}(1-u^2)^3 1_{[-1,1]}(u)$

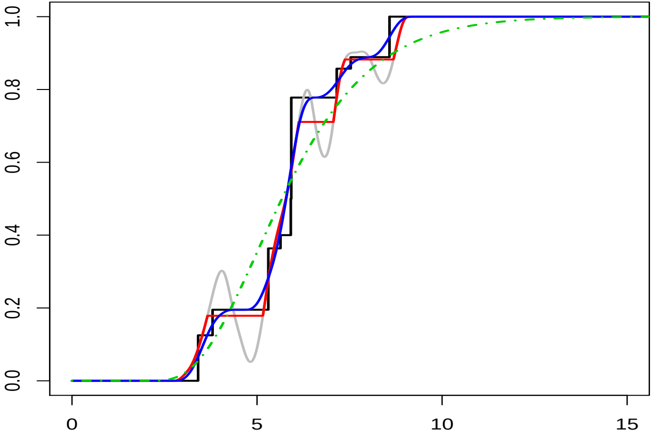


Illustration of $\hat{f}_n^{MS}$

$h = 2.0,\ K(u) = \frac{35}{32}(1-u^2)^3 1_{[-1,1]}(u)$

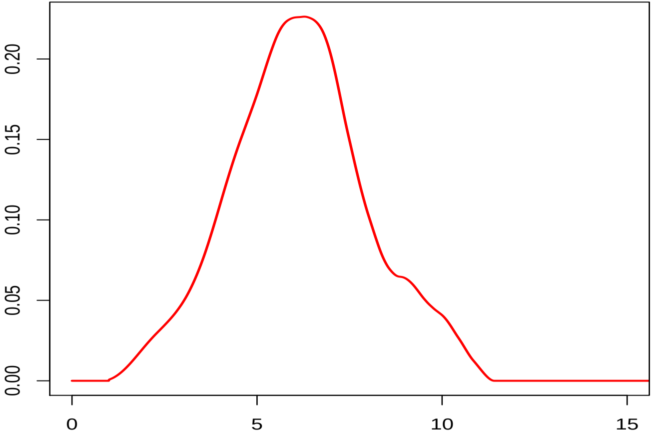


Illustration of $\hat{f}_n^{MS}$

$h = 2.0,\ K(u) = \frac{35}{32}(1-u^2)^3 1_{[-1,1]}(u)$

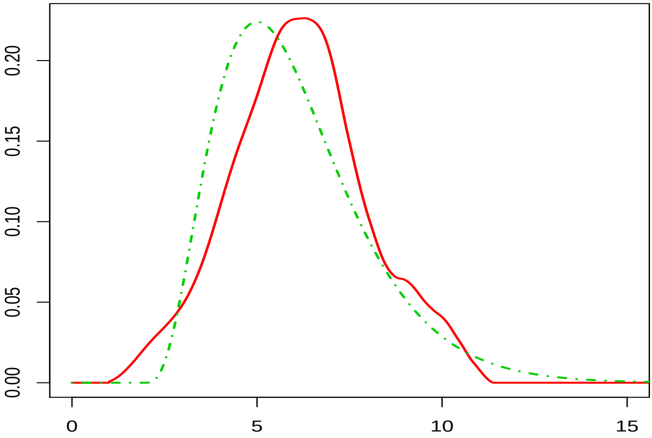


Illustration of $\hat{f}_n^{SM}$

Introduction

The Estimators

Characterization

Asymptotics
distr. function
density

Discussion

$$h = 0.8,\ K(u) = \frac{35}{32}(1-u^2)^3 1_{[-1,1]}(u)$$

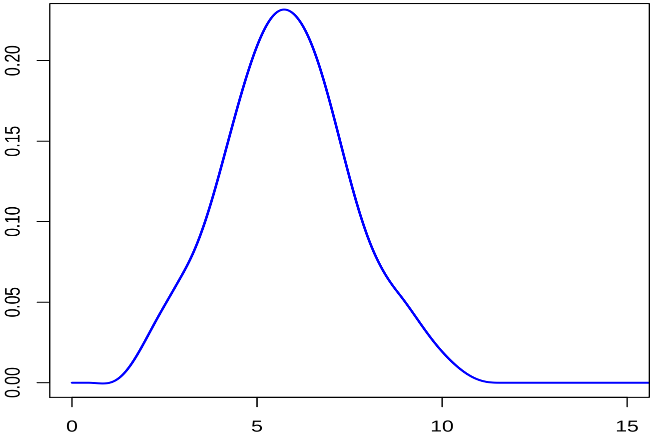


Illustration of $\hat{f}_n^{SM}$

$h = 0.8,\ K(u) = \frac{35}{32}(1-u^2)^3 1_{[-1,1]}(u)$

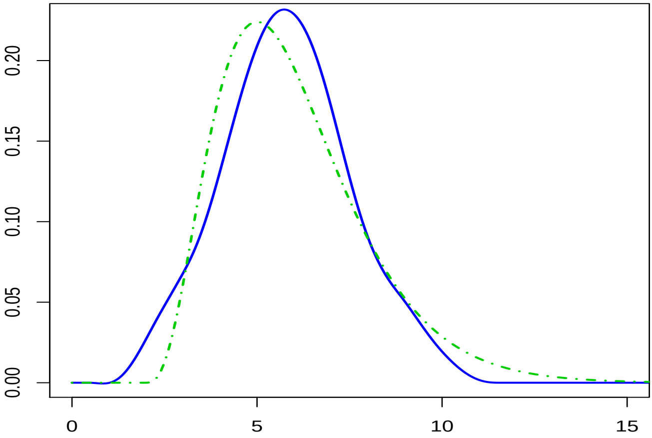
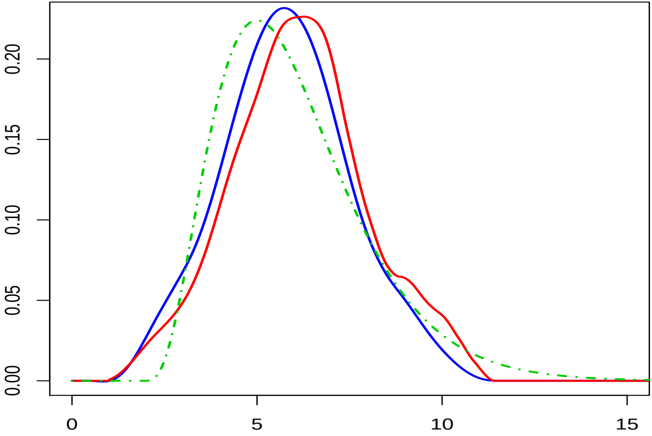


Illustration of $\hat{f}_n^{SM}$

$h = 0.8,\ K(u) = \frac{35}{32}(1-u^2)^3 1_{[-1,1]}(u)$



Overview

- Definition of the Estimators
  - Nonparametric Maximum Likelihood Estimator (NPMLE)
  - Maximum Smoothed Likelihood Estimator (MSLE)
  - Smoothed Maximum Likelihood Estimator (SMLE)
- Characterization and Illustration of the Estimators
- Asymptotic Results
  - Distribution Function
  - Density
- Discussion

Theorem (Groeneboom & Wellner(1992))

Let x be such that $0 < F_0(x), G(x) < 1$. Then we have, under certain conditions,

$$\left(\frac{F_0(x)\big(1 - F_0(x)\big) f_0(x)}{2g(x)}\right)^{-1/3} n^{1/3}\big(\hat{F}_n(x) - F_0(x)\big) \rightsquigarrow 2Z,$$

where Z is the last time where the standard two-sided Brownian motion minus the parabola $y(t) = t^2$ reaches its maximum.

Asymptotics for $\hat{F}_n^{\text{naive}}$

Theorem

Let $h_n = cn^{-\alpha}$ be the smoothing parameter. Under certain conditions it holds that $\forall\ \varepsilon > 0,\ \forall\ M > \varepsilon$ and $\forall\ \alpha \in (0, 1/3)$:

$$P\big(\hat{F}_n^{\text{naive}} \text{ is monotonically increasing on } [\varepsilon, M]\big) \longrightarrow 1.$$

Proof.

Follows from uniform consistency of estimates $\hat{g}_n(t)$, $\hat{g}_{n,1}(t)$, $\hat{g}_n'(t)$ and $\hat{g}_{n,1}'(t)$. □

Corollary

Let $h_n = cn^{-\alpha}$ be the smoothing parameter. Then for all $x > 0$ and $\forall\ \alpha \in (0, 1/3)$ it holds that the estimators $\hat{F}_n^{MS}(x)$ and $\hat{F}_n^{\text{naive}}(x)$ are asymptotically equivalent.

Asymptotics for $\hat{F}_n^{MS}$

Theorem

For fixed $x > 0$ and $h_n = cn^{-1/5}$ we have

$$n^{2/5}\big(\hat{F}_n^{MS}(x) - F_0(x)\big) \quad \rightsquigarrow \quad \mathcal{N}(\mu_F, \sigma_F^2)$$

where

$$\mu_F = \frac{1}{2}c^2 m_2(K) f_0'(x) + c^2 m_2(K)\frac{f_0(x)g'(x)}{g(x)},$$

$$\sigma_F^2 = \frac{F_0(x)\big(1 - F_0(x)\big)}{cg(x)} \int K(y)^2\, dy.$$

This also holds if we replace $\hat{F}_n^{MS}$ by $\hat{F}_n^{\mathrm{naive}}$.

Proof.

A combination of Lindeberg-Feller Central Limit Theorem and the Delta-method. □

Theorem

For fixed $x > 0$ and $h_n = cn^{-1/5}$ we have

$$n^{2/5}\big(\hat{F}_n^{SM}(x) - F_0(x)\big) \quad \rightsquigarrow \quad \mathcal{N}(\nu_F, \tau_F^2),$$

where

$$\nu_F = \frac{1}{2}c^2 m_2(K) f_0'(x)$$

$$\tau_F^2 = \frac{F_0(x)\big(1 - F_0(x)\big)}{cg(x)} \int K(y)^2 dy.$$

Proof.

Based on Empirical Process Theory. □

Theorem

For fixed $x > 0$ and $h_n = cn^{-1/7}$ we have

$$n^{2/7}\big(\hat{f}_n^{MS}(x) - f_0(x)\big) \quad \rightsquigarrow \quad \mathcal{N}(\mu_f, \sigma_f^2)$$

where

$$\begin{aligned}\mu_f \quad &= \quad \frac{1}{2}c^2 m_2(K) f_0''(x) \\ &\qquad + \frac{m_2(K)c^2}{g(x)}\left(g''(x)f_0(x) + g'(x)f_0'(x) - \frac{g'(x)^2 f_0(x)}{g(x)}\right) \\ \sigma_f^2 \quad &= \quad \frac{F_0(x)\big(1 - F_0(x)\big)}{c^3 g(x)} \int K'(y)^2\, dy.\end{aligned}$$

This also holds if we replace $\hat{f}_n^{MS}$ by $\hat{f}_n^{\text{naive}}$.

Proof.

A combination of Lindeberg-Feller Central Limit Theorem and the Delta-method. □

Asymptotics for $\hat{f}_n^{SM}$

Theorem

For fixed $x > 0$ *and* $h_n = cn^{-1/7}$ *we have*

$$n^{2/7}\big(\hat{f}_n^{SM}(x) - f_0(x)\big) \quad \rightsquigarrow \quad \mathcal{N}(\nu_f, \tau_f^2),$$

where

$$\nu_f \quad = \quad \frac{1}{2}c^2 m_2(K) f_0''(x),$$

$$\tau_f^2 \quad = \quad \frac{F_0(x)\big(1 - F_0(x)\big)}{c^3 g(x)} \int K'(y)^2 dy.$$

Proof.

Based on Empirical Process Theory. □

Overview

- Definition of the Estimators
  - Nonparametric Maximum Likelihood Estimator (NPMLE)
  - Maximum Smoothed Likelihood Estimator (MSLE)
  - Smoothed Maximum Likelihood Estimator (SMLE)
- Characterization and Illustration of the Estimators
- Asymptotic Results
  - Distribution Function
  - Density
- Discussion

Discussion

Remarks:

- Taking $h_n = 0$ in both methods for estimating F_0 yields the NPMLE $\hat{F}_n$.
- The asymptotic bias of the estimators $\hat{F}_n^{MS}$ and $\hat{F}_n^{SM}$ differ, but it does not hold that one is always better than the other. The same holds for the estimators $\hat{f}_n^{MS}$ and $\hat{f}_n^{SM}$.
  Also seen in P.N. Patil et al (1994), Some heuristics of kernel based estimators of ratio functions, *Nonparametric Statistics*, **4**, 203–209.
- The asymptotic variance of the estimators $\hat{F}_n^{MS}$ and $\hat{F}_n^{SM}$ are equal, as are the asymptotic variances of $\hat{f}_n^{MS}$ and $\hat{f}_n^{SM}$.

Future work:

- Bandwidthselection methods.
- Current Status Model with Continuous Mark Variable
- Other bivariate censoring Models under the assumption of Stochastic Ordering

Thank you for your attention!

References

Introduction

The Estimators

Characterization

Asymptotics
distr. function
density

Discussion

Groeneboom, P. & Wellner, J.A. (1992).
Information Bounds an Nonparametric Maximum Likelihood Estimation.
New York: Cambridge University Press.

Eggermont, P.P.B. & LaRiccia, V.N. (2001).
Maximum Penalized Likelihood Estimation.
New York: Springer-Verlag.

M.H. Maathuis (2003).
Nonparametric Maximum Likelihood Estimator for Bivariate Censored Data in Continuous Time.
Master's Thesis, Delft University of Technology.

Patil, P.N., Wells, M.T. & Marron, J.S. (1994).
Some heuristics of kernel based estimators of ratio functions.
Nonparametric Statistics, 4, 203–209.