Inverse problems

ETH, October 9, 2007

Piet Groeneboom

TUD/VU/ETH, Zürich/UW, Seattle

The deconvolution problem

Suppose that we have nonnegative observations $Z_1, \ldots, Z_n$ from a distribution with density

$$h_0(z) = \int g(z-x)\, dF_0(x),\ z \geq 0,$$

where g is a known decreasing continuous density on $[0, \infty)$ and F_0 is the distribution function we want to estimate.

F_0 has support, contained in $[0, \infty)$ (i.e., corresponds to nonnegative random variables).

The maximum likelihood estimator (MLE) $\hat{F}_n$ of F_0 is obtained by maximizing the log likelihood

$$\sum_{i=1}^{n} \log \int g(Z_i - x)\, dF(x),$$

over all distribution functions F.

Conjecture in part 2 of Groeneboom and Wellner (1992): at an interior point t of the support of F_0:

$$n^{1/3} \left\{ \hat{F}_n(t) - F_0(t) \right\} \longrightarrow cZ,$$

where Z is the location of the minimum of 2-sided Brownian motion plus a parabola.

An important step in proving the conjectured behavior in Groeneboom and Wellner (1992) is to write the functional

$$\int_{x\in[0,t)} g(t-x)\,d\hat{F}_n(x)$$

in the form

$$g(0)\hat{F}_n(t) - \int_{x\in[0,t)} \{g(t-x)-g(0)\}\,d\hat{F}_n(x), \tag{1}$$

and to show that a centered version of $\int_{x\in[0,t)}\{g(t-x)-g(0)\}\,d\hat{F}_n(x)$ is of lower order than $g(0)\hat{F}_n(t)$: $\int_{x\in[0,t)}\{g(t-x)-g(0)\}\,d\hat{F}_n(x)$ is a so-called smooth functional. Note that

$$g(0)\hat{F}_n(t) = \int_{x\in[0,t)} g(0)\,d\hat{F}_n(x),$$

and that the only crucial difference of the latter integral with the integral in (1) is that the integrand of the integral in (1) is continuous at $x=t$. We want in fact to prove that

$$\int_{x\in[0,t)} \{g(t-x)-g(0)\}\,d\hat{F}_n(x) - \int_{x\in[0,t)} \{g(t-x)-g(0)\}\,dF_0(x) = O_p\left(n^{-1/2}\right),$$

whereas $\hat{F}_n(t)$ itself will have the so-called "cube root" behavior.

Integral equations

Canonical approach: consider the functional

$$K_t(F) = \int_{x\in[0,t)} \{g(t-x) - g(0)\}\, dF(x),$$

and let $a_{t,F}$ solve the (adjoint, see below) equation

$$\begin{aligned}[L_F^*(a)](x) &= E\left\{a_{t,F}(Z) \mid X = x\right\} \\ &= \int_{z\ge x} a_{t,F}(z) g(z-x)\, dz = \{g(t-x) - g(0)\} 1_{[0,t)}(x) - K_t(F),\end{aligned} \tag{2}$$

where $a_{t,F}$ has to be in the range of the score operator:

$$a_{t,F}(z) = [L_F(b)](z) = E_F\left\{b_{t,F}(X) \mid X + Y = z\right\} = \frac{\int_{[0,z]} b_{t,F}(x) g(z-x)\, dF(x)}{h_F(z)}. \tag{3}$$

If we could solve these equations for $\hat{F}_n$, we would have a representation of the following form:

$$K_t(\hat{F}_n) - K_t(F_0) = \int a_{t,\hat{F}_n}(z)\, d\left(H_n - H_0\right)(z).$$

"Argument":

$$\int a_{t,\hat F_n}(z)\,dH_n(z) = \int_{x\in[0,\infty)} \frac{\int_{x\in[0,z]} g(z-x)\,b_{t,\hat F_n}(x)\,d\hat F_n(x)}{h_{\hat F_n}(z)}\,dH_n(z)$$
$$= \int_{x\in[0,\infty)}\int_{z\ge x} \frac{g(z-x)}{\hat h_n(z)}\,dH_n(z)\,b_{t,\hat F_n}(x)\,d\hat F_n(x) = \int_{x\in[0,\infty)} b_{t,\hat F_n}(x)\,d\hat F_n(x) = 0,$$

and

$$\int a_{t,\hat F_n}(z)\,dH_0(z) = \int a_{t,\hat F_n}(z)\int_0^z g(z-x)\,dF_0(x)\,dz$$
$$= \int_{x\in[0,\infty)}\int_{z\ge x} a_{t,\hat F_n}(z)g(z-x)\,dz\,dF_0(x)$$
$$= \int_{x\in[0,\infty)} \Big\{\{g(z-x)-g(0)\}1_{[0,t)}(x) - K_t(\hat F_n)\Big\}\,dF_0(x)$$
$$= \int_{x\in[0,\infty)} \{g(z-x)-g(0)\}1_{[0,t)}(x)\,dF_0(x) - K_t(\hat F_n)$$
$$= K_t(F_0) - K_t(\hat F_n).$$

Unfortunately, there is generally no $b_{t,\hat F_n}$ such that

$$a_{t,\hat F_n}(z) = E_{\hat F_n}\Big\{b_{t,\hat F_n}(X)\;\Big|\;X+Y=z\Big\} = \frac{\int_{[0,z]} b_{t,\hat F_n}(x)g(z-x)\,d\hat F_n(x)}{h_{\hat F_n}(z)}.$$

Solution (first step): We introduce a right-continuous function $B_{t,\hat{F}_n}$ such that

$$\frac{\int_{[0,z]} g(z-x)\, dB_{t,\hat{F}_n}(x)}{h_{\hat{F}_n}(z)} = a_{t,\hat{F}_n}(z), \qquad \lim_{x\to\infty} B_{t,\hat{F}_n}(x) = 0,$$

where $B_{t,\hat{F}_n}$ is no longer absolutely continuous w.r.t. $\hat{F}_n$ and try again:

$$\int a_{t,\hat{F}_n}(z)\, dH_n(z) = \int_{x\in[0,\infty)} \frac{\int_{x\in[0,z]} g(z-x)\, b_{t,\hat{F}_n}(x)\, d\hat{F}_n(x)}{h_{\hat{F}_n}(z)}\, dH_n(z)$$

$$= \int_{x\in[0,\infty)} \int_{z\geq x} \frac{g(z-x)}{\hat{h}_n(z)}\, dH_n(z)\, dB_{t,\hat{F}_n}(x) \stackrel{?}{=} 0$$

Difficulty: characterization of MLE $\hat{F}_n$ tells us that

$$\int_{z\geq x} \frac{g(z-x)}{\hat{h}_n(z)}\, dH_n(z) \quad \begin{array}{l} \geq 1, \\ = 1, \text{ if } x \text{ is a point of mass of } \hat{F}_n. \end{array}$$

Solution (second step): Introduce a function $\bar{B}_{t,\hat{F}_n}$ that is constant on the same intervals as $\hat{F}_n$ and equal to $B_{t,\hat{F}_n}$ at points of mass of $\hat{F}_n$. Then:

$$\int_{x\in[0,\infty)} \int_{z\geq x} \frac{g(z-x)}{\hat{h}_n(z)}\, dH_n(z)\, d\bar{B}_{t,\hat{F}_n}(x) = \lim_{x\to\infty} \bar{B}_{t,\hat{F}_n}(x) = \lim_{x\to\infty} B_{t,\hat{F}_n}(x) = 0,$$

and, hopefully, the following difference will be “small”:

$$\int_{x\in[0,\infty)} \int_{z\geq x} \frac{g(z-x)}{\hat{h}_n(z)}\, dH_n(z)\, d\bar{B}_{t,\hat{F}_n}(x) - \int_{x\in[0,\infty)} \int_{z\geq x} \frac{g(z-x)}{\hat{h}_n(z)}\, dH_n(z)\, dB_{t,\hat{F}_n}(x).$$

Now:

$$K_t(\hat{F}_n) - K_t(F_0) = \int_{[0,t)} \{g(t-x) - g(0)\}\, d\hat{F}_n(x) - \int_{[0,t)} \{g(t-x) - g(0)\}\, dF_0(x)$$
$$= \int a_{t,\hat{F}_n}(z)\, d\left(H_n - H_0\right)(z) + \int_{x\in[0,\infty)} \int_{z\geq x} \frac{g(z-x)}{\hat{h}_n(z)}\, dH_n(z)\, d\left(\bar{B}_{t,\hat{F}_n} - B_{t,\hat{F}_n}\right)(x).$$

Solution (third step): Prove that

$$\int a_{t,\hat{F}_n}(z)\, d\left(H_n - H_0\right)(z) = \int a_{t,F_0}(z)\, d\left(H_n - H_0\right)(z) + o_p(n^{-1/2}).$$

That's the general plan!

Example

We consider $g(x) = 4(1-x)^3 1_{[0,1]}(x)$ and F_0 the Uniform$(0,1)$ distribution. Then we get the following equation in $\phi_{t,F_0}(z) \stackrel{\text{def}}{=} \int_{x\in[0,z)} g(z-x)\,dB_{t,F_0}(x)$:

$$\int_{z=x}^{x+1} a_{t,F_0}(z)\,g(z-x)\,dz = \int_{z=x}^{x+1} \frac{\phi_{t,F_0}(z)}{h_0(z)}\,g(z-x)\,dz = \{g(t-x)-g(0)\}1_{[0,t)}(x) - K_t(F_0). \tag{4}$$

Writing $\phi = \phi_{t,F_0}$ and $B = B_{t,F_0}$ and $a(x) = \phi(x)/h_0(x)$ we get by differentiating:

$$-4a(x) + 12\int_{z=x}^{x+1} a(z)(1+x-z)^2\,dz = 12(1-t+x)^2 \cdot 1_{[0,t)}(x),\ x \neq t, \tag{5}$$

which leads to the following integral equation, using $B(1)=0$,

$$\begin{aligned}
&\frac{B(x) - 3\int_0^x (1+u-x)^2 B(u)\,du}{h_0(x)} - 3\int_{z=x}^{1} \frac{B(z)(1+x-z)^2}{h_0(z)}\,dz \\
&\qquad + 9\int_{u=0}^{x} B(u)\int_{z=x}^{1+u} \frac{(1+u-z)^2(1+x-z)^2 dz}{h_0(z)}\,du \\
&\qquad\qquad + 9\int_{u=x}^{1} B(u)\int_{z=u}^{1+x} \frac{(1+u-z)^2(1+x-z)^2\,dz}{h_0(z)}\,du \\
&= -\tfrac{3}{4}(1-t+x)^2 \cdot 1_{[0,t)}(x),\ x \neq t.
\end{aligned}$$

We can also write this integral equation in the following form:

$$
\begin{aligned}
B(x) - 3\int_0^x B(u)(1+u-x)^2\,du - 3h_0(x)\int_{u=x}^1 \frac{B(u)(1+x-u)^2}{h_0(u)}\,du \\
+ 9h_0(x)\int_{u=0}^x B(u)\int_{z=x}^{1+u} \frac{(1+u-z)^2(1+x-z)^2\,dz}{h_0(z)}\,du \\
+ 9h_0(x)\int_{u=x}^1 B(u)\int_{z=u}^{1+x} \frac{(1+u-z)^2(1+x-z)^2\,dz}{h_0(z)}\,du \\
= -\tfrac{3}{4}(1-t+x)^2 h_0(x)\cdot 1_{[0,t)}(x),\; x\neq t.
\end{aligned}
\tag{6}
$$

Introducing the notation

$$
C_{t,F_0}(x) = C(x) = \frac{B(x)}{h_0(x)},
$$

this can also be written as an integral equation in $C(x)$:

$$
\begin{aligned}
C(x) - \frac{3}{h_0(x)}\int_0^x C(u)(1+u-x)^2\,dH_0(u) - 3\int_{u=x}^1 C(u)(1+x-u)^2\,du \\
+ 9\int_{u=0}^x C(u)\int_{z=x}^{1+u} \frac{(1+u-z)^2(1+x-z)^2\,dz}{h_0(z)}\,dH_0(u) \\
+ 9\int_{u=x}^1 C(u)\int_{z=u}^{1+x} \frac{(1+u-z)^2(1+x-z)^2\,dz}{h_0(z)}\,dH_0(u) \\
= -\tfrac{3}{4}(1-t+x)^2\cdot 1_{[0,t)}(x),\; x\neq t.
\end{aligned}
\tag{7}
$$

Lemma 1. *Let B_{t,F_0} and C_{t,F_0} be the solutions of the integral equation (6) and (7), respectively.*

(i) B_{t,F_0} *is non-positive, bounded and continuous on* $[0,1]$ *and* $B_{t,F_0}(0) = B_{t,F_0}(1) = 0$*. Moreover,* B_{t,F_0} *has a bounded derivative at each point* $x \in (0,1) \setminus \{t\}$*, a jump of size* $\frac{3}{4}h_0(t)$ *at* t*, a finite right derivative at* $x = 0$ *and a left derivative, equal to zero, at* $x = 1$*.*

(ii) C_{t,F_0} *is non-positive and bounded on* $(0,1)$ *with a bounded right limit at* 0 *and a left limit, equal to zero, at* 1*. Moreover,* C_{t,F_0} *has a bounded derivative at each point* $x \in (0,1) \setminus \{t\}$*, a jump of size* $3/4$ *at* t*, a finite right derivative at* $x = 0$ *and a left derivative, equal to zero, at* $x = 1$*.*

(iii) a_{t,F_0} *is bounded on* $(0,2)$ *with a bounded right limit at* 0 *and a left limit, equal to zero, at* 2*. Moreover,* a_{t,F_0} *has a bounded derivative at each point* $z \in (0,2) \setminus \{t\}$*, a jump of size* $3/2$ *at* t*, and finite right and left derivatives at* $z = 0$ *and* $z = 2$*, respectively.*

The current status model

"Hidden space" variables are (T_i, X_i), $T_i, X_i \in \mathbb{R}$, observations are: (T_i, Δ_i).

X_i is independent of T_i, $\Delta_i = 1_{\{X_i \le T_i\}}$. The X_i are (unobservable) "failure times".

(Relevant part of) Log likelihood for the distribution function F of X_i:

$$\sum_{i=1}^{n} \left\{ \Delta_i \log F(T_i) + (1 - \Delta_i) \log\left(1 - F(T_i)\right) \right\}. \tag{8}$$

Define the empirical processes:

$$V_{n1}(t) = n^{-1} \sum_{T_i \le t} \Delta_i, \qquad V_{n2}(t) = n^{-1} \sum_{T_i \le t} (1 - \Delta_i), \qquad t \in \mathbb{R},$$

Then the log likelihood (8) for F, divided by n, can be written:

$$\int \log F(u)\, dV_{n1}(u) + \int \log\{1 - F(u)\}\, dV_{n2}(u). \tag{9}$$

How can we determine the local behavior?

Problem: Unlike in $\sqrt{n}$-asymptotics, we do not have global convergence of the (rescaled) log likelihood process.

The situation is therefore fundamentally different from (1-dimensional) right-censoring, where for example the Kaplan-Meier estimator converges at $\sqrt{n}$-rate and maximizes a process which converges globally after rescaling.

But in the current situation we are lucky: **convex minorant interpretation of the MLE**.

Proposition 1. *Let H_n be the* **greatest convex minorant** *of the (so-called)* **cusum diagram** (or **cumulative sum diagram**), *consisting of the set of points*

$$\mathcal{P}_n = \left\{ \left(\mathbb{G}_n(t), V_{n1}(t) \right), \, t \in \mathbb{R} \right\}, \, V_{n1}(t) = n^{-1} \sum_{i=1}^{n} \Delta_i 1_{(-\infty,t]}(T_i) \tag{10}$$

where $\mathbb{G}_n(t) = n^{-1} \sum_{i=1}^{n} 1_{(-\infty,t]}(T_i)$ is the empirical distribution function of the observation times $T_1, \ldots, T_n$.

Then $\hat{F}_n$ is an MLE if and only if, at each observation point $t = T_i$, $\hat{F}_n(t)$ **is the left derivative of** H_n *at $\mathbb{G}_n(t)$. $\hat{F}_n$ is uniquely determined at each observation point T_i.*

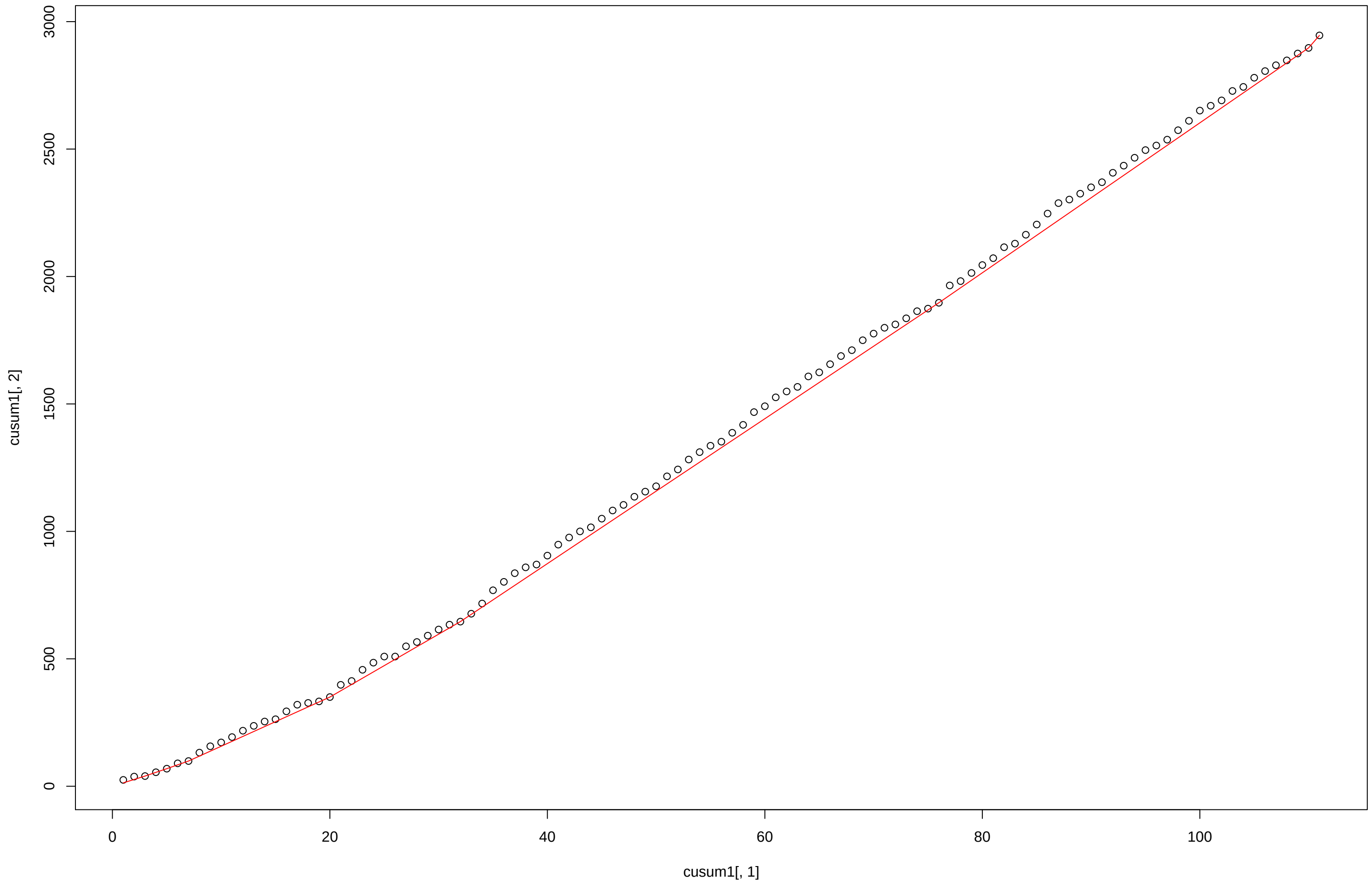

Figure 1: Cusum diagram. Simulation for $n = 20$. Observation df G of the T_i and df F of the X_i are both uniform.

Local asymptotic behavior

Define

$$W_n(t) = n^{-1} \sum_{i=1}^{n} \{\Delta_i - F_0(T_i)\} \left\{1_{\{T_i \le t\}} - 1_{\{T_i \le t_0\}}\right\}, \, t \in \mathbb{R}. \tag{11}$$

Let t_0 be such that $0 < F_0(t_0), G(t_0) < 1$, and let F_0 and G be continuously differentiable at t_0, with strictly positive derivatives $f_0(t_0)$ and $g(t_0)$, respectively.
Then we have a "Kim and Pollard (1990)-type lemma":

$$W_n(t) = O_p\left(n^{-2/3}\right) + o_p\left((t - t_0)^2\right), \text{ uniformly for } |t - t_0| \le \delta. \tag{12}$$

After rescaling, the MLE $\hat{F}_n$ is the **slope of the convex minorant** of the process

$$U_n(t) \stackrel{\text{def}}{=} n^{2/3} W_n\left(t_0 + n^{-1/3} t\right) + n^{-1/3} \sum_{i=1}^{n} \{F_0(T_i) - F_0(t_0\} \left\{1_{\{T_i \le t\}} - 1_{\{T_i \le t_0\}}\right\}, \, t \in \mathbb{R},$$

which converges to two-sided (scaled) Brownian motion with a parabolic drift. We can localize, due to the fact that, for large $|t|$, the drift in the process U_n is dominated by the parabolic drift of

$$n^{-1/3} \sum_{i=1}^{n} \{F_0(T_i) - F_0(t_0\} \left\{1_{\{T_i \le t\}} - 1_{\{T_i \le t_0\}}\right\} \sim \tfrac{1}{2} f_0(t_0) t^2, \, n \to \infty.$$

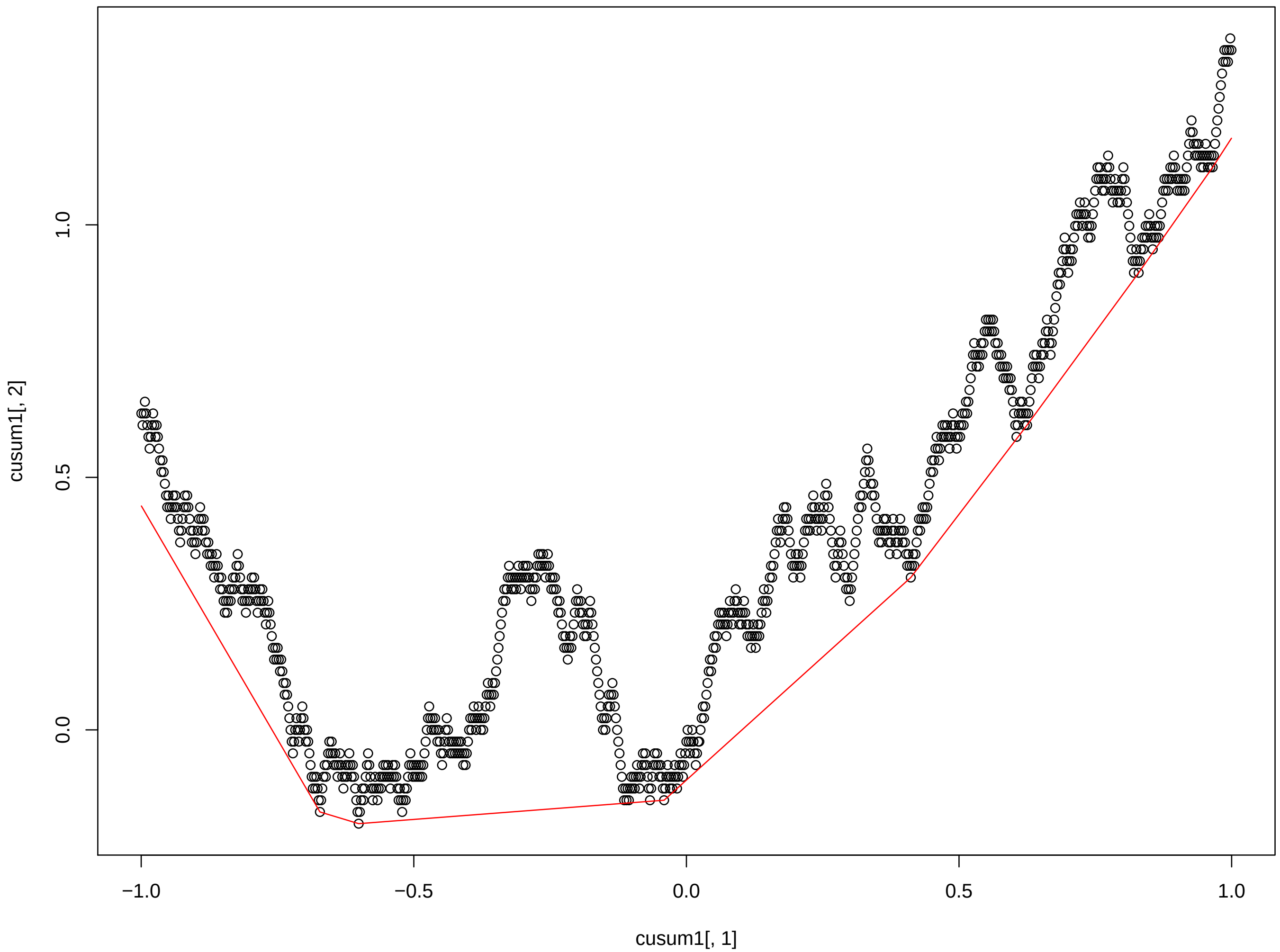

Figure 2: **Locally rescaled cusum process** $\left(n^{1/3}\left(\mathbb{G}_n(t_0+n^{1/3}t)-\mathbb{G}_n(t_0)\right), U_n(t)\right)$, **with convex minorant, for** $n^{1/3}|\mathbb{G}_n(t_0+n^{1/3}t)-\mathbb{G}_n(t_0)| \leq 1$ **and** $t_0 = 0.5$. Simulation with $n = 10,000$. Observation df G of the T_i and df F of the X_i are both **uniform**.

Local limit distribution of MLE $\hat{F}_n$

Using the "Kim and Pollard (1990)-type lemma":

$$W_n(t) = O_p\left(n^{-2/3}\right) + o_p\left((t-t_0)^2\right), \text{ uniformly for } |t-t_0| \le \delta,$$

we can "localize" the convex minorant and hence its derivative process, yielding the MLE $\hat{F}_n$.

Sketch of derivation of local limit distribution:

1. The localized cusum diagram

   $$\left(n^{1/3}\left(\mathbb{G}_n(t_0+n^{1/3}t)-t_0\right), U_n(t)\right)$$

   converges in distribution to the Brownian motion cusum diagram:

   $$(g(t_0)t, U(t)), \text{ where } U(t) = \sqrt{g(t_0)F_0(t_0)\{1-F_0(t_0)\}}\,W(t) + \tfrac{1}{2}f_0(t_0)g(t_0)t^2,\ t \in \mathbb{R},$$

   and where W is two-sided Brownian motion.

2. Continuous mapping theorem: Convex minorant of localized cusum diagram converges in distribution to convex minorant of Brownian cusum diagram.

3. Left-derivative of convex minorant of localized cusum diagram converges in distribution (in Skohorod topology) to left-derivative of convex minorant of Brownian cusum diagram and $n^{1/3}\{\hat{F}_n(t_0)-F_0(t_0)\}$ **is left-derivative of convex minorant of localized cusum diagram at zero**.

Competing risk model

Generalization of the current status model to the situation where there are more failure causes.

Hidden space variables are (T_i, X_i, Y_i), Y_i is the failure cause.

Observations are: $(T_i, \Delta_{i1}, \ldots, \Delta_{iK})$, $\Delta_{ik} = 1_{\{X_i \le T_i, Y_i = k\}}$. Define:

$$V_{nk}(t) = n^{-1} \sum_{i=1}^{n} \Delta_{ik} 1_{(-\infty,t]}(T_i),\ V_{n,K+1}(t) = n^{-1} \sum_{i=1}^{n} (1 - \Delta_{i+})\, 1_{(-\infty,t]}(T_i),\ \Delta_{i+} = \sum_{k=1}^{K} \Delta_{ik},$$

We want to estimate the subdistribution functions F_{0k}:

$$F_{0k}(t) = P\left\{X \le t, Y = k\right\},\ k = 1, \ldots, K.$$

The (relevant part of the) log likelihood for $F = (F_1, \ldots, F_K)$, divided by n, is:

$$\sum_{k=1}^{K} \int \log F_k(u)\, dV_{nk}(u) + \int \log\{1 - F_+(u)\}\, dV_{n,K+1}(u), \qquad F_+ = \sum_{k=1}^{K} F_k.$$

The MLE (maximum likelihood estimator) $\hat{F}_n = (\hat{F}_{n1}, \ldots, \hat{F}_{nK})$ can only be computed iteratively.

No direct convex minorant interpretation, as with the MLE for current status data.

Self-induced characterization

The MLE $\hat{F}_{nk}$ can be characterized as the left derivative of the greatest convex minorant of the self-induced cusum diagram

$$\mathcal{P}_{nk} = \left\{ \left(G_{\hat{F}_{n+}}(t), V_{nk}(t) \right),\, t \in \mathbb{R} \right\},\ V_{nk}(t) = n^{-1} \sum_{i=1}^{n} \Delta_{ik} 1_{(-\infty,t]}(T_i), \tag{13}$$

for $k = 1, \ldots, K$, where

$$G_{\hat{F}_{n+}}(t) = n^{-1} \sum_{i=1}^{n} \frac{1 - \Delta_{i+}}{1 - \hat{F}_{n+}(T_i)} 1_{(-\infty,t]}(T_i),\ t < T_{(n)}.$$

Compare with ordinary current status, where $\hat{F}_n$ is the left derivative of the greatest convex minorant of the (not self-induced) cusum diagram

$$\mathcal{P}_n = \left\{ \left(\mathbb{G}_n(t), V_{n1}(t) \right),\, t \in \mathbb{R} \right\},\ V_{n1}(t) = n^{-1} \sum_{i=1}^{n} \Delta_i 1_{(-\infty,t]}(T_i). \tag{14}$$

Note:

$$G_{\hat{F}_{n+}}(t) = \mathbb{G}_n(t) + n^{-1} \sum_{i=1}^{n} \frac{\hat{F}_{n+}(T_i) - \Delta_{i+}}{1 - \hat{F}_{n+}(T_i)} 1_{(-\infty,t]}(T_i),\ t < T_{(n)}.$$

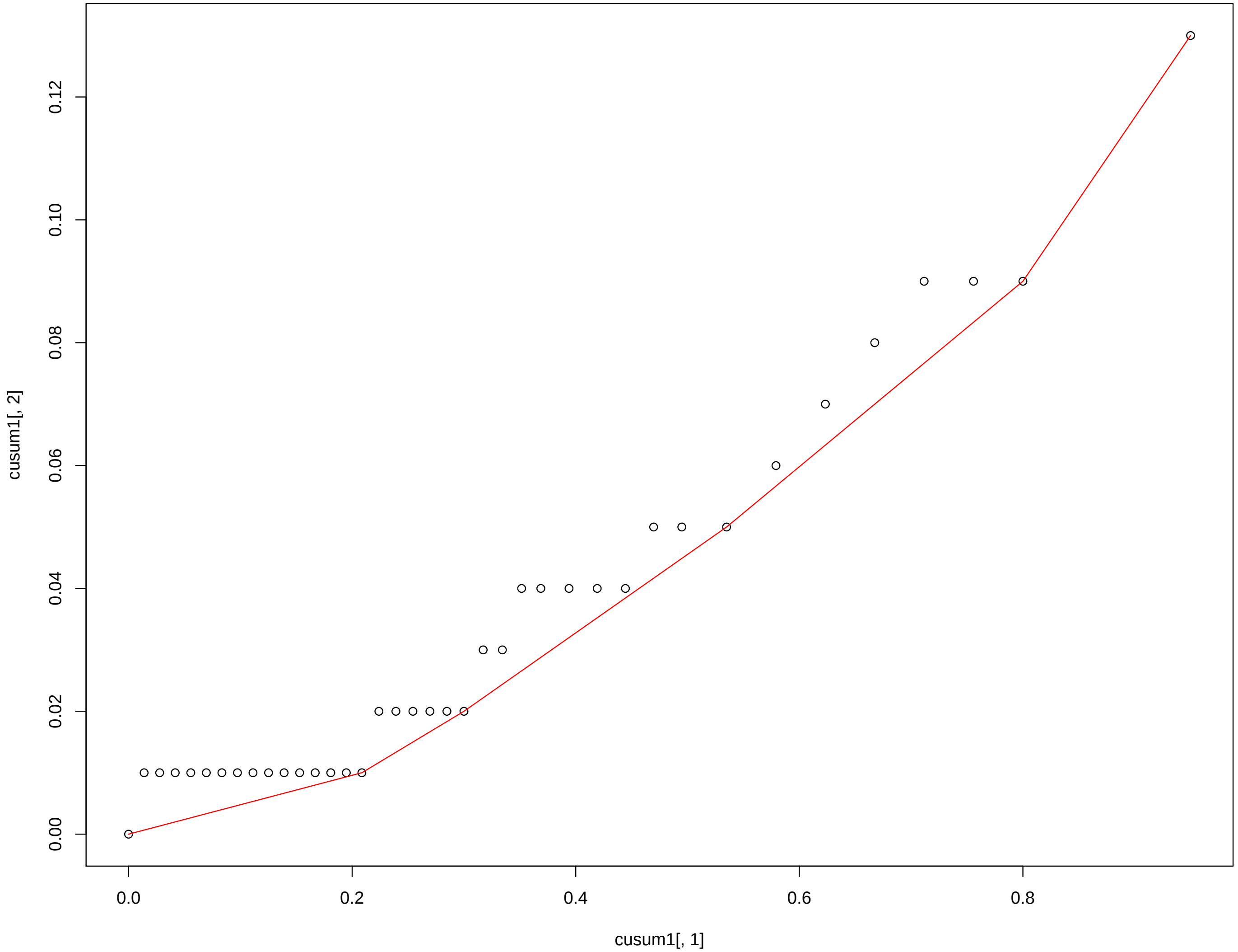

Figure 3: **Cusum diagram** $\{(G_{\hat{F}_{n+}}(t), V_{n1}(t)),\ t \in \mathbb{R}\}$. Simulation for $n = 100$, $K = 2$. $F_{0k}(t) = (k/3)\{1 - e^{-kt}\}$, $T \sim \text{Unif}(0, 1.5)$.

Some history of work on the local rate

End of 2004: The following fact was proved.

Lemma 2. *Let* $\hat F_n = (\hat F_{n1}, \ldots, \hat F_{nK})$ *be the MLE of* $F_0 = (F_{01}, \ldots, F_{0K})$, *and let* $\hat F_{n+} = \sum_{k=1}^K \hat F_{nk}$ *and, similarly,* $F_{0+} = \sum_{k=1}^K F_{0k}$. *Moreover, let, for a* $\delta \in (0,1)$, $\big[t_0 - \delta, t_0 + \delta\big]$ *be an interval on which the components* F_{0k} *have continuous derivatives staying away from zero. Then there exists for each* $\varepsilon > 0$ *and* $M > 0$ *an* $M_1 > 0$ *so that*

$$\mathbb{P}\left\{\sup_{t\in[-M,M]} n^{1/3}\left|\hat F_{n+}\big(t_0 + n^{-1/3}t\big) - F_{0+}\big(t_0\big)\right| > M_1\right\} < \varepsilon,\ k = 1, \ldots, K.$$

This is not enough! To get localization of the $\hat F_{nk}$, we need that for t outside a neighborhood of order $O(n^{-1/3})$ of t_0 we can replace $G_{\hat F_{n+}}(t)$ by

$$G_{F_{0+}}(t) \stackrel{\text{def}}{=} \mathbb{G}_n(t) + n^{-1}\sum_{i=1}^n \frac{F_{0+}(T_i) - \Delta_{i+}}{1 - F_{0+}(T_i)} 1_{(-\infty,t]}(T_i)\,,$$

up to terms of order $o_p((t-t_0)^2)$ in the self-induced cusum diagram:

$$\mathcal{P}_{nk} = \left\{\big(G_{\hat F_{n+}}(t), V_{nk}(t)\big),\ t \in \mathbb{R}\right\},\ V_{nk}(t) = n^{-1}\sum_{i=1}^n \Delta_{ik} 1_{(-\infty,t]}(T_i), \tag{15}$$

removing the self-inducedness of the coordinate $G_{\hat F_{n+}}$.

Solution in September 2005: strengthening of Kim-Pollard-type lemma.

Lemma 3. (**Global to local lemma for $\hat{F}_{n+}$**) *Let $\hat{F}_n$ be the MLE and let, for a $\delta \in (0,1)$, $\left[t_0 - 2\sqrt{\delta}, t_0 + 2\sqrt{\delta}\right]$ be an interval on which the components F_{0k} have continuous derivatives staying away from zero. Then, for all $t \in [t_0 - \delta, t_0 + \delta]$ we have:*

$$\int_{\{|u-t_0|<|t-t_0|\}} \frac{|\hat{F}_{n+}(u) - F_{0+}(u|}{1 - \hat{F}_{n+}(u)}\, d\mathbb{G}_n(u) = n^{-1/6} O_p\left(n^{-1/2} \vee |t-t_0|^{3/2}\right), \text{ uniformly in } t \in [t_0-\delta, t_0+\delta].$$

Note: $n^{-1/6} O_p\left(n^{-1/2} \vee |t-t_0|^{3/2}\right) = O_p(n^{-2/3})$ if $|t-t_0| = O_p(n^{-1/3})$.

Corollary 1. (**Tightness of $n^{1/3}\{\hat{F}_n(t_0 + n^{-1/3}t) - F_0(t_0)\}$**) *Let the conditions of Lemma 3 be satisfied. Then:*

(i) (Replacement of $G_{\hat{F}_{n+}}$ by $G_{F_{0+}}$)

$$G_{\hat{F}_{n+}}(t) = G_{F_{0+}}(t) + n^{-1/6} O_p\left(n^{-1/2} \vee |t-t_0|^{3/2}\right), \text{ uniformly in } t \in [t_0 - \delta, t_0 + \delta].$$

(ii) For each $\varepsilon > 0$ and $M > 0$ there exists an $M_1 > 0$ so that

$$\mathbb{P}\left\{ \sup_{t \in [-M,M]} n^{1/3} \left| \hat{F}_{nk}\left(t_0 + n^{-1/3}t\right) - F_{0k}\left(t_0\right) \right| > M_1 \right\} < \varepsilon,\ k = 1, \ldots, K.$$

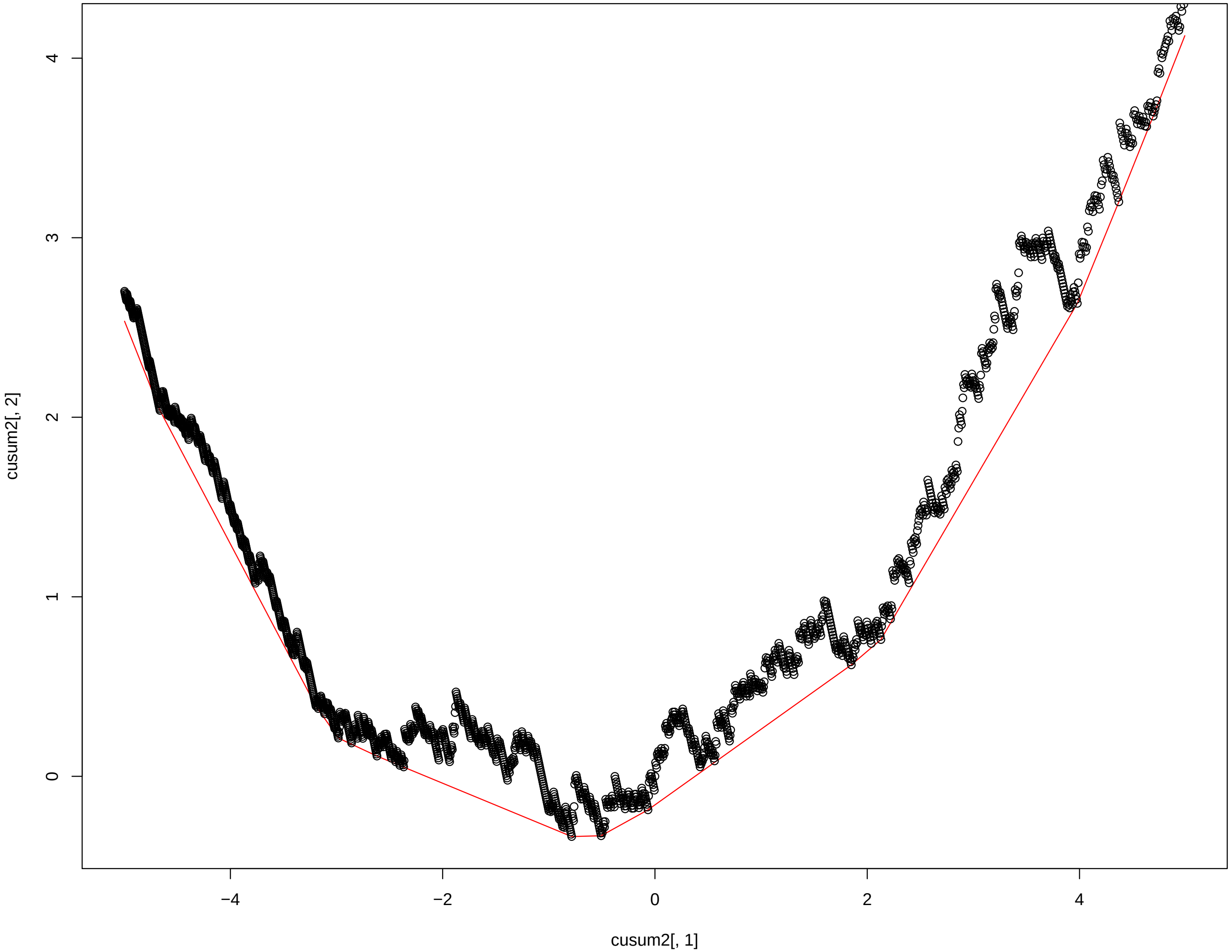

Figure 4: Localized cusum diagram $\left\{\left(n^{1/3}(G_{\hat{F}_{n+}}(t) - G_{\hat{F}_{n+}}(t_0)), n^{2/3}\left\{V_{n1}(t) - V_{n1}(t_0) - \int_{t_0}^{t} F_{01}(t_0)\, dG_{\hat{F}_{n+}}(u)\right\}\right), t \in \mathbb{R}\right\}$, $t_0 = 0.5, n = 10,000$. Red curve: $n^{2/3}\int_{t_0}^{t}\{\hat{F}_{n1}(t_0) - F_{01}(u)\}\, dG_{\hat{F}_{n+}}(u)$. If $t < t_0$: $\int_{t_0}^{t}\{\hat{F}_{n1}(u) - F_{01}(t_0)\}\, dG_{\hat{F}_{n+}}(u) \stackrel{\text{def}}{=} -\int_{[t,t_0]}\{\hat{F}_{n1}(u) - F_{01}(t_0)\}\, dG_{\hat{F}_{n+}}(u)$.

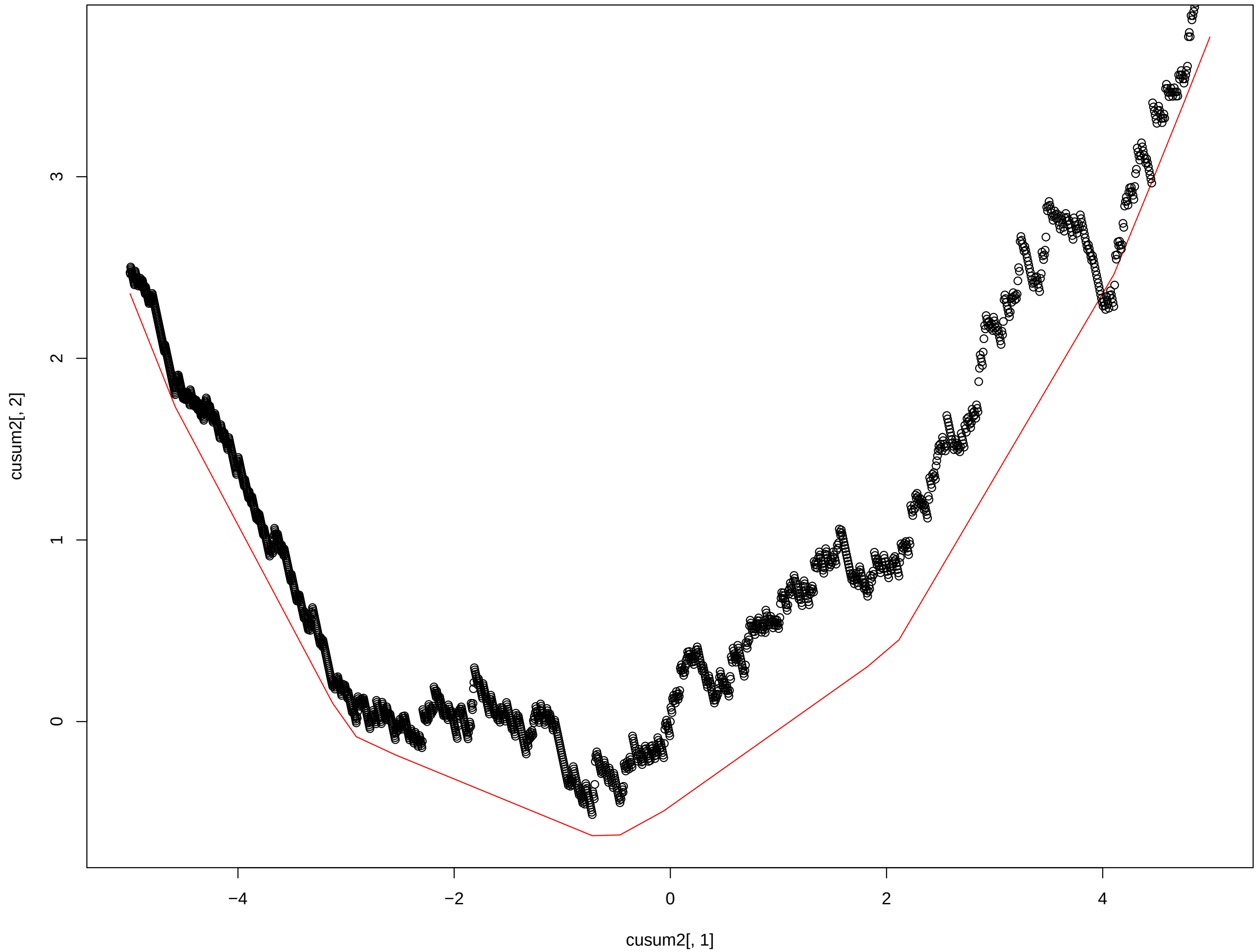

Figure 5: Localized cusum diagram with $G_{\hat{F}_{n+}}$ replaced by $G_{F_{0+}}$, $t_0 = 0.5, n = 10,000$. Red curve: $n^{2/3}\int_{t_0}^{t}\{\hat{F}_{n1}(u) - F_{01}(t_0)\}\,dG_{F_{0+}}(u)$. If $t < t_0$: $\int_{t_0}^{t}\{\hat{F}_{n1}(u) - F_{01}(t_0)\}\,dG_{F_{0+}}(u) \stackrel{\text{def}}{=} -\int_{[t,t_0]}\{\hat{F}_{n1}(u) - F_{01}(t_0)\}\,dG_{F_{0+}}(u)$.

Summary of the global to local argument

The MLE maximizes a global criterion. To extract the local limit behavior from this, we have to use some kind of characterization of the solution, for example a convex duality criterion.

1. In the case of simple current status data, this leads to a convex minorant characterization, which can be used for the determining the local behavior of the MLE.

2. In the case of competing risk with current status data, this leads to a self-induced convex minorant characterization, involving the sum $\hat{F}_{n+}$ of the individual MLE estimators $\hat{F}_{nk}$ for the several subdistribution functions F_{0k}. To get localization of the $\hat{F}_{nk}$, we need that for t outside a neighborhood of order $O(n^{-1/3})$ of t_0 we can replace $G_{\hat{F}_{n+}}(t)$ by

$$G_{F_{0+}}(t) \stackrel{\text{def}}{=} \mathbb{G}_n(t) + n^{-1} \sum_{i=1}^{n} \frac{F_{0+}(T_i) - \Delta_{i+}}{1 - F_{0+}(T_i)} 1_{(-\infty,t]}(T_i)\,,$$

up to terms of order $o_p((t-t_0)^2)$ in the self-induced cusum diagram:

$$\mathcal{P}_{nk} = \left\{ \left(G_{\hat{F}_{n+}}(t), V_{nk}(t)\right),\, t \in \mathbb{R} \right\},\, V_{nk}(t) = n^{-1} \sum_{i=1}^{n} \Delta_{ik} 1_{(-\infty,t]}(T_i), \tag{16}$$

to get rid of the self-inducedness of the coordinate $G_{\hat{F}_{n+}}$ in the tightness argument.
This is accomplished by the global to local lemma.

Limit distribution for competing risk

- First prove uniqueness of the limiting process, using tightness argument (Hardest part!)
- Localize characterization of limit process.
- Take subsequences of localized processes, based on a samples of size n, on $[-m, m]$. By tightness (using local rate result) there is a further subsequence that converges to some limit. Using a diagonal argument, it follows that there is a limit on $\mathbb{R}$. Here we go from local to global!
- By the continuous mapping theorem the limit must satisfy the limit characterization on $[-m, m]$ for each $m \in \mathbb{N}$.
- Letting $m \to \infty$ gives existence of the limiting process (almost for free!)
- By uniqueness of the limiting process, all subsequences converge to the same limit

References

Geskus, R.B. and Groeneboom, P. (1996). Asymptotically optimal estimation of smooth functionals for interval censoring, part 1. Statistica Neerlandica, **50**, 69-88.

Groeneboom, P. and Wellner, J.A. (1992). *Information bounds and nonparametric maximum likelihood estimation*, Birkhäuser Verlag.

Groeneboom, P., Maathuis, M.H. and Wellner, J.A. (2007). Current status data with competing risks: Consistency and rates of convergence of the MLE. To appear in the Annals of Statistics.

Groeneboom, P., Maathuis, M.H. and Wellner, J.A. (2007). Current status data with competing risks: Limiting distribution of the MLE. To appear in the Annals of Statistics.

Kress R. (1989). *Linear integral equations*, Applied Mathematical Sciences vol. 82, Springer Verlag, New York.