Nonparametric estimation under Shape Restrictions

Jon A. Wellner

University of Washington, Seattle

Statistical Seminar, Frejus, France
August 30 - September 3, 2010
Outline: Five Lectures on Shape Restrictions

- **L1**: Monotone functions: maximum likelihood and least squares
- **L2**: Optimality of the MLE of a monotone density (and comparisons?)
- **L3**: Estimation of convex and $k$–monotone density functions
- **L4**: Estimation of log-concave densities: $d = 1$ and beyond
- **L5**: More on higher dimensions and some open problems
Outline: Lecture 1

- A: Maximum likelihood and least squares estimators (and more?)
- B: Switching: a simple key result
- C: Limit theory via switching and argmax continuous mapping
- D: Complements: Pollard’s localization method ??
- E: Other nonparametric function estimation problems ??
A. Maximum likelihood, monotone density

• Model: \( \mathcal{D} \equiv \) all monotone decreasing densities (wrt Lebesgue measure) on \( \mathbb{R}^+ = (0, \infty) \).

• Observations: \( X_1, \ldots, X_n \) i.i.d. \( f_0 \in \mathcal{D} \).

• MLE: \( \hat{f}_n \equiv \arg\max_{f \in \mathcal{D}} \{ \sum_{i=1}^{n} \log f(X_i) \} \)

• LSE: \( \tilde{f}_n \equiv \arg\min_{f \in \mathcal{D}} \psi_n(f) \)

where

\[
\psi_n(f) \equiv \frac{1}{2} \int_{0}^{\infty} f^2(x) \, dx - \int_{0}^{\infty} f(x) \, d\mathbb{F}_n(x) \\
= \frac{1}{2} \left\{ \int_{0}^{\infty} (f^2(x) - f_n(x))^2 \, dx - \int_{0}^{\infty} f_n^2(x) \, dx \right\}
\]

if \( \mathbb{F}_n \) had density \( f_n \) (which it doesn’t, of course!).
A. Maximum likelihood, monotone density

Theorem. (a) \( \hat{f}_n = \tilde{f}_n \) exists and is unique. It is a piecewise constant function with jumps (down) only at the order statistics. (b) The MLE \( \hat{f}_n \) is characterized by the “Fenchel” conditions

\[
\mathbb{F}_n(x) \leq \hat{F}_n(x) \equiv \int_0^x \hat{f}_n(t) dt \quad \text{for all } x \geq 0, \text{ and}
\]

\[
\mathbb{F}_n(x) = \hat{F}_n(x) \text{ if and only if } \hat{f}_n(x^-) > \hat{f}_n(x^+).
\]

The equality condition in the last display can be rewritten as

\[
\int_0^{\infty} (\hat{F}_n(x) - \mathbb{F}_n(x)) df_n(x) = 0.
\]

(c) Geometrically, \( \hat{f}_n \) is the left-derivative at \( x \) of the least concave majorant \( \hat{F}_n \) of \( \mathbb{F}_n \).
A. Maximum likelihood, monotone density
A. Maximum likelihood, monotone density
A. Maximum likelihood, monotone density

Proof; Existence and Uniqueness: The log-likelihood function (divided by $n$) is $L_n(f) = \frac{1}{n} \log f = n^{-1} \sum_{i=1}^{n} \log f(X_i)$. If we define $\tilde{f}$ by $\tilde{f}(x) = C \sum_{i=1}^{n} f(X_i) 1_{(X_{i-1}, X_i]}(x)$ where $C$ is a normalizing constant to make $\int_{0}^{\infty} \tilde{f}(x) dx = 1$, then

$$L_n(\tilde{f}) = \log C + L_n(f) \geq L_n(f)$$

since

$$1 = \int_{0}^{\infty} \tilde{f}(x) dx = C \sum_{i=1}^{n} (X_i - X_{i-1}) f(X_i) \leq C \int_{0}^{X(n)} f(x) dx \leq C.$$ 

Thus the MLE $\hat{f}_n$ can be taken to be a histogram type estimator with breaks only at the order statistics.

Existence follows since we can restrict the maximization of $L_n$ to the compact set

$$\mathcal{D}_M \equiv \{ f \in \mathcal{D} : f \text{ a histogram, } f(0) \leq M, f(M) = 0 \}$$

for $M = \max\{1/X(1), 2X(n)\}$. 

Statistical Seminar, Fréjus
A. Maximum likelihood, monotone density

Proof; Characterization: Let $\mathcal{M} = \{ f : f(x) \geq 0 \text{ for all } x \geq 0, \ f \downarrow \}$. Then $\mathcal{D} \subset \mathcal{M}$ and $\mathcal{M}$ is a convex cone. We replace maximization of the log-likelihood

$$P_n \log f = n^{-1} \sum_{i=1}^{n} \log f(X_i) = \int_{0}^{\infty} \log f(x) d\mathbb{F}_n(x)$$

over $\mathcal{D}$ by minimization of

$$\ell_n(f) \equiv -P_n \log f + \int_{0}^{\infty} f(x) dx \text{ over } \mathcal{M}.$$ 

Suppose $\hat{f}_n$ minimizes $-P_n \log f$ over $\mathcal{D}$. Then $\hat{f}_n$ minimizes $\ell_n(f)$ over $\mathcal{M}$. To see this, let $g \in \mathcal{M}$ with $\int_{0}^{\infty} g(x) dx = c \in (0, \infty)$. Since $g/c \in \mathcal{D}$

$$\ell_n(g) - \ell_n(\hat{f}_n) = -P_n \log(g/c) - \log c + c + P_n \log \hat{f}_n - 1$$

$$= \ell_n(g/c) - \ell_n(\hat{f}_n - \log c - 1 + c$$

$$\geq 0 + 0 = 0$$

since $g/c \in \mathcal{D}$ and $c - 1 \geq \log c$. Equality holds if $g = \hat{f}_n$. Thus $\hat{f}_n$ maximizes $\ell_n$ over $\mathcal{M}$. 
A. Maximum likelihood, monotone density

Now for \( g \in \mathcal{M} \) and \( \epsilon > 0 \) consider

\[
\ell_n(\hat{f}_n + \epsilon g) \geq \ell_n(\hat{f}_n).
\]

Thus

\[
0 \leq \lim_{\epsilon \downarrow 0} \frac{\ell_n(\hat{f}_n + \epsilon g) - \ell_n(\hat{f}_n)}{\epsilon}
= -\int_{0}^{\infty} \frac{g(x)}{\hat{f}_n(x)} d\mathbb{F}_n(x) + \int_{0}^{\infty} g(x) dx
= -\int_{0}^{\infty} \frac{1_{[0,y]}(x)}{\hat{f}_n(x)} d\mathbb{F}_n(x) + y \quad \text{for all } y > 0
\]

by taking \( g(x) = 1_{[0,y]}(x) \)

\[
= y - \int_{0}^{y} \frac{1}{\hat{f}_n(x)} d\mathbb{F}_n(x)
= \int_{0}^{y} \frac{1}{\hat{f}_n} d(\hat{F}_n - \mathbb{F}_n). \tag{1}
\]
A. Maximum likelihood, monotone density

If \( y \) satisfies \( \hat{f}_n(y-) > \hat{f}_n(y+) \), then the function \( \hat{f}_n + \epsilon 1_{[0,y]} \in \mathcal{M} \) for \( \epsilon < 0 \) and \( |\epsilon| \) sufficiently small. Repeating the argument for \( \epsilon < 0 \) and these values of \( y \) yields

\[
0 = \int_0^y \frac{1}{\hat{f}_n} d(\hat{F}_n - F_n) \quad \text{if} \quad \hat{f}_n(y-) > \hat{f}_n(y+). \tag{2}
\]

Since \( \hat{f}_n \) is piecewise constant, the inequalities and equalities in (1) and (2) can be rewritten as claimed:

\[
\mathbb{P}_n(x) \leq \hat{F}_n(x) = \int_0^x \hat{f}_n(t)dt \quad \text{for all} \quad x \geq 0, \quad \text{and}
\]

\[
\mathbb{P}_n(x) = \hat{F}_n(x) \quad \text{if and only if} \quad \hat{f}_n(x-) > \hat{f}_n(x+).
\]

Now consider the LSE \( \tilde{f}_n \). Suppose that \( \tilde{f}_n \) minimizes

\[
\psi_n(f) = \frac{1}{2} \int_0^\infty f^2(x)dx - \int_0^\infty f d\mathbb{P}_n
\]

over \( \mathcal{M} \).
A. Maximum likelihood, monotone density

Then for $g \in \mathcal{M}$ and $\epsilon > 0$ we have $\psi_n(\tilde{f}_n + \epsilon g) \geq \psi_n(\tilde{f}_n)$ and hence

$$0 \leq \lim_{\epsilon \downarrow 0} \frac{\psi_n(\tilde{f}_n + \epsilon g) - \psi_n(\tilde{f}_n)}{\epsilon}$$

$$= \int_{0}^{\infty} g(x) \tilde{f}_n(x) dx - \int_{0}^{\infty} g d\mathbb{F}_n = \int_{0}^{\infty} g d(\tilde{F}_n - \mathbb{F}_n)$$

$$= \int_{0}^{y} d(\tilde{F}_n - \mathbb{F}_n) = \tilde{F}_n(y) - \mathbb{F}_n(y) \text{ for all } y > 0 \quad (3)$$

by choosing $g(x) = 1_{[0,y]}(x)$ for $x \geq 0$, $y > 0$. If $\tilde{f}_n(y-) > \tilde{f}_n(y+)$, then $\tilde{f}_n + \epsilon 1_{[0,y]} \in \mathcal{M}$ for $\epsilon < 0$ with $|\epsilon|$ small, so repeating the argument for $\epsilon < 0$ and these $y$'s yields

$$\tilde{F}_n(y) - \mathbb{F}_n(y) = 0 \text{ if } \tilde{f}_n(y-) > \tilde{f}_n(y+). \quad (4)$$

But (3) and (4) give exactly the same characterization of $\tilde{f}_n$ derived above for $\hat{f}_n$. Thus $\tilde{f}_n = \hat{f}_n$ in this case.
B. Switching: a simple key result

- Groeneboom (1985), Prakasa Rao (1969)?
- Introduce first in the context of $\hat{f}_n$
- More general version.

**Switching for $\hat{f}_n$:** Define

$$\hat{s}_n(a) \equiv \arg\max_{s \geq 0} \{F_n(s) - as\}, \quad a > 0$$

$$\equiv \sup\{s \geq 0 : F_n(s) - as = \sup_{z \geq 0} (F_n(z) - az)\}.$$  

Then for each fixed $t \in (0, \infty)$ and $a > 0$

$$\{\hat{f}_n(t) < a\} = \{\hat{s}_n(a) < t\}.$$
B. Switching: a simple key result
B. Switching: a simple key result

More general result: Suppose $\Phi : D \subset \mathbb{R} \rightarrow \mathbb{R}$ where $D$ is closed. Let

$$\tilde{\Phi}(x) \equiv \text{least concave majorant of } \Phi$$

$$= \inf \{g(x) \mid g : D \rightarrow \mathbb{R}, \ g \text{ closed, } g \text{ concave} \}.$$ 

Let $\hat{\phi}_L$ and $\hat{\phi}_R$ denote the left and right derivatives of $\tilde{\Phi}$. Define

$$\kappa_L(y) \equiv \arg\max_x \{\Phi(x) - yx\}$$

$$= \inf \{x \in D : \Phi(x) - yx = \sup_{z \in D} (\Phi(z) - yz)\},$$

$$\kappa_R(y) \equiv \arg\max_x \{\Phi(x) - yx\}$$

$$= \sup \{x \in D : \Phi(x) - yx = \sup_{z \in D} (\Phi(z) - yz)\}.$$
B. Switching: a simple key result

**Theorem.** Suppose that $\Phi$ is a proper upper-semicontinuous real-valued function defined on a closed subset $D \subset \mathbb{R}$. Then $\hat{\Phi}$ is proper if and only if $\Phi \leq l$ for some linear function $l$ on $D$. Furthermore, if $\text{conv} (\text{hypo}(\Phi))$ is closed, then the functions $\kappa_L$ and $\kappa_R$ are well defined and the following switching relations hold:

\[
\hat{\phi}_L(x) < y \quad \text{if and only if} \quad \kappa_R(y) < x;
\]
\[
\hat{\phi}_R(x) \leq y \quad \text{if and only if} \quad \kappa_L(y) \leq x.
\]


We will apply this theorem with $\Phi$ taken to be various random processes, including:

- $\Phi = \mathbb{U}$, a Brownian bridge process on $[0,1]$.
- $\Phi = aW(h) - bh^2$ for $a, b > 0$ and $W$ two-sided Brownian motion.
Reminder:

\[ \text{hypo}(f) = \{(x, \alpha) \in \mathbb{R}^d \times R : \alpha \leq f(x)\}, \]
\[ \text{conv}(C) = \left\{ \sum_{i=1}^{k} \lambda_i x_i : x_i \in C, \lambda_i \geq 0, \sum_{i=1}^{k} \lambda_i = 1, k \geq 0 \right\}. \]

\( f \) is upper semicontinuous at all \( x \in \mathbb{R}^d \) if and only if \( \text{hypo}(f) \) is closed.
C. Limit theory via switching and argmax CM

Two illustrative cases:

- **Case 1:** \( f_0(x) = 1_{[0,1]}(x) \) (degenerate mixing, \( G' = \delta_1 \)).
- **Case 2:** \( f_0 \) with \( f_0(x_0) > 0, f'_0(x_0) < 0 \). (Strictly decreasing at \( x_0 \)).

**Case 1:** Groeneboom (1983), Groeneboom and Pyke (1983). If \( f_0(x) = 1_{[0,1]}(x) \), then for \( 0 < x_0 < 1 \),

\[
S_n(x_0) \equiv \sqrt{n}(\hat{f}_n(x_0) - f_0(x_0)) \to_d S(x_0)
\]

where \( S \) is the left-derivative of the least concave majorant \( C \) of a standard Brownian bridge process \( U \) on \( [0,1] \).
C. Limit theory via switching and argmax CM

Proof, Case 1: By the switching relation

\[
P(\sqrt{n}(\hat{f}_n(x_0) - f_0(x_0)) < t) \\
= P(\hat{f}_n(x_0) < f_0(x_0) + n^{-1/2}t) \\
= P(\hat{s}_n(f_0(x_0) + n^{-1/2}t) < x_0) \\
= P(\text{argmax}_h\{F_n(x_0 + h) - (f_0(x_0) + n^{-1/2}t)(x_0 + h)\} < 0) \\
= P(\text{argmax}_h\mathbb{Z}_n(h) < 0) \quad (5)
\]

where, since \( f_0(x_0) = 1 \) implies that \( xf_0(x_0) = x_0 = F(x_0) \),

\[
\mathbb{Z}_n(h) \equiv n^{1/2}(F_n(x_0 + h) - F(x_0) - hf_0(x_0) - t(x_0 + h)n^{-1/2}) \\
= n^{1/2}(F_n(x_0 + h) - F(x_0 + h)) \\
+ n^{1/2}(F(x_0 + h) - F(x_0) - hf_0(x_0)) - t(x_0 + h) \\
= \mathbb{U}_n(x_0 + h) - t(x_0 + h) \\
\sim \mathbb{U}(x_0 + h) - t(x_0 + h)
\]

where \( \mathbb{U}_n \equiv \sqrt{n}(\mathbb{F}_n - F) \) denotes the uniform empirical process and \( \mathbb{U} \) denotes a Brownian bridge process.
Thus by the (argmax) continuous mapping theorem it follows that the right side of (5) converges to

\[ P(\text{argmax}_h \{ U(x_0 + h) - t(X_0 + h) \} < 0) = P(\text{argmax}_s \{ U(s) - ts \} < x_0) = P(S(x_0) < t) \]

by the general version of the switching relation. Hence

\[ \sqrt{n}(\hat{f}_n(x_0) - f_0(x_0)) \to_d S(x_0). \]

This one-dimensional convergence extends straightforwardly to convergence of the finite-dimensional distributions, and (by monotonicity) to convergence in the Skorokhod topology on \( D[a, 1-a] \) for each fixed \( a \in (0, 1/2) \).

**Exercise 1.** \( S_n \rightsquigarrow S \) in \( L_1([0, 1], \lambda) \) with \( \lambda = \text{Lebesgue measure} \); this also holds in \( L_p([0, 1], \lambda) \) for \( 1 \leq p < 2 \), but not in \( L_2([0, 1], \lambda) \).
C. Limit theory via switching and argmax CM
C. Limit theory via switching and argmax CM
**C. Limit theory via switching and argmax CM**

**Case 2:** Prakasa Rao (1969), Groeneboom (1985). If \( f_0(x_0) > 0, \) \( f'_0(x_0) < 0, \) and \( f'_0 \) is continuous at \( x_0, \) then

\[
S_n(x_0, t) \equiv n^{1/3}(\hat{f}_n(x_0 + n^{-1/3}t) - f_0(x_0)) \rightarrow_d (2^{-1}f_0(x_0)|f'_0(x_0)|)^{1/3}S(t)
\]

where \( S \) is the left-derivative of the least concave majorant \( C \) of \( W(t) - t^2 \) and \( W \) is a standard two-sided Brownian motion process starting at 0. In particular:

\[
S_n(x_0) \equiv n^{1/3}(\hat{f}_n(x_0) - f_0(x_0)) \rightarrow_d (2^{-1}f_0(x_0)|f'_0(x_0)|)^{1/3}S(0).
\]

**Proof, Case 2:** By the switching relation
C. Limit theory via switching and argmax CM

\[
P(n^{1/3}(\hat{f}_n(x_0 + n^{-1/3}t) - f(x_0)) < y) \\
= P(\hat{f}_n(x_0 + n^{-1/3}t) < f(x_0) + yn^{-1/3}) \\
= P(\hat{s}_n(f(x_0) + yn^{-1/3}) < x_0 + n^{-1/3}t) \\
= P(\text{argmax}_v\{F_n(v) - (f(x_0) + n^{-1/3}y)v\} < x_0 + n^{-1/3}t)
\]

Now we change variables \( v = x_0 + n^{-1/3}h \) in the argument of \( F_n \) and center and scale to find that the right side in the last display equals

\[
P(\text{argmax}_h\{F_n(x_0 + n^{-1/3}h) - (f(x_0) + n^{-1/3}y)(x_0 + n^{-1/3}h)\} < t) \\
= P\left(\text{argmax}_h\{F_n(x_0 + n^{-1/3}h) - F_n(x_0) - (F(x_0 + n^{-1/3}h) - F(x_0)) \\
+ F(x_0 + n^{-1/3}h) - F(x_0) - f(x_0)n^{-1/3}h - n^{-2/3}yh\} < t\right).
\]

(6)

Now the stochastic term in (6) satisfies
\[ n^{2/3} \left\{ \mathbb{F}_n(x_0 + n^{-1/3}h) - \mathbb{F}_n(x_0) - (F(x_0 + n^{-1/3}h) - F(x_0)) \right\} \]
\[ \overset{d}{=} n^{2/3 - 1/2} \left\{ \mathbb{U}_n(F(x_0 + n^{-1/3}h)) - \mathbb{U}_n(F(x_0)) \right\} \]
\[ = n^{1/(2 \cdot 3)} \left\{ \mathbb{U}(F(x_0 + n^{-1/3}h)) - \mathbb{U}(F(x_0)) \right\} + o_p(1) \quad \text{by KMT} \]
\[ \quad \text{or by Theorems 2.11.22 or 2.11.23} \]
\[ \overset{d}{=} n^{1/6} W(f(x_0)n^{-1/3}h) + o_p(1) \]
\[ \overset{d}{=} \sqrt{f(x_0)} W(h) + o_p(1) \]

where \( W \) is a standard two-sided Brownian motion process starting from 0. On the other hand, with \( \delta_n \equiv n^{-1/3} \),
\[ n^{2/3} \left( F(x_0 + n^{-1/3}) - F(x_0) - f(x_0)n^{-1/3}h \right) \]
\[ = \delta_n^{-2} (F(x_0 + \delta_n h) - F(x_0) - f(x_0)\delta_n h) \]
\[ \rightarrow -b|h|^2 \quad \text{with} \quad b = |f'(x_0)|/2 \]

by our hypotheses, while \( n^{2/3}n^{-1/3}n^{-1/3}h = n^0 h = h \).
Thus it follows that the last probability above converges to
\[ P \left( \arg\max_h \left\{ \sqrt{f(x_0)}W(h) - b|h|^2 - yh \right\} < t \right) \]
\[ = P(S_{a,b}(t) < y) \quad \text{by switching again} \]

where
\[ S_{a,b}(t) = \text{slope at } t \text{ of the least concave majorant of} \]
\[ aW(h) - bh^2 \equiv \sqrt{f_0(x_0)W(h)} - |f'_0(x_0)||h|^2/2 \]
\[ \overset{d}{=} |2^{-1}f_0(x_0)f'_0(x_0)|S(t). \]

**Exercise 2.** Prove the equality in distribution in the last display.
Exercise 3. Let
\[ S_n(x_0, t) \equiv n^{1/3}(\hat{f}_n(x_0 + n^{-1/3}t) - f(x_0)). \]
Show that with \( y_0 \neq x_0 \) and the hypotheses of Case 2 satisfied at both \( x_0 \) and \( y_0 \), we have
\[
\left( \begin{array}{c}
S_n(x_0, \cdot) \\
S_n(y_0, \cdot)
\end{array} \right) \rightsquigarrow \left( \begin{array}{c}
\tilde{S}_{a,b} \\
\tilde{S}_{\tilde{a},\tilde{b}}
\end{array} \right) \quad \text{in} \quad D[-M, M]^2
\]
for every \( M > 0 \) where \( a = \sqrt{f(x_0)}, \tilde{a} = \sqrt{f(y_0)}, b = |f'(x_0)|/2, \)
\( \tilde{b} = |f'(y_0)|/2, \) and \( S_{a,b}, \tilde{S}_{\tilde{a},\tilde{b}} \) are the left-derivatives of the least concave majorant of \( aW(h) - bh^2 \) and \( \tilde{a}\tilde{W} - \tilde{b}h^2 \) and where \( W \) and \( \tilde{W} \) are independent two-sided Brownian motion processes.
E. Other monotone function problems

- Monotone hazard (rate) function
- Regression function
- Distribution function for interval censoring model
- Cumulative mean function, panel count data
- Sub-distribution functions, competing risks with interval censored data

**Monotone hazard function:**

- Model: \( \mathcal{H} \equiv \) all monotone increasing (or decreasing) hazard rates (wrt Lebesgue measure) on \( \mathbb{R}^+ = (0, \infty) \).

\[
h(t) = \frac{f(t)}{1 - F(t)}; \quad f(t) = h(t)\exp \left(- \int_0^t h(s)ds \right) \equiv h(t)\exp (-H(t))
\]

- Observations: \( X_1, \ldots, X_n \) i.i.d. \( f_0 \) with \( h_0 \in \mathcal{H} \).
- MLE: \( \hat{f}_n \equiv \arg\max_{h \in \mathcal{H}} \left\{ \sum_{i=1}^n \{ \log h(X_i) - H(X_i) \} \right\} \)
Monotone regression:

- **Model:** \( Y = r(x) + \epsilon \) where

  \[ r \in \mathcal{M} \equiv \{ \text{all monotone (increasing) functions from } D \text{ to } \mathbb{R} \} \]

  \( E(\epsilon) = 0, \ Var(\epsilon) < \infty. \)

- **Observations:** \( \{(x_{n,i}, Y_{n,i}) : i = 1, \ldots, n\} \) where \( Y_{n,i} = r_0(x_{n,i}) + \epsilon_{n,i} \) for some \( r_0 \in \mathcal{M} \) and \( x_{n,1} \leq \ldots \leq x_{n,n}. \)

- **LSE (=MLE for Gaussian \( \epsilon \)'s):**

  \[
  \hat{r}_n \equiv \arg\min_{r \in \mathcal{M}_n} \frac{1}{2} \sum_{i=1}^{n} (Y_{n,i} - r(x_{n,i}))^2
  \]

  where \( \mathcal{M}_n \subset \mathcal{M} \) is the subclass of monotone functions which are linear between successive \( x_{n,i} \)'s and the left and right of the range of the \( x_{n,i} \)'s.
Interval censoring case 1 = Current status data:

- Model: \( X \sim F \) on \( \mathbb{R}^+ \), \( Y \sim G \) on \( \mathbb{R}^+ \) independent, \( F \in \mathcal{F} \equiv \{ \text{all distribution functions on } \mathbb{R}^+ \} \).

Observe \( (Y, \Delta) \equiv (Y, 1_{X \leq Y}) \), so that

\[
(\Delta|Y) \sim \text{Bernoulli}(F(Y)).
\]

Thus the density of \( (Y, \Delta) \) with respect to \( G \times \text{counting measure on } \{0, 1\} \) is

\[
p(y, \delta; F) = F(y)^\delta (1 - F(y))^{1-\delta}.
\]

- Observations: \( \{(Y_i, \Delta_i) : i = 1, \ldots, n\} \) i.i.d. as \((Y, \Delta)\).

- MLE:

\[
\hat{F}_n = \arg\max_{F \in \mathcal{F}} \left\{ \mathbb{P}_n(\Delta \log F + (1 - \Delta) \log (1 - F)) \right\}.
\]
E. Other monotone function problems

Panel count data:

Competing risks data with current status observations:
See Groeneboom, Maathuis and W (2008a, 2008b)