Non-Regular Examples

Yilun Li

Sep 23rd, 2021

Table of Contents

A Change-Point Model

Model

- Suppose we have i.i.d. observations of $\mathbf{X} = (Y, Z)$, where

$$Y = \alpha I_{(Z \leq \zeta)} + \beta I_{(Z > \zeta)} + \epsilon. \quad (1)$$

- Notations:
  - $\boldsymbol{\gamma} = (\alpha, \beta)^T \in \mathbb{R}^2$
  - Unknown parameter vector $\boldsymbol{\theta} = (\boldsymbol{\gamma}^T, \zeta)^T = (\alpha, \beta, \zeta)^T$
  - True parameter vector $\boldsymbol{\theta}_0 = (\boldsymbol{\gamma}_0^T, \zeta_0)^T = (\alpha_0, \beta_0, \zeta_0)^T$.
- Assumptions:
  - Z and ϵ are independent
  - ϵ is continuous, with mean $E\epsilon = 0$ and variance $E\epsilon^2 = \sigma^2 < \infty$
  - ζ is known to lie in a bounded interval $[a, b]$
  - $\alpha_0 \neq \beta_0$, so that ζ_0 is identifiable
  - Z has a bounded and strictly positive density f over $[a, b]$ with $P(Z < a) > 0$ and $P(Z > b) > 0$.

Model

- Aim: To estimate $\boldsymbol{\theta}$ with least squares, i.e., to maximize $M_n(\boldsymbol{\theta}) = \mathbb{P}_n m_{\boldsymbol{\theta}}$, where

$$m_{\boldsymbol{\theta}}(\mathbf{x}) = -\left[y - \alpha I_{(z \le \zeta)} - \beta I_{(z > \zeta)}\right]^2. \tag{2}$$

$$\Rightarrow M_n(\boldsymbol{\theta}) = -\frac{1}{n}\sum_{i=1}^{n}\left[Y_i - \alpha I_{(Z_i \le \zeta)} - \beta I_{(Z_i > \zeta)}\right]^2 \tag{3}$$

- Denote $\hat{\boldsymbol{\theta}}_n = (\hat{\gamma}_n^T, \hat{\zeta}_n)^T = (\hat{\alpha}_n, \hat{\beta}_n, \hat{\zeta}_n)^T$ as the maximizer of $M_n(\boldsymbol{\theta})$.
- Main steps:
  - To verify that $\hat{\gamma}_n = O_P(1)$. Since we are not assuming that γ is bounded, we first need to prove that $\hat{\gamma}_n$ is bounded in probability, which implies the existence of $\hat{\gamma}_n$.
  - To establish consistency of all parameter estimates, and their rates of convergence.
  - To derive the joint limiting distribution of the parameter estimates.

- By definition,

$$-\mathbb{P}_n\epsilon^2 = -\frac{1}{n}\sum_{i=1}^{n}\epsilon^2(\mathbf{X}_i) = -\frac{1}{n}\sum_{i=1}^{n}\left[Y_i - \alpha_0 I_{(Z_i\le\zeta_0)} - \beta_0 I_{(Z_i>\zeta_0)}\right]^2 = M_n(\boldsymbol{\theta}_0) \le M_n(\hat{\boldsymbol{\theta}}_n) \tag{4}$$

- Also, since the value of Z and ζ can be partitioned into four disjoint sets: $\{Z \le \zeta \wedge \zeta_0\}$, $\{\zeta < Z \le \zeta_0\}$, $\{\zeta_0 < Z \le \zeta\}$, and $\{Z > \zeta \vee \zeta_0\}$, we can derive that

$$\begin{aligned} M_n(\hat{\boldsymbol{\theta}}_n) &= -\frac{1}{n}\sum_{i=1}^{n}\Big[(\epsilon+\alpha_0-\hat{\alpha}_n)I_{(Z_i\le\hat{\zeta}_n\wedge\zeta_0)} + (\epsilon+\alpha_0-\hat{\beta}_n)I_{(\hat{\zeta}_n<Z_i\le\zeta_0)} \\ &\quad +(\epsilon+\beta_0-\hat{\alpha}_n)I_{(\zeta_0<Z_i\le\hat{\zeta}_n)} + (\epsilon+\beta_0-\hat{\beta}_n)I_{(Z_i>\hat{\zeta}_n\vee\zeta_0)}\Big]^2 \\ &= -\frac{1}{n}\sum_{i=1}^{n}\Big[(\epsilon+\alpha_0-\hat{\alpha}_n)^2 I_{(Z_i\le\hat{\zeta}_n\wedge\zeta_0)} + (\epsilon+\alpha_0-\hat{\beta}_n)^2 I_{(\hat{\zeta}_n<Z_i\le\zeta_0)} \\ &\quad +(\epsilon+\beta_0-\hat{\alpha}_n)^2 I_{(\zeta_0<Z_i\le\hat{\zeta}_n)} + (\epsilon+\beta_0-\hat{\beta}_n)^2 I_{(Z_i>\hat{\zeta}_n\vee\zeta_0)}\Big] \\ &\le -\mathbb{P}_n\left[(\epsilon+\alpha_0-\hat{\alpha}_n)^2 I_{(Z<a)} + (\epsilon+\beta_0-\hat{\beta}_n)^2 I_{(Z>b)}\right] \end{aligned} \tag{5}$$

Existence

- Thus, by (4) and (5),

$$\mathbb{P}_n\epsilon^2 \geq \mathbb{P}_n\left[\epsilon^2 I_{(Z<a)} + \epsilon^2 I_{(Z>b)} + (\alpha_0 - \hat{\alpha}_n)^2 I_{(Z<a)} + (\beta_0 - \hat{\beta}_n)^2 I_{(Z>b)}\right.$$
$$\left. + 2\epsilon(\alpha_0 - \hat{\alpha}_n) I_{(Z<a)} + 2\epsilon(\beta_0 - \hat{\beta}_n) I_{(Z>b)}\right],$$

which implies that

$$\begin{aligned} &(\alpha_0 - \hat{\alpha}_n)^2 \mathbb{P}_n I_{(Z<a)} + (\beta_0 - \hat{\beta}_n)^2 \mathbb{P}_n I_{(Z>b)} \\ &\leq \mathbb{P}_n\left[\epsilon^2 I_{(a\leq Z\leq b)}\right] + 2(\hat{\alpha}_n - \alpha_0)\mathbb{P}_n\left[\epsilon I_{(Z<a)}\right] + 2(\hat{\beta}_n - \beta_0)\mathbb{P}_n\left[\epsilon I_{(Z>b)}\right] \\ &= O_P(1)\left(1 + ||\hat{\gamma}_n - \gamma_0||\right). \end{aligned} \tag{6}$$

Existence

- Thus,

$$\begin{aligned}(\alpha_0 - \hat{\alpha}_n)^2 &= O_P(1)\,(1 + ||\hat{\gamma}_n - \gamma_0||)\\ (\beta_0 - \hat{\beta}_n)^2 &= O_P(1)\,(1 + ||\hat{\gamma}_n - \gamma_0||)\\ \Rightarrow ||\hat{\gamma}_n - \gamma_0||^2 &= O_P(1)\,(1 + ||\hat{\gamma}_n - \gamma_0||)\\ \Rightarrow ||\hat{\gamma}_n - \gamma_0|| &= O_P(1) \qquad (7)\end{aligned}$$

- So $\hat{\gamma}_n$ is bounded in probability.
- Recall that $\hat{\zeta}_n$ lies in the interval $[a, b]$ by assumption.
- Thus all the parameter estimates are bounded in probability and therefore exist.

Consistency

- To establish consistency, we will use the *Argmax Theorem*.
- Recall (Theorem 14.1, *Argmax Theorem*): Let M_n, M be stochastic processes indexed by a metric space H. If
  (1) $M_n \rightsquigarrow M$ in $\ell^\infty(K)$ for every compact $K \subseteq H$;
  (2) almost all sample paths $h \mapsto M(h)$ are upper semi-continuous and possess a unique maximum at a (random) point $\hat{h}$, which as a random map in H is tight;
  (3) the sequence $\{\hat{h}_n\}$ is uniformly tight and satisfies

$$M_n(\hat{h}_n) \geq \sup_{h \in H} M_n(h) - o_P(1), \tag{8}$$

  then $\hat{h}_n \rightsquigarrow \hat{h}$ in H.

Consistency

And we'll prove the consistency of $\hat{\boldsymbol{\theta}}_n$ with the following steps:

- Step 1: Prove that $M_n \rightsquigarrow M$ in $\ell^\infty(K)$ for every compact $K \subseteq H = \mathbb{R}^2 \times [a, b]$, where $M(\boldsymbol{\theta}) = Pm_{\boldsymbol{\theta}}$.
- Step 2: Prove that $\boldsymbol{\theta} \mapsto M(\boldsymbol{\theta})$ is upper semi-continuous with a unique maximum at $\boldsymbol{\theta}_0$.
- Then, since $\hat{\boldsymbol{\theta}}_n$ is the maximizer of $M_n(\boldsymbol{\theta})$, and $\hat{\boldsymbol{\theta}}_n = O_P(1)$ by the discussion in the Existence section, the *Argmax Theorem* yields that $\hat{\boldsymbol{\theta}}_n \rightsquigarrow \boldsymbol{\theta}_0$ as desired.

- We first verify that for any fixed compact $K \subseteq H$, $\mathcal{F}_K = \{m_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in K\}$ is a GC-class of functions, or equivalently,

$$||\mathbb{P}_n - P||_{\mathcal{F}_K} = \sup_{f \in \mathcal{F}_K} |\mathbb{P}_n f - Pf| \to 0 \quad (a.s.^*). \tag{9}$$

- By definition of $m_{\boldsymbol{\theta}}$,

$$\begin{aligned} m_{\boldsymbol{\theta}}(X) &= -\left(Y - \alpha I_{(Z \le \zeta)} - \beta I_{(Z > \zeta)}\right)^2 \\ &= -(\epsilon + \alpha_0 - \alpha)^2 I_{(Z \le \zeta \wedge \zeta_0)} - (\epsilon + \alpha_0 - \beta)^2 I_{(\zeta < Z \le \zeta_0)} \\ &\quad - (\epsilon + \beta_0 - \alpha)^2 I_{(\zeta_0 < Z \le \zeta)} - (\epsilon + \beta_0 - \beta)^2 I_{(Z > \zeta \vee \zeta_0)}. \end{aligned} \tag{10}$$

Consistency

Step 1: Prove that $M_n \rightsquigarrow M$ in $\ell^\infty(K)$ for every compact $K \subseteq H = \mathbb{R}^2 \times [a, b]$, where $M(\boldsymbol{\theta}) = Pm_{\boldsymbol{\theta}}$

- Using the definitions, the compactness of K, and strong law of large numbers, it can be verified that

$$\mathcal{F}_{K1} = \{(\epsilon + \alpha_0 - \alpha)^2 : \boldsymbol{\theta} = (\alpha, \beta, \zeta)^T \in K\}$$

$$\mathcal{F}_{K2} = \{I_{(Z \leq \zeta \wedge \zeta_0)} : \boldsymbol{\theta} = (\alpha, \beta, \zeta)^T \in K\}$$

  are both GC-class of functions.

- Similar arguments reveal that the remaining components of the sum are also GC-classes. Thus, with the preservation results, $\mathcal{F}_K = \{m_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in K\}$ is a GC-class of functions.

$$\Rightarrow \sup_{\boldsymbol{\theta} \in K} |M_n(\boldsymbol{\theta}) - M(\boldsymbol{\theta})| = \sup_{\boldsymbol{\theta} \in K} |\mathbb{P}_n m_{\boldsymbol{\theta}} - Pm_{\boldsymbol{\theta}}| \to 0 \quad (a.s.)$$

$$\Rightarrow M_n \rightsquigarrow M$$

  for any compact $K \subseteq H$.

Consistency

Step 2: Prove that $\boldsymbol{\theta} \mapsto M(\boldsymbol{\theta})$ is upper semi-continuous with a unique maximum at $\boldsymbol{\theta}_0$.

- By definition, for any $\boldsymbol{\theta}$,

$$\begin{aligned} M(\boldsymbol{\theta}) = Pm_{\boldsymbol{\theta}} &= -\int \left[Y - \alpha I_{(Z \le \zeta)} - \beta I_{(Z > \zeta)}\right]^2 dP \\ &= -P\epsilon^2 - (\alpha_0 - \alpha)^2 P(Z \le \zeta \wedge \zeta_0) - (\alpha_0 - \beta)^2 P(\zeta < Z \le \zeta_0) \\ &\quad - (\beta_0 - \alpha)^2 P(\zeta_0 < Z \le \zeta) - (\beta_0 - \beta)^2 P(Z > \zeta \vee \zeta_0), \end{aligned} \tag{11}$$

$$\begin{aligned} &\Rightarrow M(\boldsymbol{\theta}_0) = -P\epsilon^2 \\ &\Rightarrow M(\boldsymbol{\theta}_0) \ge M(\boldsymbol{\theta}) \; (\forall \boldsymbol{\theta}) \end{aligned} \tag{12}$$

with equality holds if and only if

$$\begin{aligned} 0 &= (\alpha_0 - \alpha)^2 P(Z \le \zeta \wedge \zeta_0) = (\alpha_0 - \beta)^2 P(\zeta < Z \le \zeta_0) \\ &= (\beta_0 - \alpha)^2 P(\zeta_0 < Z \le \zeta) = (\beta_0 - \beta)^2 P(Z > \zeta \vee \zeta_0) \\ &\Leftrightarrow \alpha = \alpha_0, \beta = \beta_0, \zeta = \zeta_0 \\ &\Leftrightarrow \boldsymbol{\theta} = \boldsymbol{\theta}_0. \end{aligned} \tag{13}$$

- Note: Here, we've used the assumptions that $\alpha_0 \neq \beta_0$, $P(Z < a) > 0$, $P(Z > b) > 0$, and the density of Z is strictly positive on $[a, b]$.
- Thus, M has a unique maximum at $\boldsymbol{\theta}_0$.

Consistency

Step 2: Prove that $\boldsymbol{\theta} \mapsto M(\boldsymbol{\theta})$ is upper semi-continuous with a unique maximum at $\boldsymbol{\theta}_0$.

- Also, with (11) and the facts that
  - $(\alpha_0 - \alpha)^2$, $(\alpha_0 - \beta)^2$, $(\beta_0 - \alpha)^2$, $(\beta_0 - \beta)^2$ are continuous functions of $\boldsymbol{\theta}$;
  - the density of Z is bounded and strictly positive on $[a, b]$,

  we can conclude that M is continuous.
- Now the conditions of the *Argmax Theorem* are all verified, and the desired consistency follows.

- Recall (Theorem 14.4): Let M_n be a sequence of stochastic processes indexed by a semi-metric space (Θ, d) and $M : \Theta \to \mathbb{R}$ be a deterministic function. Suppose
  - (A) for every $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}_0$, there exists a $c_1 > 0$ such that $M(\boldsymbol{\theta}) - M(\boldsymbol{\theta}_0) \leq -c_1 \tilde{d}^2(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$, where $\tilde{d} : \Theta \times \Theta \to [0, \infty)$ satisfies $\tilde{d}(\boldsymbol{\theta}_n, \boldsymbol{\theta}_0) \to 0$ whenever $d(\boldsymbol{\theta}_n, \boldsymbol{\theta}_0) \to 0$.
  - (B) for all sufficiently large n and sufficiently small $\delta > 0$, the centered process $(M_n - M)$ satisfies

$$E^* \left[\sup_{\tilde{d}(\boldsymbol{\theta}, \boldsymbol{\theta}_0) < \delta} \sqrt{n} |(M_n - M)(\boldsymbol{\theta}) - (M_n - M)(\boldsymbol{\theta}_0)| \right] \leq c_2 \phi_n(\delta) \tag{14}$$

    for some $c_2 < \infty$ and functions ϕ_n such that $\delta \mapsto \delta^{-\eta} \phi_n(\delta)$ is decreasing for some $\eta < 2$ not depending on n.
  - (C) $r_n^2 \phi_n(r_n^{-1}) \leq c_3 \sqrt{n}$ for every n and some $c_3 < \infty$.
  - (D) the sequence $\{\hat{\boldsymbol{\theta}}_n\}$ satisfies $M_n(\hat{\boldsymbol{\theta}}_n) \geq \sup_{\boldsymbol{\theta} \in \Theta} M_n(\boldsymbol{\theta}) - O_P(r_n^{-2})$ and converges to $\boldsymbol{\theta}_0$ in outer probability,

  then $r_n \tilde{d}(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) = O_P(1)$.

- Let $\tilde{d}(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = ||\boldsymbol{\gamma} - \boldsymbol{\gamma}_0|| + \sqrt{|\zeta - \zeta_0|}$.
- By equation (11),

$$\begin{aligned} M(\boldsymbol{\theta}) - M(\boldsymbol{\theta}_0) &= -(\alpha_0 - \alpha)^2 P(Z \le \zeta \wedge \zeta_0) - (\alpha_0 - \beta)^2 P(\zeta < Z \le \zeta_0) \\ &\quad - (\beta_0 - \alpha)^2 P(\zeta_0 < Z \le \zeta) - (\beta_0 - \beta)^2 P(Z > \zeta \vee \zeta_0) \\ &\le -(\alpha_0 - \alpha)^2 P(Z < a) - (\beta_0 - \beta)^2 P(Z > b) \\ &\quad - (\alpha_0 - \beta)^2 P(\zeta < Z \le \zeta_0) - (\beta_0 - \alpha)^2 P(\zeta_0 < Z \le \zeta) \end{aligned}$$

- Since $\alpha_0 \neq \beta_0$, and the density of Z is bounded and strictly positive on $[a, b]$, we can prove that for sufficiently small $\delta > 0$, if $0 < \tilde{d}(\boldsymbol{\theta}, \boldsymbol{\theta}_0) < \delta$, then
    - $(\alpha_0 - \beta)^2$ and $(\beta_0 - \alpha)^2$ has a strictly positive lower bound;
    - $P(\zeta_0 < Z \le \zeta) \ge \tilde{\lambda}|\zeta - \zeta_0|$ or $P(\zeta < Z \le \zeta_0) \ge \tilde{\lambda}|\zeta - \zeta_0|$ for some $0 < \tilde{\lambda} < \infty$.

  Thus, $\exists \lambda \in (0, \infty)$ s.t.

$$M(\boldsymbol{\theta}) - M(\boldsymbol{\theta}_0) \le -(\alpha_0 - \alpha)^2 P(Z < a) - (\beta_0 - \beta)^2 P(Z > b) - \lambda|\zeta - \zeta_0|. \tag{15}$$

- Let $\mu = \min\{P(Z < a), P(Z > b)\} > 0$,

$$\Rightarrow M(\boldsymbol{\theta}) - M(\boldsymbol{\theta}_0) \le -\mu||\boldsymbol{\gamma} - \boldsymbol{\gamma}_0||^2 - \lambda|\zeta - \zeta_0|, \tag{16}$$

$$\Rightarrow M(\boldsymbol{\theta}) - M(\boldsymbol{\theta}_0) \le -\frac{\lambda\mu}{\lambda + \mu}\tilde{d}^2(\boldsymbol{\theta}, \boldsymbol{\theta}_0). \tag{17}$$

- So condition (A) of (Theorem 14.4) is verified.

- Now consider the class of functions $M_\delta = \{m_{\boldsymbol{\theta}} - m_{\boldsymbol{\theta}_0} : \tilde{d}(\boldsymbol{\theta}, \boldsymbol{\theta}_0) < \delta\}$. Thus, to verify condition (B), it suffices to prove $E^*||\mathbb{G}_n||_{M_\delta} \leq c_2\phi_n(\delta)$ for some c_2 and ϕ_n, where $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P)$.
- By definition, we have

$$
\begin{aligned}
m_{\boldsymbol{\theta}} - m_{\boldsymbol{\theta}_0} &= 2\epsilon(\alpha - \alpha_0)I_{(Z \leq \zeta \wedge \zeta_0)} + 2\epsilon(\beta - \beta_0)I_{(Z > \zeta \vee \zeta_0)} \\
&\quad + 2\epsilon(\beta - \alpha_0)I_{(\zeta < Z \leq \zeta_0)} + 2\epsilon(\alpha - \beta_0)I_{(\zeta_0 < Z \leq \zeta)} \\
&\quad - (\alpha_0 - \alpha)^2 I_{(Z \leq \zeta \wedge \zeta_0)} - (\beta_0 - \beta)^2 I_{(Z > \zeta \vee \zeta_0)} \\
&\quad - (\alpha_0 - \beta)^2 I_{(\zeta < Z \leq \zeta_0)} - (\beta_0 - \alpha)^2 I_{(\zeta_0 < Z \leq \zeta)} \\
&= A_1(\boldsymbol{\theta}) + A_2(\boldsymbol{\theta}) + B_1(\boldsymbol{\theta}) + B_2(\boldsymbol{\theta}) \\
&\quad - C_1(\boldsymbol{\theta}) - C_2(\boldsymbol{\theta}) - D_1(\boldsymbol{\theta}) - D_2(\boldsymbol{\theta}) \qquad (18)
\end{aligned}
$$

- With Lemma 8.17, it can be proved that

$$E^*\left[\sup_{\tilde{d}(\boldsymbol{\theta},\boldsymbol{\theta}_0)<\delta}\left|\mathbb{G}_n A_1(\boldsymbol{\theta})\right|\right] \lesssim \delta, \tag{19}$$

  and similar conclusion holds for $A_2(\boldsymbol{\theta})$. Similar conclusions also hold for $C_1(\boldsymbol{\theta})$ and $C_2(\boldsymbol{\theta})$, except that the upper bounds will be δ^2.

- With Theorem 11.2, it can be proved that

$$E^*\left[\sup_{\tilde{d}(\boldsymbol{\theta},\boldsymbol{\theta}_0)<\delta}\left|\mathbb{G}_n B_1(\boldsymbol{\theta})\right|\right] \lesssim \delta^2, \tag{20}$$

  and similar conclusions hold for $B_2(\boldsymbol{\theta})$, $D_1(\boldsymbol{\theta})$, $D_2(\boldsymbol{\theta})$.

- Thus, $E^*||\mathbb{G}_n||_{M_\delta} \lesssim \delta$. And condition (B) holds with $\phi_n(\delta) = \delta$.

- Let $r_n = \sqrt{n}$, so $r_n^2 \phi_n(r_n^{-1}) = \sqrt{n}$ and condition (C) holds.
- Since $\hat{\boldsymbol{\theta}}_n$ is the maximizer of M_n and is a consistent estimate of $\boldsymbol{\theta}_0$, the condition (D) holds.
- Thus, by (Theorem 14.4),

$$
\begin{aligned}
& r_n \tilde{d}(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) = \sqrt{n}\left(||\hat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}_0|| + \sqrt{|\hat{\zeta}_n - \zeta_0|} \right) = O_P(1) \\
\Rightarrow & \begin{cases} \sqrt{n}||\hat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}_0|| = O_P(1) \\ n|\hat{\zeta}_n - \zeta_0| = O_P(1) \end{cases}
\end{aligned}
\tag{21}
$$

- Convergence of $\hat{\zeta}_n$ is faster than $(\hat{\alpha}_n, \hat{\beta}_n)$.

- It can be proved that

$$\hat{h}_n \triangleq \left(\sqrt{n}(\hat{\alpha}_n - \alpha_0), \sqrt{n}(\hat{\beta}_n - \beta_0), n(\hat{\zeta}_n - \zeta_0)\right)^T \rightsquigarrow \tilde{h} = (\tilde{h}_1, \tilde{h}_2, \tilde{h}_3)^T \tag{22}$$

where

(1) $\tilde{h}_1$, $\tilde{h}_2$, $\tilde{h}_3$ are mutually independent;

(2) $\tilde{h}_1$ and $\tilde{h}_2$ are Gaussian with mean zero and respective variance $\sigma^2 P^{-1}(Z \leq \zeta_0)$ and $\sigma^2 P^{-1}(Z > \zeta_0)$;

(3) $\tilde{h}_3$ is the smallest argmax of $Q(h_3)$ for a certain two-sided Poisson process.

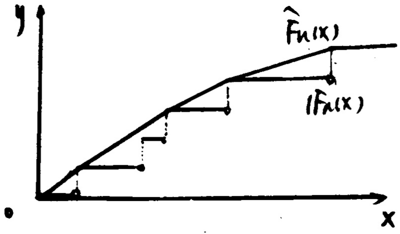
Monotone Density Estimation

Model

- Let X_1, X_2, ..., X_n be a sample from a Lebesgue density f on $[0, \infty)$, which is known to be decreasing. For any fixed $t > 0$, we assume that f is differentiable at t with $-\infty < f'(t) < 0$. So the probability distribution function F satisfies $F'' = f' < 0$, which implies that F is concave.
- For a general function g, the least concave majorant of g is defined as the smallest concave function h such that $h \geq g$.
- Let $\mathbb{F}_n$ be the empirical distribution function, and let $\hat{F}_n$ denote the least concave majorant of $\mathbb{F}_n$.

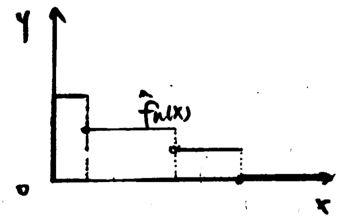
Model

- We can construct $\hat{F}_n$ by imagining a string tied at $(0,0)$ which is pulled tight over the top of the function graph $\{(x, \mathbb{F}_n(x)) : x \geq 0\}$. So the slope of each piecewise linear segment will be non-increasing, and $\hat{F}_n$ will touch $\mathbb{F}_n$ at $(0,0)$ and $\{(x_{t_j}, \mathbb{F}_n(x_{t_j})) : j = 0, 1, ..., k\}$ where x_{t_k} is the largest observation. For $x > x_{t_k}$, define $\hat{F}_n(x) = 1$.
- Thus, $\hat{F}_n$ is continuous.



Model

- Let $\hat{f}_n$ be the left-derivative of $\hat{F}_n$.
- So $\hat{f}_n$ is a non-increasing step function.
- It is also proved that $\hat{f}_n$ is the MLE of f in section 3.1 of *Grenander (1956)*.



- In the following discussion, we will
  - (*i*) establish the consistency of $\hat{f}_n(t)$;
  - (*ii*) verify that the rate of convergence of $\hat{f}_n$ is $n^{1/3}$;
  - (*iii*) derive the weak convergence of $n^{1/3}\left[\hat{f}_n(t) - f(t)\right]$.

- Lemma 14.7 (*Marshall's lemma*): Under the give conditions,

$$\sup_{t\geq 0} |\hat{F}_n(t) - F(t)| \leq \sup_{t\geq 0} |\mathbb{F}_n(t) - F(t)|. \tag{23}$$

- For any fixed $t \geq 0$ and any $\delta \in (0, t)$, by definition of $\hat{f}_n$,

$$\delta^{-1}\left[\hat{F}_n(t+\delta) - \hat{F}_n(t)\right] \leq \hat{f}_n(t) \leq \delta^{-1}\left[\hat{F}_n(t) - \hat{F}_n(t-\delta)\right] \tag{24}$$

- By *Marshall's lemma*, as $n \to \infty$, the lower bound converges to $\delta^{-1}[F(t+\delta) - F(t)]$ $(a.s.)$, and the upper bound converges to $\delta^{-1}[F(t) - F(t-\delta)]$ $(a.s.)$.
- Since δ is arbitrary and F is differentiable, let $\delta \to 0$ so we have $\hat{f}_n(t) \to f(t)$ $(a.s.)$.

- We consider an inverse transformation and define

$$\hat{s}_n(a) = \arg\max_{s \geq 0} \{\mathbb{F}_n(s) - as\} \tag{25}$$

where the largest value is selected when multiple maximizers exist. We will focus on the stochastic process $\{\hat{s}_n(a) : a > 0\}$ in the following discussion.

- It can be proved that for any $t \geq 0$ and $a > 0$,

$$\hat{f}_n(t) \leq a \Leftrightarrow \hat{s}_n(a) \leq t, \tag{26}$$

so $\hat{s}_n$ can be viewed as a sort of inverse of $\hat{f}_n$.

- Thus, for any fixed x,

$$
\begin{aligned}
& P\left(n^{1/3}\left[\hat{f}_n(t) - f(t)\right] \leq x\right) \\
&= P\left(\hat{f}_n(t) \leq f(t) + xn^{-1/3}\right) = P\left(\hat{s}_n(f(t) + xn^{-1/3}) \leq t\right) \\
&= P\left(\arg\max_{s \geq 0}\{\mathbb{F}_n(s) - (f(t) + xn^{-1/3})s\} \leq t\right) \\
&= P\left(\arg\max_{g \geq -t}\{\mathbb{F}_n(t+g) - (f(t) + xn^{-1/3})(t+g)\} \leq 0\right) \\
&= P\left(\hat{g}_n \leq 0\right), \qquad (27)
\end{aligned}
$$

where

$$
\begin{aligned}
\hat{g}_n &= \arg\max_{g \geq -t}\{\mathbb{F}_n(t+g) - (f(t) + xn^{-1/3})(t+g)\} \\
&= \arg\max_{g \geq -t}\{\mathbb{F}_n(t+g) - \mathbb{F}_n(t) - f(t)g - xgn^{-1/3}\}. \qquad (28)
\end{aligned}
$$

Next, we use Theorem 14.4 to prove that $n^{1/3}\hat{g}_n = O_P(1)$.

- For $g > -t$, let

$$M_n(g) = \mathbb{F}_n(t+g) - \mathbb{F}_n(t) - f(t)g - xgn^{-1/3} \quad (29)$$

$$M(g) = F(t+g) - F(t) - f(t)g \quad (30)$$

and define $\tilde{d}(g_1, g_2) = d(g_1, g_2) = |g_1 - g_2|$. Then by (28), $\hat{g}_n = \arg\max_{g \geq -t} M_n(g)$.

(A) By definition, the maximizer of M is $g_0 \triangleq \arg\max_{g \geq -t} M(g) = 0$, with $M(g_0) = 0$.

- Also, with Taylor Expansion, for g that is sufficiently close to g_0, we have

$$M(g) = \frac{1}{2}g^2 f'(t) + o(g^2) \lesssim -g^2 \quad (31)$$

$$\Rightarrow M(g) - M(g_0) \lesssim -g^2 = -\tilde{d}^2(g, g_0).$$

(B) By definition, $M_n(g_0) = M(g_0) = 0$. Thus,

$$\begin{aligned} &E^*\left(\sup_{\tilde{d}(g,g_0)<\delta} \sqrt{n}\left|(M_n - M)(g) - (M_n - M)(g_0)\right|\right) \\ &= E^*\left(\sup_{|g|<\delta} \left|\mathbb{G}_n I_{(X\le t+g)} - \mathbb{G}_n I_{(X\le t)} - xgn^{1/6}\right|\right) \\ &\le E^*\left(\sup_{|g|<\delta} \left|\mathbb{G}_n I_{(X\le t+g)} - \mathbb{G}_n I_{(X\le t)}\right|\right) + x\delta n^{1/6} \end{aligned} \tag{32}$$

- By (Theorem 11.2), it can be proved that

$$E^*\left(\sup_{|g|<\delta} \left|\mathbb{G}_n I_{(X\le t+g)} - \mathbb{G}_n I_{(X\le t)}\right|\right) \lesssim \delta^{1/2} \tag{33}$$

- Let $\phi_n(\delta) = \delta^{1/2} + x\delta n^{1/6}$, so

$$E^*\left(\sup_{\tilde{d}(g,g_0)<\delta} \sqrt{n}\left|(M_n - M)(g) - (M_n - M)(g_0)\right|\right) \lesssim \phi_n(\delta). \tag{34}$$

(C) Let $r_n = n^{1/3}$, so $r_n^2 \phi_n(r_n^{-1}) = O(\sqrt{n})$.

(D) Firstly, since $\hat{g}_n$ is the maximizer of M_n, we have $M_n(\hat{g}_n) \geq \sup_g M_n(g) - O_P(r_n^{-2})$.

- In addition,
  - with (29) and (30), for any compact K
$$\sup_{g \in K} |M_n(g) - M(g)| \to_p 0$$
    so $M_n \rightsquigarrow M$;
  - M is continuous with unique maximizer $g_0 = 0$;
  - It can also be verified that $\hat{g}_n = O_P(1)$,

  so by *Argmax Theorem*, $\hat{g}_n \rightsquigarrow g_0 = 0$, which implies that $\hat{g}_n \to_p g_0$.
- Thus, by Theorem 14.4, $r_n \tilde{d}(\hat{g}_n, g_0) = n^{1/3} |\hat{g}_n| = O_P(1)$.

- Let $\hat{h}_n = n^{1/3}\hat{g}_n$ be the maximizer of $h \mapsto M_n(n^{-1/3}h)$. Since the maximum of a function does not change when the function is multiplied by a constant, $\hat{h}_n$ is also the argmax of $h \mapsto n^{2/3}M_n(n^{-1/3}h)$.

$$\begin{aligned} & n^{2/3}M_n(n^{-1/3}h) \\ = & n^{2/3}\left[\mathbb{F}_n(t+n^{-1/3}h) - \mathbb{F}_n(t) - f(t)n^{-1/3}h - xhn^{-2/3}\right] \\ = & n^{2/3}\left(\mathbb{P}_n - P\right) I_{(t<X\leq t+n^{-1/3}h)} + n^{2/3}\left[F(t+n^{-1/3}h) - F(t) - f(t)n^{-1/3}h\right] - xh \end{aligned} \quad (35)$$

- It can be proved that
  - $n^{2/3}M_n(n^{-1/3}h) \rightsquigarrow \mathbb{H}(h) = \sqrt{f(t)}\mathbb{Z}(h) + \frac{1}{2}h^2 f'(t) - xh$, where $\mathbb{Z}$ is a two-sided Brownian Motion;
  - $\mathbb{H}(h)$ has a unique maximizer

$$\hat{h} = \left(\frac{4f(t)}{|f'(t)|^2}\right)^{1/3} \arg\max_h\{\mathbb{Z}(h) - h^2\} + \frac{x}{f'(t)} \quad (36)$$

- Since $\hat{h}_n = n^{1/3}\hat{g}_n = O_P(1)$, the *Argmax Theorem* implies $\hat{h}_n \rightsquigarrow \hat{h}$.

Rate of Convergence and Weak Convergence Results

$$
\begin{aligned}
\Rightarrow P\left(n^{1/3}\left[\hat{f}_n(t) - f(t)\right] \le x\right) &= P(\hat{g}_n \le 0) = P(\hat{h}_n \le 0) \\
&\to P(\hat{h} \le 0) \\
&= P(|4f(t)f'(t)|^{1/3} \arg\max_h \{\mathbb{Z}(h) - h^2\} \le x). \qquad (37)
\end{aligned}
$$

- Since x is arbitrary, we conclude that

$$
n^{1/3}\left[\hat{f}_n(t) - f(t)\right] \rightsquigarrow |4f(t)f'(t)|^{1/3} \arg\max_h \{\mathbb{Z}(h) - h^2\}. \qquad (38)
$$