Maximum likelihood in an infinite-dimensional exponential family

David Pollard

Yale University

July 2010
Seattle
In honor of Jon Wellner's 65th birthday

Based on:
Dou, Pollard, and Zhou (2010) "Functional regression for general exponential families"
arXiv:1001.3742v1 [math.ST] 21 Jan 2010

Slides for this talk at: www.stat.yale.edu/~pollard/Talks

1. Problem: estimate unknown β

- Observe independent $(y_1, x_1), (y_2, x_2), \ldots$ with $y_i \mid x_i \sim Q_{\theta_i}$ and (nonrandom or condition) $x_i' = (x_{i1}, x_{i2}, \ldots)$ and $\theta_i = \langle x_i, \beta \rangle$
- $$\frac{dQ_\theta}{dQ_0} = \exp\left(y\theta - \Psi(\theta)\right) \qquad \theta \in \mathbb{R}$$
- Ψ convex with $\Psi(0) = 0$

1. Problem: estimate unknown β

- Observe independent $(y_1, x_1), (y_2, x_2), \ldots$ with $y_i \mid x_i \sim Q_{\theta_i}$ and (nonrandom or condition) $x_i' = (x_{i1}, x_{i2}, \ldots)$ and $\theta_i = \langle x_i, \beta \rangle$
- $$\frac{dQ_\theta}{dQ_0} = \exp\left(y\theta - \Psi(\theta)\right) \qquad \theta \in \mathbb{R}$$
- Ψ convex with $\Psi(0) = 0$

Example: for $Q_\theta = \mathrm{Bernoulli}\left(\frac{e^\theta}{1+e^\theta}\right)$ on $\{0, 1\}$

$$\Psi(\theta) = \log\left(\frac{1+e^\theta}{2}\right)$$

Example: for $Q_\theta = \mathrm{Poisson}\left(e^\theta\right)$ on $\{0, 1, 2, \ldots\}$

$$\Psi(\theta) = e^\theta - 1$$

1. Problem: estimate unknown β

- Observe independent $(y_1, x_1), (y_2, x_2), \ldots$ with $y_i \mid x_i \sim Q_{\theta_i}$ and (nonrandom or condition) $x_i' = (x_{i1}, x_{i2}, \ldots)$ and $\theta_i = \langle x_i, \beta \rangle$
- $$\frac{dQ_\theta}{dQ_0} = \exp\left(y\theta - \Psi(\theta)\right) \qquad \theta \in \mathbb{R}$$
- Ψ convex with $\Psi(0) = 0$
- MLE: maximize a concave function of b?

$$\widehat{b} \stackrel{?}{=} \operatorname*{argmax}_{b \in ?} \textstyle\sum_{i \leq n} \Big(y_i \langle x_i, b \rangle - \Psi(\langle x_i, b \rangle) \Big)$$

Component of $\widehat{b}$ in $\mathrm{span}\{x_1, \ldots, x_n\}^\perp$ is unconstrained.

2. Simplify: $x_i \in \mathbb{R}^N$, with $N = N_n$?

- MLE

$$\widehat{b}_n = \operatorname*{argmax}_{b \in \mathbb{R}^N} \sum_{i \leq n} \left(y_i x_i' b - \Psi(x_i' b) \right) \qquad \text{(concave)}$$

2. Simplify: $x_i \in \mathbb{R}^N$, with $N = N_n$?

- MLE

$$\widehat{b}_n = \underset{b \in \mathbb{R}^N}{\operatorname{argmax}} \sum_{i \le n} \left(y_i x_i' b - \Psi(x_i' b) \right) \qquad \text{(concave)}$$

-

$$J_n = \sum_{i \le n} x_i x_i' \ddot{\Psi}(\theta_i) \qquad \text{(information matrix)}$$

2. Simplify: $x_i \in \mathbb{R}^N$, with $N = N_n$?

- MLE

$$\widehat{b}_n = \operatorname*{argmax}_{b \in \mathbb{R}^N} \sum_{i \le n} \left(y_i x_i' b - \Psi(x_i' b) \right) \qquad \text{(concave)}$$

-

$$J_n = \sum_{i \le n} x_i x_i' \ddot{\Psi}(\theta_i) \qquad \text{(information matrix)}$$

- Reparametrize: $w_i = J_n^{-1/2} x_i$ so that

$$\sum_{i \le n} w_i w_i' \ddot{\Psi}(\theta_i) = I_N$$

2. Simplify: $x_i \in \mathbb{R}^N$, with $N = N_n$?

- MLE

$$\widehat{b}_n = \underset{b \in \mathbb{R}^N}{\text{argmax}} \sum_{i \le n} \left(y_i x_i' b - \Psi(x_i' b) \right) \qquad \text{(concave)}$$

-

$$J_n = \sum_{i \le n} x_i x_i' \ddot{\Psi}(\theta_i) \qquad \text{(information matrix)}$$

- Reparametrize: $w_i = J_n^{-1/2} x_i$ so that

$$\sum_{i \le n} w_i w_i' \ddot{\Psi}(\theta_i) = I_N$$

- $\widehat{b}_n = \beta + J_n^{-1/2} \widehat{t}_n$ where $\widehat{t}_n = \underset{t \in \mathbb{R}^N}{\text{argmax}} L_n(t)$ for

$$L_n(t) = \sum_{i \le n} \left(y_i w_i' t - \Psi(\theta_i + w_i' t) \right) \qquad \text{(concave)}$$

3. Why reparametrize?

- [from last slide] $\widehat{t}_n = \operatorname*{argmax}_{t \in \mathbb{R}^N} L_n(t)$ where $L_n(t) = \sum_{i \le n} \Big(y_i w_i' t - \Psi(\theta_i + w_i' t) \Big)$

3. Why reparametrize?

- [from last slide] $\widehat{t}_n = \operatorname{argmax}_{t \in \mathbb{R}^N} L_n(t)$ where

  $L_n(t) = \sum_{i \le n} \Big(y_i w_i' t - \Psi(\theta_i + w_i' t) \Big)$
- $T_n = \sum_{i \le n} w_i \Big(y_i - \dot{\Psi}(\theta_i) \Big)$ so that $\mathbb{P} T_n = 0$ and $\operatorname{var}(T_n) = I_N$

3. Why reparametrize?

- [from last slide] $\widehat{t}_n = \operatorname*{argmax}_{t \in \mathbb{R}^N} L_n(t)$ where

  $L_n(t) = \sum_{i \le n} \Big(y_i w_i' t - \Psi(\theta_i + w_i' t) \Big)$
- $T_n = \sum_{i \le n} w_i \Big(y_i - \dot{\Psi}(\theta_i) \Big)$ so that $\mathbb{P} T_n = 0$ and $\mathrm{var}(T_n) = I_N$
-

$$
\begin{aligned}
\dot{L}(t) &= \frac{\partial}{\partial t} L_n(t) \\
&= \sum_{i \le n} \Big(y_i - \dot{\Psi}(\theta_i) \Big) w_i' - \sum_{i \le n} \Big(\dot{\Psi}(\theta_i + w_i' t) - \dot{\Psi}(\theta_i) \Big) w_i' \\
&= T_n' - t' \sum_{i \le n} w_i w_i' \ddot{\Psi}(\theta_i) - R_n(t)' \\
&= T_n' - t' - \text{small?}
\end{aligned}
$$

3. Why reparametrize?

- [from last slide] $\widehat{t}_n = \operatorname*{argmax}_{t \in \mathbb{R}^N} L_n(t)$ where

  $L_n(t) = \sum_{i \le n} \Big(y_i w_i' t - \Psi(\theta_i + w_i' t) \Big)$
- $T_n = \sum_{i \le n} w_i \Big(y_i - \dot{\Psi}(\theta_i) \Big)$ so that $\mathbb{P} T_n = 0$ and $\mathrm{var}(T_n) = I_N$
-

$$
\begin{aligned}
\dot{L}(t) &= \frac{\partial}{\partial t} L_n(t) \\
&= \sum_{i \le n} \Big(y_i - \dot{\Psi}(\theta_i) \Big) w_i' - \sum_{i \le n} \Big(\dot{\Psi}(\theta_i + w_i' t) - \dot{\Psi}(\theta_i) \Big) w_i' \\
&= T_n' - t' \sum_{i \le n} w_i w_i' \ddot{\Psi}(\theta_i) - R_n(t)' \\
&= T_n' - t' - \text{small?}
\end{aligned}
$$

3. Why reparametrize?

- [from last slide] $\widehat{t}_n = \operatorname*{argmax}_{t \in \mathbb{R}^N} L_n(t)$ where

$$L_n(t) = \sum_{i \le n} \Big(y_i w_i' t - \Psi(\theta_i + w_i' t) \Big)$$

- $T_n = \sum_{i \le n} w_i \Big(y_i - \dot{\Psi}(\theta_i) \Big)$ so that $\mathbb{P} T_n = 0$ and $\operatorname{var}(T_n) = I_N$
-

$$\begin{aligned}
\dot{L}(t) &= \frac{\partial}{\partial t} L_n(t) \\
&= \sum_{i \le n} \Big(y_i - \dot{\Psi}(\theta_i) \Big) w_i' - \sum_{i \le n} \Big(\dot{\Psi}(\theta_i + w_i' t) - \dot{\Psi}(\theta_i) \Big) w_i' \\
&= T_n' - t' \sum_{i \le n} w_i w_i' \ddot{\Psi}(\theta_i) - R_n(t)' \\
&= T_n' - t' - \text{small?}
\end{aligned}$$

3. Why reparametrize?

- [from last slide] $\widehat{t}_n = \operatorname*{argmax}_{t \in \mathbb{R}^N} L_n(t)$ where

  $L_n(t) = \sum_{i \le n} \Big(y_i w_i' t - \Psi(\theta_i + w_i' t) \Big)$
- $T_n = \sum_{i \le n} w_i \Big(y_i - \dot{\Psi}(\theta_i) \Big)$ so that $\mathbb{P} T_n = 0$ and $\mathrm{var}(T_n) = I_N$
-

$$
\begin{aligned}
\dot{L}(t) &= \frac{\partial}{\partial t} L_n(t) \\
&= \sum_{i \le n} \Big(y_i - \dot{\Psi}(\theta_i) \Big) w_i' - \sum_{i \le n} \Big(\dot{\Psi}(\theta_i + w_i' t) - \dot{\Psi}(\theta_i) \Big) w_i' \\
&= T_n' - t' \sum_{i \le n} w_i w_i' \ddot{\Psi}(\theta_i) - R_n(t)' \\
&= T_n' - t' - \mathrm{small?}
\end{aligned}
$$

- Should have $\widehat{t}_n \approx T_n$ (even when $N_n \to \infty$?)

4. Remainder term in $\dot{L}_n(t) = T_n' - t' - R_n(t)'$

- Assume

$$|\dddot{\Psi}(\theta + h)| \leq \ddot{\Psi}(\theta) G(|h|) \qquad \text{for all } \theta \text{ and } h$$

where G is increasing with $G(0) \geq 1$

4. Remainder term in $\dot{L}_n(t) = T_n' - t' - R_n(t)'$

- Assume

$$|\dddot{\Psi}(\theta + h)| \le \ddot{\Psi}(\theta) G(|h|) \qquad \text{for all } \theta \text{ and } h$$

where G is increasing with $G(0) \ge 1$

- Then

$$|\dot{\Psi}(\theta + h) - \dot{\Psi}(\theta) - h\ddot{\Psi}(\theta)| \le \frac{1}{2} h^2 \ddot{\Psi}(\theta) G(|h|)$$

4. Remainder term in $\dot{L}_n(t) = T_n' - t' - R_n(t)'$

- Assume

$$|\dddot{\Psi}(\theta + h)| \leq \ddot{\Psi}(\theta) G(|h|) \qquad \text{for all } \theta \text{ and } h$$

where G is increasing with $G(0) \geq 1$

- Then

$$|\dot{\Psi}(\theta + h) - \dot{\Psi}(\theta) - h\ddot{\Psi}(\theta)| \leq \frac{1}{2} h^2 \ddot{\Psi}(\theta) G(|h|)$$

- With $M_n := \max_{i \leq n} |w_i|$,

$$|R_n(t)| = |\sum_{i \leq n} \left(\dot{\Psi}(\theta_i + w_i' t) - \dot{\Psi}(\theta_i) - w_i' t \ddot{\Psi}(\theta_i) \right) w_i'|$$
$$\leq \frac{1}{2} M_n G(M_n |t|) t' \sum_{i \leq n} w_i w_i' \ddot{\Psi}(\theta_i) t$$

4. Remainder term in $\dot{L}_n(t) = T_n' - t' - R_n(t)'$

- Assume

$$|\dddot{\Psi}(\theta + h)| \le \ddot{\Psi}(\theta) G(|h|) \qquad \text{for all } \theta \text{ and } h$$

where G is increasing with $G(0) \ge 1$

- Then

$$|\dot{\Psi}(\theta + h) - \dot{\Psi}(\theta) - h\ddot{\Psi}(\theta)| \le \frac{1}{2} h^2 \ddot{\Psi}(\theta) G(|h|)$$

- With $M_n := \max_{i \le n} |w_i|$,

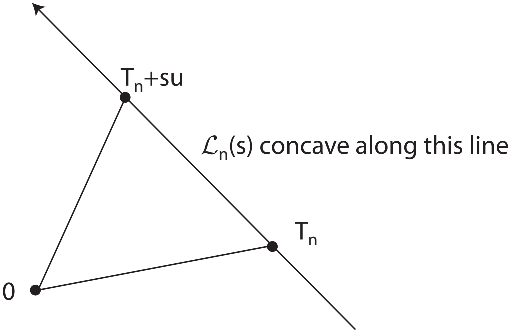
$$\begin{aligned} |R_n(t)| &= |\sum_{i \le n} \left(\dot{\Psi}(\theta_i + w_i' t) - \dot{\Psi}(\theta_i) - w_i' t \ddot{\Psi}(\theta_i) \right) w_i'| \\ &\le \frac{1}{2} M_n G(M_n |t|) t' \sum_{i \le n} w_i w_i' \ddot{\Psi}(\theta_i) t = \frac{1}{2} M_n G(M_n |t|) |t|^2 \end{aligned}$$

5. Directional derivatives for $\dot{L}_n(t) = T_n' - t' - R_n(t)'$

- Define concave $\mathcal{L}_n(s) = L_n(T_n + su)$ for fixed unit vector u.

5. Directional derivatives for $\dot{L}_n(t) = T_n' - t' - R_n(t)'$

- Define concave $\mathcal{L}_n(s) = L_n(T_n + su)$ for fixed unit vector u.
- $$\dot{\mathcal{L}}_n(s) := \frac{d}{ds}\mathcal{L}_n(s) = \dot{L}_n(T_n + su)u = -s - R_n(T_n + su)'u$$



5. Directional derivatives for $\dot{L}_n(t) = T_n' - t' - R_n(t)'$

- Define concave $\mathcal{L}_n(s) = L_n(T_n + su)$ for fixed unit vector u.
- $$\dot{\mathcal{L}}_n(s) := \frac{d}{ds}\mathcal{L}_n(s) = \dot{L}_n(T_n + su)u = -s - R_n(T_n + su)'u$$
- With $M_n := \max_{i \le n} |w_i|$ and small $\epsilon > 0$,

$$|R_n(T_n \pm \epsilon u)'u| \le \frac{1}{2} M_n G(M_n |T_n \pm \epsilon u|) |T_n \pm \epsilon u|^2$$

5. Directional derivatives for $\dot{L}_n(t) = T_n' - t' - R_n(t)'$

- Define concave $\mathcal{L}_n(s) = L_n(T_n + su)$ for fixed unit vector u.
- $$\dot{\mathcal{L}}_n(s) := \frac{d}{ds}\mathcal{L}_n(s) = \dot{L}_n(T_n + su)u = -s - R_n(T_n + su)'u$$
- With $M_n := \max_{i \le n} |w_i|$ and small $\epsilon > 0$,

$$|R_n(T_n \pm \epsilon u)'u| \le \frac{1}{2} M_n G(M_n |T_n \pm \epsilon u|) |T_n \pm \epsilon u|^2$$
$$\le \frac{\epsilon^2}{4G(1)N} G(1) \left(\frac{N}{\epsilon} + \epsilon^2 \right) < \frac{\epsilon}{2}$$

if

$$M_n \le \frac{\epsilon^2}{2G(1)N} \qquad \text{and} \qquad |T_n|^2 \le N/\epsilon$$

6. Maximization of concave $\mathfrak{L}_n(s) = L_n(T_n + su)$

- $\mathbb{P}\{|T_n|^2 > N/\epsilon\} < \epsilon$ because $\mathbb{P}|T_n|^2 = N$

6. Maximization of concave $\mathfrak{L}_n(s) = L_n(T_n + su)$

- $\mathbb{P}\{|T_n|^2 > N/\epsilon\} < \epsilon$ because $\mathbb{P}|T_n|^2 = N$
- If

$$M_n := \max_{i \le n} |w_i| \le \frac{\epsilon^2}{2G(1)N} \qquad \text{and} \qquad |T_n|^2 \le N/\epsilon$$

then

$$\dot{\mathfrak{L}}_n(-\epsilon) \ge \epsilon - \epsilon/2 > 0 > -\epsilon + \epsilon/2 \ge \dot{\mathfrak{L}}_n(+\epsilon) \qquad \text{for all } |u| = 1$$

6. Maximization of concave $\mathcal{L}_n(s) = L_n(T_n + su)$

- $\mathbb{P}\{|T_n|^2 > N/\epsilon\} < \epsilon$ because $\mathbb{P}|T_n|^2 = N$
- If

$$M_n := \max_{i \le n} |w_i| \le \frac{\epsilon^2}{2G(1)N} \qquad \text{and} \qquad |T_n|^2 \le N/\epsilon$$

then

$$\dot{\mathcal{L}}_n(-\epsilon) \ge \epsilon - \epsilon/2 > 0 > -\epsilon + \epsilon/2 \ge \dot{\mathcal{L}}_n(+\epsilon) \qquad \text{for all } |u| = 1$$

and

$$|J_n^{1/2}(\widehat{b}_n - \beta) - T_n| = |\widehat{t}_n - T_n| \le \epsilon$$

7. General case: $\theta_i = \langle x_i, \beta \rangle$ with $x_i \in \mathbb{R}^{\mathbb{N}}$

- Truncate $z = (z_1, z_2, \dots) = \pi_N z + \pi_N^{\perp} z$:

$$\pi_N z = (z_1, \dots, z_N, 0, 0, \dots), \qquad \pi_N^{\perp} z = (0, \dots, 0, z_{N+1}, z_{N+2}, \dots)$$

7. General case: $\theta_i = \langle x_i, \beta \rangle$ with $x_i \in \mathbb{R}^{\mathbb{N}}$

- Truncate $z = (z_1, z_2, \dots) = \pi_N z + \pi_N^{\perp} z$:

  $\pi_N z = (z_1, \dots, z_N, 0, 0, \dots), \qquad \pi_N^{\perp} z = (0, \dots, 0, z_{N+1}, z_{N+2}, \dots)$

- $$\widehat{b}_{n,N} = \underset{\pi_N^{\perp} b = 0}{\operatorname{argmax}} \sum_{i \le n} \Big(y_i \langle x_i, b \rangle - \Psi(\langle x_i, b \rangle) \Big)$$

7. General case: $\theta_i = \langle x_i, \beta \rangle$ with $x_i \in \mathbb{R}^{\mathbb{N}}$

- Truncate $z = (z_1, z_2, \dots) = \pi_N z + \pi_N^{\perp} z$:

  $\pi_N z = (z_1, \dots, z_N, 0, 0, \dots), \qquad \pi_N^{\perp} z = (0, \dots, 0, z_{N+1}, z_{N+2}, \dots)$

- $$\widehat{b}_{n,N} = \underset{\pi_N^{\perp} b = 0}{\operatorname{argmax}} \sum_{i \le n} \Big(y_i \langle x_i, b \rangle - \Psi(\langle x_i, b \rangle) \Big)$$

  $\theta_i^* = \langle x_i, \pi_N \beta \rangle = \sum_{j \le N} \beta_j x_{i,j}$ and $\delta_i = \theta_i - \theta_i^*$

7. General case: $\theta_i = \langle x_i, \beta \rangle$ with $x_i \in \mathbb{R}^{\mathbb{N}}$

- $$\widehat{b}_{n,N} = \underset{\pi_N^{\perp} b = 0}{\operatorname{argmax}} \sum_{i \leq n} \Big(y_i \langle x_i, b \rangle - \Psi(\langle x_i, b \rangle) \Big)$$

  $\theta_i^* = \langle x_i, \pi_N \beta \rangle = \sum_{j \leq N} \beta_j x_{i,j}$ and $\delta_i = \theta_i - \theta_i^*$

- (Conditional) distributions for $y_1, \ldots, y_n$:

  $$\mathbb{P}_n = \otimes_{i \leq n} Q_{\theta_i}, \qquad \mathbb{Q}_{n,N} = \otimes_{i \leq n} Q_{\theta_i^*}$$

  Bias term δ_i for $\widehat{b}_{n,N}$ under $\mathbb{P}_n$; no bias under $\mathbb{Q}_{n,N}$

7. General case: $\theta_i = \langle x_i, \beta \rangle$ with $x_i \in \mathbb{R}^{\mathbb{N}}$

- $$\widehat{b}_{n,N} = \underset{\pi_N^{\perp} b = 0}{\operatorname{argmax}} \sum_{i \le n} \Big(y_i \langle x_i, b \rangle - \Psi(\langle x_i, b \rangle) \Big)$$

- (Conditional) distributions for $y_1, \ldots, y_n$:

  $$\mathbb{P}_n = \otimes_{i \le n} Q_{\theta_i}, \qquad \mathbb{Q}_{n,N} = \otimes_{i \le n} Q_{\theta_i^*}$$

  Bias term δ_i for $\widehat{b}_{n,N}$ under $\mathbb{P}_n$; no bias under $\mathbb{Q}_{n,N}$
- Theory from N_n case (as before) gives

  $$\mathbb{Q}_{n,N}\{\widehat{b}_{n,N} \approx \pi_N \beta\} \approx 1 \qquad \text{if} \ldots$$

7. General case: $\theta_i = \langle x_i, \beta \rangle$ with $x_i \in \mathbb{R}^{\mathbb{N}}$

- (Conditional) distributions for $y_1, \dots, y_n$:

$$\mathbb{P}_n = \otimes_{i \le n} Q_{\theta_i}, \qquad \mathbb{Q}_{n,N} = \otimes_{i \le n} Q_{\theta_i^*}$$

  Bias term δ_i for $\widehat{b}_{n,N}$ under $\mathbb{P}_n$; no bias under $\mathbb{Q}_{n,N}$

- Theory from N_n case (as before) gives

$$\mathbb{Q}_{n,N}\{\widehat{b}_{n,N} \approx \pi_N \beta\} \approx 1 \qquad \text{if} \dots$$

- Assumptions on N_n and on decay of β_j's needed to make $\|\mathbb{P}_n - \mathbb{Q}_{n,N}\|_{\mathrm{TV}} \to 0$ and $|\pi_N^{\perp}\beta|$ suitably small.

7. General case: $\theta_i = \langle x_i, \beta \rangle$ with $x_i \in \mathbb{R}^{\mathbb{N}}$

- (Conditional) distributions for $y_1, \ldots, y_n$:

$$\mathbb{P}_n = \otimes_{i \le n} Q_{\theta_i}, \qquad \mathbb{Q}_{n,N} = \otimes_{i \le n} Q_{\theta_i^*}$$

  Bias term δ_i for $\widehat{b}_{n,N}$ under $\mathbb{P}_n$; no bias under $\mathbb{Q}_{n,N}$

- Theory from N_n case (as before) gives

$$\mathbb{Q}_{n,N}\{\widehat{b}_{n,N} \approx \pi_N \beta\} \approx 1 \qquad \text{if } \ldots$$

- Assumptions on N_n and on decay of β_j's needed to make $\|\mathbb{P}_n - \mathbb{Q}_{n,N}\|_{\mathrm{TV}} \to 0$ and $|\pi_N^{\perp} \beta|$ suitably small.

-

$$\begin{aligned} H^2(\mathbb{P}_n, \mathbb{Q}_{n,N}) &= \textstyle\sum_{i \le n} H^2(Q_{\theta_i}, Q_{\theta_i^*}) \\ &\le \textstyle\sum_{i \le n} \delta_i^2 \ddot{\Psi}(\theta_i)\,(1 + |\delta_i|)\, G(|\delta_i|) \end{aligned}$$

7. General case: $\theta_i = \langle x_i, \beta \rangle$ with $x_i \in \mathbb{R}^{\mathbb{N}}$

- (Conditional) distributions for $y_1, \dots, y_n$:

$$\mathbb{P}_n = \otimes_{i \le n} Q_{\theta_i}, \qquad \mathbb{Q}_{n,N} = \otimes_{i \le n} Q_{\theta_i^*}$$

  Bias term δ_i for $\widehat{b}_{n,N}$ under $\mathbb{P}_n$; no bias under $\mathbb{Q}_{n,N}$

- Theory from N_n case (as before) gives

$$\mathbb{Q}_{n,N}\{\widehat{b}_{n,N} \approx \pi_N \beta\} \approx 1 \qquad \text{if} \dots$$

- Assumptions on N_n and on decay of β_j's needed to make $\|\mathbb{P}_n - \mathbb{Q}_{n,N}\|_{\text{TV}} \to 0$ and $|\pi_N^{\perp}\beta|$ suitably small.
- Deduce $\mathbb{P}_n\{\widehat{b}_{n,N} \approx \pi_N \beta\} \approx 1$ if ....

References

Hall, P. and J. L. Horowitz (2007).
Methodology and convergence rates for functional linear regression.
Annals of Statistics 35(1), 70–91.

Hjort, N. L. and D. Pollard (1993).
Asymptotics for minimisers of convex processes.
Technical report, Yale University.