Bayesian, Misspecified, Parametric

IDEA lab

Department of Statistics, Seoul National University

May 20, 2026

Preliminaries

Notation

- $X_1, \cdots, X_n \overset{i.i.d}{\sim} G$ and P_θ, $\theta \in \Theta \subseteq \mathbb{R}^d$ is a misspecified model.
- g and p_θ are densities of each G and P_θ.
- $K(\theta) := \mathrm{KL}(g, p_\theta)$ and $\Theta_G = \{\theta_G \in \Theta | \theta_G = \mathrm{argmin}_{\theta \in \Theta} K(\theta)\}$.
- Let $\theta \sim \pi(\theta)$ be the prior distribution and $\pi(\theta | X_1, \cdots, X_n) \propto \pi(\theta) \prod_{i=1}^n p_\theta(X_i)$ be the posterior distribution under the misspecified model.
- For a loss function $L : \Theta \times \Theta \to [0, \infty)$,

$$\hat{\theta}_n = \mathrm{argmin}_{t \in \theta} \mathbb{E}_{\theta | X_{1:n}} [L(t, \theta) | X_1, \cdots, X_n] \tag{1}$$

  is called a pseudo-Bayes estimator.

Goal: Strong consistency and asymptotic normality of the pseudo-Bayes estimator.

Conditions

Assumption

A1. Θ is a closed convex set in $\mathbb{R}^d$ with a nonempty interior, the density $p_\theta(x)$ is bounded over $\Theta \times \mathbb{R}^k$ and its support is the same for all $\theta \in \Theta$.
A2. For all $\theta \in \Theta$, there is a sphere $S[\theta, \eta(\theta)]$ such that

$$E \sup \left\{ \left| \log \left(\frac{g(X)}{p_t(X)} \right) \right| ; t \in S[\theta, \eta(\theta)] \right\} < \infty. \tag{2}$$

A3. For all fixed $x \in \mathbb{R}^k$, the density $p_\theta(x)$ has a continuous derivative $p'_\theta(x)$ w.r.t. θ and there are positive constants c, b_0 with

$$\int ||[p_\theta(x)]^{-1} p'_\theta(x)||^{4(d+1)} p_\theta(x) dx < c(1 + ||\theta||^{b_0}) \tag{3}$$

for all $\theta \in \Theta$.

Theorem

Assumption (Continued)

A4. For some positive constant b_1,

$$Q(\theta) = \int [p_\theta(x)g(x)]^{1/2}dx < c||\theta||^{-b_1}, \quad \theta \in \Theta. \tag{4}$$

A5. There are positive constants b_2, b_3 so that for all $\theta \in \Theta$ and $r > 0$ it holds that

$$0 < \pi(S(\theta, r)) \leq cr^{b_2}(1 + (||\theta|| + r)^{b_3}). \tag{5}$$

Theorem

Under the assumption A1-A5, following holds for all $p > 0$.
$\mathbb{E}_{\theta|X_{1:n}}(d_G^p(\theta)|X_1, \cdots, X_n) \overset{a.s.}{\to} 0$ *where* $d_G(\theta) = \min\{||\theta - t|| \mid t \in \Theta_G\}$.

Theorem

Assumption (Additional)

A6. Let $L : \Theta \times \Theta \to [0, \infty)$ be a measurable loss function with $L(\theta, \theta) = 0$, c_1, c_2, c_3, b_4, b_5 be positive constants with

$$(c_1||t - \theta||^{b_4}) \wedge c_2 \leq L(t, \theta) \leq c_3||t - \theta||^{b_5} \tag{6}$$

for all $t, \theta \in \Theta$.

Theorem

If θ_G is unique, it holds under the assumption A1-A6 that for all pssudo-Bayes estimators $\hat{\theta}_n$ w.r.t. the loss function L,

$$\hat{\theta}_n \overset{a.s.}{\to} \theta_G. \tag{7}$$

Asymptotic Normality

- Under some regularity conditions regarding the $l(x, \theta) := \log \frac{p_\theta(x)}{g(x)}$ and the loss function L, the author shows the asymptotic normality of the pseudo-Bayes estimators.

Theorem

If θ_G is unique, the pseudo-Bayes estimator $\hat{\theta}_n$ is asymptotically normal.

$$\sqrt{n}(\hat{\theta}_n - \theta_G) \xrightarrow{d} N(0, \Sigma) \tag{8}$$

where, $\Sigma = L_2^{-1} L_1 M^{-1} I_G M^{-1} (L_2^{-1} L_1)^\top$, $I_G := E\left[l'(X, \theta_G) l'(X, \theta_G)^\top\right]$, $M := -El''(X, \theta_G)$, $L_1 = L^{(1,1)}(\theta_G, \theta_G)$ and $L_2 = L^{(2,0)}(\theta_G, \theta_G)$. Note that $l'(x, \theta_G) = \frac{\partial}{\partial\theta} l(x, \theta_G)$ and $l''(x, \theta_G) = \frac{\partial^2}{\partial\theta^2} l(x, \theta_G)$.

Specified Model

- Now, think about specified model for comparison.

$$X_1, \cdots, X_n \overset{i.i.d}{\sim} P_{\theta_0} \text{ for some true } \theta_0 \in \Theta. \tag{9}$$

- I.e., we consider this as $g(x) = p_{\theta_0}(x)$ and $\Theta_G = \{\theta_0\}$.
- In this case, note that $I_0 = E\left[l'(X, \theta_0) l'(X, \theta_0)^\top\right] = -El''(X, \theta_0) = M$.
- Posterior consistency:

$$\pi\left(\{\theta : ||\theta - \theta_0|| > \epsilon\} | X_1, \cdots, X_n\right) \overset{a.s.}{\to} 0 \tag{10}$$

for any $\epsilon > 0$.

Proof Comparison

Specified model
Let $r_n(t) = \mathbb{E}_{\theta|X_{1:n}}[L(t,\theta)|X_1,\cdots,X_n]$. From the definition of $\hat{\theta}_n$, $r_n'(\hat{\theta}_n) = 0$. Then, do Talyor expansion $r_n'(t)$ at $t = \theta_0$,

$$0 = r_n'(\hat{\theta}_n) \approx r_n'(\theta_0) + r_n''(\theta_0)(\hat{\theta}_n - \theta_0).$$

From this, $\sqrt{n}(\hat{\theta}_n - \theta_0) \approx -[r_n''(\theta_0)]^{-1}\sqrt{n}r_n'(\theta_0)$.
First,
$r_n''(\theta_0) =$
$\mathbb{E}_{\theta|X_{1:n}}\left[L^{(2,0)}(\theta_0,\theta)|X_1,\cdots,X_n\right] \overset{a.s.}{\to}$
$L^{(2,0)}(\theta_0,\theta_0)$ by the posterior consistency.
Second,

$$\sqrt{n}r_n'(\theta_0) = \sqrt{n}\mathbb{E}_\theta\left[L^{(1,0)}(\theta_0,\theta)|X_{1:n}\right]$$
$$\approx \sqrt{n}\mathbb{E}_\theta[L^{(1,0)}(\theta_0,\theta_0)$$
$$+ L^{(1,1)}(\theta_0,\theta_0)(\theta-\theta_0)|X_{1:n}]$$
$$= L^{(1,1)}(\theta_0,\theta_0)\mathbb{E}_\theta\left[\sqrt{n}(\theta-\theta_0)|X_{1:n}\right].$$

Misspecified model
Let $r_n(t) = \mathbb{E}_{\theta|X_{1:n}}[L(t,\theta)|X_1,\cdots,X_n]$. From the definition of $\hat{\theta}_n$, $r_n'(\hat{\theta}_n) = 0$. Then, do Talyor expansion $r_n'(t)$ at $t = \theta_G$,

$$0 = r_n'(\hat{\theta}_n) \approx r_n'(\theta_G) + r_n''(\theta_G)(\hat{\theta}_n - \theta_G).$$

From this, $\sqrt{n}(\hat{\theta}_n - \theta_G) \approx -[r_n''(\theta_G)]^{-1}\sqrt{n}r_n'(\theta_G)$.
First,
$r_n''(\theta_G) =$
$\mathbb{E}_{\theta|X_{1:n}}\left[L^{(2,0)}(\theta_G,\theta)|X_1,\cdots,X_n\right] \overset{a.s.}{\to}$
$L^{(2,0)}(\theta_G,\theta_G)$ by the above theorem.
Second,

$$\sqrt{n}r_n'(\theta_G) = \sqrt{n}\mathbb{E}_\theta\left[L^{(1,0)}(\theta_G,\theta)|X_{1:n}\right]$$
$$\approx \sqrt{n}\mathbb{E}_\theta[L^{(1,0)}(\theta_G,\theta_G)$$
$$+ L^{(1,1)}(\theta_G,\theta_G)(\theta-\theta_G)|X_{1:n}]$$
$$= L^{(1,1)}(\theta_G,\theta_G)\mathbb{E}_\theta\left[\sqrt{n}(\theta-\theta_G)|X_{1:n}\right].$$

Proof Comparison

Now then, let $t = \sqrt{n}(\theta - \theta_0)$ and obtain the distribution of $t|X_{1:n}$.
From change of variable, we can know that

$$\pi(t|X_{1:n}) \propto \pi\left(\theta_0 + \frac{t}{\sqrt{n}}\right) \frac{Z_n\left(\theta_0 + \frac{t}{\sqrt{n}}\right)}{Z_n(\theta_0)}.$$

So that,

$$\begin{aligned} &\log \pi(t|X_{1:n}) \\ \approx &\log Z_n\left(\theta_0 + \frac{t}{\sqrt{n}}\right) - \log Z_n(\theta_0) + \text{Constant}. \end{aligned}$$

Applying second order Taylor expansion to $Z_n(\theta)$,

$$\begin{aligned} &\log Z_n\left(\theta_0 + \frac{t}{\sqrt{n}}\right) - \log Z_n(\theta_0) \\ \approx &\frac{t^\top}{\sqrt{n}} \sum_i^n l'(X_i, \theta_0) - \frac{1}{2}\frac{t^\top}{\sqrt{n}}\left[\sum_{i=1}^n -l''(X_i, \theta_0)\right]\frac{t}{\sqrt{n}}. \end{aligned}$$

Now then, let $t = \sqrt{n}(\theta - \theta_G)$ and obtain the distribution of $t|X_{1:n}$.
From change of variable, we can know that

$$\pi(t|X_{1:n}) \propto \pi\left(\theta_G + \frac{t}{\sqrt{n}}\right) \frac{Z_n\left(\theta_G + \frac{t}{\sqrt{n}}\right)}{Z_n(\theta_G)}.$$

So that,

$$\begin{aligned} &\log \pi(t|X_{1:n}) \\ \approx &\log Z_n\left(\theta_G + \frac{t}{\sqrt{n}}\right) - \log Z_n(\theta_G) + \text{Constant}. \end{aligned}$$

Applying second order Taylor expansion to $Z_n(\theta)$,

$$\begin{aligned} &\log Z_n\left(\theta_G + \frac{t}{\sqrt{n}}\right) - \log Z_n(\theta_G) \\ \approx &\frac{t^\top}{\sqrt{n}} \sum_i^n l'(X_i, \theta_G) - \frac{1}{2}\frac{t^\top}{\sqrt{n}}\left[\sum_{i=1}^n -l''(X_i, \theta_G)\right]\frac{t}{\sqrt{n}} \end{aligned}$$

Proof Comparison

Then,

$$\log \pi(t|X_{1:n}) \approx t^\top S_n - \frac{1}{2} t^\top M_n t + \text{Constant},$$

where $S_n = \frac{1}{\sqrt{n}} \sum_i^n l'(X_i, \theta_0)$ and
$M_n = \frac{1}{n} \left[\sum_{i=1}^n -l''(X_i, \theta_0)\right]$.
From this formula, we can obtain
$\pi(t|X_{1:n}) \approx N(M_n^{-1} S_n, M_n^{-1})$. So that,

$$\mathbb{E}_\theta \left[\sqrt{n}(\theta - \theta_0)|X_{1:n}\right] \approx M_n^{-1} S_n \xrightarrow{d} M^{-1} N(0, I_0).$$

So that, $\mathbb{E}_\theta \left[\sqrt{n}(\theta - \theta_0)|X_{1:n}\right] \xrightarrow{d} N(0, M^{-1})$.
By combine above results,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \left[L^{(2,0)}(\theta_0, \theta_0)\right]^{-1} L^{(1,1)}(\theta_0, \theta_0) N(0, M^{-1}).$$

Then,

$$\log \pi(t|X_{1:n}) \approx t^\top S_n - \frac{1}{2} t^\top M_n t + \text{Constant},$$

where $S_n = \frac{1}{\sqrt{n}} \sum_i^n l'(X_i, \theta_G)$ and
$M_n = \frac{1}{n} \left[\sum_{i=1}^n -l''(X_i, \theta_G)\right]$.
From this formula, we can obtain
$\pi(t|X_{1:n}) \approx N(M_n^{-1} S_n, M_n^{-1})$. So that,

$$\mathbb{E}_\theta \left[\sqrt{n}(\theta - \theta_G)|X_{1:n}\right] \approx M_n^{-1} S_n \xrightarrow{d} M^{-1} N(0, I_G).$$

By combine above results,

$$\sqrt{n}(\hat{\theta}_n - \theta_G) \xrightarrow{d} \left[L^{(2,0)}(\theta_G, \theta_G)\right]^{-1} L^{(1,1)}(\theta_G, \theta_G) M^{-1} N(0, I_G).$$