MA204: Mathematical Statistics

Tutorial 5

T5.1 Bayesian Estimator

Step 1 Given an i.i.d. sample $\mathbf{x} = (X_1, \ldots, X_n)^\top$, determine the joint pdf of $\mathbf{x}$ and θ,

$$f(x_1, \ldots, x_n, \theta) = \text{Likelihood} \times \text{Prior} = \left\{ \prod_{i=1}^{n} f(x_i \mid \theta) \right\} \times \pi(\theta).$$

Step 2 Determine the posterior density of θ (i.e., the conditional density of θ given $X_i = x_i$ for $i = 1, \ldots, n$),

$$p(\theta \mid x_1, \ldots, x_n) = \frac{f(x_1, \ldots, x_n, \theta)}{\int_\Theta f(x_1, \ldots, x_n, \theta)\, \mathrm{d}\theta} \propto \text{Likelihood} \times \text{Prior}.$$

Step 3 The Bayesian estimate of θ (i.e., the conditional expectation of θ) is defined by

$$E(\theta \mid x_1, \ldots, x_n) = \int_\Theta \theta \cdot p(\theta \mid x_1, \ldots, x_n)\, \mathrm{d}\theta.$$

Example T5.1 (A normal population with known variance). Let $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$, where the variance σ^2 is known. Assume that the prior distribution of μ is $N(\mu_0, \sigma_0^2)$. Show that the posterior distribution of μ is $N(\mu^*, \sigma^{2*})$, where

$$\mu^* = \frac{n\sigma_0^2 \bar{x} + \sigma^2 \mu_0}{n\sigma_0^2 + \sigma^2}, \quad \sigma^{2*} = \frac{\sigma_0^2 \sigma^2}{n\sigma_0^2 + \sigma^2},$$

and $\bar{x}$ is the sample mean.

Proof: The likelihood function is

$$f(x_1, \ldots, x_n \mid \mu) = \prod_{i=1}^{n} f(x_i \mid \mu) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left\{ -\sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2} \right\}.$$

Since the prior density function of μ is

$$\pi(\mu) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left\{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right\}, \quad -\infty < \mu < \infty,$$

the posterior density function of μ is

$$\begin{aligned}
& p(\mu \mid x_1, \ldots, x_n) \propto f(x_1, \ldots, x_n \mid \mu) \times \pi(\mu) \\
= \;& \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left\{-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}\right\} \times \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left\{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right\} \\
\propto \;& \exp\left\{-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}\right\} \times \exp\left\{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right\} \\
= \;& \exp\left\{-\sum_{i=1}^{n} \frac{(x_i-\bar{x})^2 + (\bar{x}-\mu)^2}{2\sigma^2}\right\} \times \exp\left\{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right\} \\
= \;& \exp\left\{-\sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{2\sigma^2}\right\} \times \exp\left\{-\frac{1}{2}\left[\frac{n(\mu-\bar{x})^2}{\sigma^2} + \frac{(\mu-\mu_0)^2}{\sigma_0^2}\right]\right\} \\
\propto \;& \exp\left\{-\frac{1}{2}\left[\frac{n(\mu-\bar{x})^2}{\sigma^2} + \frac{(\mu-\mu_0)^2}{\sigma_0^2}\right]\right\} \\
= \;& \exp\left\{-\frac{1}{2}\left[\frac{n\sigma_0^2\mu^2 - 2n\sigma_0^2\mu\bar{x} + n\sigma_0^2\bar{x}^2 + \sigma^2\mu^2 - 2\sigma^2\mu\mu_0 + \sigma^2\mu_0^2}{\sigma^2\sigma_0^2}\right]\right\} \\
\propto \;& \exp\left\{-\frac{1}{2}\left[\frac{(n\sigma_0^2+\sigma^2)\mu^2 - 2(n\sigma_0^2\bar{x} + \sigma^2\mu_0)\mu}{\sigma^2\sigma_0^2}\right]\right\} \\
= \;& \exp\left\{-\frac{1}{2}\left[\frac{\mu^2 - 2\dfrac{n\sigma_0^2\bar{x}+\sigma^2\mu_0}{n\sigma_0^2+\sigma^2}\mu}{\dfrac{\sigma^2\sigma_0^2}{n\sigma_0^2+\sigma^2}}\right]\right\} \\
\propto \;& \exp\left\{-\frac{\left(\mu - \dfrac{n\sigma_0^2\bar{x}+\sigma^2\mu_0}{n\sigma_0^2+\sigma^2}\right)^2}{2\cdot\dfrac{\sigma^2\sigma_0^2}{n\sigma_0^2+\sigma^2}}\right\} = \exp\left\{-\frac{(\mu-\mu^*)^2}{2\sigma^{2*}}\right\}.
\end{aligned}$$

From the kernel of $p(\mu \mid x_1, \ldots, x_n)$, we know that the posterior distribution of μ is a normal distribution with mean μ^* and variance σ^{2*}. $\|$

Example T5.2 (Beta-binomial distribution). Let $\theta \sim \text{Beta}(\alpha, \beta)$ and $X|\theta \sim \text{Binomial}(n, \theta)$, then X is said to follow the beta-binomial distribution, denoted by $X \sim \text{BBinomial}(n, \alpha, \beta)$ with pmf

$$\text{BBinomial}(x|n, \alpha, \beta) = \binom{n}{x} \frac{B(x+\alpha, n-x+\beta)}{B(\alpha, \beta)},$$

for $x = 0, 1, \ldots, n$, where $n > 0$ is an integer and $\alpha, \beta > 0$.

Proof: The joint distribution of (X, θ) is

$$f(x, \theta) = f(\theta) \times f(x|\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)} \times \binom{n}{x} \theta^x (1-\theta)^{n-x},$$

so that the marginal distribution (i.e., pmf) of X is

$$\begin{aligned}
\text{BBinomial}(x|n, \alpha, \beta) &= \int_0^1 f(x, \theta)\mathrm{d}\theta \\
&= \binom{n}{x} \frac{1}{B(\alpha, \beta)} \int_0^1 \theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}\mathrm{d}\theta \\
&= \binom{n}{x} \frac{B(x+\alpha, n-x+\beta)}{B(\alpha, \beta)}, \quad x = 0, 1, \ldots, n.
\end{aligned}$$

Understanding Example 3.11: In Bayesian statistics, that $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Bernoulli}(\theta)$ should be understood as $X_1|\theta, \ldots, X_n|\theta \overset{\text{iid}}{\sim} \text{Bernoulli}(\theta)$. We now consider the distribution of $\mathbf{x} = (X_1, \ldots, X_n)^\top$, which is given by

$$\begin{aligned}
f(\boldsymbol{x}) &= f(x_1, \ldots, x_n) = \int_0^1 f(\boldsymbol{x}, \theta)\mathrm{d}\theta \\
&= \frac{1}{B(\alpha, \beta)} \int_0^1 \theta^{x_+ + \alpha - 1}(1-\theta)^{n-x_+ + \beta - 1}\mathrm{d}\theta \\
&= \frac{B(x_+ + \alpha, n - x_+ + \beta)}{B(\alpha, \beta)}, \quad x_i \in \{0, 1\},\ i = 1, \ldots, n,
\end{aligned}$$

where $x_+ = \sum_{i=1}^n x_i \in \{0, 1, \ldots, n\}$.

Define $X_+ = \sum_{i=1}^n X_i$, then $X_+|\theta \sim \text{Binomial}(n, \theta)$. Since $\theta \sim \text{Beta}(\alpha, \beta)$, we have $X_+ \sim \text{BBinomial}(n, \alpha, \beta)$. ||

Example T5.3 (Gamma-Poisson (mixture) distribution). Let $\theta \sim \text{Gamma}(a, b)$ and $X|\theta \sim \text{Poisson}(\theta)$, then X is said to follow the gamma-Poisson (mixture) distribution, denoted by $X \sim \text{GPoisson}(a, b)$ with pmf

$$\text{GPoisson}(x|a, b) = \frac{\Gamma(x+a)}{x!\Gamma(a)} \left(\frac{b}{b+1}\right)^a \left(\frac{1}{b+1}\right)^x, \tag{T5.1}$$

for $x = 0, 1, \ldots, \infty$, where the shape parameter $a > 0$ and the rate parameter $b > 0$.

— We have $E(X) = a/b \hat{=} \mu$ and $\text{Var}(X) = a(b+1)/b^2 = \mu(1 + 1/b) > \mu$, so the gamma-Poisson distribution is over-dispersed.

— In (T5.1), let $b/(b+1) = p$ and $a = r$, the gamma-Poisson reduces to the negative-binomial distribution or the Polya distribution (after George Pólya), denoted by $X \sim \text{NBinomial}(r, p)$ with pmf

$$\text{NBinomial}(x|r, p) = \frac{\Gamma(x+r)}{x!\Gamma(r)} p^r (1-p)^x, \quad x = 0, 1, \ldots, \infty,$$

where $r > 0$ is a **real number** and $p \in (0, 1)$.

Proof: The joint distribution of (X, θ) is

$$f(x, \theta) = f(\theta) \times f(x|\theta) = \frac{b^a \cdot \theta^{a-1} \mathrm{e}^{-b\theta}}{\Gamma(a)} \times \frac{\theta^x \mathrm{e}^{-\theta}}{x!},$$

so that the marginal distribution (i.e., pmf) of X is

$$\begin{aligned} \text{GPoisson}(x|a, b) &= \int_0^\infty f(x, \theta) \mathrm{d}\theta = \frac{b^a}{x!\Gamma(a)} \int_0^\infty \theta^{x+a-1} \mathrm{e}^{-(b+1)\theta} \mathrm{d}\theta \\ &= \frac{b^a}{x!\Gamma(a)} \cdot \frac{\Gamma(x+a)}{(b+1)^{x+a}} \\ &= \frac{\Gamma(x+a)}{x!\Gamma(a)} \left(\frac{b}{b+1}\right)^a \left(\frac{1}{b+1}\right)^x, \quad x = 0, 1, \ldots, \infty. \end{aligned}$$

Understanding Example 3.12: In Bayesian statistics, that $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Poisson}(\theta)$ should be understood as $X_1|\theta, \ldots, X_n|\theta \overset{\text{iid}}{\sim} \text{Poisson}(\theta)$. We now consider the distribution of $\mathbf{x} = (X_1, \ldots, X_n)^\top$, which is given by

$$
\begin{aligned}
f(\boldsymbol{x}) &= f(x_1, \ldots, x_n) = \int_0^\infty f(\boldsymbol{x}, \theta)\mathrm{d}\theta \\
&= \frac{b^a}{\Gamma(a)\prod_{i=1}^n x_i!}\int_0^\infty \theta^{a+x_+-1}\mathrm{e}^{-(b+n)\theta}\mathrm{d}\theta \\
&= \frac{b^a}{\Gamma(a)\prod_{i=1}^n x_i!} \times \frac{\Gamma(a+x_+)}{(b+n)^{a+x_+}} \\
&= \frac{\Gamma(a+x_+)}{\Gamma(a)\prod_{i=1}^n x_i!}\left(\frac{b}{b+n}\right)^a\left(\frac{1}{b+n}\right)^{x_+}, \quad x_i \in \{0, 1, \ldots, \infty\},\ i = 1, \ldots, n,
\end{aligned}
$$

where $x_+ = \sum_{i=1}^n x_i \in \{0, 1, \ldots, \infty\}$.

Define $X_+ = \sum_{i=1}^n X_i$, then $X_+|\theta \sim \text{Poisson}(n\theta)$. Since $\theta \sim \text{Gamma}(a, b)$, we have $n\theta \sim \text{Gamma}(a, b/n)$ and hence $X_+ \sim \text{GPoisson}(a, b/n)$. $\|$

T5.2 Asymptotic Efficiency of MLE

A sequence of estimators $\{W_n\}_{n=1}^\infty$ is said to be *asymptotically efficient* for a parameter $\tau(\theta)$, if $\sqrt{n}[W_n - \tau(\theta)] \overset{\text{L}}{\to} N(0, v(\theta))$, where

$$
v(\theta) = \frac{[\tau'(\theta)]^2}{I_n(\theta)} \quad \text{and} \quad I_n(\theta) = \text{Var}_{\mathbf{x}}\left(\frac{\mathrm{d}\log L(\theta; \mathbf{x})}{\mathrm{d}\theta}\right),
$$

i.e., the asymptotic variance of $\sqrt{n}W_n$ achieves the *Cramér–Rao lower bound.*

Example T5.4 (Asymptotic efficiency of MLEs). Let $X_1, \ldots, X_n$ be a random sample with pdf $f(x;\theta)$, and $\hat{\theta}$ be the MLE of θ. We assume that $f(x;\theta)$ satisfies the following regularity conditions:

(C1) The parameter is identifiable, i.e., if $\theta \neq \theta^*$, then $f(x;\theta) \neq f(x;\theta^*)$.

(C2) The density $f(x;\theta)$ is differentiable with respect to θ inside its support.

(C3) The parameter space Θ contains an open set ω of which the true parameter value θ_0 is an interior point.

Let $\mathbf{x} = (X_1, \ldots, X_n)^\top$, show that

(a) $\frac{1}{\sqrt{n}}\ell'(\theta_0; \mathbf{x}) = \sqrt{n} \cdot \frac{1}{n}\sum_{i=1}^{n} W_i \xrightarrow{\text{L}} N(0, I(\theta_0))$, where $W_i = \mathrm{d}\log f(X_i; \theta)/\mathrm{d}\theta\,|_{\theta=\theta_0}$ has mean 0 and variance $I(\theta_0)$.

(b) $-\frac{1}{n}\ell''(\theta_0; \mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} W_i^2 - \frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{d}^2 f(X_i; \theta)/\mathrm{d}\theta^2\,|_{\theta=\theta_0}}{f(X_i; \theta_0)}$, and the expectations of W_i^2 and $\frac{\mathrm{d}^2 f(X_i; \theta)/\mathrm{d}\theta^2\,|_{\theta=\theta_0}}{f(X_i; \theta_0)}$ equal to $I(\theta_0)$ and 0, respectively, for $i = 1, \ldots, n$. Furthermore, we have $-\frac{1}{n}\ell''(\theta_0; \mathbf{x}) \xrightarrow{\text{P}} I(\theta_0)$.

(c) $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\text{L}} N(0, v(\theta))$, where $v(\theta)$ is the Cramér–Rao lower bound, i.e., $\hat{\theta}$ is an asymptotically efficient estimator of θ.

Proof: (a) It is easy to verify that

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\ell'(\theta_0; \mathbf{x}) &= \frac{1}{\sqrt{n}}\left.\frac{\mathrm{d}\ell(\theta; \mathbf{x})}{\mathrm{d}\theta}\right|_{\theta=\theta_0} = \frac{1}{\sqrt{n}}\frac{\mathrm{d}}{\mathrm{d}\theta}\left[\sum_{i=1}^{n}\log f(X_i; \theta)\right]\Bigg|_{\theta=\theta_0} \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\left.\frac{\mathrm{d}\log f(X_i; \theta)}{\mathrm{d}\theta}\right|_{\theta=\theta_0}\right] = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} W_i\right), \\
E(W_i) &= E\left(\left.\frac{\mathrm{d}\log f(X_i; \theta)}{\mathrm{d}\theta}\right|_{\theta=\theta_0}\right) \\
&= \int_{\mathbb{R}}\left[\left.\frac{\mathrm{d}\log f(x_i; \theta)}{\mathrm{d}\theta}\right|_{\theta=\theta_0} \times f(x_i; \theta_0)\right]\mathrm{d}x_i \\
&= \int_{\mathbb{R}}\left[\left.\frac{\mathrm{d}f(x_i; \theta)}{\mathrm{d}\theta}\right|_{\theta=\theta_0}\right]\mathrm{d}x_i = \frac{\mathrm{d}}{\mathrm{d}\theta}\left[\int_{\mathbb{R}} f(x_i; \theta)\,\mathrm{d}x_i\right]\Bigg|_{\theta=\theta_0} = 0, \quad \text{and}
\end{aligned}
$$

$$\mathrm{Var}(W_i) \;=\; \mathrm{Var}\left(\frac{\mathrm{d}\log f(X_i;\theta)}{\mathrm{d}\theta}\bigg|_{\theta=\theta_0} \right) = I(\theta_0).$$

By the Central Limit theorem, we have

$$\sqrt{n}\left[\overline{W} - E(W_i)\right] \xrightarrow{\mathrm{L}} N(0, \mathrm{Var}(W_i)),$$

and

$$\frac{1}{\sqrt{n}}\ell'(\theta_0;\mathbf{x}) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} W_i\right) \xrightarrow{\mathrm{L}} N(0, I(\theta_0)).$$

(b) We have

$$\begin{aligned}
\ell''(\theta;\mathbf{x}) \;&=\; \frac{\mathrm{d}^2\ell(\theta;\mathbf{x})}{\mathrm{d}\theta^2} = \frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\left[\sum_{i=1}^{n}\log f(X_i;\theta)\right] \\
&=\; \sum_{i=1}^{n}\frac{\mathrm{d}^2\log f(X_i;\theta)}{\mathrm{d}\theta^2} = \sum_{i=1}^{n}\frac{\mathrm{d}}{\mathrm{d}\theta}\left[\frac{\mathrm{d}\,f(X_i;\theta)/\mathrm{d}\theta}{f(X_i;\theta)}\right] \\
&=\; \sum_{i=1}^{n}\frac{\mathrm{d}^2\,f(X_i;\theta)/\mathrm{d}\theta^2 \times f(X_i;\theta) - [\mathrm{d}\,f(X_i;\theta)/\mathrm{d}\theta]^2}{[f(X_i;\theta)]^2} \\
&=\; \sum_{i=1}^{n}\frac{\mathrm{d}^2\,f(X_i;\theta)/\mathrm{d}\theta^2}{f(X_i;\theta)} - \sum_{i=1}^{n}\left[\frac{\mathrm{d}\log f(X_i;\theta)}{\mathrm{d}\theta}\right]^2,
\end{aligned}$$

so that

$$\begin{aligned}
-\frac{1}{n}\ell''(\theta_0;\mathbf{x}) \;&=\; -\frac{1}{n}\left\{\sum_{i=1}^{n}\frac{\mathrm{d}^2\,f(X_i;\theta)/\mathrm{d}\theta^2}{f(X_i;\theta)} - \sum_{i=1}^{n}\left[\frac{\mathrm{d}\log f(X_i;\theta)}{\mathrm{d}\theta}\right]^2\right\}\bigg|_{\theta=\theta_0} \\
&=\; -\frac{1}{n}\left\{\sum_{i=1}^{n}\frac{\mathrm{d}^2\,f(X_i;\theta)/\mathrm{d}\theta^2\,|_{\theta=\theta_0}}{f(X_i;\theta_0)} - \sum_{i=1}^{n}\left[\frac{\mathrm{d}\log f(X_i;\theta)}{\mathrm{d}\theta}\bigg|_{\theta=\theta_0}\right]^2\right\} \\
&=\; \frac{1}{n}\sum_{i=1}^{n}W_i^2 - \frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{d}^2 f(X_i;\theta)/\mathrm{d}\theta^2\,|_{\theta=\theta_0}}{f(X_i;\theta_0)}.
\end{aligned}$$

For $i = 1, \ldots, n$, we obtain

$$\begin{aligned} E(W_i^2) &= \mathrm{Var}(W_i) + [E(W_i)]^2 = I(\theta_0) \quad \text{and} \\ E\left[\frac{\mathrm{d}^2 f(X_i;\theta)/\mathrm{d}\theta^2 \,|_{\theta=\theta_0}}{f(X_i;\theta_0)}\right] &= \int_{\mathbb{R}} \left[\frac{\mathrm{d}^2 f(x_i;\theta)/\mathrm{d}\theta^2 \,|_{\theta=\theta_0}}{f(x_i;\theta_0)} \times f(x_i;\theta_0)\right] \mathrm{d}x_i \\ &= \frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\left[\int_{\mathbb{R}} f(x_i;\theta)\mathrm{d}x_i\right]\Bigg|_{\theta=\theta_0} = 0. \end{aligned}$$

By the weak law of large number, we obtain

$$\frac{1}{n}\sum_{i=1}^{n} W_i^2 \xrightarrow{\mathrm{P}} I(\theta_0) \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n} \frac{\mathrm{d}^2 f(X_i;\theta)/\mathrm{d}\theta^2 \,|_{\theta=\theta_0}}{f(X_i;\theta_0)} \xrightarrow{\mathrm{P}} 0.$$

Thus,

$$-\frac{1}{n}\ell''(\theta_0;\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} W_i^2 - \frac{1}{n}\sum_{i=1}^{n} \frac{\mathrm{d}^2 f(X_i;\theta)/\mathrm{d}\theta^2 \,|_{\theta=\theta_0}}{f(X_i;\theta_0)} \xrightarrow{\mathrm{P}} I(\theta_0).$$

(c) Consider the first order Taylor expansion of $\ell'(\theta, \mathbf{x})$ around θ_0, we have

$$\ell'(\theta, \mathbf{x}) \approx \ell'(\theta_0, \mathbf{x}) + (\theta - \theta_0)\ell''(\theta_0, \mathbf{x}).$$

Note that $\ell'(\hat{\theta}, \mathbf{x}) = 0$ by definition. Therefore, by substituting $\theta = \hat{\theta}$, we obtain

$$\sqrt{n}(\hat{\theta} - \theta_0) \approx \frac{\ell'(\theta_0, \mathbf{x})/\sqrt{n}}{-\ell''(\theta_0, \mathbf{x})/n}.$$

Thus, by the result in (a) and (b), we can get

$$\frac{\ell'(\theta_0, \mathbf{x})/\sqrt{n}}{-\ell''(\theta_0, \mathbf{x})/n} \xrightarrow{\mathrm{L}} \frac{1}{I(\theta_0)} N(0, I(\theta_0)) = N\left(0, \frac{1}{I(\theta_0)}\right).$$

Now, by replacing θ_0 with θ, we can conclude that

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathrm{L}} N\left(0, \frac{1}{I(\theta)}\right) = N(0, v(\theta)). \qquad \|$$