Southern University of Science and Technology
Department of Mathematics

MA204: Mathematical Statistics

Tutorial 5: Examples/Solutions

A. Bayesian Estimator

(a) Given a random sample $X_1, \ldots, X_n$, determine the joint pdf of $X_1, \ldots, X_n$ and θ,

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

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

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

(c) 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 $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$, 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

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

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

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

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*}. $\|$

B. Asymptotic Efficiency

A sequence of estimators W_n is said to be **asymptotically efficient** for a parameter $\tau(\theta)$, if $\sqrt{n}[W_n - \tau(\theta)] \xrightarrow{\text{L}} N(0, \upsilon(\theta))$, where

$$\upsilon(\theta) = \frac{[\tau'(\theta)]^2}{I_n(\theta)} \quad \text{and} \quad I_n(\theta) = \operatorname{Var}_{\boldsymbol{X}}\left(\frac{\mathrm{d}\log L(\theta; \boldsymbol{X})}{\mathrm{d}\theta}\right).$$

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

Example T5.2 (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:

(A1) The parameter is identifiable, i.e., if $\theta \neq \theta^*$, then $f(x;\theta) \neq f(x;\theta^*)$.

(A2) The density $f(x;\theta)$ is differentiable with respect to θ inside its support.

(A3) The parameter space Θ contains an open set ω of which the true parameter value θ_0 is an interior point.

Let $\boldsymbol{X} = (X_1, \ldots, X_n)^{\mathrm{T}}$, show that

(a) $\dfrac{1}{\sqrt{n}}\ell'(\theta_0; \boldsymbol{X}) = \sqrt{n} \cdot \dfrac{1}{n}\displaystyle\sum_{i=1}^n W_i \xrightarrow{\text{L}} N(0, I(\theta_0))$, 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) $-\dfrac{1}{n}\ell''(\theta_0; \boldsymbol{X}) = \dfrac{1}{n}\displaystyle\sum_{i=1}^n W_i^2 - \dfrac{1}{n}\sum_{i=1}^n \dfrac{\mathrm{d}^2 f(X_i;\theta)/\,\mathrm{d}\theta^2\,|_{\theta=\theta_0}}{f(X_i;\theta_0)}$, and the expectations of W_i^2 and $\dfrac{\mathrm{d}^2 f(X_i;\theta)/\,\mathrm{d}\theta^2\,|_{\theta=\theta_0}}{f(X_i;\theta_0)}$ equal to $I(\theta_0)$ and 0, respectively, for $i = 1, \ldots, n$. Furthermore, we have $-\dfrac{1}{n}\ell''(\theta_0; \boldsymbol{X}) \xrightarrow{\text{P}} I(\theta_0)$.

(c) $\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{\text{L}} N(0, \upsilon(\theta))$, where $\upsilon(\theta)$ 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

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

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

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

(b) $\ell''(\theta;\boldsymbol{X}) = \dfrac{\mathrm{d}^2\ell(\theta;\boldsymbol{X})}{\mathrm{d}\theta^2}$

$$= \frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\left[\sum_{i=1}^{n}\log f(X_i;\theta)\right]$$

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

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

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

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

By the weak law of large number, we obtain that

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

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

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

$$
\ell'(\theta,\boldsymbol{X}) \approx \ell'(\theta_0,\boldsymbol{X}) + (\theta-\theta_0)\ell''(\theta_0,\boldsymbol{X}).
$$

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

$$
\sqrt{n}(\widehat{\theta}-\theta_0) \approx \frac{\ell'(\theta_0,\boldsymbol{X})/\sqrt{n}}{-\ell''(\theta_0,\boldsymbol{X})/n}.
$$

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

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

Now, replace θ_0 with θ, we can conclude that

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