STAT6201: Theoretical Statistics I

Satvik Saha

Homework 3

1. Let $X_1, \dots, X_n \overset{iid}{\sim} f_\theta$, with

$$f_\theta(x) = \frac{1}{2} e^{-|x-\theta|}$$

for $\theta \in \mathbb{R}$. Finding $\hat{\theta}_{\mathrm{MLE}}$ amounts to maximizing the log-likelihood, given by

$$\log f(\theta \mid \boldsymbol{X}) = -\sum_{i=1}^{n} |X_i - \theta|.$$

Thus, we set $\hat{\theta}_{\mathrm{MLE}}$ to be the minimizer of $\sum_{i=1}^{n} |X_i - \theta|$, whence $\hat{\theta}_{\mathrm{MLE}}(\boldsymbol{X}) = \mathrm{median}(\boldsymbol{X})$.

We prove that the median does indeed minimize $g(\theta) = \|\boldsymbol{X} - \theta\mathbf{1}\|_1$. It is clear that g increases rightwards of $\max(\boldsymbol{x})$ and leftwards of $\min(\boldsymbol{x})$, so g is indeed minimized somewhere in between. Observe that we can write

$$\begin{aligned} g(\theta) &= \sum_{i\colon x_i < \theta} (\theta - x_i) + \sum_{i\colon x_i \geq \theta} (x_i - \theta) \\ &= \sum_{i\colon x_i \geq \theta} x_i - \sum_{i\colon x_i < \theta} x_i + \left(\sum_{i\colon x_i < \theta} 1 - \sum_{i\colon x_i \geq \theta} 1 \right) \theta. \end{aligned}$$

Thus, g is linear on each piece $(x_{(i)}, x_{(i+1)})$, with slope

$$d(\theta) = \sum_{i\colon x_i < \theta} 1 - \sum_{i\colon x_i \geq \theta} 1.$$

When $n = 2k+1$ is odd, note that g is decreasing on $(x_{(k)}, x_{(k+1)})$ and increasing on $(x_{(k+1)}, x_{(k+2)})$, which means that g must attain its minimum at $x_{(k+1)}$.

Similarly, when $n = 2k$ is even, note that g is constant on $(x_{(k)}, x_{(k+1)})$, decreasing in the intervals before that, and increasing in the intervals after that, which means that g must attain its minimum on $(x_{(k)}, x_{(k+1)})$.

This shows that g is minimized at the median of $x_1, \dots, x_n$, as desired.

Now, we must show that $\hat{\theta}_{\mathrm{MLE}}$ is consistent, i.e. that $\hat{\theta}_{\mathrm{MLE}} \xrightarrow{p} \theta$. Note that our median $\hat{\theta}_{\mathrm{MLE}}$ may be described as $\hat{F}_n^{-1}(\frac{1}{2})$. The Glivenko-Cantelli Theorem guarantees that $\hat{F}_n \xrightarrow{a.s.} F$ uniformly. Then,

$$0 \leq |\hat{F}_n(\hat{\theta}_{MLE}) - F(\hat{\theta}_{\mathrm{MLE}})| \leq \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \xrightarrow{a.s.} 0.$$

It follows that $F(\hat{\theta}_{\mathrm{MLE}}) \xrightarrow{a.s.} \frac{1}{2}$, whence $\hat{\theta}_{\mathrm{MLE}} \xrightarrow{a.s.} F^{-1}(\frac{1}{2}) = \theta$ via the Continuous Mapping Theorem.

2. (a) Let $X \mid \theta \sim f_\theta$, and $\theta \sim \pi$ be proper. Further let $\delta(X)$ be unbiased for $g(\theta)$, and consider the squared error loss. Set $\eta(X) = \mathbb{E}_{\theta|X}[g(\theta) \mid X]$. Then,

$$\begin{aligned} R_B(\delta, g) &= \mathbb{E}\left[(\delta(X) - g(\theta))^2\right] \\ &= \mathbb{E}_X\left[\mathbb{E}_{\theta|X}\left[(\delta(X) - g(\theta))^2\right]\right] \\ &= \mathbb{E}_X\left[(\delta(X) - \eta(X))^2\right] + \mathbb{E}_X\left[\mathbb{E}_{\theta|X}\left[(\eta(X) - g(\theta))^2\right]\right] \\ &= \mathbb{E}_X\left[(\delta(X) - \eta(X))^2\right] + R_B(\eta, g) \\ &\geq R_B(\eta, g). \end{aligned}$$

Thus, $\delta(X)$ can be Bayes for $g(\theta)$ only if it is equal to $\eta(X)$, almost everywhere with respect to the marginal $f(x)$. But then, unbiasedness of $\delta(X)$ gives

$$g(\theta) = \mathbb{E}_{X|\theta}[\delta(X)] = \mathbb{E}_{X|\theta}[\mathbb{E}_{\theta'|X}[g(\theta')]],$$

whence

$$\begin{aligned} R_B(\delta, g) &= \mathbb{E}\left[(\delta(X) - g(\theta))^2\right] \\ &= \mathbb{E}[\delta(X)^2] - 2\mathbb{E}[\delta(X)g(\theta)] + \mathbb{E}[g(\theta)^2] \\ &= \mathbb{E}[\delta(X)^2] - \mathbb{E}_X[\mathbb{E}_{\theta|X}[\delta(X)g(\theta)]] - \mathbb{E}_\theta[\mathbb{E}_{X|\theta}[\delta(X)g(\theta)]] + \mathbb{E}[g(\theta)^2] \\ &= \mathbb{E}[\delta(X)^2] - \mathbb{E}_X[\delta(X) \cdot \delta(X)] - \mathbb{E}_\theta[g(\theta) \cdot g(\theta)] + \mathbb{E}[g(\theta)^2] \\ &= 0. \end{aligned}$$

(b) Now, suppose that

$$f_\theta(x) = \frac{1}{\theta} e^{-x/\theta}\, \mathbf{1}_{(0,\infty)}(x),$$

for $\theta \in (0, \infty)$, and that $\pi(\theta) = \theta^{-2}$. Then, we clearly have $\mathbb{E}_{X|\theta}[X] = \theta$ (by definition of f_θ). Furthermore,

$$\pi(\theta \mid x) \propto f_\theta(x)\pi(\theta) = \frac{1}{\theta^3} e^{-x/\theta}\, \mathbf{1}_{(0,\infty)}(x).$$

Now,

$$\int_0^\infty \theta^{-k} e^{-x/\theta}\, d\theta = \Gamma(k-1)x^{k-1},$$

whence the Bayes estimator under the squared error loss is

$$\delta_\pi(X) = \mathbb{E}_{\theta|X}[\theta \mid X] = \frac{\int_0^\infty \theta\pi(\theta \mid X)\, d\theta}{\int_0^\infty \pi(\theta \mid X)\, d\theta} = \frac{\Gamma(1)X}{\Gamma(2)} = X.$$

Remark: We have used the inverse gamma formula

$$\int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{t^{\alpha+1}} e^{-\beta/t}\, dt = 1.$$

3. Let $X_i \mid (\mu, \sigma^2) \overset{iid}{\sim} N(\mu, \sigma^2)$ and $\pi(\mu, \sigma^2) \propto \sigma^{-2}$, with $\mu \in \mathbb{R}$.

(a) We can compute the posterior distribution

$$\begin{aligned} \pi(\mu, \sigma^2 \mid x) &\propto f(x \mid \mu, \sigma^2)\, \pi(\mu, \sigma^2) \\ &\propto (2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu)^2} \cdot (\sigma^2)^{-1} \\ &\propto (\sigma^2)^{-n/2-1} e^{-n(\bar{x}-\mu)^2/2\sigma^2} e^{-(n-1)S^2/2\sigma^2}. \end{aligned}$$

Integrating out μ gives

$$\pi(\sigma^2 \mid x) \propto (\sigma^2)^{-n/2-1}(2\pi\sigma^2/n)^{1/2}e^{-(n-1)S^2/2\sigma^2} \propto (\sigma^2)^{-(n+1)/2}e^{-(n-1)S^2/2\sigma^2}.$$

Thus, $\sigma^2 \mid X \sim \text{InvGamma}((n-1)/2, (n-1)S^2/2)$, whence $\sigma^{-2} \mid X \sim \text{Gamma}((n-1)/2, (n-1)S^2/2)$. It follows that $(n-1)S^2\sigma^{-2} \mid X \sim \text{Gamma}((n-1)/2, 1/2) \sim \chi^2_{n-1}$.

(b) This time, integrating out σ^2 from the posterior (using the inverse gamma formula) gives

$$\pi(\mu \mid x) \propto \left[n(\mu-\bar{x})^2 + (n-1)S^2\right]^{-n/2} \propto \left[1 + \frac{(\sqrt{n}(\mu-\bar{x})/S)^2}{n-1}\right]^{-\frac{((n-1)+1)}{2}}.$$

This immediately gives $\sqrt{n}(\mu-\bar{x})/S \mid X \sim t_{n-1}$.

4. Let $X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, 1)$, and let $\mathscr{D} = \{\delta_{a,b} \colon \delta_{a,b}(X) = a\bar{X} + b; a, b \in \mathbb{R}\}$ be a class of estimators for $\theta \in \mathbb{R}$. Further suppose that we are working with the squared error loss.

   (a) Suppose that $\delta_b(X) = \bar{X} + b$ is a Bayes estimator for θ with respect to some proper prior π. Then, note that

$$\begin{aligned} R_B(\delta_b, \mathrm{id}_{\mathbb{R}}) &= \mathbb{E}[(\bar{X} - \theta + b)^2] \\ &= \mathbb{E}[(\bar{X} - \theta)^2] + b^2 + 2b\cancel{\mathbb{E}[\bar{X} - \theta]} \\ &= \mathbb{E}_\theta[\mathrm{var}_{X|\theta}(\bar{X})] + b^2 \\ &= \frac{1}{n} + b^2. \end{aligned}$$

   Minimality forces $b = 0$. However, our estimator $\delta_0(X) = \bar{X}$ is unbiased for θ under a proper prior, yet it has Bayes risk $1/n > 0$, contradicting the result from Problem 2.

   (b) Consider $\theta \sim N(\mu, \tau^2)$, so

$$\begin{aligned} \pi(\theta \mid x) &\propto f(x \mid \theta)\,\pi(\theta) \\ &\propto e^{-\sum_{i=1}^n (x_i-\theta)^2/2} \cdot (\tau^2)^{-1/2}e^{-(\theta-\mu)^2/2\tau^2} \\ &\propto e^{-n(\theta-\bar{x})^2/2}\,e^{-(\theta-\mu)^2/2\tau^2} \\ &\sim N\left(\alpha\bar{x} + (1-\alpha)\mu, \alpha\right), \end{aligned}$$

   where $\alpha = n\tau^2/(1+n\tau^2)$. Thus, the Bayes estimator for θ will be the posterior mean $\alpha\bar{X} + (1-\alpha)\mu$, which is of the form $\delta_{a,b}$ with $a = \alpha$, $b = (1-\alpha)\mu$.
   With this, given $a \in (0,1)$ and $b \in \mathbb{R}$, the prior $\theta \sim N(\mu, \tau^2)$ with

$$\mu = \frac{b}{1-a}, \qquad \tau^2 = \frac{a}{n(1-a)}$$

   yields the Bayes estimator $\delta_{a,b}$, as desired.

5. Write $Y \mid \theta \sim N(X\theta, \sigma^2 I_n)$, and $\theta \sim N(0, \Sigma)$, where

$$X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} \tau_1^2 & 0 \\ 0 & \tau_2^2 \end{bmatrix}.$$

Then,

$$
\begin{aligned}
\pi(\theta \mid y) &\propto f(y \mid \theta)\, \pi(\theta) \\
&\propto (2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}\|y - X\theta\|^2} \cdot (\tau_1^2\tau_2^2)^{-1/2} e^{-\beta_0^2/2\tau_1^2} e^{-\beta_1^2/2\tau_2^2} \\
&\propto \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left[y_i^2 + \beta_0^2 + \beta_1^2 x_i^2 - 2\beta_0 y_i - 2\beta_1 x_i y_i - \cancel{2\beta_0\beta_1 x_i} \right] - \frac{\beta_0^2}{2\tau_1^2} - \frac{\beta_1^2}{2\tau_2^2} \right) \\
&\propto \exp\left(-\left(\frac{n}{2\sigma^2} + \frac{1}{2\tau_1^2} \right) \beta_0^2 + \frac{\sum_i y_i}{\sigma^2} \beta_0 \right) \cdot \exp\left(-\left(\frac{\sum_i x_i^2}{2\sigma^2} + \frac{1}{2\tau_2^2} \right) \beta_1^2 + \frac{\sum_i x_i y_i}{\sigma^2} \beta_1 \right),
\end{aligned}
$$

which normally distributed. The Bayes estimator for θ under the squared error loss can be recovered as the posterior mean, whence

$$
\hat{\beta}_0 = \left(n + \frac{\sigma^2}{\tau_1^2} \right)^{-1} \sum_{i=1}^{n} y_i, \qquad \hat{\beta}_1 = \left(\sum_{i=1}^{n} x_i^2 + \frac{\sigma^2}{\tau_2^2} \right)^{-1} \sum_{i=1}^{n} x_i y_i.
$$

6. Let $X \mid \theta \sim f_\theta$ and $\theta \sim \pi$. We assume that

   (i) There exists an estimator δ_0 with finite Bayes risk. It follows that a Bayes estimator for $g(\theta)$ exists with finite Bayes risk.

   (ii) There exists an estimator δ_π satisfying (uniquely) for each x,

   $$\delta_\pi(x) = \arg\min_y \mathbb{E}_{\theta|X}\left[\ell(y, g(\theta)) \mid X = x\right] \quad \text{a.e. } f,$$

   where f is the marginal density of X.

   (iii) For every θ, a.e. f implies a.e. f_θ. Then, uniqueness a.e. f will force uniqueness a.e. f_θ for all θ.

   With this, suppose that δ is a Bayes estimator of θ which achieves minimum Bayes risk $R_B^* = R_B(\delta, g) \le R_B(\delta_0, g) < \infty$. Then, by construction of δ_π, we have

   $$
   \begin{aligned}
   R_B(\delta, g) &= \mathbb{E}_X[\mathbb{E}_{\theta|X}[\ell(\delta(X), g(\theta)) \mid X]] \\
   &\ge \mathbb{E}_X[\mathbb{E}_{\theta|X}[\ell(\delta_\pi(X), g(\theta)) \mid X]] \\
   &= R_B(\delta_\pi, g).
   \end{aligned}
   $$

   Minimality forces $R_B(\delta_\pi, g) = R_B(\delta, g)$, hence the equality

   $$\mathbb{E}_{\theta|X}[\ell(\delta(X), g(\theta)) \mid X = x] = \mathbb{E}_{\theta|X}[\ell(\delta_\pi(X), g(\theta)) \mid X = x] \quad \text{a.e. } f.$$

   But δ_π achieves this (minimum) value a.e. f *uniquely*, forcing $\delta = \delta_\pi$ a.e. f. Our assumption forces $\delta = \delta_\pi$ a.e. f_θ for all θ, whence δ_π is the unique Bayes estimator.

   Remark: The fact that a Bayes estimator exists in the first place follows by considering δ_π as a candidate and using the above string of inequalities to conclude that it must be Bayes.