Mathematical Statistics Solutions: HW 5

October 19, 2022

Exercise 1

Part (a)

For any $h(\cdot)$ and $A(\cdot)$ that satisfy

$$p_\theta(x) = h(x)e^{\theta T(x) - A(\theta)}, \tag{1}$$

we notice that, for any constant $c \in \mathbb{R}$, $e^c h(\cdot)$ and $A(\cdot) + c$ also satisfy Equation (1). Therefore, h and A are not unique.

Part (b)

The density of normal distribution $\mathcal{N}(\theta, 1)$ is:

$$p_\theta(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(x-\theta)^2}{2}\right\} = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} \exp\left\{\theta x - \frac{\theta^2}{2}\right\}.$$

Therefore, it forms an exponential family with $h(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}$, $T(x) = x$, $A(\theta) = \frac{\theta^2}{2}$.

Part (c)

We rewrite the density as

$$p_\theta(x) = (\theta - 1)x^{-\theta} = \exp\left\{-\theta \log x + \log(\theta - 1)\right\}, \ \theta > 1.$$

Therefore, it forms an exponential family with $h(x) = 1$, $T(x) = -\log x$ and $A(\theta) = -\log(\theta - 1)$.

Part (d)

For any exponential family, since $p_\theta(x)$ is a probability density function, we have

$$\int p_\theta(x) dx = \int h(x) e^{\theta T(x) - A(\theta)} dx = 1;$$

solving for $A(\theta)$ yields

$$A(\theta) = \log \int h(x) e^{\theta T(x)} dx.$$

Straightforward calculation gives

$$A'(\theta) = \frac{\int h(x)T(x)e^{\theta T(x)}dx}{\int h(x)e^{\theta T(x)}dx} = \int h(x)T(x)e^{\theta T(x) - A(\theta)}dx = \mathbb{E}_\theta T(X),$$

where we use the fact that $\int h(x)e^{\theta T(x)}dx = e^{A(\theta)}$ in the second equality.

Part (e)

In an exponential family, the log likelihood is

$$l(\theta|X_1,\dots,X_n) = \sum_{i=1}^n \log h(X_i) + \theta \sum_{i=1}^n T(x_i) - nA(\theta).$$

Since $A(\theta)$ is convex, θ is a maximum likelihood estimator if and only if it satisfies the first order condition, i.e.,

$$\frac{d}{d\theta} l(\theta|X_1,\dots,X_n) = 0 \iff \sum_{i=1}^n T(x_i) = nA'(\theta) \iff \frac{1}{n}\sum_{i=1}^n T(x_i) = \mathbb{E}_\theta T(X),$$

where we use the result from Part (e) in the second equivalence.

Exercise 2

Part (a)

Let p_θ denote the probability density of the vector $(X_1,\dots,X_n)$ under $P_\theta = \mathsf{Unif}([0,\theta])$. Then, the likelihood is

$$p_\theta(X_1,\dots,X_n) = \mathbb{1}\{0 \le X_i \le \theta \text{ for all } i = 1,\dots,n\}\theta^{-n}.$$

This is maximized by choosing θ as small as possible so long as the first indicator function is not zero. This is achieved by $\hat\theta = \max_{i=1}^n X_i$, which is therefore the MLE.

Part (b)

We have $\mathbb{E}X_1 = \theta/2$, so the method-of-moments estimator $\hat\theta$ solves

$$\frac{\hat\theta}{2} = \frac{1}{n}\sum_{i=1}^n X_i = \bar X.$$

Thus, the estimator is $\hat\theta = 2\bar X$.

Part (c)

Fix some small $\epsilon > 0$. With positive probability, we have $0 \le X_i \le \epsilon$ for all $i = 1,\dots,n-1$, while $X_n \ge (1-\epsilon)\theta$. On such a sample, when $n \ge 3$, the method-of-moments estimator satisfies

$$\hat\theta \le \frac{2}{n}((n-1)\epsilon + \theta) = 2\left(1 - \frac{1}{n-1}\right)\epsilon + \frac{2}{n}\theta \le 2\epsilon + \frac{2}{3}\theta.$$

In particular, for ϵ sufficiently small, we have $\hat\theta < (1-\epsilon)\theta$. On the other hand, since $X_n \ge (1-\epsilon)\theta$, such a sample is never generated under $P_{\hat\theta}$.

Part (d)

For the MLE, we have

$$\begin{aligned}\mathbb{P}_\theta\left[\left|\theta - \max_{i=1}^n X_i\right| \ge \epsilon\right] &= \mathbb{P}_\theta\left[X_i \le \theta - \epsilon \text{ for all } i = 1,\dots,n\right] \\ &= \left(\frac{\theta-\epsilon}{\theta}\right)^n && \text{(by independence)} \\ &= \left(1 - \frac{\epsilon}{\theta}\right)^n.\end{aligned}$$

Since the quantity being exponentiated is in $(0,1)$, this will tend to zero as $n \to \infty$ for any $\epsilon > 0$.

For the method-of-moments estimator, since $\mathbb{E}X_i = \theta/2$, we have $\bar X \xrightarrow{p} \theta/2$ by the weak law of large numbers, and so $\hat\theta = 2\bar X \xrightarrow{p} \theta$.

Exercise 3

Part (a)

$$
\begin{aligned}
s(\theta|\omega) &:= \sum_{i=1}^{n} s(\theta|x_i) \\
&= \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log p_\theta(x_i) \\
&= \frac{\partial}{\partial\theta} \sum_{i=1}^{n} \log p_\theta(x_i) \quad \text{(linearity)} \\
&= \frac{\partial}{\partial\theta} \log \left(\prod_{i=1}^{n} p_\theta(x_i) \right) \\
&= \frac{\frac{\partial}{\partial\theta} \prod_{i=1}^{n} p_\theta(x_i)}{\prod_{i=1}^{n} p_\theta(x_i)} \quad \text{(chain rule)}\,.
\end{aligned}
$$

Part (b)

We make use of the hint and part **(a)**,

$$
\begin{aligned}
\mathbb{E}_\theta\left[f(\omega)s(\theta|\omega)\right] &= \int f(x_1,\ldots,x_n) s(\theta|x_1,\ldots,x_n) p_\theta(x_1,\ldots,x_n) \\
&= \int f(x_1,\ldots,x_n) s(\theta|x_1,\ldots,x_n) \prod_{i=1}^{n} p_\theta(x_i) \\
&= \int f(x_1,\ldots,x_n) \frac{\frac{\partial}{\partial\theta} \prod_{i=1}^{n} p_\theta(x_i)}{\prod_{i=1}^{n} p_\theta(x_i)} \prod_{i=1}^{n} p_\theta(x_i) \quad \text{by } \textbf{(a)} \\
&= \int f(x_1,\ldots,x_n) \frac{\partial}{\partial\theta} \prod_{i=1}^{n} p_\theta(x_i) \\
&= \frac{\partial}{\partial\theta} \int f(x_1,\ldots,x_n) \prod_{i=1}^{n} p_\theta(x_i) \quad \text{(interchange integration and diff.)} \\
&= \frac{\partial}{\partial\theta} \mathbb{E}_\theta[f(\omega)]\,.
\end{aligned}
$$

Part (c)

For the first part, we take $f(\omega) \equiv 1$, which means $\mathbb{E}_\theta[f(\omega)] = 1$. Using part **(b)**,

$$
\mathbb{E}_\theta[s(\theta|\omega)] = \mathbb{E}_\theta[1 \cdot s(\theta|\omega)] = \frac{\partial}{\partial\theta} 1 = 0\,.
$$

For the second part, we take $f(\omega) = \hat{\theta}(\omega)$ which is unbiased, meaning $\mathbb{E}_\theta[f(\omega)] = \mathbb{E}_\theta[\hat{\theta}(\omega)] = \theta$. Again, using **(b)**,

$$
\mathbb{E}_\theta[\hat{\theta}(\omega)s(\theta|\omega)] = \frac{\partial}{\partial\theta} \mathbb{E}_\theta[\hat{\theta}(\omega)] = \frac{\partial}{\partial\theta} \theta = 1\,.
$$

Part (d)

The expectation of a non-negative quantity is non-negative, so we have

$$
0 \le \mathbb{E}_\theta[(\lambda(\hat{\theta}(\omega) - \theta) - s(\theta|\omega))^2] = \mathbb{E}_\theta[(\lambda(\hat{\theta}(\omega) - \mathbb{E}_\theta[\hat{\theta}(\omega)]) - s(\theta|\omega))^2]\,.
$$

Expanding the square, noting that $\mathbb{E}_\theta[s(\theta|\omega)] = 0$, $\mathbb{E}_\theta[\hat{\theta}(\omega)s(\theta|\omega)] = 1$ and that $\hat{\theta}$ is unbiased i.e. $\mathbb{E}_\theta[\hat{\theta}(\omega)] = \theta$

$$\begin{aligned}
\mathbb{E}_\theta[(\lambda(\hat{\theta}(\omega) - \mathbb{E}_\theta[\hat{\theta}(\omega)]) - s(\theta|\omega))^2] &= \mathbb{E}_\theta[(\lambda(\hat{\theta}(\omega) - \mathbb{E}_\theta[\hat{\theta}(\omega)])^2] + \mathbb{E}_\theta[s(\theta|\omega)^2] - 2\lambda\mathbb{E}_\theta[\hat{\theta}(\omega) - \mathbb{E}_\theta[\hat{\theta}(\omega)]s(\theta|\omega)] \\
&= \lambda^2\mathbb{E}_\theta[(\hat{\theta}(\omega) - \mathbb{E}_\theta[\hat{\theta}(\omega)])^2] + \mathrm{Var}_\theta(s(\theta|\omega)) - 2\lambda\mathbb{E}_\theta[\hat{\theta}(\omega)s(\theta|\omega)] - 2\lambda\theta\mathbb{E}_\theta[s(\theta|\omega)] \\
&= \lambda^2\mathrm{Var}_\theta(\hat{\theta}(\omega)) + \mathrm{Var}_\theta(s(\theta|\omega)) - 2\lambda \cdot 1 - 2\lambda \cdot 0 \\
&= \lambda^2\mathrm{Var}_\theta(\hat{\theta}(\omega)) + \mathrm{Var}_\theta(s(\theta|\omega)) - 2\lambda\,.
\end{aligned}$$

Taking $\lambda = \mathrm{Var}_\theta(s(\theta|\omega))$ and rearranging,

$$0 \le \mathrm{Var}_\theta(s(\theta|\omega))^2\mathrm{Var}_\theta(\hat{\theta}(\omega)) - \mathrm{Var}_\theta(s(\theta|\omega)) \implies \frac{1}{\mathrm{Var}_\theta(s(\theta|\omega))} \le \mathrm{Var}_\theta(\hat{\theta}(\omega))\,.$$

By independence, we have that

$$\mathrm{Var}_\theta(s(\theta|\omega)) = \mathrm{Var}_\theta\left(\sum_{i=1}^n s(\theta|X_i)\right) = n\mathrm{Var}_\theta(s(\theta|X_1))\,,$$

which completes the claim.

Part (e) (Optional)

Recall that the variance quantity is written

$$\sigma^2(\rho, \theta) = \frac{\mathbb{E}\left(\frac{\partial}{\partial\theta}\rho(X,\theta)\right)^2}{\left(\mathbb{E}\frac{\partial^2}{\partial\theta^2}\rho(X,\theta)\right)^2} \tag{2}$$

Let $\rho(X, \theta) = -\ell(X, \theta) = -\log(p_\theta(X))$. The first derivative is $-s(\theta|X)$, and we compute the second via chain rule

$$\frac{\partial^2}{\partial\theta^2}\rho(X,\theta) = \frac{\partial}{\partial\theta}\left(\frac{-\frac{\partial}{\partial\theta}p_\theta(X)}{p_\theta(X)}\right) = -\frac{\frac{\partial^2}{\partial\theta^2}p_\theta(X)p_\theta(X) - (\frac{\partial}{\partial\theta}p_\theta(X))^2}{p_\theta(X)^2} = -\frac{\frac{\partial^2}{\partial\theta^2}p_\theta(X)}{p_\theta(X)} + (s(\theta|X))^2$$

From this we see that the numerator is

$$\mathbb{E}\left(\frac{\partial}{\partial\theta}\rho(X,\theta)\right)^2 = \mathbb{E}\left(-s(\theta|X)\right)^2 = \mathrm{Var}_\theta(s(\theta|X))\,,$$

and the square-root of the denominator is

$$\begin{aligned}
\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\rho(X,\theta)\right] &= \mathbb{E}\left[-\frac{\frac{\partial^2}{\partial\theta^2}p_\theta(X)}{p_\theta(X)} + (s(\theta|X))^2\right] \\
&= \left[\int\left(-\frac{\frac{\partial^2}{\partial\theta^2}p_\theta(X)}{p_\theta(X)}\right)p_\theta(X)\right] + \mathrm{Var}_\theta(s(\theta|X)) \\
&= \left[\int\frac{\partial^2}{\partial\theta^2}p_\theta(X)\right] + \mathrm{Var}_\theta(s(\theta|X)) \\
&= \left[\frac{\partial^2}{\partial\theta^2}\int p_\theta(X)\right] + \mathrm{Var}_\theta(s(\theta|X)) \\
&= \left[\frac{\partial^2}{\partial\theta^2}1\right] + \mathrm{Var}_\theta(s(\theta|X)) = 0 + \mathrm{Var}_\theta(s(\theta|X)).
\end{aligned}$$

And so the denominator is

$$\left(\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\rho(X,\theta)\right]\right)^2 = (\mathrm{Var}_\theta(s(\theta|X))^2.$$

Thus, the asymptotic variance is

$$\sigma^2(\rho, \theta) = \frac{\mathbb{E}\left(\frac{\partial}{\partial\theta}\rho(X,\theta)\right)^2}{\left(\mathbb{E}\frac{\partial^2}{\partial\theta^2}\rho(X,\theta)\right)^2} = \frac{\mathrm{Var}_\theta(s(\theta|X))}{\mathrm{Var}_\theta(s(\theta|X))^2} = \frac{1}{I(\theta)}\,, \tag{3}$$

as desired.

Exercise 4

Part (a)

Taking the derivative of $\ell(\beta, \sigma^2)$, the log-likelihood indicated above, with respect to β, we find

$$\frac{\partial \ell}{\partial \beta} = \frac{1}{\sigma^2}\sum_{i=1}^n x_i Y_i - \frac{\beta}{\sigma^2}\sum_{i=1}^n x_i^2,$$

whereby, setting this to zero, the MLE is

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}.$$

Part (b)

Since $\mathbb{E}Y_i = \beta x_i$, the expectation is

$$\mathbb{E}\hat{\beta} = \frac{1}{\sum_{i=1}^n x_i^2}\sum_{i=1}^n x_i \mathbb{E}Y_i = \frac{1}{\sum_{i=1}^n x_i^2}\sum_{i=1}^n \beta x_i^2 = \beta.$$

Since $\mathsf{Var}Y_i = \mathsf{Var}\epsilon_i = \sigma^2$, the variance is

$$\mathsf{Var}\hat{\beta} = \frac{1}{(\sum_{i=1}^n x_i^2)^2}\sum_{i=1}^n x_i^2 \mathsf{Var}Y_i = \frac{\sigma^2}{\sum_{i=1}^n x_i^2}.$$

Part (c)

By Chebyshev's inequality, we have

$$\mathbb{P}[|\hat{\beta} - \beta| \geq t] \leq \frac{\mathsf{Var}\hat{\beta}}{t}.$$

If $\sum_{i=1}^n x_i^2 \to \infty$, then by the calculation from the previous part $\mathsf{Var}\hat{\beta} \to 0$, whereby for any fixed t the above probability tends to zero. Thus, $\hat{\beta} \xrightarrow{p} \beta$, so the MLE is consistent.

Part (d)

Suppose $x_1 = 1$ and $x_i = 0$ for all $i \geq 2$. Then, $\sum_{i=1}^n x_i^2 = 1$ for all $n \geq 1$, and $\hat{\beta} = Y_1$. But, $Y_1 \sim \mathcal{N}(\beta, \sigma^2)$, so $\mathbb{P}[|\hat{\beta} - \beta| \geq t] = \mathbb{P}[|g| \geq t]$ for $g \sim \mathcal{N}(0, \sigma^2)$, which is a positive constant independent of n. In particular, $\hat{\beta}$ does not converge in probability to β, so the MLE is not consistent.