Math 541a Homework 5

Qilin Ye

September 22, 2022

Problem 1

Let $X, Y, Z : \Omega \to \mathbb{R}$ be discrete or continuous random variables. Let A be the range of Y. Define $g : A \to \mathbb{R}$ by $g(y) := \mathbb{E}(X \mid Y = y)$ for any $y \in A$. We then define the **conditional expectation** of X given Y, denoted $\mathbb{E}(X \mid Y)$, to be the random variable $g(Y)$.

(i) Let X, Y be random variables such that (X, Y) is uniform distributed on the triangle given by $\{(x, y) \in \mathbb{R}^2 : x \geqslant 0, y \geqslant 0, x + y \leqslant 1\}$. Show that $\mathbb{E}(X \mid Y) = (1 - Y)/2$.

(ii) Prove the following version of the Total Expectation Theorem:

$$\mathbb{E}(\mathbb{E}(X \mid Y)) = \mathbb{E}(X).$$

(iii) Show the following

$$\mathbb{E}(X \mid X) = X \qquad \text{and} \qquad \mathbb{E}(X + Y \mid Z) = \mathbb{E}(X \mid Z) + \mathbb{E}(Y \mid Z).$$

(iv) If Z is independent of X and Y, show that

$$\mathbb{E}(X \mid Y, Z) = \mathbb{E}(X \mid Y).$$

(v) If Z is independent of X and Y, show that

$$\mathbb{E}(X \mid Y, z) = E(X \mid Y).$$

Proof. (i) For $y \in [0, 1]$, note that $X \mid Y = y$ is uniformly distributed on $[0, 1 - y]$, so $\mathbb{E}(X \mid Y = y) = (1 - y)/2$. Therefore by definition $\mathbb{E}(X \mid Y) = (1 - Y)/2$.

(ii) For the continuous case:

$$\begin{aligned}\mathbb{E}(\mathbb{E}(X\mid Y)) &= \int_{-\infty}^{\infty} f_Y(y)\mathbb{E}(X\mid Y=y)\,\mathrm{d}y\\ &= \int_{-\infty}^{\infty} f_Y(y)\left[\int_{-\infty}^{\infty} x f_{X|Y}(x\mid y)\,\mathrm{d}x\right]\mathrm{d}y\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X|Y}(x,y) f_Y(y)\,\mathrm{d}x\,\mathrm{d}y\\ &= \int_{-\infty}^{\infty} x\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}y\,\mathrm{d}x\\ &= \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x = \mathbb{E}X.\end{aligned}$$

For the discrete case:

$$\begin{aligned}\mathbb{E}(\mathbb{E}(X\mid Y)) = \mathbb{E}(\sum_x x\mathbb{P}(X=x\mid Y=y)) &= \sum_y(\sum_x x\mathbb{P}(X=x\mid Y=y))\mathbb{P}(Y=y)\\ &= \sum_x\sum_y x\mathbb{P}(X=x\mid Y=y)\mathbb{P}(Y=y)\\ &= \sum_x x(\sum_y \mathbb{P}(X=x,Y=y)) = \sum_x x\mathbb{P}(X=x) = \mathbb{E}X.\end{aligned}$$

(iii) The first claim is trivial, as $\mathbb{E}(X\mid X=x)=x$. The continuous case for the second equation:

$$\begin{aligned}\mathbb{E}(X+Y\mid Z=z) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y) f_{X+Y|Z}(x+y\mid z)\,\mathrm{d}x\,\mathrm{d}y\\ &= \iint_{\mathbb{R}^2} x f_{X+Y|Z}(x+y\mid z)\,\mathrm{d}x\mathrm{d}y + \iint_{\mathbb{R}^2} y f_{X+Y|Z}(x+y\mid z)\,\mathrm{d}x\mathrm{d}y\\ &= \int_{-\infty}^{\infty} x\int_{-\infty}^{\infty} f_{X+Y|Z}(x+y\mid z)\mathrm{d}y\,\mathrm{d}x + \int_{-\infty}^{\infty} y\int_{-\infty}^{\infty} f_{X+Y|Z}(x+y\mid z)\,\mathrm{d}x\mathrm{d}y\\ &= \int_{-\infty}^{\infty} x f_{X|Z}(x\mid z)\,\mathrm{d}x + \int_{-\infty}^{\infty} y f_{Y|Z}(y\mid z)\,\mathrm{d}y = \mathbb{E}(X\mid Z=z)+\mathbb{E}(Y\mid Z=z).\end{aligned}$$

The discrete case for the second equation:

$$\begin{aligned}\mathbb{E}(X+Y\mid Z=z) &= \sum_x\sum_y(x+y)\mathbb{P}(X=x,Y=y\mid Z=z)\\ &= \sum_x x\sum_y \mathbb{P}(X=x,Y=y\mid Z=z) + \sum_y y\sum_x \mathbb{P}(X=x,Y=y\mid Z=z)\\ &= \sum_x \mathbb{P}(X=x\mid Z=z) + \sum_y y\mathbb{P}(Y=y\mid Z=z) = \mathbb{E}(X\mid Z=z)+\mathbb{E}(Y\mid Z=z).\end{aligned}$$

(iv) If Z is independent of X and Y then (assuming they are continuous)

$$f_{X|(Y,Z)}(x\mid(y,z)) = \frac{f_{X,Y,Z}(x,y,z)}{f_{Y,Z}(y,z)} = \frac{f_{X,Y}(x,y)}{f_Y(y)},$$

so

$$\mathbb{E}(X\mid(Y,Z)=(y,z)) = \int_{-\infty}^{\infty} x f_{X|(Y,Z)}(x\mid(y,z)\,\mathrm{d}x = \int_{-\infty}^{\infty} x f_{X|Y}(x\mid y)\,\mathrm{d}x = \mathbb{E}(X\mid Y=y).$$

(v) If Y,Z are independent, then (assuming all variables are continuous),

$$\begin{aligned}\mathbb{E}(X\mid(Y,Z)=(y,z)) &= \int_{-\infty}^{\infty} x f_{X|Y,Z}(x\mid y,z)\,\mathrm{d}x\\ &= \int_{-\infty}^{\infty} x\cdot\frac{f_{X,Y,Z}(x,y,z)}{f_{Y,Z}(y,z)}\,\mathrm{d}x\\ &= \int_{-\infty}^{\infty} x\cdot\frac{f_{X,Y}(x,y)}{f_Y(y)}\,\mathrm{d}x = \mathbb{E}(X\mid Y=y).\end{aligned}$$ □

Problem 2

Prove Jensen's inequality for the conditional expectation. Let $X, Y : \Omega \to \mathbb{R}$ be random variables that are either both discrete or both continuous. Let $\varphi : \mathbb{R} \to \mathbb{R}$ be convex. Show that

$$\varphi(\mathbb{E}(X \mid Y)) \leqslant \mathbb{E}(\varphi(X) \mid Y)$$

and that the equality can be attained if and only if X is constant on any set where Y is constant.

Proof. Since φ is convex, there exists constant c and a linear function $L(x) = c(x - \mathbb{E}X) + \varphi(\mathbb{E}X)$. Then $L(X) \leqslant \varphi(X)$ implies $\mathbb{E}(L(X) \mid Y) \leqslant \mathbb{E}(\varphi(X) \mid Y)$ by the very definition of expectation. Then

$$\sup \mathbb{E}(L(X) \mid Y) = \sup L(\mathbb{E}(X \mid Y)) \leqslant \varphi(\mathbb{E}(X \mid Y)) \leqslant \mathbb{E}(\varphi(X) \mid Y)$$

where the supremum is taken over all linear functions with $LL(X) \leqslant \varphi(X)$. □

Problem 3

Let Y, Z be statistics and suppose Z is sufficient for $\{f_\theta : \theta \in \Theta\}$. Show that $W := \mathbb{E}_\theta(Y \mid Z)$ does not depend on θ. That is, there is a function $t : \mathbb{R}^n \to \mathbb{R}$ that does not depend on θ such that $W = t(X)$.

Proof. Let $W := g(Z)$ where $g(z) := \mathbb{E}(Y \mid Z = z) = \int_{-\infty}^{\infty} y f_\theta(y \mid Z = z)\, \mathrm{d}y$. By sufficiency $f_\theta(x \mid Z = z)$ does not depend on θ, so W does not depend on θ either. □

Problem 4

Let $X_1, \ldots, X_n$ be a random sample of size n so that X_1 is a sample from the uniform distribution on $[\theta - 1/2, \theta + 1/2]$ where $\theta \in \mathbb{R}$ is unknown.

(1) Show that $(X_{(1)}, X_{(n)})$ is minimal sufficient but not complete.

(2) The sample mean $\overline{X}$ might seem to be a reasonable estimator for θ but it is not a function of the minimal sufficient statistic so it is not so good. Find an unbiased estimator for θ with smaller variance than $\overline{X}$. Examine the ratio of variances for $\overline{X}$ and your estimator.

Proof. (1) Sufficiency follows from factorization because the joint likelihood is $1_{\theta-1/2 \leqslant X_{(1)} \leqslant X_{(n)} \leqslant \theta+1/2}$. Minimal sufficiency follows from the characterization since for $x_1, \ldots, x_n, y_1, \ldots, y_n$,

$$f(x_1, \ldots, x_n) = f(y_1, \ldots, y_n) \qquad \text{for all } \theta \in \mathbb{R}$$

if and only if $x_{(1)} = y_{(1)}$ and $x_{(n)} = y_{(n)}$.

(2) The variance of sample mean is $1/3 - 1/4 = 1/12$.

We consider $Y := ((X_{(1)} + X_{(2)})/2$ which is unbiased for θ due to symmetry. This estimator certainly has smaller variance than $\overline{X}$ because both $X_{(1)}, X_{(n)}$ are part of $\overline{X}$ whereas $\overline{X}$ contains more random data for $n > 2$, thereby increasing its variance. □

Problem 5

Let $X_1, \ldots, X_n$ be a random sample of size n from an exponential distribution with unknown parameter $\theta > 0$, i.e., the PDF of X_1 is $\theta e^{-x\theta}\chi_{x>0}$. Suppose we want to estimate the mean

$$g(\theta) := \frac{1}{\theta}.$$

(1) Find the UMVU for $g(\theta)$. (Hint: Cramér-Rao.)

(2) Show that $\sqrt{X_1 X_2}$ has smaller mean squared error than the UMVU, i.e.,

$$\mathbb{E}(\sqrt{X_1 X_2} - 1/\theta)^2$$

is less than that of the UMVU.

(3) Find an estimator with even smaller mean square error than $\sqrt{X_1 X_2}$ for all $\theta \in \Theta$.

Solution. (1) Claim: the sample mean $\frac{1}{n}\sum_{i=1}^{n} X_i$ is the UMVU. In this case the UMVU is simply $\overline{X} := (X_1 + X_2)/2$. The variance of $\overline{X}$ is $\mathrm{var}(X_1)/n = 1/(n\theta^2)$. (In this case it's just $1/(2\theta^2)$.) We now compute the Fisher information $I_X(1/\theta)$. Let $\lambda := 1/\theta$. Assuming $x_i > 0$,

$$\begin{aligned}\frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}\log f_\lambda(X) &= \frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}\log\left(\prod_{i=1}^{n}\lambda^{-1}e^{-x_i/\lambda}\right) = \frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}\left(\sum_{i=1}^{n}\log(1/\lambda) - x_i/\lambda\right)\\ &= \frac{\mathrm{d}}{\mathrm{d}\lambda}\left[-\frac{n}{\lambda} + \frac{n\overline{x}}{\lambda^2}\right] = \frac{n}{\lambda^2} - \frac{2n\overline{x}}{\lambda^3}.\end{aligned}$$

Therefore $I_\lambda(1/\theta) = -\mathbb{E}[n/\lambda^2 - (2n\overline{X})/\lambda^3] = n/\lambda^2 = n\theta^2$. Indeed we have

$$\mathrm{var}_\lambda(1/\theta) = \frac{1}{I_X(1/\theta)},$$

so Cramér-Rao shows the sample mean is the UMVU.

(2) Note that by independence

$$\mathbb{E}(\sqrt{X_1X_2} - 1/\theta)^2 = \mathbb{E}X_1X_2 - \frac{2}{\theta}\mathbb{E}\sqrt{X_1X_2} + \frac{1}{\theta^2} = (\mathbb{E}X_1)^2 - \frac{2}{\theta}(\mathbb{E}\sqrt{X_1})^2 + \frac{1}{\theta^2} = \frac{2}{\theta^2} - \frac{2}{\theta}(\mathbb{E}\sqrt{X_1})^2. \tag{1}$$

It remains to compute $\mathbb{E}\sqrt{X_1} = \int_0^\infty \sqrt{x}\theta e^{-x\theta}\,\mathrm{d}x = \theta\int_0^\infty \sqrt{x}e^{-x\theta}\,\mathrm{d}x$. Let

$$\begin{aligned} u &= \sqrt{x} & \mathrm{d}v &= e^{-x\theta}\mathrm{d}x\\ \mathrm{d}u &= \mathrm{d}x/(2\sqrt{x}) & v &= -e^{-x\theta}/\theta.\end{aligned}$$

Then

$$\int_0^\infty \sqrt{x}e^{-x\theta}\,\mathrm{d}x = -\frac{\sqrt{x}e^{-x\theta}}{\theta}\bigg|_{x=0}^{\infty} + \int_0^\infty \frac{e^{-x\theta}}{2\theta\sqrt{x}}\,\mathrm{d}x. \tag{2}$$

Letting $s := \sqrt{\theta}\sqrt{x}$ so that $\mathrm{d}s = \frac{\sqrt{\theta}}{2\sqrt{x}}\mathrm{d}x$, we have

$$\int_0^\infty \frac{e^{-x\theta}}{2\theta\sqrt{x}}\,\mathrm{d}x = \int_{x=0}^{x=\infty}\frac{e^{-x\theta}}{2\theta\sqrt{x}}\frac{2\sqrt{x}}{\sqrt{\theta}}\,\mathrm{d}s = \theta^{-3/2}\int_0^\infty e^{-s^2}\,\mathrm{d}s. \tag{3}$$

By a well-known result that $\int_{-\infty}^{\infty} e^{-s^2/2}\ \mathrm{d}s = \sqrt{2\pi}$ we know $\int_0^{\infty} e^{-s^2/2}\ \mathrm{d}s = \sqrt{\pi/2}$ (this is related to a Gaussian PDF; for proof, see here). Another simple u-substitution suggests $\int_0^{\infty} e^{-s^2}\ \mathrm{d}s = \sqrt{\pi}/2$. Thus (3) becomes $\theta^{-3/2}\sqrt{\pi}/2$, and putting this back to (2) we obtain

$$\int_0^{\infty} \sqrt{x}e^{-x\theta}\ \mathrm{d}x = 0 + \theta^{-3/2}\frac{\sqrt{\pi}}{2}.$$

Therefore,

$$\mathbb{E}\sqrt{X_1} = \theta\int_0^{\infty} \sqrt{x}e^{-x\theta}\ \mathrm{d}x = \frac{\sqrt{\pi}}{2\sqrt{\theta}}.$$

Finally, putting everything into (1), we have

$$\mathbb{E}(\sqrt{X_1X_2} - 1/\theta)^2 = \frac{2}{\theta^2} - \frac{2}{\theta}\cdot\left(\frac{\sqrt{\pi}}{2\sqrt{\theta}}\right)^2 = \frac{2}{\theta^2} - \frac{\pi}{2\theta^2} = \frac{4-\pi}{2\theta^2} < \frac{1}{2\theta^2} = \mathrm{var}(\overline{X}).$$

(3) To replace $\sqrt{X_1X_2}$ by $t\sqrt{X_1X_2}$, the MSE becomes

$$\frac{t^2}{\theta^2} - \frac{2t}{\theta}\frac{\pi}{4\theta} + \frac{1}{\theta^2} = \frac{1}{\theta^2}(t^2 - t(\pi/2) + 1)$$

which can be minimized when $t = \pi/4$. Hence $\pi\sqrt{X_1X_2}/4$ is a better estimator in terms of MSE.