MATH 408 Quiz 4 Solution Sketch

Qilin Ye

March 23, 2022

Problem 1

Let $X_1, \dots, X_n$ 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 $\sqrt{X_1X_2}$ has smaller mean squared error than the UMVU, i.e.,

$$\mathbb{E}(\sqrt{X_1X_2} - 1/\theta)^2$$

is less than that of the UMVU.

(3) Does (2) contradict the definition of UMVU?

(4) (Optional) Find an estimator with even smaller mean square error than $\sqrt{X_1X_2}$ 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 $\text{var}(X_1)/n = 1/(n\theta^2)$. (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

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

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

(2) *Heilman said this question is best solved by brute force computation.* Note that by independence

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

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

$$u = \sqrt{x} \qquad \mathrm{d}v = e^{-x\theta}\mathrm{d}x$$
$$\mathrm{d}u = \mathrm{d}x/(2\sqrt{x}) \quad v = -e^{-x\theta}/\theta.$$

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

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

(3) No, because $\sqrt{X_1X_2}$ is not unbiased a priori: since $\sqrt{\cdot}$ is not affine, Jensen's inequality (for concave functions) becomes strict in this case:

$$\mathbb{E}\sqrt{X_1X_2} < \sqrt{\mathbb{E}X_1X_2} = \frac{1}{\theta}.$$

(Alternatively, part 2 suggests $\mathbb{E}\sqrt{X_1X_2} = (\mathbb{E}\sqrt{X_1})^2 = (\sqrt{\pi}/(2\sqrt{\theta}))^2 = (\pi/4)(1/\theta) < 1/\theta$.)

Problem 2

Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. Let $x \in \mathbb{R}^n$ be a local minimum of f. Show that x is a global minimum.

If f is strictly convex, show there exists at most one global minimum of f.

Now suppose additionally that f is C^1 and $x \in \mathbb{R}^n$ with $\nabla f(x) = 0$. Show that x is a global minimum of f.

Proof. Let x^* be a local minimum. That is, there exists $r > 0$ such that x^* is the minimum on the ϵ-neighborhood of x:

$$f(x^*) \leqslant f(y) \text{ for all } y \text{ with } \|y - x^*\| < r.$$

Thus, for all $y \in \mathbb{R}^n$, there exists $\epsilon > 0$ sufficiently small so that $x^* + \epsilon(y - x^*)$ is in the r-neighborhood of x^*. Then

$$\begin{aligned} f(x^*) &\leqslant f(x^* + \epsilon(y - x^*)) \\ &= f(\epsilon y + (1-\epsilon)x^*) \\ &\leqslant \epsilon f(y) + (1-\epsilon) f(x^*). \end{aligned}$$

This implies $\epsilon f(x^*) \leqslant \epsilon f(y)$. Since $\epsilon > 0$, $f(x^*) \leqslant f(y)$, showing that x^* is a global minimum.

Now suppose f is strictly convex. If x_1, x_2 are global minima and $x_1 \neq x_2$, then by convexity $(x_1 + x_2)/2$ is a third point at which f takes a strictly smaller value. Contradiction. Hence global minimum in this case is unique.

Finally, if $\nabla f(x) = 0$ then x is a critical point. *By a well-known result (see e.g. here, p3 Theorem 2)*

$$f \in C^1 \text{ and } f \text{ convex} \iff f(y) \geqslant f(x) + \nabla f(x)^T (y - x) \text{ for all } x, y \in \mathbb{R}^n,$$

if $\nabla f(x) = 0$ then we obtain $f(y) \geqslant f(x) + 0$ for all $y \in \mathbb{R}^n$.That is, x is a global minimum. □

Problem 3

Let A be a real $m \times n$ matrix. Let $x \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$. Show that $f : \mathbb{R}^n \to \mathbb{R}$ defined by $f(x) = \|Ax - b\|^2/2$ is convex. Moreover show that

$$\nabla f(x) = A^T(Ax - b) \qquad \nabla^2 f(x) = A^T A.$$

Proof. Let $x, y \in \mathbb{R}^n$ and le $\lambda \in (0,1)$. We want to show

$$f(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda) f(y).$$

2 times the LHS is

$$\begin{aligned} \|A(\lambda x + (1-\lambda)y) - b\|^2 &= \|\lambda Ax + (1-\lambda)Ay - b\|^2 \\ &= \|\lambda(Ax - b) + (1-\lambda)(Ay - b)\|^2 \\ &= \lambda^2\|Ax - b\|^2 + (1-\lambda)^2\|Ay - b\|^2 + 2\lambda(1-\lambda)\langle Ax - b, Ay - b\rangle \\ &\overset{*}{\leqslant} \lambda^2\|Ax - b\|^2 + (1-\lambda)^2\|Ay - b\|^2 + 2\lambda(1-\lambda)\frac{\|Ax - b\|^2 + \|Ay - b\|^2}{2} \\ &= \lambda^2\|Ax - b\|^2 + (1-\lambda)^2\|Ay - b\|^2 + \lambda(1-\lambda)(\|Ax - b\|^2 + \|Ay - b\|^2) \\ &= \lambda\|Ax - b\|^2 + (1-\lambda)\|Ay - b\|^2 \end{aligned}$$

(The starred $\leqslant$ is because $a^2 - 2ab + b^2 \geqslant 0$ so $ab \leqslant (a^2 + b^2)/2$. Cauchy-Schwarz then gives $\langle u, v\rangle \leqslant \|u\|\|v\| \leqslant (\|u\|^2 + \|v\|^2)/2$.) 2 times the RHS is

$$\lambda\|Ax - b\|^2 + (1-\lambda)\|Ay - b\|^2.$$

Thus f is convex.

We write $f(x) := \frac{1}{2}((Ax - b)^T(Ax - b)) = \frac{1}{2}\cdot((Ax)^T Ax - 2b^T Ax + \|b\|^2)$. Then differentiating w.r.t. x gives

$$\nabla f(x) = x^T A^T A - b^T A = A^T Ax - A^T b = A^T(Ax - b).$$

Differentiating once more gives $\nabla^2 f(x) = A^T A$. □

Problem 4

Let $f_1, \dots, f_n : \mathbb{R} \to \mathbb{R}$ be strictly convex and define $g : \mathbb{R}^n \to \mathbb{R}$ by

$$g(x_1, \dots, x_n) := \sum_{i=1}^{n} f(x_i).$$

Show that g is strictly convex.

Proof. Let $x = (x_1, \dots, x_n), y = (y_1, \dots, y_n) \in \mathbb{R}^n$ and $\lambda \in (0, 1)$ be given. Then

$$\begin{aligned} g(\lambda x + (1-\lambda)y) &= g(\lambda x_1 + (1-\lambda)y_1, \dots, \lambda x_n + (1-\lambda)y_n) \\ &= \sum_{i=1}^{n} f(\lambda x_i + (1-\lambda)y_i) \\ &< \sum_{i=1}^{n} [\lambda f(x_i) + (1-\lambda) f(y_i)] && \text{(by strict convexity of } f\text{)} \\ &= \sum_{i=1}^{n} \lambda f(x_i) + \sum_{i=1}^{n} (1-\lambda) f(y_i) = \lambda g(x) + (1-\lambda) g(y). \end{aligned}$$

□

Problem 5

Let $f : \mathbb{R}^n \to \mathbb{R}$. Suppose for any fixed $1 \leqslant i \leqslant n$ and any fixed $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$, the function

$$x_i \mapsto f(x_1, \dots, x_n)$$

is strictly convex. Prove that f has at most one global minimum.

Proof. Suppose there exist two distinct global minima $x, y \in \mathbb{R}^n$. It follows that $x_i \neq y_i$ for some i. But then for the i^{th} component, if we fix $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$, the map

$$x \mapsto f(x_1, \dots, x_n)$$

is a strictly convex function with two distinct global minima. This is impossible by problem 2(b). Hence f has at most one global minimum. □

Problem 6

Let $X_1, \dots, X_n$ be a random sample of size n from a Poisson distribution with unknown parameter $\lambda > 0$ (so $\mathbb{P}(X_1 = k) = e^{-\lambda}\lambda^k / k!$ for $k \in \mathbb{N}$).

(1) Find an MLE for λ.

(2) Find an MLE for $e^{-\lambda}$.

(3) How do your results compare to the previous homework, where we found two different estimators for $e^{-\lambda}$ (one from the method of moments, and the other by applying the Rao-Blackwell Theorem)?

Solution. (1) The likelihood function $\ell(\lambda)$ is given by

$$\ell(\lambda) = \prod_{i=1}^{n} f_\lambda(x_i) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{(x_i)!}.$$

Taking log gives

$$\log \ell(\lambda) = -n\lambda + \sum_{i=1}^{n} x_i \log \lambda - \sum_{i=1}^{n} \log(x_i!).$$

Hence

$$\ell'(\lambda) = -n + \lambda^{-1} \sum_{i=1}^{n} x_i.$$

Setting $\ell'(\lambda)$ to 0, we see that the critical is the sample mean $\frac{1}{n}\sum_{i=1}^{n} x_i$. Verify that this is a maximum:

$$\ell''(\theta) = -\lambda^{-2} \sum_{i=1}^{n} x_i \leqslant 0.$$

Hence $\frac{1}{n}\sum_{i=1}^{n} X_i$ is an MLE estimator for λ.

(2) By functional equivariance of MLE (proposition 4.45), $\exp\left(-\frac{1}{n}\sum_{i=1}^{n} X_i\right)$ is an MLE for $e^{-\lambda}$.

(3) It does agree with the estimator obtained from MoM.

Problem 7

Let $X_1, \dots, X_n$ be a random sample of size n from a Gamma distribution with unknown $\alpha > 0$ and known $\beta > 0$.

(1) Try to find an MLE for α.

(2) Using a computer, after fixing some possible values of $X_1, \dots, X_n$, find an MLE for α using any computational optimization method you want to use. Can you guarantee that you have found the global maximum?

Solution. The likelihood function is given by

$$\ell(\theta) = \prod_{i=1}^{n} \frac{x_i^{\alpha-1}\exp(-x_i/\beta)}{\beta^\alpha \Gamma(\alpha)} = \frac{1}{\beta^{n\alpha}\Gamma(\alpha)^n} \cdot \prod_{i=1}^{n} x_i^{\alpha-1} \cdot \exp(-\sum_{i=1}^{n} x_i/\beta).$$

The log-likelihood is therefore

$$\log \ell(\theta) = -n\alpha \log \beta - n \log \Gamma(\alpha) + (\alpha - 1)\sum_{i=1}^{n} \log x_i - \sum_{i=1}^{n} x_i/\beta.$$

Setting the derivative to 0, we need

$$\frac{\mathrm{d}}{\mathrm{d}\theta} \log \ell(\theta) = \sum_{i=1}^{n} \log x_i - n \log \beta - n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} = 0.$$

...fuck!