

Homework 7: Fisher Information and Bayesian Statistics

Due Thursday, October 31, 2019

This homework will cement your understanding of the material in Rice Chapter 8, Sections 8.5 through 8.8. The assignment is due before the end of class on the date listed above. You may discuss problems with your classmates, but you must turn in your own writeup, and list all students with whom you discussed the assignment. You must show **all your work** in order to receive full credit. Stop by office hours if you need additional help or hints.

Problem 1 (The Pareto distribution). The Pareto distribution is often a good model for the distribution of wealth or other resources in a society. It has a density

$$f(x; x_0, \theta) = \begin{cases} \theta x_0^\theta x^{-(\theta+1)} & \text{if } x \geq x_0 \\ 0 & \text{otherwise,} \end{cases}$$

where $x_0 > 0$ and $\theta > 1$.

- (a) Show that $f(x; x_0, \theta)$ indeed defines a probability density.
- (b) Derive a method of moments estimate for θ under the assumption that x_0 is known.
- (c) Derive the MLE for θ under the assumption that x_0 is known.
- (d) Find the Fisher information for θ under the assumption that x_0 is known. You may assume that the density of the Pareto distribution obeys the “suitable regularity conditions” discussed in lecture.
- (e) Use part (d) to construct an approximate $100(1 - \alpha)\%$ confidence interval for θ .

Problem 2. In the previous homework, you derived a method-of-moments estimator and the MLE for the Rayleigh distribution, which has density

$$f(x; \theta) = \begin{cases} \frac{x}{\theta^2} \exp \left\{ \frac{-x^2}{2\theta^2} \right\} & \text{if } x \geq 0 \\ 0 & \text{otherwise,} \end{cases}$$

where $\theta > 0$.

- (a) Derive the Fisher information for θ . You may assume that the density of the Rayleigh distribution obeys the “suitable regularity conditions” discussed in lecture.
- (b) Letting $\hat{\theta}_{\text{MLE}}$ be the MLE for θ that we found in last week’s homework, use part (a) to derive the limiting distribution for $\sqrt{n}(\hat{\theta}_{\text{MLE}} - \theta)$.
- (c) Use part (b) to construct an approximate $100(1 - \alpha)\%$ confidence interval for θ .

Problem 3 (Bernoulli random variables). Let X_1 be a Bernoulli random variable with success parameter p .

- (a) Show that the Fisher information for p is $I(p) = 1/p(1-p)$.
- (b) Verify the “suitable regularity conditions” by showing that

$$-\mathbb{E}_p \frac{\partial^2}{\partial p^2} \log f(X_1; p) = \frac{1}{p(1-p)}.$$

- (c) Observe that in the case of the Bernoulli (like the Poisson we saw in lecture), $I(p)$ is the inverse of the variance. One interpretation of the Fisher information is that it measures “how much we learn” about the true value of the parameter by observing a single draw from the distribution. In light of the fact that $I(p)$ is the inverse of the variance, does this interpretation make sense? **Hint:** think about the case where $p \approx 1/2$ and compare to the cases where $p \approx 0$ or $p \approx 1$. **Note:** no need to do any math for this subproblem; a sentence or two is plenty.

Problem 4. In lecture, we said that provided we can move the derivative past the expectation, we have

$$\mathbb{E}_\theta \frac{\partial}{\partial \theta} \log f(X_1; \theta) = 0.$$

Show that this is true.

Problem 5 (Uniform random variables, revisited). Let $X_1, X_2, \dots, X_n$ be i.i.d. uniform on $[0, \theta]$ for some $\theta > 0$. In the previous homework, we derived a method-of-moments estimator and MLE for θ . We alluded in lecture to the fact that this distribution does not obey the “suitable regularity conditions” related to the Fisher information. In this homework, we’ll explore this setup a bit more.

Recall that the uniform has density $f(x; \theta) = 1/\theta$ if $x \in [0, \theta]$ and $f(x; \theta) = 0$ otherwise.

- (a) Show that for $0 < x < \theta$,

$$\frac{\partial}{\partial \theta} \log f(x; \theta) = \frac{-1}{\theta}.$$

- (b) Use part (a) to show that

$$\mathbb{E}_\theta \left(\frac{\partial}{\partial \theta} \log f(x; \theta) \right)^2 = \frac{1}{\theta^2}.$$

- (c) We showed in lecture that the maximum likelihood estimator was given by $\hat{\theta}_{\text{MLE}} = \max\{X_1, X_2, \dots, X_n\}$. Show that the density of $\hat{\theta}_{\text{MLE}}$ is given by

$$g_\theta(t) = \begin{cases} \frac{nt^{n-1}}{\theta^n} & \text{if } 0 \leq t \leq \theta \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

Hint: note that $\hat{\theta}_{\text{MLE}} \leq t$ if and only if $X_i \leq t$ for all $i = 1, 2, \dots, n$. Use this fact along with the independence of the X_i to obtain the CDF for $\hat{\theta}_{\text{MLE}}$, then differentiate to get the density.

- (d) Let $Y = (n+1)\hat{\theta}_{\text{MLE}}/n$. Use the previous part to show that Y is an unbiased estimator for θ . **Hint:** we have

$$\mathbb{E}_\theta Y = \frac{n+1}{n} \int_0^\theta t g_\theta(t) dt,$$

where g_θ is as defined in Equation (1).

- (e) Let Y be as defined in the previous part. Show that

$$\text{Var } Y = \frac{\theta^2}{n(n+2)}.$$

Hint: you already computed $\mathbb{E}Y$ above, so it's enough to compute $\mathbb{E}Y^2$.

- (f) Above, we showed that the Fisher information for θ was $1/\theta^2$. Explain, in light of the Cramér-Rao lower bound, why the variance we found in the previous part seems surprising. You may take it as given that $\hat{\theta}_{\text{MLE}}$ is sufficient for θ .
- (g) Of course, the Cramér-Rao lower bound hasn't actually been violated. The problem is that the interchange of derivative and integral is not allowed. To see why this is so, let $h : \mathbb{R} \rightarrow \mathbb{R}$ be any function, and consider

$$\frac{\partial}{\partial \theta} \mathbb{E}h(X_1) = \frac{\partial}{\partial \theta} \int_0^\theta h(x) f(x; \theta) dx = \frac{\partial}{\partial \theta} \int_0^\theta \frac{h(x) dx}{\theta}.$$

Recall Leibniz' Rule from calculus, which says that if functions $q(x, \theta)$, $a(\theta)$ and $b(\theta)$ are differentiable with respect to θ , then

$$\frac{\partial}{\partial \theta} \int_{a(\theta)}^{b(\theta)} q(x, \theta) dx = q(b(\theta), \theta) \frac{\partial}{\partial \theta} b(\theta) - q(a(\theta), \theta) \frac{\partial}{\partial \theta} a(\theta) + \int_{a(\theta)}^{b(\theta)} \frac{\partial}{\partial \theta} q(x, \theta) dx.$$

Take $a(\theta) = 0$, $b(\theta) = \theta$ and $q(\theta, x) = h(x)f(x; \theta)$ and use Leibniz' rule to show that

$$\frac{\partial}{\partial \theta} \mathbb{E}h(X_1) = \frac{h(\theta)}{\theta} + \int_0^\theta h(x) \left(\frac{\partial}{\partial \theta} \frac{1}{\theta} \right) dx.$$

- (h) Compare the previous part with

$$\int_0^\theta h(x) \left(\frac{\partial}{\partial \theta} f(x; \theta) \right) dx = \int_0^\theta h(x) \left(\frac{\partial}{\partial \theta} \frac{1}{\theta} \right) dx.$$

Argue why interchange of differentiation and integration fails here unless $h(\theta)/\theta = 0$ for all $\theta > 0$.

- (i) We showed previously that the MOM estimator for θ is $\hat{\theta}_{\text{MoM}} = 2\bar{X}$, where $\bar{X}$ denotes the sample mean. Can you think of anything wrong with this estimator? **Hint:** think about the possibility that $\max_i X_i > 2\bar{X}$.

Problem 6 (Sufficiency and the factorization theorem). Let $X_1, X_2, \dots, X_n$ be drawn i.i.d. from a Poisson distribution with parameter λ .

- (a) Show that $\bar{X} = n^{-1} \sum_{i=1}^n X_i$ is sufficient for λ by proving that the distribution of $X_1, X_2, \dots, X_n$ is independent of λ conditioned on $\bar{X}$.
- (b) Show that X_1 is *not* sufficient for λ .
- (c) Use the factorization theorem (Theorem A in Rice Section 8.8.1) to show (again) that $\bar{X}$ is sufficient for λ . **Hint:** you may find it easier to consider $n\bar{X} = \sum_i X_i$.