Homework 5

Aleck Zhao

October 12, 2016

1. We stated the *Cramer-Rao lower bound* in lecture; namely, that if $T = g(X_1, X_2, \cdots, X_n)$ is an unbiased estimate for a parameter θ based on iid observations X_i from a sufficiently smooth density f_θ, then the variance of T satisfies the following lower bound:

$$\mathrm{Var}(T) \geq \frac{1}{nI(\theta)}$$

where $I(\theta)$ is the Fisher information.

(a) Let

$$Z = \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(X_i|\theta)$$

Show that $E[Z] = 0$.

Proof. We have

$$\begin{aligned} E[Z] &= E\left[\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(X_i|\theta)\right] = \sum_{i=1}^{n} E\left[\frac{\partial}{\partial\theta} \log f(X_i|\theta)\right] \\ &= nE\left[\frac{\partial}{\partial\theta} \log f(X|\theta)\right] = nE\left[\frac{\frac{\partial}{\partial\theta} f(X|\theta)}{f(X|\theta)}\right] \end{aligned}$$

where in the last step we invoke the chain rule. Treating the entire expression within the expectation as a random variable with density $f(x|\theta)$, this is

$$\begin{aligned} nE\left[\frac{\frac{\partial}{\partial\theta} f(X|\theta)}{f(X|\theta)}\right] &= n\int \frac{\frac{\partial}{\partial\theta} f(x|\theta)}{f(x|\theta)} f(x|\theta)\, dx \\ &= n\int \frac{\partial}{\partial\theta} f(x|\theta)\, dx \\ &\implies n\frac{\partial}{\partial\theta}\int f(x|\theta)\, dx \\ &= n\frac{\partial}{\partial\theta} 1 = 0 \end{aligned}$$

as desired. ∎

(b) Use the fact that $E[Z] = 0$ to prove that

$$\mathrm{Cov}(Z, T) \leq \sqrt{\mathrm{Var}(Z)\mathrm{Var}(T)}$$

Proof. ∎

¡++¿

(c) Compute the variance of Z.

(d) Show that $\text{Cov}(Z, T) = 1$.

2. Let $X_1, \cdots, X_n$ be iid uniform on $[0, \theta]$.

a. Find the method of moments estimate of θ and its mean and variance.

Solution. We have $\mu_1 = E[X] = \theta/2$, so the method of moments estimate of θ is $\hat{\theta} = 2\hat{\mu}_1$. Then,

$$\begin{aligned} E[\hat{\theta}] &= E[2\hat{\mu}_1] = 2E[\hat{\mu}_1] = 2(\theta/2) = \theta \\ \text{Var}(\hat{\theta}) &= \text{Var}(2\hat{\mu}_1) = 4\text{Var}(\hat{\mu}_1) \\ &= 4 \cdot \frac{\text{Var}(X)}{n} \end{aligned}$$

where

$$\begin{aligned} \text{Var}(X) &= E[X^2] - (E[X])^2 = E[X^2] - \left(\frac{\theta}{2}\right)^2 \\ &= \int_0^\theta x^2 \frac{1}{\theta}\, dx - \frac{\theta^2}{4} \\ &= \frac{\theta^2}{3} - \frac{\theta^2}{4} \\ &= \frac{\theta^2}{12} \end{aligned}$$

so the variance of the estimate is

$$\text{Var}(\hat{\theta}) = \frac{4}{n} \cdot \frac{\theta^2}{12} = \frac{\theta^2}{3n}.$$

□

b. Find the MLE of θ.

Solution. The likelihood function is

$$f(X_1, X_2, \cdots, X_n|\theta) = \prod_{i=1}^n f(X_i|\theta) = \prod_{i=1}^n \frac{1}{\theta} = \frac{1}{\theta^n}$$

Clearly, $\theta \geq X_i$ for all X_i, and since $1/\theta^n$ is decreasing with respect to θ, the MLE is $\max(X_1, X_2, \cdots, X_n)$.

□

c. Find the probability density of the MLE, and calculate its mean and variance. Compare the variance, the bias, and the MSE to those of the method of moments estimate.

Solution. Consider the probability $P(\max(X_1, X_2, \cdots, X_n) \leq x$. This is equivalent to

$$P(X_1, X_2, \cdots, X_n \leq x) = \prod_{i=1}^n P(X_i \leq x) = \prod_{i=1}^n \frac{x}{\theta} = \left(\frac{x}{\theta}\right)^n$$

since the X_i are iid uniform. Then the distribution of the MLE is the derivative of this with respect to x, which is

$$f(x) = \frac{nx^{n-1}}{\theta^n}$$

where x ranges from 0 to θ. Then

$$
\begin{aligned}
E[\hat{\theta}] &= \int_0^\theta x \frac{nx^{n-1}}{\theta^n}\, dx = \frac{n\theta}{n+1} \\
\mathrm{Var}(\hat{\theta}) &= E[\hat{\theta}^2] - (E[\hat{\theta}])^2 = E[\hat{\theta}^2] - \left(\frac{n\theta}{n+1}\right)^2 \\
&= \int_0^\theta x^2 \frac{nx^{n-1}}{\theta^n}\, dx - \left(\frac{n\theta}{n+1}\right)^2 \\
&= \frac{n\theta^2}{n+2} - \frac{n^2\theta^2}{(n+1)^2} \\
&= \frac{n\theta^2}{(n+1)^2(n+2)}
\end{aligned}
$$

Clearly, these are different from the mean and variance of the method of moments estimators.

□

d. Find a modification of the MLE that renders it unbiased.

Solution. We can just let

$$\hat{\theta}_2 = \frac{n+1}{n} \max\left(X_1, X_2, \cdots, X_n\right)$$

be the modified MLE, so

$$E[\hat{\theta}_2] = E\left[\frac{n+1}{n}\hat{\theta}\right] = \frac{n+1}{n} \cdot \frac{n\theta}{n+1} = \theta$$

which is unbiased, as desired.

□

3. Let X_i be iid uniform on $[0, \theta]$. Let $\hat{\theta}_n$ be the MLE for θ that you obtained from the previous exercise.

a) Show that $P(\hat{\theta}_n - \theta > \varepsilon) = 0$ for any $\varepsilon > 0$.

b) For any $\varepsilon > 0$, determine an explicit expression for the probability

$$P\left(|\hat{\theta} - \theta| > \varepsilon\right)$$

c) Compute, for any $\varepsilon > 0$, the limit

$$P\left(\sqrt{n}(\hat{\theta}_n - \theta)| > \varepsilon\right)$$

as $n \to \infty$.

d) What do your previous answers suggest about the asymptotic distribution of

$$\sqrt{n}(\hat{\theta}_n - \theta)?$$

In particular, does this still look approximately normal?

Chapter 8: Estimation of Parameters and Fitting of Probability Distributions

58. If gene frequencies are in equilibrium, the genotypes AA, Aa, and aa occur with probabilities $(1-\theta)^2, 2\theta(1-\theta)$, and θ^2, respectively. Data on a sample of 190 people: 10 with Hp1-1, 68 with Hp1-2, 112 with Hp2-2.

    a. Find the MLE of θ.

    Solution. The likelihood function is the product

    $$\prod_{i=1}^{10}(1-\theta)^2\prod_{j=1}^{68}2\theta(1-\theta)\prod_{k=1}^{112}\theta^2 = 2^{68}\theta^{292}(1-\theta)^{88}$$

    and the log-likelihood is

    $$\ell(\theta) = \log\left[2^{68}\theta^{292}(1-\theta)^{88}\right] = 86\log 2 + 292\log\theta + 88\log(1-\theta)$$

    Taking the derivative with respect to θ and setting equal to 0, we have

    $$\begin{aligned}\frac{\partial}{\partial\theta}\ell(\theta) &= \frac{292}{\theta} - \frac{88}{1-\theta} = 0\\ \implies \hat{\theta} &= \frac{73}{95}\end{aligned}$$

    is the MLE.

    □

    b. Find the asymptotic variance of the MLE.

    c. Find an approximate 99% confidence interval for θ.

    d. Use the bootstrap to find the approximate standard deviation of the MLE and compare to the result of part b).

    e. Use the bootstrap to find an approximate 99% confidence interval and compare to part c).

30. The exponential distribution if $f(x;\lambda) = \lambda e^{-\lambda x}$ and $E[X] = \lambda^{-1}$. The CDF is $F(x) = P(X \leq x) = 1 - e^{-\lambda x}$. Three observations are made by an instrument that reports $x_1 = 5, x_2 = 3$, but x_3 is too large for the instrument to measure and it only reports that $x_3 > 10$.

    a. What is the likelihood function?

    Solution. We have

    $$P(X_3 > 10) = 1 - P(X_3 \leq 10) = 1 - (1 - e^{-10\lambda}) = e^{-10\lambda}.$$

    Then the likelihood function is given by

    $$f(5)f(3)P(X_3 > 10) = (\lambda e^{-5\lambda})(\lambda e^{-3\lambda})e^{-10\lambda} = \lambda^2 e^{-15\lambda}.$$

    □

    b. What is the MLE of λ?

Solution. The log-likelihood is

$$\log\left(\lambda^2 e^{-15\lambda}\right) = 2\log\lambda - 15\lambda.$$

Taking the derivative with respect to λ and setting equal to 0, we have

$$\begin{aligned}\frac{\partial}{\partial\lambda}\left(2\log\lambda - 15\lambda\right) &= \frac{2}{\lambda} - 15 = 0 \\ \implies \hat{\lambda} &= \frac{2}{15}\end{aligned}$$

is the MLE.

□

31. George spins a coin three times and observes no heads. He then gives the coin to Hilary. She spins it until the first head occurs, and ends up spinning it four times total. Let θ denote the probability the coin comes up heads.

   a. What is the likelihood of θ?

   Solution. Let $p(x) = \theta^x(1-\theta)^{1-x}$ be the PMF for flipping a coin, where 1 represents H, and 0 represents T. Then the likelihood is the joint distribution of 7 iid flips, which is

$$\prod_{i=1}^{7} \theta^{X_i}(1-\theta)^{1-X_i}$$

□

   b. What is the MLE of θ?

   Solution. We are given that X_1 through X_6 are 0, and X_7 is 1 from the sample. The log-likelihood function is

$$\begin{aligned}\ell(\theta) &= \log\left(\prod_{i=1}^{7}\theta^{X_i}(1-\theta)^{1-X_i}\right) \\ &= \sum_{i=1}^{7}\log\left[\theta^{X_i}(1-\theta)^{1-X_i}\right] \\ &= \sum_{i=1}^{7}\left[X_i\log\theta + (1-X_i)\log(1-\theta)\right] \\ &= \log\theta\sum_{i=1}^{7}X_i + \log(1-\theta)\sum_{i=1}^{7}(1-X_i)\end{aligned}$$

   so evaluating with the sample data, we have

$$\ell(\theta) = \log\theta + 6\log(1-\theta)$$

   and taking the derivative with respect to θ and setting equal to 0, we have

$$\begin{aligned}\frac{\partial}{\partial\theta}\ell(\theta) &= \frac{\partial}{\partial\theta}\left[\log\theta + 6\log(1-\theta)\right] \\ &= \frac{1}{\theta} - \frac{6}{1-\theta} = 0 \\ \implies \hat{\theta} &= \frac{1}{7}\end{aligned}$$

   is the MLE.

□

34. Suppose that $X_1, X_2, \cdots, X_n$ are iid $N(\mu_0, \sigma_0^2)$ and μ and σ^2 are estimated by the method of maximum likelihood, with resulting estimates $\hat{\mu}$ and $\hat{\sigma}^2$. Suppose the bootstrap is used to estimate the sampling distribution of $\hat{\mu}$.

   a. Explain why the bootstrap estimate of the distribution of $\hat{\mu}$ is $N\left(\hat{\mu}, \frac{\hat{\sigma}^2}{n}\right)$.

   b. Explain why the bootstrap estimate of the distribution of $\hat{\mu} - \mu$ is $N\left(0, \frac{\hat{\sigma}^2}{n}\right)$.

   c. According to the result of the previous part, what is the from of the bootstrap confidence interval for μ, and how does it compare to the exact confidence interval based on the t distribution?

50. Let $X_1, \cdots, X_n$ be an iid sample from a Rayleigh distribution with parameter $\theta > 0$:

$$f(x|\theta) = \frac{x}{\theta^2} e^{-x^2/2\theta^2}, \quad x \geq 0$$

   a. Find the method of moments estimate of θ.

   Solution. We have

$$\mu_1 = E[X] = \int_0^\infty x \frac{x}{\theta^2} e^{-x^2/2\theta^2} = \theta\sqrt{\frac{\pi}{2}}$$

   according to Wolfram, so the method of moments estimate of θ is

$$\hat{\theta} = \sqrt{\frac{2}{\pi}}\hat{\mu}_1.$$

□

   b. Find the MLE of θ.

   Solution. The log-likelihood function is

$$\begin{aligned}
\ell(\theta) &= \sum_{i=1}^n \log f(X_i|\theta) = \sum_{i=1}^n \log\left(\frac{X_i}{\theta^2} e^{-X_i^2/2\theta^2}\right) \\
&= \sum_{i=1}^n \left(\log X_i - \log\theta - \frac{X_i^2}{2\theta^2}\right) \\
&= -n\log\theta + \sum_{i=1}^n \log X_i - \frac{1}{2\theta^2}\sum_{i=1}^n X_i^2
\end{aligned}$$

   so the partial with respect to θ and setting equal to 0, we have

$$\begin{aligned}
\frac{\partial}{\partial\theta}\ell(\theta) &= \frac{\partial}{\partial\theta}\left(-n\log\theta + \sum_{i=1}^n \log X_i - \frac{1}{2\theta^2}\sum_{i=1}^n X_i^2\right) \\
&= -\frac{n}{\theta} + \frac{1}{\theta^3}\sum_{i=1}^n X_i^2 = 0 \\
\implies \hat{\theta} &= \sqrt{\frac{1}{n}\sum_{i=1}^n X_i^2}
\end{aligned}$$

   is the MLE.

□

   c. Find the asymptotic variance of the MLE.

Solution. We have

$$\begin{aligned}
I(\theta) &= E\left[\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right)^2\right] \\
&= E\left[\left(\frac{\partial}{\partial\theta}\left(\log X - \log\theta - \frac{X^2}{2\theta^2}\right)\right)^2\right] \\
&= E\left[\left(-\frac{1}{\theta} + \frac{X^2}{\theta^3}\right)^2\right] \\
&= \frac{1}{\theta^2} - \frac{2}{\theta^4}E[X^2] + \frac{1}{\theta^6}E[X^4]
\end{aligned}$$

where

$$\begin{aligned}
E[X^2] &= \int_0^\infty x^2 \frac{x}{\theta^2} e^{-x^2/2\theta^2}\, dx = 2\theta^2 \\
E[X^4] &= \int_0^\infty x^4 \frac{x}{\theta^2} e^{-x^2/2\theta^2}\, dx = 8\theta^4
\end{aligned}$$

according to Wolfram, so the Fisher information is given by

$$I(\theta) = \frac{1}{\theta^2} - \frac{2}{\theta^4}(2\theta^2) + \frac{1}{\theta^6}(8\theta^4) = \frac{7}{\theta^2}$$

so the asymptotic variance is given by

$$\frac{1}{nI(\theta)} = \frac{\theta^2}{7n}$$

□

73. Find a sufficient statistic for the Rayleigh density

$$f(x|\theta) = \frac{x}{\theta^2} e^{-x^2/(2\theta^2)}, \quad x \geq 0$$

68. Let $X_1, \cdots, X_n$ be an iid sample from a Poisson distribution with mean λ and let $T = \sum_{i=1}^n X_i$.

    a. Show that the distribution of $X_1, \cdots, X_n$ given T is independent of λ, and conclude that T is sufficient for λ.
    b. Show that X_1 is not sufficient.
    c. Use Theorem A of section 8.8.1 to show that T is sufficient. Identify the functions g and h of that theorem.

69. Use the factorization theorem to conclude that $T = \sum_{i=1}^n X_i$ is a sufficient statistic when the X_i are an iid sample from a geometric distribution.

70. Use the factorization theorem to find a sufficient statistic for the exponential distribution.

71. Let $X_1, \cdots, X_n$ be an iid sample from a distribution with the density function

$$f(x|\theta) = \frac{\theta}{(1+x)^{\theta+1}}, \quad 0 < \theta < \infty; \quad 0 \leq x < \infty$$

    Find a sufficient statistic for θ.

72. Show that $\prod_{i=1}^n X_i$ and $\sum_{i=1}^n X_i$ are sufficient statistics for the gamma distribution.