Statistics GIDP
Ph.D. Qualifying Exam
Theory

Jan 11, 2016, 9:00am-1:00pm

Instructions: Provide answers on the supplied pads of paper; write on only one side of each sheet. Complete exactly 5 of the 6 problems. Turn in only those sheets you wish to have graded. Stay calm and do your best; good luck.

1. Hidden inside each box of Primate brand breakfast cereal is a small plastic figure of an animal: an ape or a baboon. Suppose a fraction p of the very large population of cereal boxes contain apes, a fraction $q = 1 - p$ contain baboons. Let X be the minimal number of boxes you need to buy to collect both figures.

   (a) Find $P(X = 3)$.

   (b) Find $E(X)$.

2. Two random variables X and Y are jointly normal and marginally standard normal with correlation ρ.

   (a) Show that $|X - Y| = \max\{X, Y\} + \max\{-X, -Y\}$.

   (b) Find $E(\max\{X, Y\})$.

3. Let X_1,...,X_n be independent and identically distribution random variables with mean 1 and finite variance. Show that $\frac{2}{n(n+1)} \sum_{j=1}^{n} jX_j \to 1$ in probability.

4. Let X_1,...,X_n be independent and identically distributed observations from a normal distribution $N(\theta, \theta^2)$, with $\theta > 0$ unknown.

   (a) Show that $T = (\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2)$ is sufficient for θ.

   (b) Find two unbiased estimators based on T that are unbiased for θ^2.

   (c) Is T complete? Justify your answer.

   (d) Suppose that we test $H_0 : \theta = 1$ vs $H_1 : \theta > 1$ by using the rule

   $$\text{Reject } H_0 \text{ if } \sum_{i=1}^{n} X_i > c.$$

   Find the value of c to give a test of size α. Assume $\alpha > 0$ is given.

5. Let $X_1, \cdots, X_n$ be independent and identically distributed random variables with the probability density function (pdf)

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

where $\theta > 0$ is an unknown parameter. Note that

$$E(X_1) = \theta,\ E(X_1^2) = \left(\frac{\pi}{2}\right)\theta^2,\ E(X_1^4) = \left(\frac{3\pi^2}{4}\right)\theta^4.$$

(a) Derive the Cramér-Rao lower bound for variance of an unbiased estimator of θ^2.

(b) Find the maximum likelihood estimator (MLE) for θ^2.

(c) Let $Y = \frac{2}{\pi\theta^2}\sum_{i=1}^n X_i^2$. Specify the distribution of Y.

(d) Construct a 95% confidence interval for θ^2.

6. Let X_1,...,X_n are independent and identically distributed samples with the pdf

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

where θ is an unknown parameter. Let $X_{(1)} = \min_{1\leq i\leq n} X_i$ and $X_{(n)} = \max_{1\leq i\leq n} X_i$. You may use the fact that $X_{(1)}$ is a complete statistic in the following questions.

(a) Provide with justification a minimum sufficient statistic for θ.

(b) Derive the method-of-moment estimator (MME) of θ. Call the estimator $\hat{\theta}_{MME}$.

(c) Derive the maximum likehood estimator (MLE) of θ. Call the estimator $\hat{\theta}_{MLE}$.

(d) Between $\hat{\theta}_{MME}$ and $\hat{\theta}_{MLE}$, which is a better estimator for θ in terms of the mean squared error (MSE)? Justify your answer.

(e) Find an unbiased estimator of θ which is better than both $\hat{\theta}_{MME}$ and $\hat{\theta}_{MLE}$.

Solutions

1. (a) $P(X=3) = P(aab) + P(bba) = p^2q + q^2p = pq$.

   (b) It is easy to see $P(X=x) = pq(p^{x-2} + q^{x-2})$ for $x \geq 2$. So $E(X) = ...$

2. (a) $|X-Y| = \max\{X,Y\} - \min\{X,Y\} = \max\{X,Y\} + \max\{-X,-Y\}$.

   (b) By symmetry, $\max\{X,Y\}$ and $\max\{-X,-Y\}$ have the same distribution. So $E(\max\{X,Y\}) = \frac{1}{2}E(|X-Y|)$. $X-Y$ is normal with mean 0 and variance $2-2\rho$. $E(\max\{X,Y\}) = \frac{1}{2}E(|X-Y|) = \sqrt{(1-\rho)/\pi}$.

3. Simply check that the mean goes to 1 and variance goes to 0.

4. (a) The joint density of $X_1, ..., X_n$ is

$$f(x|\theta) = \frac{1}{(\theta\sqrt{2\pi})^n} \exp\left[-\frac{1}{2\theta^2}\sum_{i=1}^n x_i^2 + \frac{1}{\theta}\sum_{i=1}^n x_i\right] \exp\{-\frac{n}{2}\}.$$

   By factorization theorem, $T = (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ is sufficient for θ. Define $T_1 = \sum_{i=1}^n X_i$ and $T_2 = \sum_{i=1}^n X_i^2$.

   (b) The sample variance is always unbiased for variance. So

$$g_1(T) = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n-1} = \frac{1}{n}[T_2 - \frac{T_1^2}{n}].$$

   Secondly, consider $\bar{X}^2 = (T_1/n)^2$. Note that

$$E(\bar{X}^2) = E(\bar{X})^2 + Var(\bar{X}) = \theta^2 + \theta^2/n = \frac{n+1}{n}\theta^2.$$

   So the second unbiased estimator is

$$g_2(T) = \frac{n}{n+1}\bar{X}^2 = \frac{T_1^2}{n(n+1)}$$

   (c) T is not complete since

$$E[g_1(T) - g_2(T)] = 0,$$

   but $g_1 - g_2$ is not a zero function.

   (d) Suppose that I were to test

$$H_0 : \theta = 1 \text{ vs } H_1 : \theta > 1$$

   ans use a test of the form:

$$\text{Reject if } \sum_{i=1}^n X_i \geq c.$$

   Note that $\sum_{i=1}^n X_i \sim N(n\theta, n\theta^2)$. The size of the test

$$\sup_{\theta=1} P(\sum_{i=1}^n X_i \geq c) = P(\sum_{i=1}^n X_i \geq c|\theta = 1) = P(\frac{\sum_{i=1}^n X_i - n}{\sqrt{n}} \geq \frac{c-n}{\sqrt{n}}) = \alpha,$$

   which leads to $c = n + z_\alpha\sqrt{n}$.

5. (a) Note that

$$s_1(x,\theta) = \frac{\partial}{\partial\theta}\log f(x;\theta) = -\frac{1}{\theta} + \frac{2}{\pi\theta^3}x^2.$$

Then

$$I_1(\theta) = \mathrm{Var}(s_1) = \frac{4}{\pi^2\theta^6}\mathrm{Var}(X^2).$$

Since $Var(X^2) = E(X^4) - [E(X^2)]^2 = \left(\frac{3\pi^2}{4}\right)\theta^4 - \left(\frac{\pi}{2}\right)^2\theta^4 = \frac{\pi^2\theta^4}{2}$, we have

$$I_1(\theta) = \frac{4}{\pi^2\theta^6}\frac{\pi^2}{2}\theta^4 = \frac{2}{\theta^2}.$$

So the Cramér lower variance bound for θ^2 is

$$\frac{(2\theta)^2}{I_n(\theta)} = \frac{(2\theta)^2}{2n/\theta^2} = \frac{2\theta^4}{n}.$$

(b) It is easy to show that $\hat\theta_{MLE} = \sqrt{\frac{2}{n\pi}\sum_{i=1}^n x_i^2}$ and $\hat\theta^2_{MLE} = \frac{2}{n\pi}\sum_{i=1}^n x_i^2$.

(c) Let $W_1 = \frac{2}{\pi\theta^2}X_1^2$. Then the pdf of W_1 is

$$f_{W_1}(w) = \sqrt{\frac{1}{2\pi w}}\exp\{-\frac{w}{2}\}I(w>0) = \frac{1}{2^{\frac{1}{2}}\Gamma(\frac{1}{2})}e^{-w/2}w^{\frac{1}{2}-1}I(w>0),$$

which is χ_1^2. Then $Y = \sum_{i=1}^n W_i^2 \sim \chi_n^2$.

(d) Using Y as a pivotal, then we look for a and b such that

$$P(a \le Y \le b) = P(a \le \frac{2}{\pi\theta^2}\sum_{i=1}^n X_i^2 \le b) = 95\%.$$

Setting $a = \chi^2_{n,0.975}$ and $b = \chi^2_{n,0.025}$, where $\chi_{n,\alpha}$ is $(1-\alpha)$ quantile of χ_n^2 distribution. Then 95% CI for θ^2 is $[\frac{2\sum_{i=1}^n X_i^2}{\pi\chi^2_{n,0.025}}, \frac{2\sum_{i=1}^n X_i^2}{\pi\chi^2_{n,0.975}}]$.

6. (a) Let $\mathbf{X} = (X_1,\ldots,X_n)$. Its joint pdf is

$$f(\mathbf{x}|\theta) = e^{n\theta}I(\theta \le X_{(1)})e^{-\sum_{i=1}^n x_i},$$

where $X_{(1)} = \min_{1\le i\le n} X_i$. By factorization theorem, $X_{(1)}$ is a sufficient statistic. Since it is also complete, it is a minimum sufficient statistic.

(b) Let $\bar X = E(X_1) = \theta + 1$, $\bar X$ is the sample mean. Therefore $\hat\theta_{MME} = \bar X - 1$.

(c) The likelihood function is

$$L(\theta) = e^{n\theta}I(\theta \le X_{(1)})e^{-\sum_{i=1}^n x_i}$$

is increasing in $\theta \in (-\infty, X_{(1)}]$; and is equal to zero for $\theta \in (X_{(1)}, \infty)$. Therefore $\hat\theta_{MLE} = X_{(1)}$.

(d) $\hat{\theta}_{MME}$ is unbiased. $Var(X_i) = 1$, so $Var(\hat{\theta}_{MME}) = Var(\bar{X}) = \frac{1}{n}$.So the MSE of $\hat{\theta}_{MME} = \frac{1}{n}$. The density of $X_{(1)}$ is $f_{X_{(1)}}(x|\theta) = ne^{n(\theta-x)}I(x \geq \theta)$, we have $E(X_{(1)}) = \theta + \frac{1}{n}$ and $Var(X_{(1)}) = \frac{1}{n^2}$. Therefore, the MSE of $\hat{\theta}_{MLE}$ is $\frac{2}{n^2}$. If $n = 1$, then $\hat{\theta}_{MME}$ is better; if $n = 2$, two estimators are the same; if $n > 2$, the MLE is better.

(e) Consider $\hat{\theta} = X_{(1)} - \frac{1}{n}$. Its MSE is $\frac{1}{n}$, smaller than both $\hat{\theta}_{MME}$ and $\hat{\theta}_{MLE}$ for $n > 1$. When $n = 1$, $\hat{\theta}$ is better than MLE and is the same as MME.