Friday, August 12, 2008

Mathematical Statistics Preliminary Examination

Statistics Group, Department of Mathematics and Statistics, Auburn University

Name: ____

1. It is a closed-book and in-class exam.
2. One page (letter size, 8.5-by-11in) cheat sheet is allowed.
3. Calculator is allowed. No laptop (or equivalent).
4. Show your work to receive full credits. *Highlight your final answer.*
5. Solve any **five** problems out of the seven problems.
6. Total points are **50**. Each question is worth **10** points.
7. If you work out more than five problems, your score is the sum of five highest points.
8. Time: **180** minutes. (8:30am - 11:30am, Friday, August 12, 2011)

1	2	3	4	5	6	7	Total

1. Suppose that X_1, X_2, and X_3 are mutually independent uniform$(0,1)$. Calculate

$$P(\max\{X_1, X_2\} \geq X_3).$$

You need to calculate it rigorously using density functions to get full points.

2. Let $U_1, \ldots, U_n$ be a random sample from the uniform$(0,1)$ distribution. Let $U_{(s)}$ be the sth order statistic. Determine $E(U_{(s)})$ and $Var(U_{(s)})$.

3. Suppose a population consists of the integers $1, \ldots, N$. Suppose we draw a sample using *binomial sampling* in the following manner: we flip a fair coin N times (once for each of integer in the population) and the integer is included if the coin lands heads. Let W be the sum of the integers in the sample. Compute $E(W)$ and $Var(W)$.

4. Let t_n be a sequence of real-valued estimators such that for $\theta \in \mathbb{R}$, the expectation and variance of t_n exist and $E(t_n) \to \theta$ and $Var(t_n) \to 0$ as $n \to \infty$. Show that t_n is consistent for θ.

5. Suppose that $X_1, \ldots, X_n$ are iid with density

$$f(x, \theta) = \frac{3}{4\theta}\Big(1 - \frac{x^2}{\theta^2}\Big), \qquad |x| \leq \theta \text{ and } \theta > 0.$$

   (a) Find an estimate $\hat{\theta}$ using the method of moments.

   (b) Find the asymptotic distribution of $\hat{\theta}$ as $n \to \infty$. (Hint: use the Central Limit Theorem and delta method)

6. Suppose that $X_1, \ldots, X_n$ are iid $N(\mu, \sigma^2)$ random variables. Find the uniformly minimum variance unbiased estimator (UMVUE) of μ^2/σ^2. Hint: for any function h and a chi-squared random variable χ^2_d, we have $E[h(\chi^2_p)] = p \cdot E[h(\chi^2_{p+2})/\chi^2_{p+2}]$.

7. Suppose that $X_1, \ldots, X_n$ are iid $N(\mu, \sigma^2)$ and $Y_1, \ldots, Y_m$ are iid $N(\theta, \sigma^2)$. Derive the likelihood ratio test for

$$H_0 : \mu = 2\theta + 1, \quad H_a : \mu \neq 2\theta + 1$$

Note that the two samples have different means but the same variance. To get full points, you should try your best to simplify the rejection region.

Solutions

1. The probability is

$$\begin{aligned}&P(\max\{X_1, X_2\} \geq X_3)\\&= P(X_1 \geq X_3) + P(X_1 < X_3, X_2 \geq X_3)\\&= E[P(X_1 \geq X_3|X_3)] + E[P(X_1 < X_3|X_3)P(X_2 \geq X_3|X_3)]\end{aligned}$$

For $X \sim$ uniform(0, 1) and $0 \leq t \leq 1$, we have

$$P(X < t) = t, \qquad P(X \geq t) = 1 - t$$

Therefore,

$$\begin{aligned}P(\max\{X_1, X_2\} \geq X_3) &= E(1 - X_3) + E[X_3(1 - X_3)]\\&= 1 - E(X_3^2) = \frac{2}{3}.\end{aligned}$$

Alternative approach: We may argue that with equal probability X_i is the largest number. Since uniform(0, 1) is a continuous distribution, we do not need to consider the case where some of X_i are equal. Therefore,

$$P(\max\{X_1, X_2\} \geq X_3) = P(X_1 \text{ is largest}) + P(X_2 \text{ is largest}) = \frac{2}{3}.$$

This second solution will not get full points.

2. Recall that the pdf of the sth order statistic for a random sample from a distribution F with density f is

$$n\binom{n-1}{s-1}[F(x)]^{s-1}[1 - F(x)]^{n-s}f(x)\ .$$

For the uniform$(0, 1)$ distribution, this becomes

$$n\binom{n-1}{s-1}x^{s-1}(1 - x)^{n-s}\ , \quad 0 < x < 1\ .$$

This is the density of the beta distribution $B(s, n - s + 1)$. Thus

$$E(U_s) = \frac{s}{n+1} \quad \text{and} \quad Var(U_{(s)}) = \frac{s(n - s + 1)}{(n + 1)^2(n + 2)}\ .$$

3. The sum of the values in the sample is

$$W = 1I_1 + 2I_2 + \cdots + NI_N\ ,$$

where I_j is 1 or 0 as j is included in the sample for $j = 1, \ldots, N$. But

$$E(I_j) = P(I_j = 1) = \frac{1}{2}$$

and

$$Var(I_j) = E(I_j^2) - E^2(I_j) = E(I_j) - E^2(I_j) = \frac{1}{4}\ .$$

Thus

$$E(W) = \frac{1}{2}\sum_{j=1}^{N} j = \frac{N(N+1)}{4}$$

and, since $I_1, \ldots, I_N$ are independent,

$$Var(W) = \frac{1}{4}\sum_{j=1}^{N} j^2 = \frac{N(N+1)(2N+1)}{24}\ .$$

4. By Chebychev's inequality, for any $\epsilon > 0$, we have

$$0 \le P(|t_n - \theta| \ge \epsilon) \le \frac{E(\{t_n - \theta\}^2)}{\epsilon^2} = \frac{V(t_n) + \{E(t_n) - \theta\}^2}{\epsilon^2}\ .$$

Since $E(t_n) \to \theta$ and $V(t_n) \to 0$ as $n \to \infty$, we have $P(|t_n - \theta| \ge \epsilon) \to 0$ as $n \to \infty$.

5. (a) The expectation is

$$E(X) = \int_{-\theta}^{\theta} x f(x) dx = 0$$

The second moment is

$$E(X^2) = \int_{-\theta}^{\theta} x^2 f(x) dx = \frac{1}{5}\theta^2$$

Therefore, we can match the the second moment with its sample counterpart, and

$$\hat{\theta} = \sqrt{\frac{5}{n}\sum_{i=1}^{n} x_i^2}$$

(b) According to the Central Limit Theorem, we have

$$\sqrt{n}\Big(n^{-1}\sum x_i^2 - \mu\Big) \sim N(0, \sigma^2), \quad \text{as } n \to \infty$$

where $\mu = E(X^2)$ and $\sigma^2 = var(X^2)$. Noticing that $\theta = \sqrt{5\mu}$, we have

$$\sqrt{n}(\hat{\theta} - \theta) \sim N(0, \sigma^2(\sqrt{5}/(2\sqrt{\mu}))^2), \quad \text{as } n \to \infty.$$

We can calculate σ^2 as follows,

$$\sigma^2 = E(X^4) - E(X^2)^2 = \frac{3}{35}\theta^4 - \frac{1}{25}\theta^4 = \frac{8}{175}\theta^4$$

Therefore,

$$\sqrt{n}(\hat{\theta} - \theta) \sim N(0, (2/7)\theta^2), \quad \text{as } n \to \infty.$$

6. We know that the sample mean $\bar{X} \sim N(\mu, \sigma^2/n)$ and the sample variance S^2 have $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$. Additionally $\bar{X}$ is independent of S^2. It is also known that $(\bar{x}, s^2)$ is a sufficient and complete statistic. Try to calculate the expectation of $\bar{X}^2/S^2$.

$$\begin{aligned} E(\bar{X}^2/S^2) &= E(\bar{X}^2)E(1/S^2) = (\mu^2 + \sigma^2/n) \cdot E(\frac{n-1}{\sigma^2 \chi^{n-1}}) \\ &= (\mu^2 + \sigma^2/n)\,\frac{n-1}{\sigma^2}\frac{1}{n-3} = \frac{n-1}{n-3}(\frac{\mu^2}{\sigma^2} + \frac{1}{n}) \end{aligned}$$

Therefore,

$$E(\frac{n-3}{n-1}\frac{\bar{X}^2}{S^2} - \frac{1}{n}) = \frac{\mu^2}{\sigma^2}$$

The UMVUE of μ^2/σ^2 is $\frac{n-3}{n-1}\frac{\bar{X}^2}{S^2} - \frac{1}{n}$.

7. The log-likelihood function is

$$\ell(\mu, \theta, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2 - \frac{m}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{m}(y_i - \theta)^2$$

The unrestricted MLEs can be solved from

$$\begin{aligned} \frac{\partial \ell}{\partial \mu} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0 \\ \frac{\partial \ell}{\partial \theta} &= \frac{1}{\sigma^2}\sum_{i=1}^{m}(y_i - \theta) = 0 \\ \frac{\partial \ell}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 - \frac{m}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{m}(y_i - \theta)^2 = 0 \end{aligned}$$

Therefore, the unrestricted MLEs are

$$\hat{\mu} = \bar{x}, \quad \hat{\theta} = \bar{y}, \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{i=1}^{m}(y_i - \bar{y})^2}{m+n}$$

If we let $\mu = 2\theta + 1$, then the likelihood becomes

$$\ell(\theta, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - 2\theta - 1)^2 - \frac{m}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{m}(y_i - \theta)^2$$

The restricted MLEs can be solved from

$$\begin{aligned} \frac{\partial \ell}{\partial \theta} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}2(x_i - 2\theta - 1) + \frac{1}{\sigma^2}\sum_{i=1}^{m}(y_i - \theta) = 0 \\ \frac{\partial \ell}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - 2\theta - 1)^2 - \frac{m}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{m}(y_i - \theta)^2 \end{aligned}$$

Therefore, the restricted MLEs are

$$\tilde{\theta} = \frac{2n\bar{x} + m\bar{y} - 2n}{4n + m}, \quad \tilde{\mu} = 2\tilde{\theta} + 1, \quad \tilde{\sigma}^2 = \frac{\sum_{i=1}^{n}(x_i - \tilde{x})^2 + \sum_{i=1}^{m}(y_i - \tilde{y})^2}{m+n}$$

The likelihood ratio test statistic is

$$\Lambda = \frac{\ell(\tilde{\mu}, \tilde{\theta}, \tilde{\sigma}^2)}{\ell(\hat{\mu}, \hat{\theta}, \hat{\sigma}^2)} = (\hat{\sigma}^2/\tilde{\sigma}^2)^{m+n}$$

Therefore, the rejection region is

$$\hat{\sigma}^2/\tilde{\sigma}^2 < c$$

where c is a constant in $(0, 1)$ and satisfies $P(\hat{\sigma}^2/\tilde{\sigma}^2 < c) = \alpha$ if H_0 is true.