Test 4 – Math 3070 – Fall 2003

Name: ________________________________

SHOW ALL YOUR WORK.

	Value	Score
Problem 1	9	
Problem 2	7	
Problem 3	7	
Problem 4	8	
Problem 5	9	
Problem 6	10	
Total	50	

Problem 1 (9 points). Let X_1, X_2 be a random sample of size 2, from a distribution with mean μ and variance σ^2.

(a) (2 points) Suppose c is a constant satisfying $0 < c < 1$. Is the estimator $cX_1 + (1-c)X_2$ unbiased for the population mean μ? Prove your answer.

(b) (4 points) What is the variance of $cX_1 + (1-c)X_2$?

(c) (3 points) What value of c mimizes the variance of $cX_1 + (1-c)X_2$?

Solution. We have

$$E(cX_1 + (1-c)X_2) = cE(X_1) + (1-c)E(X_2) = c\mu + (1-c)\mu = \mu(c + 1 - c) = \mu\,,$$

so $cX_1 + (1-c)X_2$ is an unbiased estimator. Also,

$$V(cX_1 + (1-c)X_2) = [c^2 + (1-c)^2]\sigma^2 = [2c^2 - 2c + 1]\sigma^2\,.$$

Thus, to minimize variance, we need to minimize $2c^2 - 2c + 1$. Differentiating, setting equal to zero, and solving for c gives $c = 1/2$. The second derivative is 4, so $c = 1/2$ is a minimum. □

Problem 2 (7 points). Let $X_1, \ldots, X_n$ be a random sample from the probability distribution with probability density function

$$f(x) = \begin{cases} \frac{1}{\lambda} e^{-\frac{1}{\lambda}x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0\,. \end{cases}$$

(a) (5 points) Derive the maximum likelihood estimator of λ. *Hint:* Recall that the joint probability density for $(X_1, \ldots, X_n)$ is $f(x_1)f(x_2)\cdots f(x_n)$.

(b) (2 points) Is this estimator unbiased? Prove your answer. *Hint:* $\int_0^\infty x \frac{1}{\lambda} e^{-\frac{1}{\lambda}x} dx = \lambda$.

Solution. The joint density of $X_1, \ldots, X_n$ is

$$f(x_1, \ldots, x_n\,;\,\lambda) = \prod_{i=1}^n \frac{1}{\lambda} e^{-\frac{1}{\lambda}x_i} = \lambda^{-n} e^{-\frac{1}{\lambda}\sum_{i=1}^n x_i}\,.$$

We want to maximize this as a function of λ. Taking logs gives

$$\log f(x_1, \ldots, x_n\,;\,\lambda) = -n \log \lambda - \frac{1}{\lambda} \sum_{i=1}^n x_i\,.$$

Differentiating and setting equal to 0 gives

$$\begin{aligned} 0 &= -\frac{n}{\lambda} + \frac{1}{\lambda^2} \sum_{i=1}^n x_i \\ &= -\lambda + \frac{1}{n} \sum_{i=1}^n x_i\,. \end{aligned}$$

Solving yields $\lambda = \bar{x}$. We must check that this is indeed a global maximum. We use the second derivative test:

$$\begin{aligned} \frac{d^2}{d\lambda^2} \log f(x_1, \ldots, x_n\,;\,\lambda) &= \frac{n}{\lambda^2} - \frac{2}{\lambda^3} \sum_{i=1}^n x_i \\ \frac{d^2}{d\lambda^2} \log f(x_1, \ldots, x_n\,;\,\lambda)\Big|_{\lambda=\bar{x}} &= \frac{n}{\bar{x}^2} - \frac{2}{\bar{x}^3} n\bar{x} \\ &= -\frac{n}{\bar{x}^2} < 0\,. \end{aligned}$$

Thus $\bar{x}$ is a local maximum. Checking that the function tends to $-\infty$ as $\lambda \to 0$ and $\lambda \to \infty$ shows that $\bar{x}$ is a global maximum. We conclude that the M.L.E. is $\hat{\lambda} = \bar{X}$.

We have $E(\bar{X}) = E(X_1) = \lambda$, so this estimator is unbiased. $\square$

Problem 3 (7 points). Suppose that $X_1, \ldots, X_{10}$ is a random sample from a Normal distribution. What is the probability that the random interval $(\bar{X} - 0.58S, \bar{X} + 0.58S)$ contains μ.

Solution. We know that $\bar{X} \pm S t_{\frac{\alpha}{2},9}/\sqrt{10}$ is a $(1-\alpha) \times 100\%$ confidence interval. so solving

$$0.58S = S\frac{t_{\frac{\alpha}{2},9}}{\sqrt{10}}$$

for $t_{\frac{\alpha}{2},10}$ yields

$$t_{\frac{\alpha}{2},9} = 0.58\sqrt{10} = 1.834\,.$$

But the area to the right of 1.834 on the t-curve with 9 degrees of freedom is 0.05. Thus $\alpha/2 = 0.05$ and $\alpha = 0.1$. Thus the coverage probability is 0.9.

□

Problem 4 (8 points). Suppose you want to estimate a proportion p in a large population, based on a sample of size n and the sample proportion $\hat{p}$. Assume that $np > 50$ and $n(1-p) > 50$.

(a) (2 points) What is the formula for an approximate 95% confidence interval for p?

(b) (6 points) What is the length of the interval you gave above? For what value of $\hat{p}$ is this interval the largest.

You would like this interval to be at most length 0.01. How big must n be so that this interval is always, regardless of the value of $\hat{p}$, no more than length 0.01?

Solution. We use the formula for an approximate 95% confidence interval

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

The length of the interval is then

$$3.92\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\,.$$

We note that $\hat{p}(1-\hat{p})$ is maximized at $p = 0.5$, so it is never more than 0.25. Thus the length of the interval is at most

$$1.96\sqrt{n}\,.$$

Solving $1.96n^{-1/2} = 0.01$ for n yields $n = 38416$. That is for $n \geq 38416$, the interval length is guaranteed to be at most $1.96/\sqrt{38416} = 0.01$. □

Problem 5 (9 points). Let $X_1, \ldots, X_n$ be a random sample from a Uniform $[0, \theta]$ distribution:

$$f(x) = \begin{cases} \frac{1}{\theta} & \text{if } 0 \leq x \leq \theta\,, \\ 0 & \text{if } x < 0 \text{ or } x > \theta\,. \end{cases}$$

(a) (1 point) Is $\bar{X}$ an unbiased estimator of θ?

(b) (2 points) If not, find an unbiased estimator of θ based on the statistic $\bar{X}$.

(c) (2 points) Compute the variance of the unbiased estimator you found in (b). *Hint:* The variance of a Uniform(A, B) random variable is $(B - A)^2/12$.

(d) (3 points) Consider the statistic

$$\hat{\theta} = \max\{X_1, X_2, \ldots, X_n\}\,.$$

We computed in class that $\hat{\theta}$ has probability density function

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

We furthermore saw that the statistic $\left(\frac{n+1}{n}\right)\hat{\theta}$ is unbiased for θ. Compute the variance of $\left(\frac{n+1}{n}\right)\hat{\theta}$.

(e) (1 point) Which is a better estimator, your answer in (b) or $\left(\frac{n+1}{n}\right)\hat{\theta}$? Justify your answer.

Solution. We have $E(\bar{X}) = E(X_1) = \theta/2$. Thus $\bar{X}$ is not unbiased. But the estimator $2\bar{X}$ is unbiased.

We have

$$V(2\bar{X}) = 4V(\bar{X}) = 4\frac{\sigma^2}{n} = 4\frac{\theta^2/12}{n} = \frac{3\theta^2}{n}\,.$$

We have

$$E(\hat{\theta}^2) = \int_0^\theta t^2 \frac{nt^{n-1}}{\theta^n}\,dt = \int_0^\theta \frac{nt^{n+1}}{\theta^n}\,dt = \frac{nt^{n+2}}{(n+2)\theta^n}\Big|_0^\theta = \frac{n}{n+2}\theta^2$$

Thus

$$\begin{aligned} V(\hat{\theta}) &= E(\hat{\theta}^2) - \left[E(\hat{\theta})\right]^2 \\ &= \frac{n}{n+2}\theta^2 - \frac{n^2}{(n+1)^2}\theta^2 \\ &= \frac{n}{(n+2)(n+1)^2}\theta^2 \,. \end{aligned}$$

Thus

$$\begin{aligned} V\left(\frac{n+1}{n}\hat{\theta}\right) &= \left(\frac{(n+1)^2}{n^2}\right) V(\hat{\theta}) \\ &= \left(\frac{(n+1)^2}{n^2}\right) \frac{n}{(n+2)(n+1)^2}\theta^2 \\ &= \frac{1}{n(n+2)}\theta^2 \,. \end{aligned}$$

□

Problem 6 (10 points). Say if each of the following answers is True or False. **You must include a short justification for each of your answers. Credit will not be given without a justification.** Each is worth 2 points.

(a) $\bar{X}$ is always an unbiased estimator of the population mean.

(b) Suppose I take a sample of size 10 from a Normal distribution with variance 1 and find that $\bar{X} = 11$. Then the probability that the confidence interval $(11 - 1.96/\sqrt{10}, 11 + 1.96/\sqrt{10})$ contains the population mean μ is 0.95.

(c) For any pair of random variables X and Y, and constants a and b,

$$V(aX + bY) = a^2V(X) + b^2V(Y).$$

(d) Let $\hat{p}$ be a sample proportion from a sample of size n. Then $V(\hat{p}) = np(1-p)$, where p is the population proportion.

(e) If X_1 and X_2 are independent and identically distributed, then always $X_1 + X_2$ is approximately Normally distributed.

Solution. (a) Yes, $E(\bar{X}) = \mu$ always, where μ is the population mean.

(b) No. The interval given is a fixed interval which either contains μ or doesn't contain it. The probability statement only applies to the random interval before the sample is collected.

(c) No. This only holds for independent X and Y.

(d) No. $V(\hat{p}) = p(1-p)/n$.

(e) No. $X_1 + X_2 + \cdots + X_n$ is only approximately normal if n is large, or the X_i are close to normal themselves.

□