Ma 494 — Theoretical Statistics

Final — May 10, 2010

Prof. Sawyer — Washington University

Seven problems on two pages. Closed book and closed notes. Two $8\frac{1}{2} \times 11$ sheets of paper with notes on both sides and a calculator are allowed. Parts of a problem may be weighted according to difficulty.

1. Let $X = (X_1, \ldots, X_n)$ be independent with density $f(x, \theta) = \theta e^{-\theta e^x} e^x$ for $\theta > 0$ and $-\infty < x < \infty$. (This is a special case of what is called the extreme-value distribution.) Find a real-valued function $S(X) = S(X_1, \ldots, X_n)$ such that $S(X)$ is a sufficient statistic for θ.

2. Let $Y_1, Y_2, \ldots, Y_n$ for $n = 40$ be independent random variables with a geometric distribution. That is, $P(Y = x) = (1-p)^{x-1}p$ for $x = 1, 2, \ldots$ where $0 < p < 1$. The values of the entire sequence are not available, but we do have the table

Value :	1	2	3	4	5	≥ 6	
Counts :	8	7	4	6	5	10	Sum : 40

Find the likelihood $L(p, K_1, \ldots, K_5, R_6)$ in terms of the counts $K_1, \ldots, K_5, R_6$ in the second row of the table and find the MLE $\widehat{p}$ of p.

3. Let $X = (X_1, X_2, \ldots, X_n)$ be an independent sample from a normal distributions $N(\mu, \sigma^2)$. Let $S_1(X) = \sum_{j=1}^n X_j$ and $S_2(X) = \sum_{j=1}^n X_j^2$. Prove that $T(X) = \big(S_1(X), S_2(X)\big)$ is a sufficient statistic for $\theta = (\mu, \sigma^2)$. (That is, the likelihood $L(\theta, X)$ can be written $g\big(T(X), \theta\big)h(X)$ where $h(X)$ depends only on X.)

4. Let $Y = (Y_1, \ldots, Y_n)$ be independent with density $f(y, \theta) = \theta/(y+1)^{\theta+1}$ for $y > 0$ and $\theta > 0$.

(i) Find the MLE $\widehat{\theta}$ of θ in terms of Y and the Fisher information for one observation of Y.

(ii) Use the central limit theorem for MLEs ("the asymptotic efficiency of an MLE") to find an approximate 95% confidence interval for θ about $\widehat{\theta}$ in terms of Y and n.

5. (i) Let $X = (X_1, X_2, \ldots, X_n)$ be an independent sample that can have density either $f(x)$ or $g(x)$. Let H_1 be the hypothesis that the density is $g(x)$ and H_0 that it is $f(x)$. A test procedure can be thought of as a subset $\mathcal{C} \subseteq R^n$ with the convention that we accept H_1 if $X \in \mathcal{C}$ and accept H_0 if $X \notin \mathcal{C}$. What is the power of this test? What is its level of significance? How do these depend on $f(x)$ and $g(x)$?

(ii) State the Neyman-Pearson Lemma. What does it say about finding powerful tests?

6. Let $X = (X_1, X_2, \ldots, X_n)$ be an independent sample from a density $f(x, \theta)$ with an unknown parameter θ. To estimate θ, we have a choice between an unbiased estimator $T_U(X)$ and a biased estimator $T_B(X)$ with $\mathrm{Var}(T_B) = 0.80\,\mathrm{Var}(T_U)$. How big can the bias of T_B (that is, $|E_\theta(T_B) - \theta|$) be as a fraction of the standard deviation of T_U and still have the quadratic-loss risk $R(\theta, T_B) < R(\theta, T_U)$? (*Hint*: Recall $R(\theta, T) = E_\theta\big((T(X) - \theta)^2\big)$.)

7. Let $X = (X_1, X_2, \ldots, X_n)$ be an independent sample that has a gamma distribution gamma$(r, 1)$ for some $r > 1$. (That is, has density $\big(1/\Gamma(r)\big)x^{r-1}e^{-x}$ for $x > 0$.)

(i) Find the most powerful test at level of significance α for $H_0 : r = 1$ against $H_1 : r = 3$. (*Hint*: By definition, a test is a subset $\mathcal{C} \subseteq R^n$ with the convention that we reject H_0 and accept H_1 if $X \in \mathcal{C}$. The test $\mathcal{C}$ should be defined in terms of a free parameter. Say how this free parameter is related to α.)

(ii) Are the tests that you found in the part UMP (uniformly most powerful) for all $r > 1$? Why?