Statistics (MA & Ph. D. Version)

Instructions to the Student

a. Answer all six questions. Each will be graded from 0 to 10.

b. Use a different booklet for each question. Write the problem number and your code number (NOT YOUR NAME) on the outside cover.

c. Keep scratch work on separate pages in the same booklet.

d. If you use a "well known" theorem in your solution to any problem, it is your responsibility to make clear which theorem you are using and to justify its use.

1. Let $X_1, \ldots, X_n$ be a sample from the $N(\mu, \sigma^2)$ density. Give a lower bound $L_n$ which is valid for all sample-sizes $n$ for the variance of unbiased estimators of $\theta = \mu/\sigma$, and find an estimator $\hat{\theta}_n$ which has finite variance for $n \geq 5$ and has variance satisfying: $\text{Var}(\hat{\theta}_n)/L_n \to 1$ as $n \to \infty$.

2. Based on a sample $M_1, M_2, \ldots, M_n$ from the Geometric($p$) distribution (with probability mass function $p_M(k) = (1-p)^{k-1}p, \quad k = 1, 2, \ldots,$), find and justify the Uniformly Minimum Variance Unbiased Estimator of the parameter $p(1-p)$.

Hint: recall that the sum $S$ of $n$ independent Geometric($p$) random variables $M_i, \quad 1 \leq i \leq n$, can be interpreted as the number of Bernoulli($p$) trials needed to accumulate $n$ occurrences of outcomes 1, and has the Negative Binomial distribution $S \sim \text{NegBin}(n, p)$, with probability mass function $p_S(k) = \binom{k}{n} p^n (1-p)^{k-n}$ for $k = n, n+1, \ldots$

3. Suppose that iid values $Y_i \sim \text{Gamma}(\alpha, \alpha \lambda)$ are observed for $i = 1, \ldots, n$, where $\alpha \geq 1$ and $\lambda > 0$ are unknown parameters, and where $\text{Gamma}(\alpha, \beta)$ has density $\beta^\alpha y^{\alpha-1} e^{-\beta y}/\Gamma(\alpha)$ for $y > 0$. (Note that some textbooks call this density $\text{Gamma}(\alpha, 1/\beta)$.)
(a). Let $\hat{\lambda}$ denote the MLE of $\lambda$ calculated as though $\alpha = 1$ were known. Find the large-sample distribution of $\sqrt{n}(\hat{\lambda} - \lambda)$ and use it to provide a 2-sided approximate 95% Confidence Interval for $\lambda$ which is asymptotically valid (i.e. has coverage converging to 0.95) as $n \to \infty$ no matter what the true value of $\alpha \geq 1$ is.

(b). How does your answer (regarding the large-sample distribution and the form of large-sample valid Confidence Interval) change if $\hat{\lambda}$ is replaced by the MLE of $\lambda$ calculated as though $\alpha = 1.1$ were known? Explain.

4. Consider i.i.d. data $X_1, X_2, \ldots, X_n$ with the double-exponential density given for all real arguments $x$ by $f(x, \theta) = (\theta/2) \exp(-\theta |x|)$.

(a). Suppose that $\theta$ has the prior density $0.1 \exp(-\theta/10)$ for $\theta > 0$. Give the form of a Bayesian optimal test procedure, i.e., a rule depending on the data $(X_1, \ldots, X_n)$ choosing one of the two actions 1 = Reject and 0 = Accept $H_0$ optimally for the Loss function $L(\theta, a) = I_{\{\theta \leq 1, a=1\}} + 2 I_{\{\theta>1, a=0\}}$.

(b). Now consider the test procedure you found in (a) from a frequentist point of view. Find its size, and explain with justification whether it has any UMP properties for testing $H_0 : \theta \leq 1$ versus $H_1 : \theta > 1$.

5. Let $Z_i$ for $1 \leq i \leq n$ be a sample from the $N(\alpha p, b p (1 - p))$ density, where $\alpha > 0$, $b > 0$ are known but $p \in (0, 1)$ is an unknown parameter.

(a). Find and justify a two-dimensional minimal sufficient statistic for the unknown parameter $p$.

(b). Show that the sufficient statistic you found in (a) is not complete.

6. Two samples of data are observed simultaneously: $X_i \sim N(\mu_1, \sigma^2)$ for $1 \leq i \leq n$ and $Y_j \sim N(\mu_2, \sigma^2)$ for $1 \leq j \leq n$, where $\mu_1, \mu_2,$ and $\sigma^2$ are unknown parameters. Based on these two-sample data, find as explicitly as you can the size 0.10 Likelihood Ratio Test of $H_0 : \mu_2 = 2\mu_1$ versus $H_1 : \mu_2 \neq 2\mu_1$: first find the form of the rejection region involving an undetermined constant, and then express the rejection region in a form so that you can evaluate the constant in terms of a quantile of a well-known distribution.