Indian Statistical Institute
Bangalore Centre
Mid-Semester Examination
B. Math. Second Year

Statistical Methods II 02.05.03

Answer as much as you can. The maximum you can score is 100
Time :- 3 hours

1. (a) For a testing of hypothesis problem define the following terms. (i) Size of a test, (ii) rejection region, (iii) Power of a test and power function and (iv) p-value.

   (b) Suppose X is a random variable with the density function $f(\theta, x)$ and you want to test $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$. Let $\rho(x) = f(\theta_1, x)/f(\theta_0, x)$. Prove the following statements.

   (i) $\exists$ a test procedure with a given size which rejects H_0 if " $\rho(x)$ is big enough". Write the statement in quotation mark explicitly in terms of the size.

   (ii) Such a test is most powerful among all tests of the same size.

   (iii) The power of this test is greater than its size.

   (c) Consider the testing problem $H_0 : \theta = \theta_0$ against $H_1 : \theta > \theta_0$ for a family of densities with monotone likelihood ratio in a statistics $T(x)$. Obtain a UMP test.

   (d) An owner of a telephone booth wants to see if it is worth keeping it open after 11 P.M. upto 12. It was found that at least 8 calls are necessary to meet the operating cost. The owner observed the no. of calls for 10 randomly selected days. Fearing that this data might be insufficient, she observed for 15 more days, selected at random. She also felt that the probability of type 1 error should not exceed 0.05. Both data sets are given below. Suggest what action to be taken based on

   (i) the first data set of size 10 and

   (ii) the pooled data set (of size 25) and using an appropriate approximation.

   Compare the conclusions and make comment.

   No. of calls for the first 10 days : 3,5,9,4,3,8,10,12,5,2. Those for the next 15 days : 2,3,7,9,6, 10,5,8,12,11, 10,13,9,8,4.

   [(3+2+4+4) + (6x2 + 4) + (7) + (12) = 48]

2. The meausurements of the diameter of a ball bearing measured by two slide callipers, C_1 and C_2, found to be different. To see whether the two callipers differ in the accuracy of their measurements, the diameter of a ball was measured m times by C_1 and n times by C_2. Derive a test procedure. Assume (i) the two callipers have the same precision and (ii) all the measurements were made by one person. [8]

3. Consider a family of densities $\{f(\theta, x)\}$with monotone likelihood ratio in a statistics $T(x)$ which is a real-valued random variable with continuous distribution.

   (a) Show that $F_\theta(t)$, the distribution function of $T(x)$, is $\downarrow$ in θ.

   (b) Consider the testing problem $H_0 : \theta \geq \tau$ against $H_1 : \theta < \tau$. Let $k(\tau)$ denote the point of separation of the rejection region from the acceptance region of a UMP test. Prove that $k(\tau)$ is $\uparrow$ in τ.

   (c) Define uniformly most accurate (UMA) upper confidence bound (UCB) of a parameter. Descibe how one can derive an UMA UCB for θ.

   (d) Explain how to find a UMA UCB for the average life of bulbs made by a certain factory.

[7 + 6 x 3 = 25]

4. (a) Let $\mathcal{U}$ denote the class of all unbiased estimators of $q(\theta)$ and $\mathcal{Z}$ denote the class of all statistics with expectation zero. Prove that $U \in \mathcal{U}$ has minimum variance in $\mathcal{U}$ iff $E(UZ) = 0 \; \forall \; Z \in \mathcal{Z}$.

   (b) Suppose $S(X)$ is sufficient for θ. Let $T(X)$ be an estimator for $q(\theta)$. Let

$$T_1(X) = E[T(X) \mid S(X)].$$

   Show that the mean square error of $T_1(X)$ is smaller than or equal to that of $T(X)$.

   (c) Suppose $S(X)$ is complete sufficient for θ. Let $S_1(X)$ be a function of $S(X)$ which is unbiased for $q(\theta)$. Then show that $S_1(X)$ is UMVUE for $q(\theta)$.

[6+4+5 = 15]

5. Consider a vector of random variables X with a $k \times 1$ vector of parameters $(\theta_1, \theta_2, \cdots \theta_k)^T$.

   (a) Consider a vector $L(r \times 1)$ of statistics, $r \leq k$. Let $E(L) = q(\theta)$ and $cov(L) = V$. Suppose $T = (T_1, T_2, \cdots T_K)^T$ is jointly sufficient for θ. Prove that $\exists \; S = (S_1, S_2, \cdots S_r)^T$ satisfying
   (i)$E(S) = q(\theta)$ and
   (ii) if $U = cov(S)$, then $V - U$ is n.n.d.
   [Hint : use Q 4(b).]

   (b) Assume (i) $\partial log f(\theta, x)/\partial\theta_i$ exists and is finite for each i and for all x and (ii) for any statistic $S(X)$ with finite expectation, it is possible to interchange differentiation and integration in the expression for $E(S(X)$.
   Let $H = (H_1, H_2, \cdots H_k)^T$, where $H_i = \partial log f(\theta, X)/\partial\theta_i, i = 1, 2, \cdots k$ and $G = ((g_{i,j}))$, where $g_{i,j} = \partial q_i(\theta)/\partial\theta_j$. Show that $E(H) = 0$ and $Cov(S, H) = G$

(c) Define

$$\mathcal{I}(\theta) = ((E[\partial^2 log f(\theta, X)/\partial\theta_i \partial\theta_j])).$$

Assuming $\mathcal{I}(\theta) = Cov(H)$ and using (b) and (c) show that $V - G(\mathcal{I}(\theta))^{-1}G^T$ is nonnegative definite (n.n.d.).

[Hint : $\Sigma = Cov((S, H)^T)$ is n.n.d, so that $M\Sigma M^T$ is n.n.d for any matrix M.]

(d) In a small town of Karnataka, the residents speak in one of the three languages : Kannada, Tamil and Marathi. It was decided to estimate the proportion of residents speaking in each of the languages. A random sample of n residents was drawn. and data on the language spoken was obtained. Show that the 'obvious' estimator for the proportions is very good with regard to the 'lower bound' of the covariance matrix.

[(3x2) + 6 x 2 + 8 = 26]

Indian Statistical Institute
Bangalore Centre
Back-paper Examination
B. Math. Second Year
Statistical Methods II
Answer as much as you can. The maximum you can score is 100
Time :- 3 hours

1. Suppose $X^{(n)}$ is a vector of i.i.d random variables $X_1, X_2, \cdots X_n$ and θ is the parameter of the common distribution.
   (a) Define an unbiased estimator for a parametric function $q(\theta)$.
   (b) Prove or disprove (with a counter example) the following statement. If $T_n = T(X^{(n)})$ is unbiased for $\phi = q(\theta)$, and g is a continuous function, then $g(T_n)$ is unbiased for $g(\phi)$.
   (c) Consider the result obtained by replacing "continuous" by "linear" in Q.1 (b). Is it true ? Prove or give counter example.
   (d) Suppose X_i follows $U(0, \theta)$. Find linear functions $T_1(X)$ of $\bar{X}$ and $T_2(X)$ of $X_{(n)}$ [the largest order statistic] which are unbiased for θ. Suppose in an experiment the observed values of T_1 is 5.3 and that of T_2 is 6.4. What will be your estimate of θ ?

   [2+4+3+(6+3)=15]

2. (a) Define sufficient statistics and minimal sufficient statistics.
   (b) Suppose $X_1, X_2, \cdots X_n$ are i.i.d random variables with common density $f(\theta, x)$. Without using any result, show that $T(X)$ is sufficient for θ in the following examples.
   (i) $f(\theta, x) = (1/\theta) I_{(0,\theta)}(x)$; $T(X) = X_{(n)}$.
   (ii)$f(\theta, x) = (1/\sqrt{(2\pi)}) exp[-(1/2)\ (x-\theta)^2]$; $T(X) = \bar{X}$.

   [4 + 5 x 2 = 14]

3. (a) Define maximum likelihood estimator (M.L.E.) for a parametric function $q(\theta)$. Show that if $T(X)$ is an M.L.E. of θ and g is 1-1, then $g(T(X))$ is also M.L.E. of $g(\theta)$.
   (b) Suppose $X_1, X_2, \cdots X_n$ are i.i.d random variables following Normal distribution with mean μ and variance ρ. Find M.L.E.s of μ and ρ. Also find an M.L.E. of the standard deviation of X_i.

   [(2+5) + 4 = 11]

4. (a) Define Fisher information. State the information inequality for the one parameter case. Under what conditions does this hold ?

   (b) Let X follows uniform distribution on $(0, \theta)$. Show that

   (i) $\log f(x, \theta)$ is differentiable for all $\theta > x$.

   (ii) $Var\ \left(\frac{\partial}{\partial \theta} \log p(X, \theta)\right) = 0$

   (iii) $\exists$ an unbiased estimator for θ which has finite variance. [Hint : Try cX, c is a constant]

   (c) In the light of the above facts, prove or disprove the following statement. "Information inequality always hold".

[(2 + 3 + 3) + (3 + 4 + 3) + 5 = 23]

5. (a) For a testing of hypothesis problem define the following terms. (i) Size of a test, (ii) rejection region, (iii) Power of a test and power function and (iv) p-value.

   (b) Suppose X is a random variable with the density function $f(\theta, x)$ and you want to test $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$. Let $\rho(x) = f(\theta_1, x)/f(\theta_0, x)$. Prove the following statements.

   (i) $\exists$ a test procedure with a given size which rejects H_0 if " $\rho(x)$ is big enough". Write the statement in " " explicitly in terms of the size.

   (ii) Such a test is most powerful among all tests of the same size.

   (c) Consider the testing problem $H_0 : \theta = \theta_0$ against $H_1 : \theta > \theta_0$ for a family of densities with monotone likelihood ratio in a statistics $T(x)$. Obtain a UMP test.

   (d) The average life of bulbs made by a factory is believed to be 30 months. A consumer suspects that it has gone down. Suggest an experiment to verify the matter. State clearly your assumptions.

[(3 x 4) + (6 x 2) + 7 + 8 = 39]

6. (a) Define uniformly most accurate (UMA) upper confidence bound (UCB) of a parameter. Descibe how one can derive an UMA UCB for θ. State clearly all the results you use.

   (b) Let S denote the no. of defective items among n items selected at random. Show how you can find an upper confidence bound (UCB) of level $1 - \alpha$ for the average no. of defective items in the population, assuming (i) n small and (ii) n large.

[6 + 8 = 14]