Indian Statistical Institute, Bangalore
B. Math (II)
Second Semester 2010-11
Semester Examination : Statistics (II)

Date: 04-05-2011 **Maximum Score 80** **Duration: 3 Hours**

1. There are 8 different eateries in the neighborhood of ISI. An eatery is open on any particular day with probability θ, $0 < \theta < 1$. For reasons of proximity and convenience etc., Partha visits either eatery 1 or eatery 2. Partha is interested in knowing $i)$ $\psi_1(\theta)$, the probability that either eatery 1 is open or eatery 2 is open on any given day and $ii)$ $\psi_2(\theta)$, the probability that exactly one of the eateries 1 and 2 is open on a given day. Let $X_i = 1(0)$ if the ith eatery is open (closed) on a given day, $1 \leq i \leq 8$. Let $X_1,\ X_2, ..., X_8$ be the random sample taken on some day indicating whether various of the eateries were open or not.

   (a) State clearly the assumptions you make.
   (b) Find $\psi_1(\theta)$ and $\psi_2(\theta)$.
   (c) Show that $T = \sum_{i=1}^{8} X_i$ is a minimal sufficient statistic for θ.
   (d) Is $T = \sum_{i=1}^{8} X_i$ complete as well? Substantiate.
   (e) Find Fisher information $I(\theta)$ contained in the sample $X_1,\ X_2, ..., X_8$ about θ.
   (f) Find an unbiased estimator for $\psi_2(\theta)$. Hence or otherwise obtain $UMVUE$ for $\psi_2(\theta)$.

   $[2 + 2 + 3 + 4 + 3 + 6 = 20]$

2. A drilling machine is used to make holes in metal sheets using shafts of different diametric specifications. Let θ be the mean diameter, measured in *mm,* of the holes drilled using one such shaft. However, the mean diametric specification θ is unknown. Let $X_1,\ X_2, ..., X_n$ denote the diameters of n holes all drilled using the given shaft. The variability σ_0^2, in the diameters of holes drilled, is an indicator of the quality of the drilling machine. Based on the prolonged use of the drilling machine we assume that σ_0^2 is known . Stating clearly the assumptions you make derive *likelihood ratio test (LRT) at level of significance* $\alpha = 0.05$, for testing the hypothesis

$$H_0 : \theta \leq 2 \ versus \ H_1 : \theta > 2.$$

   How would you report the $p - value$?

   $[2 + 10 + 2 = 14]$

3. Leaves were randomly collected from *wax-leaf ligustrum* grown in shade. Another random sample was collected from *wax-leaf ligustrum* grown in full sun. The thickness in *micrometers* of the *palisade layer* was recorded for all selected leaves of either type. Thicknesses of 7 *sun* leaves were reported as:

$$150, 100, 210, 300, 200, 210 \text{ and } 300.$$

   Thicknesses of 7 *shade* leaves were reported as:

$$120, 125, 160, 130, 200, 170 \text{ and } 200.$$

(a) State clearly the assumptions you make.

(b) Do the two types of leaves differ in thickness? Take $\alpha = 0.05$.

(c) Report the *p-value.*

(d) Find 90% confidence interval for (mean thickness of *sun* leaves $-$ mean thickness of *shade* leaves).

(e) Was it crucial to have the two samples of same size for your test?

$[2 + 8 + 2 + 4 + 2 = 18]$

4. Let $X_1,\ X_2, \cdots, X_n$ be a random sample from $N(\theta, \sigma^2)$, both $\theta \in I\!R$ and $\sigma^2 > 0$ are unknown. Let $\overline{X}_n = \frac{1}{n}\Sigma_{i=1}^n X_i$, $S_n^2 = \frac{1}{n-1}\Sigma_{i=1}^n \left(X_i - \overline{X}_n\right)^2$ and $T_n = \frac{\sqrt{n}\left(\overline{X}_n - \theta_0\right)}{S_n}$, $\theta_0 \in I\!R$ being a known value.

   (a) Obtain $\widehat{\sigma}_n^2$, *method of moments estimator* for σ^2.

   (b) Show that $\widehat{\sigma}_n^2$ is *asymptotically unbiased* for σ^2.

   (c) Show that S_n^2 is *consistent* for σ^2.

   (d) Is $\widehat{\sigma}_n^2$ *consistent* for σ^2?

   (e) Show that T_n, the likelihood ratio test statistic for testing $H_0 : \theta = \theta_0$ *versus* $H_1 : \theta \neq \theta_0$, is *asymptotically normal* under H_0.

   (f) Obtain Fisher information contained in the sample about σ^2 under H_0.

   (g) Show that $Var\left(S_n^2\right)$ does not attain *CRLB,* however, S_n^2 is *asymptotically efficient.*

$[2 + 3 + 4 + 2 + 4 + 4 + (4 + 2) = 25]$

5. Suppose we observe $X_1,\ X_2, ...,\ X_n$ and $Y_1,\ Y_2, ...,\ Y_n$, all mutually independent random variables such that $Y_i \sim Poisson\,(\beta\lambda_i)$ and $X_i \sim Poisson\,(\lambda_i)$; $\beta > 0$ and $\lambda_i > 0$; $1 \leq i \leq n$. This would model, for instance, the incidence of a disease, Y_i, where underlying rate is a function of on overall effect β and an additional factor λ_i. For example, λ_i could be a measure of population density in locality i, or perhaps health status of the population in the locality i. We do not know λ_i but get information on it through X_i.

   (a) Obtain *maximum likelihood estimators* (*mle*) based on the complete data for β and λ_i; $1 \leq i \leq n$.

   (b) If x_1 is missing calculate the *expected complete-data loglikelihood* and show that the *Expectation-Maximization* (*EM*) sequence is given by

$$\widehat{\beta}^{(r+1)} = \frac{\Sigma_{i=1}^n y_i}{\widehat{\lambda}_1^{(r)} + \Sigma_{i=2}^n x_i},\ \widehat{\lambda}_1^{(r+1)} = \frac{\widehat{\lambda}_1^{(r)} + y_1}{\widehat{\beta}^{(r+1)} + 1},\ \text{and}\ \widehat{\lambda}_j^{(r+1)} = \frac{x_j + y_j}{\widehat{\beta}^{(r+1)} + 1},\ 2 \leq j \leq n,$$

   r being a positive integer.

$[4 + 9 = 13]$