Indian Statistical Institute, Bangalore
B. Math (II)
Second Semester 2017-18
Backpaper Examination : Statistics (II)

Date: 14-06-2018 **Maximum Score 40** **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)* $\tau_1(\theta)$, the probability that either eatery 1 is open or eatery 2 is open on any given day and *ii)* $\tau_2(\theta)$, 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 $\tau_1(\theta)$ and $\tau_2(\theta)$.

   (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 $\tau_2(\theta)$. Hence or otherwise obtain *Uniformly Minimum Variance Unbiased Estimator* (*UMVUE*) for $\tau_2(\theta)$.

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

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. This amusing classical example is from von Bortkiewicz (1898). The number of fatalities that resulted from being kicked by a horse was recorded for 10 corps of Prussian cavalry over a period of 20 years, giving 200 corps-years worth of data. These data are displayed in the following table. The first column of the table gives the number of deaths per year, ranging from 0 to 4. The second column lists how many times that number of deaths was observed.

Thus, for example, in 65 of the 200 corps-years, there was one death.

No. of Deaths per Year	Observed Frequency
0	109
1	65
2	22
3	3
4	1

Carry out χ^2 *goodness of fit test* to test the hypothesis, at level of significance $\alpha = 0.05$, that the data come from *Poisson distribution.* Also report the $p - value$. [12]

4. Let X_n, Y_n, $n \geq 1$ be sequences of random variables and X be a random variable, all defined on the same probability space, such that $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} c$, where c is a finite constant. Prove that $X_n + Y_n \xrightarrow{d} X + c$. [12]