Indian Statistical Institute, Bangalore
B. Math (II)
First semester 2010-2011
Semester Examination : Statistics (I)

Date: 01-12-2010 **Maximum Score 80** **Duration: 3 Hours**

1. Let $X_1,\ X_2, ..., X_n$ be *independently and identically distributed* (*iid*) with one of the two *probability density function (pdfs).* If $\theta = 0$, then

$$\begin{aligned} f(x|\theta) &= 1 \qquad if\ 0 < x < \theta + 1 \\ &= 0 \qquad otherwise \end{aligned}$$

while if $\theta = 1$, then

$$\begin{aligned} f(x|\theta) &= \frac{1}{2\sqrt{x}} \qquad if\ 0 < x < \theta \\ &= 0 \qquad otherwise \end{aligned}$$

Find *maximum likelihood estimators (mle)* for θ.

[8]

2. Let us consider *logistic distribution* with parameters $\mu \in (-\infty, \infty)$ *and* $\beta > 0$ having *probability density function (pdf)* given by

$$f(x|\mu, \beta) = \frac{\exp\left(-\frac{(x-\mu)}{\beta}\right)}{\beta\left[1 + \exp\left(-\frac{(x-\mu)}{\beta}\right)\right]^2};\ x \in (-\infty, \infty). \tag{1}$$

   (a) Check that $f(x|\mu, \beta)$ in (1) is indeed a *pdf.*
   (b) Find mean, median and mode of the *logistic distribution* (1) .
   (c) Find a *method of moments (mom)* estimator for the parameter μ of the *logistic distribution* (1) .
   (d) If U is uniform on $(0, 1)$ then find the distribution of $W = \log \frac{U}{1-U}$.
   (e) Suppose we can draw observations from uniform distribution on $(0, 1)$. *Explain*, using the answer to (d) , how you would draw observations from the *logistic distribution* (1) .

[3 + 10 + 2 + 5 + 6 = 26]

3. Let $X_1,\ X_2, ..., X_n$ be a random sample from the population with distribution function F. Let $F_n(x) = \frac{\# X_i \leq x}{n},\ \ x \in (-\infty, \infty)$, be the *empirical distribution function.* Let $c < d$ be two given real numbers. Define $V = F_n(c)$ and $W = F_n(d)$. Find ρ_{UW}, the correlation coefficient between V and W, and interpret the result.

[12]

4. Consider the regression model

$$Y_i = \alpha + \beta x_i + \varepsilon_i;\ 1 \leq i \leq n \tag{2}$$

where $x_1,\ x_2, ..., x_n$ are to be treated as given values of x (constants); $\varepsilon_1,\ \varepsilon_2, ..., \varepsilon_n$ are random errors that are *independently and identically distributed* (*iid*) as $N\left(0, \sigma^2\right)$ and α and β are parameters of the model.

(a) Find *maximum likelihood estimators (mle)* for α and β based on $Y_1,\ Y_2, ..., Y_n$. How do they compare with *least squares estimators (lse)* for α and β.

(b) If an individual has her x value to be x_0 such that x_0 belongs to the interval $\left[\min\limits_{1\leq i\leq n} x_i,\ \max\limits_{1\leq i\leq n} x_i\right]$ then how would you predict her y value using model (2) and your answer to (a)?

(c) If $x_1,\ x_2, ..., x_n$ are to be chosen from $[-1, 1]$ then how would you choose them so that $Var\left(\widehat{\beta}_{lse}\right)$ is minimized, $\widehat{\beta}_{lse}$ being the *lse for* β?

$[(8 + 2) + 2 + 6 = 18]$

5. Let $X_1,\ X_2, ..., X_n$ be a random sample from $Uniform(0, \theta)$; $\theta > 0$. Let $X_{(1)},\ X_{(2)}, ...,$ $X_{(n)}$ be the *order statistics*. Obtain the joint density function of $V = X_{(n)}$ and $W = \frac{X_{(1)}}{X_{(n)}}$. Hence or otherwise show that $X_{(n)}$ and $\frac{X_{(1)}}{X_{(n)}}$ are independent.

[12]

6. 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]