Indian Statistical Institute, Bangalore
B. Math (II)
First semester 2014-2015
End-Semester Examination : Statistics (I)

Date: 12-11-2014 **Maximum Score 80** **Duration: 3 Hours**

1. The data below show IQ scores for 30 sixth graders.

088	102	126	095	109	099
102	151	115	097	092	107
081	119	094	090	109	099
102	117	098	093	105	084
114	122	087	094	101	081.

   (a) Make a stem and leaf plot of these data.

   (b) Find the sample median M.

   (c) Find $100p$-th percentiles for $p = 0.25$ and 0.75.

   (d) Find the first and third quartiles.

   (e) Draw the box plot and identify the outliers.

   (f) If there are outliers, decide on trimming fraction just enough to eliminate the outliers and obtain the trimmed sample. If there are no outliers, then for trimming fraction $\alpha = 0.05$, get the trimmed sample. For your trimmed sample obtain the trimmed mean $\overline{X}_{\mathrm{T}}$ and explain how to find the trimmed standard deviation S_{T}.

   (g) Between the box plot and the stem and leaf plot what do they tell us about the above data set? In very general terms what can you say about the population from which these data arrived?

$[5 + 2 + 4 + 2 + 5 + (2 + 2 + 2) + 4 = 28]$

2. Let $X_1, X_2, \cdots, X_n$ be a random sample from the *uniform*$[-\theta, \theta]$; $\theta > 0$, distribution. Find *method of moments (MOM) estimator* for θ. Is *MOM estimator* unique in this problem? Is your *MOM estimator* consistent? Also find *maximum likelihood estimator (MLE)* for θ.

$[4 + 2 + 2 + 6 = 14]$

3. Let X uniform$(-d, d)$, $d > 0$, and W uniform$(0, h)$, $h > 0$, be independent. Define $Y = X^2 + W$. Find the joint distribution of X and Y. Draw the support of (X, Y) . Let $(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)$ be a random sample from this bivariate distribution of (X, Y). Let $V = \frac{1}{n}\Sigma_{i=1}^n X_i$ and $Z = \frac{1}{n}\Sigma_{i=1}^n Y_i$. Find ρ_{VZ}, the correlation coefficient between V and Z. How does ρ_{VZ} vary with d and h?

$[6 + 6 + 4 = 16]$

[PTO]

4. Let $X_1, X_2, \cdots, X_{2n}$ be a random sample from $N(0,1)$. Let us define Y as

$$Y = \left(\sum_{i=1}^{n} X_i \right)^2 + \left(\sum_{i=1}^{n} X_{n+i} \right)^2 .$$

Find the distribution of cY, $c \neq 0$. For what values of c would cY have *chi-square distribution?* For what values of n and c would you be able to generate observations on cY using a direct method? Explain.

$[8 + 2 + 4 = 14]$

5. This amusing classical example is from von Bortkiewicz (1898). The number of fatalities that resulted from being kicked by a horse was recorded for 14 corps of Prussian cavalry over a period of 20 years, giving 280 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 91 of the 280 corps-years, there was one death.

No. of Deaths per Year	Observed Frequency
0	144
1	91
2	32
3	11
4	2

Do these data come from a *Poisson model?* Carry out a *chi-square goodness of fit test at level of significance* $\alpha = 0.05$. Also report the p value.

[16]