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

Date: 01-10-2010 **Maximum Score 80** **Duration: 3 Hours**

1. Following is the data set of daily minimum temperature at a hill station recorded in $^{\circ}F$ during the month of April.

77, 80, 82, 68, 65, 59, 61, 57, 50, 62, 61, 70, 69, 64, 67,

70, 62, 65, 65, 73, 76, 87, 80, 82, 83, 79, 79, 71, 80, 77.

   (a) Make a stem and leaf plot of these data with 2 lines per stem.
   (b) Find the sample mean $\overline{X}$.
   (c) Find the sample standard deviation s.
   (d) Find the sample median M.
   (e) Find $100p$-th percentiles for $p = 0.25$ and 0.75.
   (f) Find the first and third quartiles.
   (g) What proportion of the data lies within $\overline{X} \pm 3s$?
   (h) Draw the box plot and identify the outliers.
   (i) Decide on trimming fraction just enough to eliminate the outliers and obtain the trimmed mean $\overline{X}_{\mathrm{T}}$.
   (j) Also obtain the trimmed standard deviation s_{T}.
   (k) 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 the data arrived?

   If there are no outliers then for sub-questions (i) and (j) explain how you would compute, in general, trimmed mean $\overline{X}_{\mathrm{T}}$ and trimmed standard deviation s_{T}.

   $[4+2+2+2+4+2+2+5+2+3+4=32]$

2. Suppose the lifetime X of an electronic component is known to have *Weibull distribution* with parameter $\theta > 0$ with *probability density function (pdf)* given by $f(x|\theta) = \theta c x^{c-1} e^{-\theta x^c}$; for $x > 0$ and $c > 0$ is a constant; $\theta > 0$. Suppose n such randomly chosen components are put to survival test and that $X_1,\ X_2, ..., X_n$ are their lifetimes. Find *maximum likelihood estimator (mle)* for θ based on $X_1,\ X_2, ..., X_n$.

   [8]

[PTO]

3. *Pareto distribution* is often appropriate to model income. Let $X_1,\ X_2, ..., X_n$ be a random sample from *Pareto density* given by $f(x|\theta) = c\theta^c x^{-(c+1)}$ for $x > \theta$ and 0 otherwise, where $\theta > 0$ and c is a positive constant. Find *maximum likelihood estimator (mle)* for θ based on $X_1,\ X_2, ..., X_n$. Also obtain *method of moments (mom)* estimator for θ and check whether it is consistent.

[12]

4. Of a leading national political party a large section of its members is opposed to 'uniform civil code' arguing that the country has a very heterogenous population. A simple random sample of 300 members of the political party is drawn. **Build** a model to obtain the distribution of number of members opposed to 'uniform civil code' in the random sample. What would be a good guess for the proportion of members opposed to 'uniform civil code' in the political party? If there were 250 members in the sample opposed to 'uniform civil code' what would be the numerical value of your guess? Have you used an unbiased estimator? Find its mean squared error.

[12]

5. Let $X_1,\ X_2, ...,$ be a sequence of *i.i.d.* random variables from $\varepsilon xp\,(\lambda)\,,\ \lambda > 0$. Consider positive integers $\alpha_1, \alpha_2, ..., \alpha_L$. Define $Y_1 = \sum_{i=1}^{\alpha_1} X_i,\ Y_j = \sum_{i=1}^{\alpha_j} X_{\left(\Sigma_{k=1}^{j-1}\alpha_k\right)+i}$ for $2 \leq j \leq L$. For integers $m, n;\ 1 \leq m < n \leq L$ find the distribution of the ratio $R = \frac{Y_m}{Y_n}$. Hence find the distribution of $W = cR = c\frac{Y_m}{Y_n}$, where $c \in (-\infty, \infty)$ is a constant. Can you find conditions on c and λ so that the random variable W has $F-$ distribution? Substantiate.

[12]

6. Suppose you can draw observations on a random variable that has *normal distribution* $N(\theta, \sigma^2)$. Suppose further that you can draw a random sample from $U \sim uniform[0, 1]$. Explain how you would draw observations on a random variable T that has *Student's t-distribution* with $2n$ *degrees of freedom*, n being a positive integer.

[12]