Indian Statistical Institute
Bangalore Centre
B.Math (Hons.) II Year 2013-2014
Semestral Examination

Statistics I — Date 11.11.13

Answer as many questions as possible. The maximum you can score is 110.
The notation used have their usual meaning unless stated otherwise.
State clearly the assumptions you make and the results you use.

1. Consider n distinct real numbers $x_1, x_2, \cdots x_n$. For a real number A, define $D(A) = \sum_{i=1}^{n} \|x_i - A\|$. Show that $D(A)$ is minimum when $A =$ the median of $x_1, x_2, \cdots x_n$. [7]

2. Suppose X is a random variable following standard normal distribution.

   (a) For a fixed real number l, show that $Prob.(a \leq X \leq a + l)$ is maximum if $a = -l/2$.

   (b) Let $\mathcal{I}$ denote the class of intervals on the real line such that $Prob(X \in I) = p$, for every I in $\mathcal{I}$. Find the interval $I_0 \in \mathcal{I}$ with minimum length.

   (c) Suppose $X_1, X_2, \cdots X_n$ are i.i.d. random variables following $N(\mu, 1)$. Two persons A and B were asked to obtain a 80 % confidence interval for μ using the observed values of X_i's. A found an interval $[\bar{X}, a]$, while B found the interval $[\bar{X} - b, \bar{X} + b]$. Which one would you recommend and why ?

   [5 + 2 + 4 = 11]

3. In a study the Systolic blood pressure of the workers in a factory in the age group 30-40 were observed. The data had mean 132.8 and variance 22.3.

   (a) Estimate the probabilities that a randomly selected worker would have Systolic blood pressure (i)$>$ 150, (ii)$<$ 110 and (iii) between 100 and 130.

   [4 x 3 = 12]

4. Suppose $X \sim P(\lambda)$. Let $\phi(\lambda, m)$ denote the probability that $X \leq m$. Let $m(\lambda, \alpha)$ denote an integer m such that $\phi(\lambda, m) = \alpha$, where $\alpha \leq 0.1$.

   (a) Show that $\phi(\lambda', m(\lambda, \alpha)) \leq \alpha$ for every $\lambda' > \lambda$.

   (b) Consider the testing of hypothesis problem $H_0 : \lambda \geq \lambda_0$, against $H_1 : \lambda \leq \lambda_0$.. Show that in order to find a size α test procedure ($\alpha \leq 0.1$), it is enough to find the integer $m_0 = m(\lambda_0, \alpha)$.

   [10 + 5 = 15]

5. A pharmaceutical company wants to protect against too many vaccines becoming sterile. In a random sample of 10 vials, 2 were found to be sterile. Obtain a 95% upper confidence bound for the probability that a vaccine becomes sterile.

   [10]

6. (a) Suppose (X, Y) follows bivariate normal distribution. with parameters $\mu_x, \mu_y, \sigma_x^2, \sigma_y^2$ and ρ Define random variables U and V as follows.

$U = Y - q - c(X - p)$ and $V = X - s$.

Find conditions on p, q, s and c such that (i) $E(U) = E(V) = 0$ and (ii) U, V are independent. What would be the distribution of U if these conditions are satisfied ?

Using the above results, find the conditional distribution of Y, given $X = x$.

[7 + 3 + 9 =19]

7. A chemist wanted to see if the solubility of Plutonium in Berilium fluoride depends on temperature. He took n observations with k different temperatures ($k < n$) and noted the amounts of plutonium solved. Then he fitted a linear model.

(a) Write down a appropriate model, stating clearly al the assumptions. Derive the distribution of the sum of squares for error. Obtain a 95% confidence interval for the regression coefficient.

(b) Another chemist looked at the data and felt that a quadratic model would fit the data better. Formulate a testing of hypothesis problem, which would resolve the difference betwe the two chemists. Derive the distribution of the test statistics.

[(3 + 7 + 3) + 12 = 25]

8. Suppose $(n_1, \cdots n_k)$ follows multinomial distribution with parameters $n, p_1, \cdots p_k$. Let

$$\phi = (\sqrt{p_1}, \cdots \sqrt{p_k})' \text{ and } V = (V_1, \cdots V_k)', \text{ where } V_i = (n_i - np_i)/\sqrt{np_i},\ i = 1, \cdots k.$$

(a) Show that for any $k \times 1$ vector b, the asymptotic distribution of $b'V$ is normal with mean zero and variance $\sigma_b^2 = b'(I_k - \phi\phi')b)$.

(b) A brand of washing machine is sold in five different colours. A market researcher wants to see whether all colours are equally popular. Formulate this as a a testing of hypothesis problem. Obtain a large sample test procedure, using the result in Q(a).

(c) The data from a random sample of 300 recent sales of washing machines is given below. What conclusion does the market researcher draw ?

Colours	Red	Blue	Green	White	Brown
Numbers sold	52	40	88	55	65

[8 + (3 + 6) + 5 = 22]