INDIAN STATISTICAL INSTITUTE

B-Stat I, Second Semester, 2013-14

STATISTICAL METHODS II

Date: 25.07.14 Backpaper Examination Time: 3 hours

(Total Point 100)

1. (a) Suppose that $p_1, p_2, \ldots, p_k$ are known, real, nonnegative constants, and they satisfy the condition $\sum_{i=1}^{k} p_i = 1$; let $f_1, f_2, \ldots, f_k$ represent k probability density functions. Consider the function

$$f_{\text{M}}(x) = p_1 f_1(x) + p_2 f_2(x) + \ldots + p_k f_k(x). \tag{1}$$

   i. Show that $f_{\text{M}}(\cdot)$ is a valid probability density function (it is called the mixture of the densities $f_1, f_2, \ldots, f_k$ with mixing weights $p_1, p_2, \ldots, p_k$).

   ii. Assuming that we know how to generate random numbers from each of the component densities $f_1, f_2, \ldots, f_k$, describe how to draw a random number from the mixture density in (1).

   (b) Suppose we wish to draw a random number X from the Laplace distribution with density

$$f(x) = \frac{1}{2} e^{-|x|}, \quad -\infty < x < \infty.$$

   Suppose we know how to generate random numbers from the Uniform (0, 1) distribution. Starting from random variables distributed uniformly on (0,1), describe how you can draw a random number from the Laplace distribution presented above. [(2+8)+15=25]

2. Suppose that $X_1, \ldots, X_n$ form a random sample from a Poisson distribution with unknown mean θ and let $Y = \sum_{i=1}^{n} X_i$.

   (a) Determine the value of a constant c such that the estimator e^{-cY} is an unbiased estimator of $e^{-\theta}$.

   (b) Use the information inequality to obtain a lower bound for the variance of the unbiased estimator found in part (a).

   (c) Does the variance of the estimator in (a) attain the Crámer-Rao bound found in part (b)?

   (d) Show that the ratio of the bound obtained in part (b) and the variance obtained in part (c) goes to 1 as $n \to \infty$.

(e) Show that Y is the sufficient statistic for the model, and argue whether the estimator obtained in part (a) is a minimum varianced unbiased estimator.

[5+5+5+5+5=25]

3. Under standard notations used in class for the multiple regression, partial regression part, show the following:

(a) $Cov(X_{1.23...p}, x_{1.23...p}) = 0$.

(b) $r^2_{1.23...p} = 1 - \frac{s^2_{1.23...p}}{s^2_1}$.

(c) $b_{12.34...p} = r_{12.34...p} \times \frac{s_{1.23...p}}{s_{2.34...p}}$.

(d) $r^2_{1p.23...p} = 1 - \frac{s_{1.23...p}}{s_{1.23...p-1}}$.

(e) Suppose the correlation matrix of $(x_1, x_2, \ldots, x_p)$ satisfies $corr(x_i, x_j) = r$ for all $i, j = 1, \ldots, p$, $i \neq j$. Show that the multiple correlation is equal to

$$r^2_{1.23...p} = \frac{(p-1)r^2}{1 + (p-2)r}.$$

[5+5+5+5+5=25]

4. Consider the A-B-O blood group gene frequency data given by the vector of cell frequencies

$$\boldsymbol{y} = (n_O, n_A, n_B, n_{AB}) = (176, 182, 60, 17),$$

where the individual components of $\boldsymbol{y}$ are the frequencies of the corresponding phenotypes. The parameter vector is $\theta = (p, q, r)$, where the individual components of θ represent the proportion of the gene counts (that is, count of the alleles A, B and O), with $p + q + r = 1$. We want to find the maximum likelihood estimators of the parameters, and want to use the EM algorithm for this purpose.

(a) Explain what would be the complete data in this example.

(b) Given estimates $p^{(k)}$, $q^{(k)}$ and $r^{(k)}$ at the k-th step of the iteration, explain how the E and M steps will be performed to get the parameter estimates at the $(k+1)$-th step.

(c) Given current estimates 0.264, 0.093 and 0.643 of p, q and r, find the parameter estimates after the next iteration.

[5+10+10=25]

Indian Statistical Institute

Back Paper Examination 2013-14

B. Stat. I st Year | Vectors and Matrices II

Date : 31.7.14 | Maximum Marks : 100 | Time: 3 hrs.

1. Prove or disprove: If x, y, and z vectors in an inner product space such that $< x, y >=< x, z >$, then $y = z$. [6]

2. Prove or disprove: If $< x, y >= 0$ for all x in an inner product space, then $y = 0$. [6]

3. Prove or disprove: Every nonzero finite dimensional inner product space has an orthonormal basis. [6]

4. Prove or disprove: The adjoint of a linear operator is not unique. [6]

5. Prove or disprove: Every normal operator is diagonalisable. [6]

6. Let $A = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ and let $B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & i & 0 \\ 0 & 0 & -i \end{pmatrix}$.
Show that they are unitarily equivalent. [15]

7. Find a polar decomposition of the matrix $\begin{pmatrix} 20 & 4 & 0 \\ 0 & 0 & 1 \\ 4 & 20 & 0 \end{pmatrix}$. [15]

8. Let V be a finite dimensional vector space over a field of real or complex numbers. Show that every symmetric bilinear form on V is diagonalisable. [15]

9. Find a Jordan canonical form of $\begin{pmatrix} 3 & 1 & -2 \\ -1 & 0 & 5 \\ -1 & -10 & 4 \end{pmatrix}$. [15]

10. Show that two real symmetric matrices are congruent if and only if they have the same rank, index and signature. [10]

INDIAN STATISTICAL INSTITUTE

Backpaperl Examination: 2013-14 (Second Semester)
Bachelor of Statistics (B. Stat.) I Year
Probability Theory II

Teacher: Parthanil Roy

Date: 08/08/2014 **Maximum Marks: 100** **Duration: 3 Hours**

Note:

- **Please write your roll number on top of your answer paper.**
- **You may use any theorem proved or stated in the class but do not forget to quote the appropriate result.**
- **You are NOT allowed to use class notes, books, homework solutions, list of theorems, formulas etc. If you are caught using any, you will get a zero in this examination.**
- **Failing to follow the examination guidelines, copying in the examination, rowdyism or some other breach of discipline or unlawful/unethical behavior, etc. are regarded as unsatisfactory conduct.** ***Any student caught cheating or violating examination rules will get a zero in this examination.***

1. Buses arrive at a certain bus stop according to a Poisson process with rate $\lambda > 0$. If you take the bus from that stop, then it takes a deterministic time R, measured from the time at which you enter the bus, to arrive home. If you walk from the bus stop, then it takes a deterministic time W to arrive home. Suppose that your policy when arriving at the bus stop is to wait up to a time s, and if a bus has not yet arrived by that time then you walk home. Compute the expected time from when you arrive at the bus stop until you reach home. [12]

2. Let $Y_1 < Y_2 < Y_3 < Y_4$ be an order statistic of size 4 from the exponential distribution with parameter 1. For $i = 1, 2, 3$, define $U_i = Y_{i+1} - Y_i$. Compute the joint probability density function of U_1, U_2 and U_3. [12]

3. Let X, Y and Z be three independent standard uniform random variables and $S = X + Y + Z$. Find (a) the conditional density of S given $Y = y$ and $Z = z$ and (b) the conditional density of (Y, Z) given $S = s$. [6 + 6]

4. Suppose U and V are two independent random variables with $U, V \sim Unif(0,1)$. Compute the probability density function of $Z = U - V$. [12]

5. Let R be the range of a random sample of size n from uniform distribution on $(0,1)$. Find the probability density function of R. [12]

6. Let $Z_1, Z_2, \ldots Z_{20}$ be independent and identically distributed random variables with $Z_1 \sim N(0,1)$. Find the joint probability density function of $\sum_{i=1}^{20} Z_i^2$ and $\frac{\sum_{i=1}^{8} Z_i^2}{\sum_{i=1}^{20} Z_i^2}$. Are they independent? [10 + 2]

[P. T. O.]

7. Suppose (U, V) is a bivariate normal random vector with $E(U) = E(V) = 0$, $Var(U) = Var(V) = 1$ and $Cov(U, V) = \rho$. Compute the correlation coefficient between U^2 and V^2, and the probability $P(U > 0, V > 0)$. [3 + 7]

8. A continuous random vector (X, Y) has a joint probability density function given by

$$f_{X,Y}(x, y) = \begin{cases} 2/3 & \text{if } x > 0, y > 0, x + y < 1, \\ 4/3 & \text{if } x < 1, y < 1, x + y > 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Compute the marginal probability density function of X. [5]

(b) For $0 < x < 1$, find the conditional probability density function of Y given $X = x$. [5]

9. Suppose $T \sim Exp(\lambda)$. For any real number x, denote by $[x]$ the largest integer less that or equal to x. For example, $[2.3] = 2, [5] = 5$ and so on.

(a) Compute the probability mass function of $S = [T]$. [6]

(b) Exactly which distribution does S follow? [2]

Wish you all the best

Indian Statistical Institute
Mid-Semestral Examination: 2014-15
B. Stat (Hons.) First Year

Vectors & Matrices –I

Date: **25-08-2014** Maximum Marks: **30** Duration:**90 min.**

1. Consider a three-dimensional vector space **V(F)** over a *finite* field **F**, where **F** consists of s elements. Let **S** be the collection of all one-dimensional subspaces of **V(F)**. Find the cardinality of **S**. 7

2. Let **F** = {0, 1, 2} with addition and multiplication modulo 3. Let **S** be the subspace of $\mathbf{F}^3$ spanned by (1, 0, 0) and (1, 0, 2). Construct the quotient space $\mathbf{F}^3/\mathbf{S}$. Find a complement of the subspace **S** different from the quotient space $\mathbf{F}^3/\mathbf{S}$. 7

3. Let $\mathbf{P}_n(\mathbb{R})$ be the vector space of polynomials of degree less than or equal to n with real coefficients. For fixed $a \in \mathbb{R}$, consider the set
$S = \{f \in P_n(\mathbb{R}) : f(a) = 0\}$.
Show that S is a subspace. Find its dimension. 7

4. Let **Q** be the field of rational numbers. Consider the system of equations
$$2x + z - w = 0$$
$$y - 2z - 3w = 0$$
in unknown $x, y, z, w \in \mathbf{Q}$. Determine the dimension of the solution subspace **S** of $\mathbf{Q}^4$. Show that
$$2x + z - w = u$$
$$y - 2z - 3w = v$$
admit a solution for every u and v in **Q**. 7

5. Let $\mathbf{W}_1$ and $\mathbf{W}_2$ be subspaces of a vector space **V** such that **V** is the direct sum of $\mathbf{W}_1$ and $\mathbf{W}_2$. If Γ_1 and Γ_2 are bases for $\mathbf{W}_1$ and $\mathbf{W}_2$ respectively, show that $\Gamma_1 \cap \Gamma_2 = \varphi$ and $\Gamma_1 \cup \Gamma_2$ is a basis for V. Show that the converse is also true in the sense that if Γ_1 and Γ_2 are bases for subspaces $\mathbf{W}_1$ and $\mathbf{W}_2$ respectively, with $\Gamma_1 \cap \Gamma_2 = \varphi$ and $\Gamma_1 \cup \Gamma_2$ is a basis for **V**, then **V** is the direct sum of $\mathbf{W}_1$ and $\mathbf{W}_2$. 7

Indian Statistical Institute
Semester 1 (2014-2015)
B. Stat 1st Year
Mid-semestral Examination
Probability Theory 1

Tuesday 26.8.2014, 10:30-12:30 Total Points $5 \times 6 = 30$

Answers must be justified with clear and precise arguments. More than one uncrossed answers to the same question or part of a question will not be entertained and only the first uncrossed answer will be graded. If you use any theorem or lemma proved in class state it explicitly.

1. Let α = the number of ways of selecting 5 distinct integers from $1, 2, \ldots, 100$ no two of which are consecutive, and β = the number of ways of selecting integers $r_1, r_2, \ldots, r_5$ satisfying $1 \leq r_1 \leq r_2 \leq r_3 \leq r_4 \leq r_5 \leq 92$. Find the value of β and explain whether $\alpha < \beta$ or $\alpha = \beta$ or $\alpha > \beta$.

2. From an urn containing $(2n+1)$ tickets numbered serially starting from 1 to $(2n+1)$, three tickets are drawn at random without replacement. Find the probability that the numbers on them are in arithmetical progression.

3. A cell contains N chromosomes, between any two of which an interchange of parts may occur. If r interchanges occur (which can happen in $\binom{N}{2}^r$ distinct ways) then find the probability that exactly m chromosomes will be involved.

4. Consider a random walk with $S_0 = 0$ and ± 1 increments. Let $M_{2n} = \max\{S_0, S_1, \ldots, S_{2n}\}$ and $m_{2n} = \min\{S_0, S_1, \ldots, S_{2n}\}$. Find a closed form expression for
$$P(S_0 = 0, M_{2n} \geq a, m_{2n} \leq -b, S_{2n} = 0),$$
where a and b are two positive integers.

5. In the Ehrenfest model of diffusion there are a total of d balls of which i are in Box 1 and $(d-i)$ are in Box 2. Balls are distinguishable and numbered $1, 2, \ldots, d$. A person selects a number from $1, 2, \ldots, d$ at random and the ball with that number is transferred from its box to the other one. Suppose initially one doesn't know how many balls are in box 1, but only knows the probabilities of causes:
$$P(i \text{ balls in box } 1) = \frac{\binom{d}{i}}{2^d}, i = 0, 1, \ldots, d.$$
If the person conducts his experiment once, find the probability the after the experiment there will be k balls in box 1. (Here k can be $0, 1, \ldots, d$, but you may do the calculation for only one fixed positive value of k, $0 < k < d$.)

INDIAN STATISTICAL INSTITUTE
B. Stat. I: 2014–2015
Introduction To Programming & Data Structure
Mid-Semester Examination

Date: 28.08. 2014 Marks: 50 Time: 2 Hours

Answer any part of any question. The question is of 55 marks. The maximum marks you can get is 50. Please write all the part answers of a question at the same place.

1. (a) Write a C program, that will generate all the Fibonacci numbers less than a given integer n. The integer n should be an input to the program.
   (b) Write a C function that will return the factorial of an integer n. The value of n should be passed as an input parameter.

   $5 + 5 = 10$

2. (a) Write a C program that finds the GCD as well as the LCM of two unsigned integers and show how the function executes when the two integers are 4620 and 96900.
   (b) Will your program work properly for all pairs of unsigned integers? Explain.

   $5 + 5 = 10$

3. (a) Write a C program that can find the maximum and minimum from an unsorted integer array of n elements. [Consider that the array is already populated with integers, i.e., you do not read to write any input statement. Efficient algorithms will be given more credit.]
   (b) How many comparisons are required in your program? Explain.

   $5 + 5 = 10$

4. (a) How can a two dimensional co-ordinate be implemented in C?
   (b) Use your data structure to define a triangle and then write a C function to calculate the area of the triangle.

   $2 + (2 + 6) = 10$

5. (a) Given an unsigned integer n, write a C program that can calculate $\lfloor\sqrt{n}\rfloor$. [Do not use the *sqrt* function of C mathematical library.]
   (b) Given $n = m^2$, for a non-negative integer m, prove the correctness of the algorithm you used.

   $5 + 5 = 10$

6. Explain what happens when the following codes are executed.
   (a) `char *p, *q; while (*p++ = *q++);`
   (b)
   ```
   int i, k = 1, n = 5;
   for (i = 0; k < n+1; i = k-i) { printf("%d\n", k); k = k+i; }
   ```

   $2 + 3 = 5$

INDIAN STATISTICAL INSTITUTE

Analysis I : B. Stat 1st year
First Semester Examination: 2014-15
August 29, 2014.

Maximum Marks: 40 Maximum Time: $2\frac{1}{2}$ hrs.

Answer all questions but maximum you can score is 40.

(1) Let $f : [0, 2\pi] \to [0, 2\pi]$ be a continuous function. If $f(0) = f(2\pi)$ then prove that there exists $x \in [0, 2\pi]$ such that $f(x) = f(x + \pi)$. [4]

(2) Let $\{a_n\}$ be a sequence of nonnegative real numbers and

$$b_n = \frac{a_1 + a_2 + \ldots a_n}{n}.$$

Prove that $\sum_{n=1}^{\infty} b_n$ converges iff $a_n = 0$ for all $n \in \mathbb{N}$. [6]

(3) Given a nonempty subset A of $\mathbb{R}$ define the function

$$f_A(x) = \inf \{|x - a| \mid a \in A\}.$$

Prove that f is continuous on whole of $\mathbb{R}$. [6]

(4) Let $\{a_n\}$ be a decreasing sequence of nonnegative real numbers. If $\sum_{n=1}^{\infty} a_n$ converges prove that $\lim_{n\to\infty} na_n = 0$. What about the converse? [8]

(5) Let $\{x_n\}$ and $\{y_n\}$ be sequences converging to l and m respectively. If

$$z_n = \frac{1}{n}\sum_{k=1}^{n} x_n y_{n+1-k},$$

then prove that $\{z_n\}$ converges to lm. [10]

(6) Let $\{x_n\}$ be a sequence of real numbers converging to 1. If $d \in \mathbb{N}$ then prove that the sequence $\{|x_n|^{1/d}\}$ also converges to 1. [10]

INDIAN STATISTICAL INSTITUTE

Semester Examination, 1st Semester, 2014-15

Statistical Methods I

Total Points 100

Date: November 03, 2014 Time: 3 hours

1. The following data represent the maximum daily rainfall in inches at South Bend, Indiana over a one year period:

1.88	2.23.	2.58	2.07	2.94	2.29	3.14	2.15	1.95	2.51
2.86	1.48	1.12	2.76	3.10	2.05	2.23	1.70	1.57	2.81
1.24	3.29	1.87	1.50	2.99	3.48	2.12	4.69	2.29	2.12

   (a) Compute the five number summary for these data.

   (b) Are there any suspected outliers by the $1.5 \times$ IQR rule?

   (c) Among the measures of mean and the median, which one would you expect to be greater for this data based on the five number summary?

[8+5+5=18 points]

2. A person was asked to calculate the mean and standard deviation of 50 observations. He noticed that 4 out of 50 observations were zeroes and hence ignored them and calculated the mean and standard deviation of the remaining 46 observations as 42.96 and 13.8812. Find the correct values of mean and standard deviation of all 50 observations. [16 points]

3. Let there be k groups of data on x and y, with means $\bar{x}_i$ and $\bar{y}_i$, variances s_{xi}^2 and s_{yi}^2, and correlations r_i in the i-th group $(i = 1, 2, \ldots, k)$. Show that the correlation for the combined data is

$$r = \frac{\sum_{i=1}^{k} n_i r_i s_{xi} s_{yi} + \sum_{i=1}^{k} n_i (\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y})}{\left[\sum_{i=1}^{k} n_i s_{xi}^2 + \sum_{i=1}^{k} n_i (\bar{x}_i - \bar{x})^2\right]^{1/2} \left[\sum_{i=1}^{k} n_i s_{yi}^2 + \sum_{i=1}^{k} n_i (\bar{y}_i - \bar{y})^2\right]^{1/2}}$$

where $\bar{x}$ and $\bar{y}$ are the grand means of x and y and n_i is the number of pairs in the i-th group. [16 points]

4. Suppose that X_1 and X_2 have a joint bivariate distribution given by

$$f(x_1, x_2) = \begin{cases} 4x_1x_2 & 0 < x_1 < 1, 0 < x_2 < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Fully describe the joint distriubtion of the transformed variables $Y_1 = X_1X_2$ and $Y_2 = X_1/X_2$. [14 points]

5. Suppose that $X_1, X_2, \ldots, X_n$ form a random sample from a uniform distribution on the interval $(0, \theta)$, and that the following hypothesis are to be tested.

$$H_0 : \theta = 2.0 \quad \text{versus} \quad H_1 : \theta = 1.6.$$

Let $X_{(n)} = \max\{X_1, X_2, \ldots, X_n\}$. Suppose that our strategy is to reject the null hypothesis when $X_{(n)} \leq 1.5$.

(a) Find the size of the above test when $n = 10$.

(b) Find the power of the above test $n = 10$.

[8+8=16 points]

6. The number of alpha particles emitted by a certain mass of radium is observed for 50 disjoint time periods each of which lasted 6 seconds. The results are given in the following table. Test the hypothesis that these 50 observations form a random sample from a Poisson distribution.

Number of Particles	0	1	2	3	4	5	6	7	8	9
Frequency	1	11	10	10	10	4	0	0	3	1

[20 points]

Indian Statistical Institute
Semester 1 (2014-2015)
B. Stat 1st Year
Semestral Examination
Probability Theory I

Wednesday 5.11.2014, 10:30-1:30 Total Points: $5 \times 14 = 70$

Answers must be justified with clear and precise arguments. More than one uncrossed answers to the same question or part of a question will not be entertained and only the first uncrossed answer will be graded. If you use any theorem or result proved in class state it explicitly.

1. (a) Ten manuscripts are arranged in 30 files (3 files for each manuscript). Find the probability that 6 files selected at random do not contain an entire manuscript. 7 pts.

   (b) An urn contains n balls, each of different color, of which one is white. Two independent observers, each with probability 0.1 of telling the truth, assert that a ball drawn at random from the urn is white. Prove that the probability that the ball is, in fact, white is $(n-1)/(n+80)$. 7 pts.

2. Consider a random walk in which $S_0 = 0, S_n = X_1 + X_2 + \cdots + X_n, n \geq 1$, where $X_i, i = 1, 2, \ldots$ are iid ± 1 with equal probability. Use the reflection principle directly to obtain a closed form expression for $f_{2n} = P(S_0 = 0, S_1 \neq 0, S_2 \neq 0, \ldots, S_{2n-1} \neq 0, S_{2n} = 0), n \geq 1$. You cannot use the ballot theorem (unless you prove it also). 14 pts.

   Hint: By symmetry one can consider paths above the horizontal axis. Think where S_1 and S_{2n-1} must be for such paths.

3. Suppose n distinguishable balls are distributed at random into r boxes. Let S_r denote the number of empty boxes. Compute $Var(S_r)$ in a closed form. 14 pts.

4. (a) Two gamblers A and B start with Rs 2 each at time 0. Each time they toss a coin and whoever wins gets one rupee from the other. The game continues until one of them loses all the money. Let P_{1n}, P_{2n}, P_{3n} denote the probabilities that at time n, A has Rs. 1, 2, 3 respectively. Find these explicitly as functions of n. 7 pts.

   (b) A fair coin is tossed repeatedly and independently. Let T denote the time from the begining until the observation of the first HH. Find the expectation of T in a closed from. 7 pts.

   Note: In part (b) conditional expectation may yield the result quickly. However any method you use is welcome.

P. T. O.

5. (a) In a sequence of Bernoulli trials with $P(S) = p, P(F) = q = 1 - p$, let u_n be the probability that the pattern SF occurs for the first time at trials $n-1$ and n. Find a closed form expression for $\sum_{n=2}^{\infty} u_n s^n, 0 < s < 1$.

(b) Consider the binomial probability generating function when p depends on n as $p_n = \lambda/n$ with $\lambda > 0$, a fixed number and n is large enough so that $p_n < 1$. Write down the limit of this pgf when $n \to \infty$.

7 + 7 = 14 pts.

INDIAN STATISTICAL INSTITUTE
B. Stat. I: 2014–2015
Introduction To Programming & Data Structure
Semestral Examination

Date: 07. 11. 2014 Marks: 100 Time: 3 Hours

Answer any part of any question. The question is of 110 marks. The maximum marks you can get is 100. Please write all the part answers of a question at the same place.

1. (a) Write a C function to swap two integers that will be passed as arguments. No variable can be declared inside this function.

   (b) Given an unsigned integer (of 4 bytes) u, write an efficient C function to find the number of 1's in its binary representation.

   (c) Trace the execution of your function in (1b) on the input 0x8000AA55.

   (d) Write down the Postfix and Prefix expression of $A * B + C/D$.

   (e) Briefly explain what will be the output of the following program.

```
#include <stdio.h>
#include <stdlib.h>
int add(int a, int b){ return a*b; }
int mult(int a, int b){ return a+b; }
main(){
        int a = 5, b = 3, c;
        int (*fp)(int, int);
        fp = add; c = (*fp)(a, b); printf("%d\n", c);
        fp = mult; c = (*fp)(a, b); printf("%d\n", c);
}
```

$5+6+5+4+5=25$

2. (a) How can you implement a binary search tree using an array?

   (b) Explain why this implementation is inefficient.

   (c) Write a non recursive algorithm to implement the inorder traversal in a binary search tree implemented in an array. $2+3+10=15$

3. Write a C program that will read two matrices from a file and then multiply them; the result of the multiplication has to be written at the end of the same file. 10

P. T. O.

4. (a) What are the prototypes of **scanf** and **printf** functions in C programming language?

   (b) Write a variable argument function in C that takes the number of arguments n as the first parameter and then calculates the standard deviation of n values supplied to the function as following n arguments.

   (c) Consider that a function **func1** is provided to search for a substring in a string. You can assume the prototype of the function according to your convenience. Use the function **func1** to implement a function **func2** in C to search for a substring in a circular string. As an example, the substring "abc" is absent in the string "cxyzsdfgab", but exists when the string is considered in a circular manner.

$6 + 8 + 6 = 20$

5. (a) Briefly explain the concept of a Pseudo Random Number Generator.

   (b) Describe an algorithm to generate Pseudo Random integers between 0 to 255.

   (c) Consider that a uniform random number generator is provided as a black box. How can you use that to generate a permutation of integers uniformly at random?

$7 + 7 + 6 = 20$

6. Explain how can you use a singly linked list to implement the following.

   (a) A queue with the functionality of insertion and deletion.

   (b) A polynomial with the functionality of addition.

   Clearly explain the data structures and functionalities with the help of C codes.

$10 + 10 = 20$

Indian Statistical Institute
Semestral Examination: 2014-15
B. Stat (Hons.) First Year

Vectors & Matrices –I

Date: **11-11-2014** Maximum Marks: **70** Duration: **3 Hours**

1. Let V and W be arbitrary vector spaces over the same field F and let Γ be a basis for V. Show that for any function $f:\Gamma \to W$ there exists exactly one linear transformation $T:V \to W$ such that $T(\gamma)=f(\gamma)$ for all $\gamma \in \Gamma$.
8

2. Let x and y be two distinct vectors in a vector space $V(F)$ over a field F. Show that there exists a linear functional f in V' such that $f(x)\neq f(y)$, where V' is the dual space of $V(F)$. 5

3. Define the quotient space associated with a subspace S of $V(F)$. Show that this space is isomorphic to every complement of the subspace S. Hence find its dimension.
2+5+1

4. Let $T:V \to W$ be a linear transformation, where V and W are arbitrary vector spaces over the same field F. Construct a linear transformation $S:W \to V$ such that $TST=T$. Is it unique? 5+2

5. Define a projector from a vector space V to the subspace V_1 along the subspace V_2. Show that a linear transformation $P:V \to V$ is a projector from V along some complimentary subspace of V if and only if $P^2=P$.

2+6

6. Show that for any arbitrary matrix A, its row reduced echelon form is unique.
5

7. Let $x=(1,0,-1,2)$ and $y=(2,3,1,1)$, and let W be the subspace of R^4 spanned by x and y. Find W^a, where, W^a is the annihilator of W.
6

8. Let A be any matrix of order $m\times n$. Then show that $N(A)\cap R(A')=\{0\}$ if and only if $AA'\alpha=0$ for any $\alpha \Rightarrow A'\alpha=0$, where $N(A), R(A')$ denote the null space and range space of A and A' respectively and A' denotes the transpose of A.
6

9. Let A and B be two matrices of orders $m\times n$ and $m\times s$, respectively. Let $(A\mid B)$ be the augmented matrix of order $m\times(n+s)$. Show that
Rank $(A\mid B)$ = Rank (A) if and only if $B=AC$ for some matrix C. 6

[P.T.O.]

10. What is a Schur complement of a Matrix? Consider the following matrix A.

$$\begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{pmatrix}$$

Find the rank of A using a suitable Schur complement.

2+6

11. Use LU decomposition to solve the following system of equation:

$$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 8 \\ 2x_1 - 5x_2 + 12x_3 &= 4 \\ 2x_2 - 10x_3 &= -6 \end{aligned}$$

6

12. Obtain the rank factorisation of the matrix

$$A = \begin{pmatrix} -1 & 2 & 4 \\ 2 & -1 & 2 \\ 0 & 3 & 10 \end{pmatrix}$$

6

First Semester Examination (2014-2015)
B. Stat – 1 yr
Remedial English
100 Marks
1 ½ hours

Date : 12.11.2014

1. **Write an essay *on any one* of the following topics. Five paragraphs are expected.**
   a. **Surveys and Statistics**
   b. **Smoke : its evil effects**
   c. **The Television : its uses and abuses**

(60 marks)

2. **Fill in the blanks with appropriate prepositions :**

- **There was an irritation _______ dinner _____ his place.**
- **I was there _____ time in spite _____ traffic dislocation.**
- **There was a cat ______ the table _______ we started eating.**
- **The party broke _______ soon ________ we finished eating.**
- **I asked a friend ________ his reaction _________ the meet.**

(20 marks)

3. **Fill in the blanks with appropriate words :**

Bricks _______ _______ different ________ and sizes. Most of the ______ _______ sides, but _____ of _____ are curved. Some, _______, are _______ the ______ ______ ______ of a triangle. Not _____ can Tom _____ _____ and churches but can _____ _____ _____ differ _____ _____ often houses or churches.

(20 marks)

INDIAN STATISTICAL INSTITUTE

Analysis I : B. Stat 1st year
First Semester Examination: 2014-15
November 14th, 2014.

Maximum Marks 50 Maximum Time 3 hrs.

Answer all the questions. But maximum you can score is 50.

(1) Let $f : [0,1] \to \mathbb{R}$ be a continuous function and $x_1, x_2, \ldots, x_n$ be n distinct real numbers in [0,1]. Prove that there exists a real number $x_0 \in [0,1]$ such that $f(x_0) = \frac{1}{n}\sum_{k=1}^{n} f(x_k)$. [6]

(2) Let $f : [0,1] \to \mathbb{R}$ be a differentiable function such that $f(0) = 0$ and $f'(x) > 0$ for all $x \in (0,1]$. Prove that there exists a real number $c \in (0,1)$ such that $\frac{f'(1-c)}{f(1-c)} = \frac{f'(c)}{f(c)}$. [6]

(3) Let $f : \mathbb{R} \to \mathbb{R}$ be such that $f'''(x) > 0$ for all $x \in \mathbb{R}$. If $x_1 < x_2$ are two real numbers then prove that $f(x_2) - f(x_1) > f'\left(\frac{x_1 + x_2}{2}\right)(x_2 - x_1)$. [6]

(4) Let $f : [0,1] \to \mathbb{R}$ be twice differentiable and $f(0) = f(1)$. If $|f''(x)| \le 2$ for all $x \in (0,1)$ then prove that $|f'(x)| \le 1$ for all $x \in [0,1]$. [6]

(5) Let $f : (-1,1) \to \mathbb{R}$ be a twice differentiable function. If $f\left(\frac{1}{n}\right) = 1$ for all $n \in \mathbb{N}$ then prove that $f'(0) = 0$. Using this prove that $f''(0) = 0$. [6]

(6) Suppose that $a_n > 0$ for all $n \in \mathbb{N}$ and $\sum_{n=1}^{\infty} a_n$ converges. Prove that $\sum_{n=1}^{\infty} a_n\sqrt{a_n}$ converges. Discuss the convergence/divergence of the series $\sum_{n=1}^{\infty} \sqrt{a_n}\sin(a_n)$. [6]

(7) For $a \ge 0$ define $x_1 = a$ and $x_{n+1} = \frac{1}{5}(x_n^2 + 6)$, for all $n \in \mathbb{N}$.
  (a) Find all values of $a \in [0,\infty)$ for which $\{x_n\}$ is decreasing/increasing. [5]
  (b) Verify whether the sequence converges for $a > 3$. [3]

(8) Let $\{c_n\}$ be a sequence of real numbers such that $\sum_{n=1}^{\infty} c_n$ converges. Define $f(x) = \sum_{n=1}^{\infty} c_n x^n$, $x \in (-1,1]$. Prove that f is continuous at $x = 1$. [10]

Indian Statistical Institute
Back Paper Examination: 2014-15
B. Stat (Hons.) First Year

Vectors & Matrices –I

Date: 26 · 12 · 2014 Full Marks: 100 Duration: **3 Hours**

1. Define dimension of a vector space. Let V and W be arbitrary vector spaces over the same field F. Show that V is isomorphic to W if and only if $\dim W = \dim V$.

2+8

2. Let $T: V \to W$ be a linear transformation, where V and W are arbitrary vector spaces over the same field. Prove that $V = R(T^k) \oplus N(T^k)$, for some positive integer k, where the symbols have usual meaning.

10

3. Define a projector from a vector space V to the subspace V_1 along the subspace V_2. Show that a linear transformation $P: V \to V$ is a projector from V along some complimentary subspace of V if and only if $(I-P)$ is a projector if and only if $R(P) = N(I-P)$, where the symbols have usual meaning.

10

4. Define transpose of a linear transformation $T: V \to W$. Show that $Rank(T) = Rank(T')$.

2+8

5. Show that $Rank(A+B) = Rank(A) + Rank(B)$ if and only if $Rank(A \mid B) = Rank\begin{pmatrix} A \\ B \end{pmatrix} = Rank(A) + Rank(B)$ if and only if $Span(A) \cap Span(B) = \{0\}$ and $Span(A') \cap Span(B') = \{0\}$, where A and B are matrices of the same order.

10

6. Let A be any matrix of order $m \times n$. Then show that $N(A) \cap R(A') = \{0\}$ if and only if $AA'\alpha = 0$ for any $\alpha \Rightarrow A'\alpha = 0$, where $N(A), R(A')$ denote the null space and range space of A and A' respectively and A' denotes the transpose of A.

10

7. State and prove Frobenius's inequality for ranks involving three matrices. 10

[P.T.O.]

8. What is a Schur complement of a Matrix? Explain how you can determine rank of a matrix using Schur complement. 2+8

9. Use row reduced echelon form to solve the following system of equation:

$$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 6 \\ 2x_1 - 5x_2 + 12x_3 &= 9 \\ 2x_2 - 10x_3 &= -8 \end{aligned}$$

10

10. Obtain the LU decomposition of the matrix

$$A = \begin{pmatrix} -1 & 2 & 4 \\ 2 & -1 & 2 \\ 0 & 3 & 10 \end{pmatrix}$$

10

INDIAN STATISTICAL INSTITUTE

Analysis I : B. Stat 1st year

Back Paper Examination: 2014-15

30/~~November~~12, 2014.

Maximum Marks 100 Maximum Time 3 hrs.

(1) Assume that $\{a_0, a_1, \ldots, a_n\}$ are real numbers such that

$$\frac{a_0}{n+1} + \frac{a_1}{n} + \ldots + a_n = 0.$$

Prove that $P(x) = \sum_{k=0}^{n} a_k x^{n-k}$ has at least one real root in $(0, 1)$. [6]

(2) For $x > -1$, $x \neq 0$ prove that

$$(1+x)^\alpha > 1 + \alpha x, \text{ if } \alpha \notin (0,1) \text{ and } (1+x)^\alpha < 1 + \alpha x, \text{ if } \alpha \in (0,1).$$

[6]

(3) Determine whether the series

$$\sum_{n=1}^{\infty} (n+2)\left(1 - \cos\left(\frac{1}{n}\right)\right)$$

is convergent or divergent. [6]

(4) Let f and g be real valued continuous functions defined on $[0, 1]$. If $\inf\{f(x) \mid x \in [0,1]\} = \inf\{g(x) \mid x \in [0,1]\}$ then prove that there exists $x_0 \in [0,1]$ such that $f(x_0) = g(x_0)$. [6]

(5) Suppose f is defined in a neighbourhood of x_0 and suppose $f''(x_0)$ exists. Prove that $\lim_{h \to 0} \frac{f(x_0+h) + f(x_0-h) - 2f(x_0)}{h^2} = f''(x_0)$. [8]

(6) Let $f : [-1, 1] \to \mathbb{R}$ be a thrice differentiable function with $f(-1) = 0$, $f(1) = 1$ and $f'(0) = 0$. Prove that there exists $c \in (-1, 1)$ such that $f'''(c) \geq 3$. [8]

P.T.O

(7) For fixed positive real numbers α and β define $x_n = n^{\alpha}(1+\beta)^{-n}\sin n$, $n \in \mathbb{N}$. Does the sequence $\{x_n\}$ converge? Justify your answer. [8]

(8) Define $f(x) = \frac{e^x - e^{-x}}{2}$ for $x \in \mathbb{R}$. Prove that there exist positive constants c_1 and c_2 (independent of x) such that

$$c_1 \frac{xe^x}{1+x} \leq f(x) \leq c_2 \frac{xe^x}{1+x},$$

for all $x \geq 0$. [12]

(9) Determine all values of x for which the series $\sum_{n=2}^{\infty} \frac{x^n}{n(\log n)^2}$ converges. Give reasons for your answer. [12]

(10) Compute the limit $\lim_{x \to \infty} \left(x^2 - x^3 \sin\left(\frac{1}{x}\right)\right)$. [12]

(11) Prove that $\log(1+x) = \sum_{k=1}^{\infty} (-1)^{k-1} \frac{x^k}{k}$ for all $x \in (-1, 1)$. [16]

INDIAN STATISTICAL INSTITUTE

Back Paper Examination, 1st Semester, 2014-15

Statistical Methods I

Total Points 100

Date: December 31, 2014 Time: 3 hours

1. The following are zinc concentrations (in mg/ml) in the blood for two groups of rats. Group A received a dietary supplement of calcium and group B did not. Researchers are interested in variations in zinc level as a side effect of dietary supplementation of calcium.

   Group A: 1.31, 1.45, 1.12, 1.16, 1.30, 1.50, 1.20, 1.22, 1.42, 1.14, 1.23, 1.59, 1.10, 1.53, 1.52, 1.17, 1.49, 1.62, 1.29.

   Group B: 1.13, 1.71, 1.39, 1.15, 1.33, 1.00, 1.03, 1.68, 1.76, 1.55, 1.34, 1.47, 1.74, 1.74, 1.19, 1.15, 1.20, 1.59, 1.47.

   Draw back to back stem and leaf plots, and side by side box plots for the above data, pointing out the mean, median and the other quartiles. Comment on the comparative features of the two distributions. [20]

2. Suppose that thirty six heat lamps are connected in a greenhouse so that when one lamp fails the next one takes over immediately (only one lamp is turned on at one time). The lamps operate independently, and each has a mean lifetime of 50 hours when turned on, and a standard deviation of 4 hours. If the greenhouse is not checked for 1850 hours after the lamp system is turned on, what is the probability that the lamp will be burning when it is inspected at the end of the 1850 hour period? [15]

3. Let X and Y have joint pdf $f(x, y) = 4e^{-2(x+y)}, 0 < x < \infty, 0 < y < \infty$, and zero otherwise.

   (a) Find the CDF of $W = X + Y$.

   (b) Find the joint pdf of $U = X/Y$ and $V = X$.

   (c) Find the marginal pdf of U.

   [8+6+6=20]

4. Let X_1, X_2, X_3 and X_4 be independently distributed standard normal variables. Obtain, by using appropriate tables, the following probabilities:

(a) $P[X_1 + 2X_2 + 3X_3 \geq 6.5]$,

(b) $P[X_1 - 2X_2 + 3X_3 \leq 8]$,

(c) $P[X_1^2 + X_2^2 \geq 7.815 - X_3^2]$,

(d) $P[2X_1^2 \leq (4.303)^2(X_2^2 + X_3^2)]$,

(e) $P[X_1^2 - 19X_2^2 \leq 19X_3^2 - X_4^2]$.

[3 × 5 = 15]

5. Suppose that $X_1, X_2, \ldots, X_n$ represent an independently and identically distributed sample from the $N(\mu, 1)$ distribution.

(a) Suppose we want to test the null hypothesis $H_0 : \mu = 0$ versus $H_1 : \mu = 0.5$. We will reject the null hypothesis when $\bar{X} > b$, for some constant b (We refer to b as a constant in the sense it is independent of the actual values of $X_1, X_2, \ldots, X_n$, but it may depend on n). For $n = 25$, find the value of b so that the test has a size equal to 0.05.

(b) What sample size will be necessary for the test in part (5a) to have a power of 0.8 or higher?

[7+8=15]

6. A die is rolled 60 times, and we wish to test whether the die is a fair die. Given the following data, perform an appropriate test of hypothesis.

Face Value	1	2	3	4	5	6
Observed Frequency	8	11	5	12	15	9

[15]

Indian Statistical Institute
Semester 1 (2014-2015)
B. Stat 1st Year
Backpaper Examination
Probability Theory I

01/01/2015

Date: [illegible], Time: Total Points: $5 \times 20 = 100$

Answers must be justified with clear and precise arguments. More than one uncrossed answers to the same question or part of a question will not be entertained and only the first uncrossed answer will be graded. If you use any theorem or result proved in class state it explicitly.

1. N men run out of a men's club after a fire and each takes a coat and a hat at random. Prove that:

   (a) the probability that no one will take his own coat and hat is

   $$\sum_{k=0}^{N}(-1)^k \frac{(N-k)!}{N!k!}.$$

   (b) the probability that each man takes a wrong coat and a wrong hat is

   $$\left[\sum_{k=2}^{N}(-1)^k \frac{1}{k!}\right]^2.$$

   $10 + 10 = 20$ pts.

2. A person tosses an unbiased coin repeatedly and moves on the upper right quadrant as follows: if the result of a toss is H he moves 1 unit to the right, if it is T he moves 1 unit up. Thus after the kth toss his coordinates are the number of heads and the number of tails upto that point.

   (a) If he tosses $2n$ rimes, show that the probability that he does not touch or cross the diagonal is

   $$\frac{1 \cdot 3 \cdots (2n-1)}{2 \cdot 4 \cdots 2n}.$$

   (b) If after $2n$ tosses he has reached (n, n), find the probability that he did not touch or cross the diagonal before reaching (n, n).

   $10 + 10 = 20$ pts.

3. The probability p_n that n customers visit a supermarket in one day is $p_n = p^n q, n = 0, 1, 2, \cdots$. Two out of three customers, on the average, buy a certain type of item. The probability that an item is defective is $1/4$.

P. T. O.

(a) What is the probability that a customer buys a nondefective item?

(b) Given that k nondefective items were sold, show that the conditional probability a_n that n customers visited the shop is given by

$$a_n = \binom{n}{k} p^{n-k}(2-p)^{k+1}/2^{n+1}.$$

10 + 10 = 20 pts.

4. Supose we have two decks of n cards each numbered $1, 2, \cdots, n$. The two decks are shuffled and the cards are matched against each other. We say that a match occurs at position i if the ith card drawn from each deck has the same number. Let S_n denote the number of matches. Compute ES_n and $Var(S_n)$. 20 pts.

5. Suppose X_i are i.i.d. Bernoulli with parameter $p, 0 < p < 1$, and N is Poisson λ where N is independent of $X_i, i \geq 1$. For example, a random number N of cancer cells may develop and independently of N, each cell has a probability p of surviving a treatment, say radiation. Then $S_N = X_1+X_2+\cdots+X_N$ represents the total number of cells surviving treatment. Find the pgf of S_N and hence the pmf of S_N. 20 pts

INDIAN STATISTICAL INSTITUTE
B. Stat. I: 2014–2015
Introduction To Programming & Data Structure
Back Paper Examination

Date: 2/1/2015 Marks: 100 Time: 3 Hours

Answer any part of any question. The question is of 110 marks. The maximum marks you can get is 100. Please write all the part answers of a question at the same place.

1. (a) Write a C function to swap two characters that will be passed as arguments.
   (b) Given an unsigned integer (of 4 bytes) u, write an efficient C function to find the number of 0's in its binary representation.
   (c) Trace the execution of your function in (1b) on the input 0xA0008A55.
   (d) Write down the Postfix and Prefix expression of $A * (B + C)/D$.
   (e) Briefly explain what will be the output of the following program.

```
#include <stdio.h>
#include <stdlib.h>
int add(int a, int b){ return a/b; }
int mult(int a, int b){ return a%b; }
main(){
        int a = 5, b = 3, c;
        int (*fp)(int, int);
        fp = add; c = (*fp)(a, b); printf("%d\n", c);
        fp = mult; c = (*fp)(a, b); printf("%d\n", c);
}
```

$5 + 6 + 5 + 4 + 5 = 25$

2. (a) How can you implement a binary search tree in C programming language?
   (b) Write a C function to implement the inorder traversal in a binary search tree as implemented by you above. $7 + 8 = 15$

3. Write a C program that will read two matrices from the keyboard and then store the result of their addition in a file. 10

P. T. O.

4. (a) What are the prototypes of **scanf** and **printf** functions in C programming language?

   (b) Write a variable argument function in C that takes the number of arguments n as the first parameter and then finds the minimum one of the n values supplied to the function as following n arguments.

   (c) Write down a function in C programming language that will check whether a substring s exists in a given string t. You need to pass both s, t as arguments.

$6 + 8 + 6 = 20$

5. (a) Explain why it is important to have an initial seed in a Pseudo Random Number Generator.

   (b) Describe an algorithm to generate Pseudo Random integers between 0 to 10.

$10 + 10 = 20$

6. Explain how can you use a singly linked list to implement the following.

   (a) A stack with the functionality of push and pop.

   (b) A polynomial with the functionality of subtraction.

   Clearly explain the data structures and functionalities with the help of C codes.

$10 + 10 = 20$

Indian Statistical Institute

Vectors and Matrices II: B. Stat First Year: Mid-Semester Examination: 2014-15

Maximum Marks: 40 **Time: 2hrs** **Date: 23-02-2015**

1. Let V be an inner product space of dimension n. Let $x_1, x_2, \ldots, x_n$ be some n vectors in V having the property that for any two vectors u and v in V

$< u, v > = < u, x_1 >< x_1, v > + < u, x_2 >< x_2, v > + \ldots + < u, x_n >< x_n, v >$. Then prove that

$x_1, x_2, \ldots, x_n$ is an orthonormal basis for V. [9]

2. Let R^3 be the three -dimensional vector space equipped with the standard inner product. Let $f : R^3 \rightarrow R$ be the linear functional defined by

$f(x) = f(x_1, x_2, x_3) = 2x_1 + x_2 - x_3$ for $x = (x_1, x_2, x_3) \in R^3$.

Determine the vector $z \in R^3$ such that

$f(x) = < x, z >, x \in R^3$.

[9]

3. Let T be the linear operator in R^3 which is represented in the standard ordered basis by the matrix

$$\begin{bmatrix} -9 & 4 & 4 \\ -8 & 3 & 4 \\ -16 & 8 & 7 \end{bmatrix}$$

Prove that T is diagonalisable by exhibiting a basis of eigenvectors of T for R^3.

[9]

4. Let $f : M_{nxn}(R) \rightarrow R$ be a function defined on matrices. When is f called an n-linear and alternating function? Show that for such a f its value only changes by the sign if we change the matrix by interchanging any two rows.

[9]

5. Assume that $T : V \rightarrow V$ is self-adjoint. It is given that there exists $v \in V$ such that $\| v \| = 1$ and $\| T(v) - \lambda v \| < \varepsilon$ for some given $\varepsilon > 0$ and $\lambda \in R$. Show that T has an eigenvalue μ such that $| \lambda - \mu | < \varepsilon$.

[9]

INDIAN STATISTICAL INSTITUTE
B. Stat. I: 2014–2015
Numericl Analysis
Mid-Semester Examination

Date: 24. 02. 2015 Marks: 50 Time: $2\frac{1}{2}$ Hours

Answer any part of any question. The question is of 55 marks. The maximum marks you can get is 50. Please write all the part answers of a question at the same place.

1. Consider the function $f(x) = \sqrt{x+2} - \sqrt{x} + 1$. Discuss how one should calculate $f(22221)$ in six-decimal arithmetic to have less error. Marks : 10

2. Write down the algorithms for finding a real root of polynomials using the following methods: (a) Bisection Method, (b) Regula Falsi, (c) Modified Regula Falsi, (d) Secant Method, (e) Newton's Method.

   Marks : $15 = 5 \times 3$

3. (a) Solve the following system of linear equations for x_1, x_2, x_3, x_4, using each of the methods prescribed below. In each case, count or estimate the number of individual operations required to solve the system.

$$2x_1 + x_2 + x_3 = 4$$
$$4x_1 + 3x_2 + 3x_3 + x_4 = -3$$
$$8x_1 + 7x_2 + 9x_3 + 5x_4 = 3$$
$$6x_1 + 7x_2 + 9x_3 + 8x_4 = 2$$

   i. Using simple Gaussian elimination with *nonzero* pivoting.
   ii. Using simple Gaussian elimination with *partial* pivoting.
   iii. Using PLU factorization technique, with *nonzero* pivoting.
   iv. Using PLU factorisation technique, with *partial* pivoting.

   (b) Estimate the number of operations required to solve a system of n linear equations with n unknowns, for each of the above methods.

   Marks : $18 = (3 + 3 + 3 + 3) + 6$

4. (a) Solve the following system of linear equations for x_1, x_2, x_3, x_4, using any suitable computational method of your choice.

$$2x_1 + x_2 + x_3 = 4 + i$$
$$8x_1 + 7x_2 + 9x_3 + 5x_4 = 3 + 2i$$
$$8x_1 + 6x_2 + 6x_3 + 2x_4 = -6$$
$$6x_1 + 7x_2 + 9x_3 + 8x_4 = 2 - 3i$$

   (b) Propose an algorithm to solve a system of complex linear equations $\mathbf{A}(\mathbf{x} + i\mathbf{y}) = (\mathbf{u} + i\mathbf{v})$? Estimate the number of operations.

   (c) Propose an algorithm to solve a system of complex linear equations $(\mathbf{A} + i\mathbf{B})(\mathbf{x} + i\mathbf{y}) = (\mathbf{u} + i\mathbf{v})$? Estimate the number of operations.

   Marks : $12 = 4 + 4 + 4$

Indian Statistical Institute
Semester 2 (2014-2015)
B. Stat 1st Year
Mid-Semestral Examination
Probability Theory 2

Wednesday 25.2.2015, 2:30-4:30 PM Total Points $5 \times 6 = 30$

Answers must be justified with clear and precise arguments. More than one uncrossed answers to the same question or part of a question will not be entertained and only the first uncrossed answer will be graded. If you use any theorem or lemma proved in class state it explicitly.

1. A random point B is uniformly distributed on the circle $x^2 + (y-a)^2 = r^2$ with center at $(0, a)$. The random point $C = (X, 0)$ is the intersection of the x-axis and the line AB extended. Find the distribution function and the pdf of X.

2. A continuous random variable X has the probability density function proportional to $e^{-x}(1+x)^2$ in the range $-1 < x < \infty$. Show that the median of the distribution satisfies the relation

$$e^{m+1} = 1 + (m+2)^2.$$

3. (a) Suppose that X is a nonnegative random variable such that both EX and $E(1/X)$ are finite. Show that

$$E(1/X) \geq 1/EX.$$

(b) Suppose a nonnegative random variable X has a density f which is decreasing and is convex downwards. Determine the signs (i.e. positve or negative) of $E \sin X$ and $E \cos X$ with adequate explanation.

4. If F is the distribution function of a random variable X then find

$$\lim_{x \downarrow 0} x \int_x^\infty \frac{1}{u} dF(u).$$

(Hint: Assuming $x < 1$, divide the integral over the two intervals $(x, \sqrt{x})$ and $(\sqrt{x}, \infty)$.)

5. Let X be a random variable with df F and moment generating function $\phi(t) = \int_{-\infty}^{\infty} e^{tx} dF(x)$ which is assumed to be finite for all $t \in \mathbb{R}$.

(a) Assuming $P(X > 0) > 0$ and $P(X < 0) > 0$, what is $\lim_{t\uparrow\infty} \phi(t)$ and $\lim_{t\downarrow-\infty} \phi(t)$ and why?

(b) Draw the graph of $\phi(.)$ and indicate (with proof) whether ϕ has one minimum or can have multiple minima.

INDIAN STATISTICAL INSTITUTE

B-Stat I, Second Semester, 2014-15

STATISTICAL METHODS II

Date: 26.02.15 Semester Examination Time: 3 hours

Total Point 100

1. Given a parametric model $\{f_\theta : \theta \in \Theta\}$ of probability density functions, an unethical experimenter desires to test the following hypothesis $H_0 : \theta = \theta_0$ againt $H_1 : \theta = \theta_1$. He draws a random sample $X_1, X_2, \ldots, X_n$ from the distribution and carries out a test at size α. If this test does not reject H_0, he discards the sample, draws an independent sample afresh, and repeats the test based on the new sample. He continues drawing new samples until he obtains a sample where the test is rejected.

   (a) What is the overall size of this testing procedure?

   (b) If H_0 is true, what is the expected number of samples that the experimenter will have to draw until he rejects H_0?

   [5+15=20 points]

2. Let X be a random variable whose pmf under H_0 and H_1 are given by

x	1	2	3	4	5	6	7
$f(x\|H_0)$	0.01	0.01	0.01	0.01	0.01	0.01	0.94
$f(x\|H_1)$	0.06	0.05	0.04	0.03	0.02	0.01	0.79

   (a) What are the most powerful tests of size α for testing the above hypothesis for $\alpha = 0.01, 0.02, 0.03, 0.04$ and 0.05 respectively?

   (b) What are the powers of the above tests?

   [10+10=20 points]

3. An estimator T of θ is called the "best unbiased estimator", if $E(T) = \theta$, and $Var(T) \leq Var(T^*)$ where T^* is any other unbiased estimator of θ. Show that the best unbiased estimator is unique. (That is, if T_1 and T_2 are both best unbiased estimators of θ, they must be equal with probability 1). [Hint: Show that otherwise $(T_1 + T_2)/2$, which is also unbiased, will have a smaller variance and hence contradict the "best" qualification.] [20 points]

4. (a) Suppose that $X \sim N(\mu, \sigma^2)$. Show that $E(X^3) = \mu(\mu^2 + 3\sigma^2)$.

   (b) Suppose that $X_1, X_2, \ldots, X_n$ form a random sample from $N(\mu, 1)$. Show that $\bar{X}^2 - \frac{1}{n}$ is an unbiased estimator of μ^2.

   (c) Verify whether the variance of the estimator of μ^2 given in part (4b) attains the Cramer-Rao lower bound.

[5+5+10=20 points]

5. (a) Suppose $Y \sim N_p(\mu, \Sigma)$, where $\Sigma > 0$. Let B be real, symmetric matrix. Show that if $B\Sigma B = B$ and $\mu^T B\mu = 0$ and $Tr(B\Sigma) = k(< p)$, $Y^T BY$ has a chi-square distribution with k degrees of freedom. [Hint: Let $X = \Sigma^{-1/2}Y$. Rewrite $Y^T BY$ as a quadratic form in X.] [10 points]

   (b) Let X be $N_3(\mu, \Sigma)$ with $\mu = (-3, 1, 4)^T$ and

$$\Sigma = \begin{pmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

   Which of the following variables are independent? Explain.

   i. X_1 and X_2.

   ii. X_2 and X_2.

   iii. (X_1, X_2) and X_2.

   iv. $(X_1 + X_2)/2$ and X_3.

   v. X_2 and $X_2 - \frac{5}{2}X_1 - X_3$.

[2×5=10 points]

INDIAN STATISTICAL INSTITUTE

Analysis 2 : B. Stat 1st year
Mid-Semestral Examination: 2014-15
26, February , 2014.

Maximum Marks: 40 Maximum Time: $2\frac{1}{2}$ hrs.

Answer all the questions. But maximum you can score is 40.

(1) Let f and g be real valued continuous functions on $[a, b]$ such that

$$\int_a^b f(x)dx = \int_a^b g(x)dx.$$

Prove that there exists $c \in [a, b]$ such that $f(c) = g(c)$. [6]

(2) If $f : [0, 1] \to (0, \infty)$ is a continuous function then evaluate the limit:

$$\lim_{n\to\infty} \left(f\left(\frac{1}{n}\right) f\left(\frac{2}{n}\right) \dots f\left(\frac{n}{n}\right)\right)^{\frac{1}{n}}.$$

[6]

(3) Determine all values of $k \in \mathbb{R}$ for which the improper integral

$$\int_1^\infty \left(\frac{kt}{1+t^2} - \frac{1}{2t}\right) dt,$$

exists. [6]

(4) Let $f : [a, \infty) \to \mathbb{R}$ be such that $f \in R[a, b]$ for all $b > a$. If $\lim_{t\to\infty} |f(t)|^{\frac{1}{t}} = l$ then prove that

(a) If $l < 1$ then the improper integral $\int_1^\infty f(t)dt$ exists. [3]

(b) If $l > 1$ and $f \geq 0$ then the improper integral $\int_1^\infty f(t)dt$ does not exist. [3]

(5) Let $f : [0, 1] \to \mathbb{R}$ be a continuos function such that

$$|f(x)| \leq \int_0^x f(t)dt. \quad \text{for all } x \in [0, 1].$$

Prove that $f(x) = 0$ for all $x \in [0, 1]$. [8]

(6) Let $f : [a, b] \to (0, \infty)$ be a continuous function with $M = \text{Sup}\{f(x) \mid x \in [a, b]\}$.
Evaluate the limit $\lim_{n\to\infty} \left(\int_0^1 f(x)^n dx\right)^{1/n}$. [10]

INDIAN STATISTICAL INSTITUTE

B-Stat I, Second Semester, 2014-15

STATISTICAL METHODS II

Date: 23.04.2015 Semestral Examination Time: 3 hours

Total Point 100

1. Consider the problem of multiple linear regression where Y is the response variable, and $X_1, \ldots, X_k$ are the explanatory variables. Prove the following:

   (a) The correlation coefficient between the observed values Y_i and the predicted values $\hat{Y}_i$ (based on the above multiple linear regression fitted by the method of least squares) is r, where r is the multiple correlation coefficient between Y and $X_1, \ldots, X_k$. [You can use the form of the multiple correlation coefficient or its generalized versions derived in class.]

   (b) The proportion of variability explained by the multiple regression, defined as the ratio of the sum of squares due to regression to total sum of squares, is r^2, where r is as defined in 1(a).

   [10+10=20]

2. We are interested in simulating a random number from a Poisson(λ) distribution. Sequence of uniform (0, 1) random numbers $U_1, U_2, \ldots$ are selected, stopping at

$$N = \max\{n : \prod_{i=1}^{n} U_i > e^{-\lambda}\}.$$

   Show that this N has the required Poisson(λ) distribution. [16]

3. An estimator T is an efficient estimator for $\tau(\theta)$ if it is unbiased and if its variance attains the Cramer-Rao lower bound. Show that if T is efficient for $\tau(\theta)$, then T must be the maximum likelihood estimator of $\tau(\theta)$. [16]

4. Let ϕ_1 and ϕ_2 represent the $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ densities, and suppose that f is a two component mixture of the form

$$f(x) = (1 - \pi)\phi_1(x) + \pi\phi_2(x), \quad 0 < \pi < 1.$$

   Consider the following data, supposed to be have been generated from the above mixture.

$$-0.39, 0.12, 1.67, 3.72, 5.53.$$

(a) Write down the log likelihood based on the above data.

(b) What are the "incomplete" data and "complete" data in this context.

(c) Given current estimators of $\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1^2, \hat{\sigma}_2^2$, and $\hat{\pi}$ of $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$, and π, explain how to compute the next iterate using the Expectation-Maximization (EM) algorithm.

(d) Given the current estimates are $\hat{\mu}_1 = 4.5, \hat{\mu}_2 = 1.00, \hat{\sigma}_1^2 = 0.8, \hat{\sigma}_2^2 = 0.7, \hat{\pi} = 0.52$, find the updated estimates at the next stage. You may use the densities of each of the observations at either of the two normal distributions as given below.

Observation	$N(1.45, 0.8)$ density	$N(1.00, 0.7)$ density
-0.39	0.0100	0.0146
0.12	0.0719	0.1623
1.67	0.5876	0.3197
3.72	0.0012	1.65×10^{-7}
5.53	9.33×10^{-10}	2.24×10^{-19}

[4+4+8+8=24]

5. The strength of concrete depends on, to some extent, the method used for drying. Two different drying methods yielded the following results for independent tested specimens (in appropriate units).

$$n_1 = 7 \quad n_2 = 10$$
$$\bar{x}_1 = 3250 \quad \bar{x}_2 = 3240$$
$$s_1 = 210 \quad s_2 = 190$$

Test at level 0.05 to determine if the methods appear to produce concrete with different mean strengths.

[24]

Indian Statistical Institute

Vectors and Matrices II: B. Stat First Year: 2nd Semester Examination: 2014-15

Maximum Marks: 60 **Time: 3hrs** **Date: 27-04-2015**

1. Suppose that A and B are real matrices such that $A' = A$, $v'Av \geq 0 \ \forall v \in R^n$ and $AB + BA = 0$. Show that $AB = BA = 0$. Find nonzero A and B satisfying the conditions. (8)

2. Let A be an $n \times n$ real symmetric matrix with nonnegative entries. Prove that A has an eigenvector with nonnegative entries. (8)

3. Let Q be an $n \times n$ orthogonal matrix, and $f(t)$ be the characteristic polynomial of Q. Show that (for $t \neq 0$) $f(t) = \pm t^n f(1/t)$. (5)

4. Let A be an $n \times n$ Hermitian matrix satisfying $A^5 + A^3 + A = 3I$. Show that $A = I$. (8)

5. Consider the transformation given by the matrix

$$\begin{pmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

It is given that a rational canonical basis contains a disjoint union of cyclic bases of three and one elements. Find the form. (5)

6. Let T be a linear operator on a finite dimensional vector space V such that the characteristic polynomial splits. Show that the generalised eigenspace corresponding to an eigenvalue λ with multiplicity m is the null space of $(T - \lambda I)^m$. (8)

7. Let $H_u : V \to V$ given by $H_u(v) = v - 2 < v, u > u \ \forall v \in V$ be the Householder reflector, V being a finite dimensional real inner product space. Let w, z be two linearly independent vectors such that $\| w \| = \| z \|$. Show that there exists a unit vector $u \in V$ such that $H_u(w) = z$. (8)

8. Find the best approximation to a solution having minimum norm for the following system of linear equations.

$x_1 + x_2 + x_3 = 3$

$x_1 + x_2 - x_3 = 1$

$x_1 + x_2 + x_3 = 2$ (10)

9. Show that every operator T can be written as $S\sqrt{T^*T}$, where S is an isometry and T^* is the adjoint of T. (8)

INDIAN STATISTICAL INSTITUTE

Analysis 2 : B. Stat 1st year
End Semester Examination: 2014-15
MAY 01. 2015.

Maximum Marks: 50 Maximum Time: 3 hrs.

Answer all the questions. But maximum you can score is 50.

1.a) Let $f:[0,1]\to\mathbb{R}$ be a continuous function. Prove that, given $\epsilon>0$, there exists a step function $g:[0,1]\to\mathbb{R}$ such that $|f(x)-g(x)|<\epsilon$ for all $x\in[0,1]$. [6]

1.b) Let $f:[0,1]\to\mathbb{R}$ be a continuous function such that

$$\int_0^1 x^n f(x)dx = 0,$$

for all odd $n\in\mathbb{N}$. Prove that $f(x)=0$ for all $x\in[0,1]$. [8]

2.a) Prove that the improper integral

$$\int_1^\infty x\sin(x^4)dx,$$

converges. [6]

2.b) Define $f:[1,\infty)\to\mathbb{R}$ by defining

$$f(x)=\sum_{n=1}^{\infty}\frac{1}{n^2+x^2}.$$

Prove that there exist positive numbers C_1, C_2 (independent of x) such that

$$\frac{C_1}{x}\le f(x)\le\frac{C_2}{x},$$

for all $x\in[1,\infty)$. [8]

3. Let $f_n:[a,b]\to\mathbb{R}$ be a sequence of continuous functions such that $f_n(x)\le f_{n+1}(x)$ for all $x\in[a,b]$, $n\in\mathbb{N}$. If $\{f_n(x)\}$ converges to a continuous function $f(x)$ for all $x\in[a,b]$ then prove that $\{f_n\}$ converges uniformly to f on $[a,b]$. [8]

[P.T.O]

4.a) Compute the Fourier series of the 2π-periodic function $f(x) = |x|, |x| \leq \pi$. Use the Fourier series to prove that $\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$ and $\sum_{n \text{ odd} \geq 1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{8}$. [8]

4.b) Let f be a 2π-periodic continuous function with Fourier coefficients $\{\hat{f}(n)\}$ and let $S \subset \mathbb{Z}$ be a finite set. Define $a_n = \hat{f}(n)$ for $n \in \mathbb{Z} \setminus S$ and is zero otherwise. Are $\{a_n\}$ Fourier coefficients of a continuous 2π-periodic function? Justify your answer. [2]

5. Find the general solution of the differential equation $y'' + 6y' + 9y = f(x)$ where $f(x) = 1$, for $x \in [1, 2]$ and $f(x) = 0$ otherwise. [8]

Indian Statistical Institute
Semester 2 (2014-2015)
B. Stat 1st Year
Final Examination
Probability Theory 2

Tuesday 5.5.2015, 10:30-13:00 Total Points: $5 \times 14 = 70$

Answers must be justified with clear and precise arguments. More than one uncrossed answers to the same question or part of a question will not be entertained and only the first uncrossed answer will be graded. If you use any theorem or lemma proved in class state it explicitly.

1. If F is a distribution function then show that the following are also distribution functions (here h is a fixed positive number)

$$G(x) = \frac{1}{h}\int_{x}^{x+h} F(u)du, H(x) = \frac{1}{2h}\int_{x-h}^{x+h} F(u)du.$$

2. Suppose $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ are independent. Calculate the density $h(.)$ of $Z = X + Y$ using the formula

$$h(z) = \int_{-\infty}^{\infty} f(z-y)g(y)dy,$$

where f and g are the densities of X and Y respectively.

3. Consider the bivariate density of (X, Y) given by

$$f_{X,Y}(x,y) = C\frac{\sqrt{xy}}{x^2+y^2}$$

supported on the domain $D = \{(x,y) : x > 0, y > 0, xy < 1/2\}$. Find the constant C.

4. Suppose X and Y are jointly distributed as bivariate normal with means μ_X, μ_Y, variances σ_X^2, σ_Y^2, and correlation coefficient ρ. Prove that $\rho = \cos q\pi$ where $q = P[(X-\mu_X)(Y-\mu_Y) < 0]$.

5. Suppose X and Y are i.i.d. $U(0,1)$. Consider the pair (X, V) where $V = \max(X, Y)$. Write down the conditional distribution function $G(x, v)$ of $(X|V = v)$, and explain why for $A = (c, d], B = (c, d], 0 < c < d < 1$, one has $P(X \in A, V \in B) = \int_c^d [G(d, v) - G(c, v)] 2v dv$. Also find the conditional expectation $E(X|V = v)$.

INDIAN STATISTICAL INSTITUTE
B. Stat. I: 2014–2015
Numerical Analysis
Semestral Examination

Date: 08. 05. 2015 Marks: 100 Time: 3 Hours

Answer any five questions. Each question is of 20 marks. Please write all the part answers of a question at the same place. Scientific calculators are allowed in the examination hall.

1. Briefly describe (with suitable examples) the algorithms for finding a real root of polynomials using the following methods:
   (a) Bisection Method,
   (b) Regula Falsi,
   (c) Modified Regula Falsi,
   (d) Secant Method,
   (e) Newton's Method.

Marks : $20 = 5 \times 4$

2. Consider that the (x_i, y_i) values for a function $y = f(x)$ is given for $n + 1$ many points $i = 0$ to n.
   (a) Explain the Newton's interpolation method using forward differences.
   (b) How the expressions will look like when we consider equidistant points?

Marks : $20 = 10 + 10$

3. (a) Deduce the Lagrange's formula for interpolation.
   (b) Explain how this formula can be used for numerical differentiation.

Marks : $20 = 10 + 10$

4. (a) Deduce the Trapezoidal and Simpson's rule for numerical integration.
   (b) Provide example for each case using $\int_0^1 e^x dx$ considering the values of the function e^x at $x = 0.0, 0.2, 0.4, 0.6, 0.8, 1.0$.

Marks : $20 = 10 + 10$

5. Deduce the expression for the piecewise curves in a cubic spline.

Marks : 20

6. (a) Describe the Runge-Kutta method of order 4 and the Predictor-Corrector method for solving ordinary differential equations.

(b) Provide detailed calculation for the Runge-Kutta method considering the equation $y' = y^2$. Consider the initial value $y(0) = 1$ and you need to estimate $y(0.5)$.

Marks : 20 = 10 + 10

7. Given an arbitrary non-singular $m \times n$ real matrix A, the QR decomposition of the matrix satisfies $A = QR$, where Q is an $m \times n$ matrix with orthonormal columns, and R is an $n \times n$ upper-triangular matrix.

   (a) The first approach is to orthogonalize A by successive right multiplications with upper triangular matrices, such that $AR_1R_2 \cdots R_n = Q$, and $R = (R_1R_2 \cdots R_n)^{-1}$. Describe one such algorithm for QR decomposition.

   (b) The second approach is to triangularize A by successive left multiplications with orthonormal matrices, such that $Q_n \cdots Q_2Q_1A = R$, and $Q = (Q_n \cdots Q_2Q_1)^{-1}$. Describe one such algorithm for QR decomposition.

   (c) Which of the above two algorithms will you prefer for QR decomposition of an arbitrary non-singular $m \times n$ real matrix A? Justify.

Marks : 20 = 8 + 8 + 4

8. Given a polynomial $p(z) = z^m + a_{m-1}z^{m-1} + \cdots + a_2z^2 + a_1z + a_0$, the companion matrix A_p of $p(z)$ is defined as follows.

$$A_p = \begin{bmatrix} 0 & & & & & -a_0 \\ 1 & 0 & & & & -a_1 \\ & 1 & 0 & & & -a_2 \\ & & 1 & \ddots & & \vdots \\ & & & \ddots & 0 & -a_{m-2} \\ & & & & 1 & -a_{m-1} \end{bmatrix}$$

   (a) Is there a straight forward relationship between the roots of the polynomial $p(z)$ and the eigenvalues of the matrix A_p? Justify.

   (b) Is it possible to find the eigenvalues of A_p, for an arbitrary choice of $p(z)$, using simple row-column operations on the matrix? Justify.

   (c) Describe an algorithm of your choice to find the eigenvalues of A_p for an arbitrary $p(z)$, or to find eigenvalues of any arbitrary matrix A.

Marks : 20 = 4 + 4 + 12

9. It is a well known fact in linear algebra that any $m \times n$ real matrix A whose number of rows m is greater than or equal to its number of columns n, can be decomposed as a product $A = U\Sigma V^T$, where U is an $m \times n$ matrix with orthonormal columns, Σ is an $n \times n$ diagonal matrix containing the non-negative singular values of the matrix $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$, and V is an $n \times n$ matrix with orthonormal columns. This is known as the Singular Value Decomposition (SVD) of matrix A.

(a) Given the SVD of a square ($n \times n$) real matrix A, can you determine whether the matrix is singular or not? If the matrix is not singular, can you find the inverse A^{-1} using the SVD of A? Explain.

(b) Given the SVD of a square ($n \times n$) real singular matrix A, can you find the rank of A? Can you extend your argument to find the range and nullspace of A as well? Justify.

(c) Recall that an eigenvalue decomposition of a square ($m \times m$) symmetric real matrix M is given by $M = X\Lambda X^{-1}$, where Λ is a $m \times m$ diagonal matrix containing the eigenvalues as its diagonal elements, and X is an $m \times m$ matrix with orthonormal columns (eigenvectors). Suppose that a square ($n \times n$) real matrix A has the SVD decomposition $A = U\Sigma V^T$, where Σ is an $n \times n$ diagonal matrix containing the non-negative singular values of the matrix $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$. Using this given SVD of A, find an eigenvalue decomposition, and hence the eigenvalues, of the square ($2n \times 2n$) symmetric real matrix

$$H = \begin{bmatrix} 0 & A^T \\ A & 0 \end{bmatrix}.$$

Marks : 20 = 4 + 8 + 8

INDIAN STATISTICAL INSTITUTE

B-Stat I, Second Semester, 2014-15

STATISTICAL METHODS II

Date: 17.07.2015 Backpaper Examination Time: 3 hours

(Total Point 100)

1. (a) Define the multiple correlation coefficient $r_{1.23\ldots p}$ between x_1 and the set of variables $x_2, x_3, \ldots, x_p$. Show that

$$r^2_{1.23\ldots p} = \frac{\text{var}(X_{1.23\ldots p})}{\text{var}(x_1)}$$

where $X_{1.23\ldots p}$ is the value of x_1 predicted by the multiple regression equation of x_1 on $x_2, \ldots, x_p$.

(b) For a set of variables $x_1, x_2, \ldots, x_p$, define the partial correlation coefficient $r_{12.34\ldots p}$. For three variables x_1, x_2, x_3, show that

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1 - r_{13}^2}\sqrt{1 - r_{23}^2}},$$

where r_{ij} is the simple correlation coefficient between x_i and x_j.

[15+10=25 points]

2. (a) Suppose that you have a random observation U from the Uniform (0, 1) distribution at your disposal. Explain how you can draw a random observation from the Pareto distribution given by the density

$$f(x) = \frac{ca^c}{x^{c+1}}, \quad a, c \geq 0, \quad x \geq a.$$

What is the median of this distribution?

(b) Suppose you have a random number generator such that you can generate random numbers from any binomial distribution. Explain how you can drawn random numbers X and Y such that X is distributed as binomial (8, 2/3) and Y is distributed as binomial (18, 2/3), and the correlation between X and Y is 0.5.

[15+10=25 points]

3. Suppose that $X_1, \ldots, X_n$ form a random sample from a Poisson distribution with unknown mean θ and let $Y = \sum_{i=1}^{n} X_i$.

(a) Determine the value of a constant c such that the estimator e^{-cY} is an unbiased estimator $e^{-\theta}$.

(b) Find the variance of the estimator obtained in part (a).

(c) Use the information inequality to obtain a lower bound for the variance of the unbiased estimator found in part (a).

(d) Does the variance obtained in (b) attain the Crámer-Rao bound derived in (c)? If not, does the ratio of the bound obtained in part (c) and the variance obtained in part (b) goes to 1 as $n \to \infty$.

[7+7+6+5=25]

4. Consider the A-B-O blood group gene frequency data given by the vector of cell frequencies

$$\boldsymbol{y} = (n_O, n_A, n_B, n_{AB}) = (176, 182, 60, 17),$$

where the individual compoments of $\boldsymbol{y}$ are the frequencies of the corresponding phenotypes. The parameter vector is $\theta = (p, q, r)$, where the individual components of θ represent the proportion of the gene counts (that is count of the alleles A, B and O), with $p + q + r = 1$. We want to find the maximum likelihood estimators of the parameters, and want to use the EM algorithm for this purpose.

(a) Explain what would be the complete data in this example.

(b) Given estimates $p^{(k)}, q^{(k)}$ and $r^{(k)}$ at the k-th step of the iteration, explain how the E and M steps will be performed to get the parameter estimates at the $(k+1)$-th step.

(c) Given current estimates 0.264, 0.093 and 0.643 of p, q and r, find the parameter estimates after the next iteration.

[5+10+10=25]