LIFE CONTINGENCIES

Date: 3 September, 2012 Maximum Marks: 100 Time: 10:30 am

Duration: 3 hours

Note: (i) Desk calculators are allowed; (ii) Actuarial tables are allowed; (iii) Symbols and notations have their usual meaning. The entire question paper is for 120 marks.

1. If $P[X > x] = [1 - (x/100)]^{1/2}$, $0 \leq x \leq 100$, evaluate
   (a) ${}_{17}p_{19}$, [2]
   (b) ${}_{15}q_{36}$, [2]
   (c) ${}_{15|13}q_{36}$, [2]
   (d) μ_{36}, [2]
   (e) ~~$V[T(IA)]$.~~ $E(T(36))$ [2]

2. Calculate the following quantities from the AM92 tables ($i = 0.04$):
   (a) ${}_{3|3}q_{40}$, [2]
   (b) ${}_{3|}q_{[39]+1}$, [2]
   (c) Variance of the present value random variable whose mean is $A_{[39]}$, [2]
   (d) ${}_{10|}\ddot{a}_{40}$, [2]
   (e) $(IA)_{40:\overline{10}|}$. [2]

3. Give an expression for ${}_tp_x$ in terms of t and p_x for $0 \leq t \leq 1$ and $x = 1, 2, 3, \ldots$, when interpolation between integer ages is made
   (a) by assuming constant force of mortality, and [3]
   (b) by using the Balducci assumption. [4]

4. Prove and interpret the following relations.
   (a) $a_{x:\overline{n}|} = {}_1E_x \ddot{a}_{x+1:\overline{n}|}$. [4]
   (b) $A_{x:\overline{n}|} = v\ddot{a}_{x:\overline{n}|} - a_{x:\overline{n-1}|}$. [6]

5. Show algebraically that, under the assumption of a uniform distribution of death over the insurance year of age, $A_x^{(m)} = \frac{i}{i^{(m)}} A_x$. [6]

6. If $A_x = 0.25$, $A_{x+20} = 0.40$ and $A_{x:\overline{20}|} = 0.55$, calculate $A^1_{x:\overline{20}|}$ and $A_{x:\overline{20}|}^{\ \ 1}$. [7]

7. A 2-year term assurance policy is issued to a male aged x. The benefit amount is 100 if the life dies in the first year, and 200 if the life dies in the second year. Benefits are payable at the end of the year of death.

   (a) Give an expression for the present value random variable for the said benefit. [3]

   (b) Calculate the variance of the present value random variable assuming that $q_x = 0.025$, $q_{1+x} = 0.030$ and $i = 0.06$. [6]

8. By considering a term assurance policy as a series of one year deferred term assurance policies, show that, under the assumption of uniform distribution of death,

$$\overline{A}^1_{x:\overline{n}|} = \frac{i}{\delta} A^1_{x:\overline{n}|}.$$

   By using this relation, calculate the expected present value and variance of the present value of a term assurance of 1 payable immediately on death for a life aged 40 exact, if death occurs within 30 years.

   Basis
   Interest: 4% per annum Mortality: AM92 select [5+6]

9. Using the assumption of a uniform distribution of deaths in each year of age and the AM92 Ultimate life table with interest at the effective annual rate of 6%, calculate (a) $\overline{a}_{40}$, (b) $\overline{a}_{40:\overline{30}|}$. [4+4]

10. A life insurance company issues a 10-year decreasing term assurance benefit to a man aged 50 exact. The death benefit is 100,000 in the first year, 90,000 in the second year and decreases by 10,000 each year so that the benefit in the 10th year is 10,000. The death benefit is payable at the end of the year of death.

    Level premiums are payable monthly in advance for the term of the policy, ceasing at earlier death. Calculate the annual premium.

    Basis
    Interest: 6% per annum Mortality: AM92 select [7]

11. If $P^{1\ (12)}_{x:\overline{20}|} = 1.032 P^1_{x:\overline{20}|}$ and $P_{x:\overline{20}|} = 0.040$, what is the value of $P^{(12)}_{x:\overline{20}|}$? [6]

12. Express $A_{40} P_{40:\overline{25}|} + (1 - A_{40}) P_{40}$ as an annual benefit premium. [6]

13. Which of the following are correct formulas for ${}_{15}V_{40}^{(m)}$?

(a) $\left(P_{55}^{(m)} - P_{40}^{(m)}\right) \ddot{a}_{55}^{(m)}$

(b) $\left(1 - \frac{P_{40}^{(m)}}{P_{55}^{(m)}}\right) A_{55}$

(c) $1 - \frac{\ddot{a}_{55}^{(m)}}{\ddot{a}_{40}^{(m)}}$ [7]

14. Write the prospective formula for the benefit reserve required at the end of 5 years for a unit benefit 10-year term insurance issued to (45) on a single premium basis. [6]

15. A life insurance company issues the following policies:

- 15-year term assurances with a sum assured of Rs. 150,000 where the death benefit is payable at the end of the year of death,
- 15-year pure endowment assurances with a sum assured of Rs.75,000.

On 1 January 2002, the company sold 5,000 term assurance policies and 2,000 pure endowment policies to male lives aged 45 exact. Premiums are payable annually in advance. During the first two years, there were fifteen actual deaths from the term assurance policies and five actual deaths from the pure endowment policies.

(a) Calculate the death strain at risk for each type of policy during 2004. [5]

(b) During 2004, there were eight actual deaths from the term assurance policies written and one actual death from each of the other two types of policy written. Calculate the total mortality profit or loss to the office in the year 2004.

Basis

Interest: 4% per annum Mortality: AM92 Ultimate [5]

INDIAN STATISTICAL INSTITUTE
MID-SEMESTRAL EXAMINATION 2012-2013

M.STAT 2nd year. Advanced Design of Experiments

September 3, 2012, Total marks 40 Duration: Two hours
Answer all questions.
Keep your answers brief and to the point.

1. a) Define mutually orthogonal Latin squares.

   b) Let p be a prime number. Consider the squares A_j, $j = 1, \ldots, p-1$, constructed as follows:

$$\begin{array}{ccccc} 0 & 1 & 2 & \ldots & p-1 \\ j & 1+j & 2+j & \ldots & p-1+j \\ 2j & 1+2j & 2+2j & \ldots & p-1+2j \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ (p-1)j & 1+(p-1)j & 2+(p-1)j & \ldots & (p-1)+(p-1)j \end{array}$$

   where all entries in A_j are reduced mod p. Prove that the above squares form a set of mutually orthogonal Latin squares.

   (c) Construct two mutually orthogonal Latin squares of order 8.

   (d) Hence or otherwise, **indicate** how to construct an orthogonal array $OA(64, 8, 8, 2)$. (Actual array need not be constructed) [2+4+4+4=14]

2. a) Define a Hadamard matrix.

   b) Can there exist a Hadamard matrix of order 15? Justify your answer with a proof.

   c) Prove that the existence of a Hadamard matrix of order N is equivalent to the existence of an OA(N, $N-1$, 2, 2). [2+5+6=13]

3. a) Describe an experimental situation where you have to compare 4 treatments and you would use a row-column design. Justify your answer.

   b) Give the model for analysing data from an experiment conducted using your design in (a) above.

   c) Write down the information matrix for (a) above under the model in (b). [4+4+5=13]

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination: 2012-13 (First Semester)
Master of Statistics (M. Stat.) II Year
Advanced Probability I

Teacher: Parthanil Roy

Date– 04.09.12

Maximum Marks: 40 **Duration: 2 hours**

Note:

- Please write your name and roll number on top of your answer booklet(s).
- There are four problems with a total of 40 points. Show all your works and write explanations when needed.
- This is an open note examination. You are allowed to use your own hand-written notes (such as class notes, exercise solutions, list of theorems, formulas etc.). Please note that no printed or photocopied materials are allowed. In particular, you are not allowed to use books, photocopied class notes etc. If you are caught using any, you will get a zero in the mid-semestral examination.

1. (5 points) Let P and Q be two probability measures on a measurable space $(\Omega, \mathcal{A})$ such that for each $\epsilon > 0$ there exists $A \in \mathcal{A}$ with $Q(A) < \epsilon$ and $P(A) > 1 - \epsilon$. Show that P and Q are mutually singular.

2. (8 points) Suppose X is an integrable random variable defined on a probability space $(\Omega, \mathcal{F}, P)$ and $\{\mathcal{G}_n\}_{n\geq 1}$ is a sequence of decreasing sub σ-fields of $\mathcal{F}$ such that $Y_n := E(X|\mathcal{G}_n)$ converges (to Y, say) in $L^1(\Omega, \mathcal{F}, P)$, i.e., $E|Y_n - Y| \to 0$. Show that $Y = E(X|\bigcap_{n\geq 1}\mathcal{G}_n)$.

3. (12 points) Let $(\Omega, \mathcal{A}, P)$ be a probability space and $\mathcal{C}$ be a countably generated sub σ-field of $\mathcal{A}$. Suppose $Q(A, \omega)$ is a regular conditional probability on $\mathcal{A}$ given $\mathcal{C}$ induced by P. Show that there exists $N \in \mathcal{C}$ with $P(N) = 0$ such that for each $\omega \notin N$, $Q(\cdot, \omega)$ is concentrated on the $\mathcal{C}$-atom containing ω.

4. (15 points) Suppose for each $i \geq 1$, μ_i and ν_i are two probability measures on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that $\nu_i \ll \mu_i$. Is it true that for all $k \geq 1$, $\otimes_{i=1}^{k} \nu_i \ll \otimes_{i=1}^{k} \mu_i$? Justify your answer. Show (with an example) that $\otimes_{i=1}^{\infty} \nu_i$ and $\otimes_{i=1}^{\infty} \mu_i$ can even be mutually singular.

Wish you all the best

INDIAN STATISTICAL INSTITUTE

Mid-semester exam. (Semester I: 2012-2013)

Course Name: M. Stat. 2nd year

Subject Name: Analysis of discrete data

Date: 04 Sep., 2012, Maximum Marks: 30. Duration: 1 hr. 30 min.

Note: Answer all questions.

1. Find the sample values of the measures of association Goodman-Kruskal's τ for the following three tables.

(a)

1	3	10	6
2	3	10	7
1	6	14	12
0	1	9	11

(b)

1	6	14	12
0	1	9	11
1	3	10	6
2	3	10	7

(c)

12	1	6	14
11	0	1	9
6	1	3	10
7	2	3	10

Comment on the three values.

[6+2]

2. (a) Derive the joint asymptotic distribution of log odds ratios in a 2×4 contingency table. [7]

(b) Test for independence for the following table by both conditional and unconditional test procedures. What happens if we use large sample approximation for log odds ratio? When do you recommend large sample approximation for log odds ratio in a 2×2 table?

	Guess poured first	
Poured first	Milk	Tea
Milk	3	2
Tea	1	4

[5+5+3+2]

INDIAN STATISTICAL INSTITUTE
Mid-Semester Examination: 2012-2013, First Semester
M-Stat II and M-Math I
Set Theory and Topology

Date: 06.09.12 **Max. Marks 50** **Duration:** $2\frac{1}{2}$ **Hours**

Note: Answer all questions.
You must state clearly any result that you may be using.

1. a) Show that if X is infinite and $A \subset X$ is finite, then $X - A$ and X have the same cardinality.

   b) Let X and Y be sets such that there is a map from X onto Y. Show that $Y \leq_c X$.

   [4+4]

2. Let Y be a totally ordered set with order topology.

   a) Let $y_1,\ y_2\ \in Y$. What are the conditions under which the closed interval $[y_1, y_2]$ is an open subset of Y?

   [4]

   b) Let X be any topological space and $f,\ g : X \to Y$ be continuous.

   i) Show that $\{x : f(x) \leq g(x)\}$ is closed in X.

   ii) Let $h : X \to Y$ be the function $h(x) = \min\{f(x), g(x)\}$. Show that h is continuous.

   [5+5]

P.T.O.

3. Let $\{X_\alpha\}$ be a family of topological spaces. Let $A_\alpha \subset X_\alpha$, $\forall\ \alpha$.

a) i)Show that if A_α is closed in X_α then $\prod_\alpha A_\alpha$ is closed in $\prod_\alpha X_\alpha$ with product topology.
ii) Is it closed in $\prod_\alpha X_\alpha$ with box topology?

[3+3]

b) i) Show that $\overline{\prod_\alpha A_\alpha} = \prod_\alpha \overline{A_\alpha}$ in $\prod_\alpha X_\alpha$ with product topology.
ii) Does the equality hold in box topology?

[3+3]

4. a) Define **quotient map**. Let $p : X \to Y$ be a surjective map. Define **saturated subset** of X *with respect to* p.

Show that p is a quotient map iff p is continuous and p maps saturated open sets (with respect to p) of X to open sets of Y.

[2+2+5]

b) Let $Y = \mathbf{R} \times \{0\} \cup \{0\} \times \mathbf{R} \subset \mathbf{R^2}$. Give $\mathbf{R^2}$ Euclidean metric topology. Define $g : \mathbf{R^2} \to \mathbf{Y}$ by the equations

$$g((x,y)) = (x,0) \text{ if } x \neq 0$$
$$g((0,y)) = (0,y)$$

i) Is g continuous if Y has subspace topology?

ii) Show that in the quotient topology induced by g, the space Y is not Hausdorff.

[6+6]

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination : (2012-2013)

M.Stat. 2nd Year

TOPICS IN BAYESIAN INFERENCE

Date: 7 September, 2012 Max. Marks: 90 Duration: $2\frac{1}{2}$ Hours

Answer as many questions as you can. Maximum you can score is 90.

1. What is the difference between the Bayesian paradigm and classical inference with respect to evaluation of performance of a decision rule.

Give an example showing the paradoxical behaviour of an inference procedure based on averaging over the sample space. [4+5=9]

2. (a) Describe how the posterior distribution can be used for estimation of a real parameter. How do you measure the accuracy of an estimate?

(b) Consider the problem of estimation of a real parameter θ. Given a loss function $L(\theta, a)$, how does one find an optimum estimate in the Bayesian paradigm and in classical Statistics? [5+6=11]

3. What is a conjugate prior? Give an example to show that a conjugate prior can be interpreted as additional data. [5]

4. (a) Let $X_1, \ldots, X_n$ be i.i.d. with a common density $f(x|\theta)$ where $\theta \in R$. State the result on asymptotic normality of posterior distribution of suitably normalized and centered θ under suitable conditions on the density $f(\cdot|\theta)$ and the prior distribution.

(b) Consider i.i.d. observations with a common distribution involving an unknown real parameter θ. Assuming that the usual regularity conditions hold, find a large sample approximation to a $100(1-\alpha)$ % HPD credible interval for θ. [5+5=10]

5. Let $X_1, \ldots, X_n$ be i.i.d. $N(\theta, \sigma^2)$ variables.

(a) Consider a standard noninformative prior for (θ, σ^2) and find the corresponding $100(1-\alpha)\%$ HPD credible set for θ.

(b) Assume that σ^2 is known and consider a conjugate prior for θ. Find the posterior distribution of θ and the posterior predictive distribution of a future observation X_{n+1}. [(8)+(4+7)=19]

6. Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be independent random samples, respectively, from $N(\mu, \sigma_1^2)$ and $N(\mu, \sigma_2^2)$, where both σ_1^2 and σ_2^2 are known. Construct a $100(1-\alpha)\%$ HPD credible interval for the common mean μ assuming a uniform prior. Compare this with the frequentist $100(1-\alpha)\%$ confidence interval for μ. [8]

7. Let $X \sim N(\theta, 1)$ where θ is known to be nonnegative. Find the Bayes estimate of θ (in its simplest form) for squared error loss using the standard noninformative prior. (Express the estimate in terms of standard normal density and c.d.f.) [6]

8. Let $X_1, \ldots, X_n$ be i.i.d. $\sim N(\theta, 1)$. Consider the problem of testing $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$. A classical test rejects H_0 if $T = \sqrt{n}(\bar{X} - \theta_0)$ is large. Let t be the observed value of T. Find the P-value (in terms of t) for this problem.

Consider now the uniform prior $\pi(\theta) \equiv 1$. Find the posterior probability of H_0 (in terms of t) and compare it with the P-value. [6+7=13]

9. Consider observations $X_1, \ldots, X_n$, where

$$\begin{aligned} X_i|\theta_i &\sim N(\theta_i, \sigma^2), \quad i = 1, \ldots, n, \text{ independent} \\ \theta_i &\sim N(\mu, \tau^2), \quad i = 1, \ldots, n, \text{ independent.} \end{aligned}$$

Show that the marginal distribution of X_i is $N(\mu, \sigma^2 + \tau^2)$ and that marginally, $X_1, \ldots, X_n$ are i.i.d. (assume σ^2 to be known). [9]

10. Show that the result on asymptotic normality of the posterior distribution of $\sqrt{n}(\theta - \hat{\theta}_n)$, stated in the class, implies consistency of the posterior distribution of θ at θ_0. [10]

INDIAN STATISTICAL INSTITUTE
MID-SEMESTRAL EXAMINATION, 2012-2013
M.Stat. II year and M.S.(Q.E.) II year
Econometrics Methods/Econometric Methods II

Date: 07.09.12 Maximum Marks: 50 Time: 2 hours

Answer **question No. 1** and **any two** from the rest.
Marks allotted to each question are given within parentheses.

1. (a) Explain the nature of dependence implied by an ARCH (q) process.

   (b) Find the unconditional variance of a GARCH (p, q) process in terms of its parameters. Also discuss the theoretical implication(s) of the case when the unconditional variance is infinite.

[4+6 = 10]

2. (a) Find the unconditional fourth-order central moment of a GARCH (1,1) process and hence obtain its kurtosis coefficient as

$$\frac{3\left[1-(\alpha_1+\beta_1)^2\right]}{1-(\alpha_1+\beta_1)^2-2\alpha_1^2}$$

(*Notations have their usual meanings.*)

Is the value of kurtosis coefficient always greater than 3? Justify.

   (b) Show that for a simple ARCH – M regression model specified as

$$y_t = \xi + \delta h_t + \varepsilon_t,\ \varepsilon_t \mid \Psi_{t-1} \sim N(0, h_t),\ h_t = \alpha_0 + \alpha_1\, \varepsilon_{t-1}^2$$

where Ψ_{t-1} is the information set at $t-1$, $\alpha_0 > 0$ and $\alpha_1 \geq 0$,

$$\text{Corr}\ (y_t, y_{t-1}) = \frac{2\alpha_1^3\delta^2\alpha_0}{2\alpha_1^2\delta^2\alpha_0 + (1-\alpha_1)(1-3\alpha_1^2)}.$$

[8 + 12 = 20]

3. (a) Explain what is meant by 'leverage effect'. Is the GARCH volatility model capable of capturing this effect? Explain.

   (b) State the EGARCH model, and explain how it can incorporate this effect.

(c) In carrying out a test for the null hypothesis of 'no conditional heteroscedasticity' against the alternative of nonlinear GARCH (NGARCH) model, do you think that you would face any statistical problem(s)? Give explanations for your answer. In case your answer is in the affirmative, explain of nature of the problem.

[6 + 6 + 8 = 12]

4. (a) Suppose that a time series $\{y_t\}$ follows an ARMA (k, l) process where the error $\{\varepsilon_t\}$ follows a GARCH (p, q) process. Find the optimal h-period ahead point forecast of y_t at origin t. Does the presence of ARCH affect the way in which the point forecast is constructed? Justify your answer.

(b) Find the conditional variance of the forecast error in terms of the parameters involved.

[12 + 8 = 20]

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination: (2012-2013)

MS(QE) I & MSTAT II

Microeconomic Theory I

Date: 10.09.2012 **Maximum Marks:** 40 **Duration:** 2 hrs.

Note: Answer all questions.

(1) (a) Consider a rational preference relation R on X. Show that if $u(x) = u(y)$ implies that xIy and $u(x) > u(y)$ implies that xPy, then $u(.)$ is a utility function representing R. **(2)**

(b) Show that a choice structure $(\mathcal{B}, C(.))$ for which a rationalizing preference relation exists, satisfies the path-invariance property: For every pair $B_1, B_2 \in \mathcal{B}$ such that $B_1 \cup B_2 \in \mathcal{B}$ and $C(B_1) \cup C(B_2) \in \mathcal{B}$, we have $C(B_1 \cup B_2) = C(C(B_1) \cup C(B_2))$. **(10)**

(2) Define the weak axiom of revealed preference for the market economy. Show that if the Walrasian demand function $x(p, w)$ is homogeneous of degree zero and satisfies Walras' law, then the weak axiom of revealed preference holds if and only if it holds for all compensated price changes. **(1+13=14)**

(3) Define lexicographic preferences. Show that lexicographic preferences satisfy completeness, transitivity, strong monotonicity and strict convexity. Also show that there is no utility function that can represent lexicographic preferences. **(1+8+5=14)**

INDIAN STATISTICAL INSTITUTE

Semestral Examination

M. Stat. - II Year (Mid-Semester - I)

Graph Theory and Combinatorics

Date : 11.9.12 Maximum Marks : 50 Duration : 3:00 Hours

Note : You may answer any part of any question, but maximum you can score is 50.

1. (i) Prove that if $A = R(k, m-1)$ and $B = R(k-1, m)$ are both even then $R(k, m) \leq A + B - 1$.

   (ii) Find, with proof, the value of R(4,3).

   [10+10=20]

2. We call a hypergraph *2-colorable*, if its vertices can be assigned 2 colors so that every hyperedge contains both colors.

   Prove that any r-uniform hypergraph with less than 2^{r-1} hyperedges is 2-colorable. [20]

3. If S is a set of n points in the plane with no pair more than distance 1 apart, then the maximum number of pairs of points more than distance $1/\sqrt{2}$ apart is $\lfloor n^2/3 \rfloor$. [20]

4. Suppose that G is a triangle-free simple n-vertex graph such that every pair of nonadjacent vertices has exactly two common neighbors.

   (i) Prove that G is regular. (A graph is regular if degrees of all its vertices are same.)

   (ii) Given that G is regular of degree k, prove that the number of vertices in graph G is $1 + \binom{k+1}{2}$.

   [20]

5. Suppose $a_1 < a_2 < \ldots < a_k$ are distinct positive integers. Prove that there is a simple graph with $a_k + 1$ vertices whose *set* of distinct vertex degrees is $a_1, a_2, \ldots, a_k$. (Hint: use induction on k to construct such a graph.) [20]

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination : Semester I (2012-13)

M. Stat. II Year

Actuarial Methods

Date: 12.09.2012 Maximum marks: 50 Time: 2 hours 15 minutes

Calculator and Actuarial table can be used. Answer as many as you can. Total mark is 56.

1. The profit per client-hour made by a privately owned health centre depends on the variable cost involved. Variable cost, over which the owner of the health centre has no control, takes one of the three levels θ_1 = high, θ_2 = medium and θ_3 = low. The owner has to decide at what level to set the number of client-hours that can be either $d_1 = 16,000$, $d_2 = 13,400$ or $d_3 = 10,000$. The profit (in Rs.) per client-hour is as follows:

	θ_1	θ_2	θ_3
d_1	85	95	110
d_2	105	115	130
d_3	125	135	150

Determine the minimax solution. Given the probability distribution $p(\theta_1) = 0.1$, $p(\theta_2) = 0.6$, $p(\theta_3) = 0.3$, determine the solution based on the Bayes criterion. [3+5=8]

2. An actuarial student observes that the size of claims follows a Gamma distribution with parameters a and l having density

$$f(x; a, l) = \frac{l^a}{\Gamma(a)} e^{-lx} x^{a-1}.$$

Past experience suggests that while a is known, l is unknown and has a prior distribution that can be modeled as exponential with mean m. The actuarial student has obtained recent claims data $x_1, \cdots, x_n$, where x_i is the size of the i th claim.

(a) Derive the posterior distribution of l.

(b) Determine the Bayesian estimate of l under zero-one loss and quadratic loss.

(c) If $a = 100$, $m = 10$ and the student observes that the last 10 claims total 250, calculate the Bayesian estimate of l under both zero-one and quadratic loss.

[4+(2+2)+2=10]

3. The last ten claims (in rupees) under a particular class of insurance policy were: 1330 , 201 , 111 , 2368 , 617 , 309 , 35 , 4685 , 442 , 843.

(a) Assuming that the claims come from a log-normal distribution with parameters μ and σ^2, find the maximum likelihood estimates of these parameters using the observed data.

P.T.O.

(b) Assuming that the claims come from a Pareto distribution with parameters λ and α, use the method of moments to estimate these parameters.

(c) If the insurance company takes out reinsurance cover with an individual excess of loss of Rs. 3,000, estimate the percentage of claims that will involve the re-insurer under each of the two models above.

[4+4+4=12]

4. Describe compound Poisson distribution in the context of general insurance and derive its moment generating function using that of the claim distribution. [2+2=4]

5. (a) Mention two characteristics of an insurable risk.

(b) Describe employers' liability.

(c) Mention two differences between collective and individual risk models.

(d) Consider an XOL arrangement with retention limit M. The re-insurer, having extensive experience in this line of business, suggests that 70% of the claims are exponentially distributed with mean 4 and 30% of claims are exponentially distributed with mean 10. Determine the probability that a randomly selected claim will need to go to the re-insurer.

(e) Briefly explain model heterogeneity and model uncertainty in the context of general insurance by means of examples.

[2+2+2+2+4=12]

6. The ISI employees are covered by a group life insurance which pays a specific benefit amount (in Rs.) if an employee dies while in service. There are two categories of employees who are entitled to the following benefit amounts with corresponding probability of dying during a year:

Category	No. of employees	Benefit amount	Prob. of dying
Active	1250	50,000	0.008
Affiliated	250	20,000	0.012

Using individual risks model, calculate the mean and variance of the aggregate claim amount during a year. Find the probability that the aggregate claim amount in a given a year will exceed Rs 1000,000. What loading factor should be used to fix the premium to be 99% sure of making a profit in this portfolio? Assume a normal approximation whenever necessary.

[2+3+2+3=10]

Date: 12.9.2012
Time: 2 hours

INDIAN STATISTICAL INSTITUTE

Statistical Methods in Genetics – I
M-Stat (2nd Year) 2012-2013
Mid-Semester Examination

This paper carries 40 marks.

1. Consider the following genotype data at a biallelic locus on 200 randomly chosen individuals in each of three populations:

Genotype	Population 1	Population 2	Population 3
AA	70	80	75
AB	100	95	95
BB	30	25	30

Do the above data provide evidence that the allele frequencies at this locus differ across the three populations? [10]

2. Suppose, in every generation of a certain population, a fraction α practises self mating, while the remaining $(1-\alpha)$ fraction of the population practises random mating. The initial genotype frequencies at a biallelic locus in this population are D_0, H_0 and R_0. Examine whether the genotype frequencies reach equilibria. If so, what are the equilibrium values? If not, provide suitable justification. [15]

3(a) What is the probability that a pair of first cousins are both homozygous at an autosomal biallelic locus? [You need to show all computations explicitly]

(b) Give an example to show that the marginal effect of either mutation or selection may not result in non-trivial equilibrium values of alleles frequencies at a locus, but their joint effect may result in non-trivial equilibrium values. [10 +5]

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination: (2012–2013)

M. Stat 2nd Year

Statistical Computing

Date: 14/09/12 Marks: 30 Duration: 3 hours

Attempt all questions

1. (a) Let $X_1, \ldots, X_m$ be an *iid* sample from a normal density with mean μ and variance σ^2. Suppose for each X_i we observe $Y_i = |X_i|$ rather than X_i. Formulate an EM algorithm for estimating μ and σ^2.

(b) Let

$$X_1, \ldots, X_n \sim f(x \mid \theta) = \frac{1}{\pi\left\{1 + (x-\theta)^2\right\}}.$$

Derive an EW algorithm to obtain the MLE of θ. (*Hint: You may use the fact that the ratio of independent standard normal random variables have the standard Cauchy distribution.*)

Marks: 5+5=10

2. (i) Show that the EM algorithm is a special case of the MM algorithm.

(ii) In the context of maximum likelihood estimation in multinomial distribution show that the MM algorithm converges to the maximum likelihood estimate at a linear rate.

(iii) Develop an MM algorithm for minimizing the function

$$f(x_1, x_2) = \frac{1}{x_1^3} + \frac{3}{x_1 x_2^2} + x_1 x_2.$$

Marks: 2+3+5=10

3. (i) For nodes $x_0 < x_1 < \cdots < x_n$ and function values $f_i = f(x_i)$, develop a quadratic interpolating spline $s(x)$ satisfying

(a) $s(x)$ is a quadratic polynomial on each interval $[x_i, x_{i+1}]$,

(b) $s(x_i) = f_i$ at each node x_i,

(c) the first derivative $s'(x)$ exists and is continuous throughout the entire interval $[x_0, x_n]$.

Do you require any additional information to completely determine the spline?

(ii) If the function $f(x)$ is integrable, then show that its Fourier transform $\hat{f}(y)$ is bounded, continuous, and tends to zero as $|y|$ tends to infinity.

Marks: 7+3=10

INDIAN STATISTICAL INSTITUTE
First Semestral Examination: 2012-13
M. Stat. II Year
Topics in Bayesian Inference

Date: 16/11/2012 **Maximum Marks: 100** **Duration: 3 Hours**

This question paper carries 110 points.
Answer as many questions as you can. The maximum you can score is 100.

1. (a) Consider the problem of model selection with two competing models. Suppose we want to use noninformative priors which are improper. Describe a suitable model selection procedure. Also describe the intrinsic Bayes factor in this context.

   (b) What is an intrinsic prior in the context of nonsubjective Bayes testing ? Suppose we have observations $X_1,\ldots,X_n$. Under model M_0, X_i are i.i.d. $N(0,1)$ and under model M_1, X_i are i.i.d. $N(\theta,1), \theta \in R$. Consider the noninformative prior $g_1(\theta) \equiv 1$ for θ under M_1. Find the intrinsic prior for θ corresponding to the AIBF and show that the ratio of the AIBF and the BF with this intrinsic prior tends to one as n tends to infinity.

[12+(3+10)=25]

2. Consider the linear regression model $y = X\beta + \varepsilon$ where $y = (y_1,\ldots,y_n)'$ is the vector of observations on the "dependent" variable, $X = (x_{ij})_{n\times p}$ is of full rank, x_{ij} being the values of the nonstochastic regressor variables, $\beta = (\beta_1,\ldots,\beta_p)'$ is the vector of regression coefficients and the components of ε are independent, each following $N(0,\sigma^2)$ Consider the non informative prior $\pi(\beta,\sigma^2) \propto \frac{1}{\sigma^2}$, $\beta \in \mathrm{R}^p, \sigma^2 > 0$.

   a) Show that the marginal posterior distribution of β is a multivariate t distribution.

   b) Find a $100\ (1-\alpha)\%$ *HPD* credible set for β.

[9+14=23]

3. Consider p independent random samples, each of size n, from p normal populations $N(\theta_j,\sigma^2)$ j = 1,2,...,p. Assume σ^2 to be known. Also assume that θ_i are i.i.d. $N(\eta_1,\eta_2)$. Our problem is to estimate $\theta_1,\ldots,\theta_p$. Describe the Hierarchical Bayes and the parametric empirical Bayes approaches in this context. Derive the James-Stein estimate as a PEB estimate.

[15+7=22]

4. Consider the hierarchical Bayesian model where we have k independent random samples $(y_{i1}, y_{i2},\ldots,y_{in_i})$, $i = 1, \ldots, k$, from k normal populations,

   $y_{ij} \sim N(\theta_i,\sigma_i^2)$, $j = 1,\ldots,n$, $i = 1,\ldots,k$,

   θ_i are i.i.d. $N(\mu_\pi,\sigma_\pi^2)$

   σ_i^2 are i.i.d. Inverse – Gamma (a_1,b_1),

   $(\theta_1,\ldots,\theta_k)$ and $(\sigma_1^2,\ldots,\sigma_k^2)$ are independent and the second stage priors on μ_n and σ_π^2 are $\mu_\pi \sim \mathrm{N}(\mu_0,\sigma_0^2)$ and $\sigma_\pi^2 \sim$ Inverse – Gamma (a_2,b_2).

Assume that a_1, a_2, b_1, b_2 μ_0 and σ_0^2 are specified constants. Describe how you can find estimates of $\theta_1, \ldots, \theta_k$ using Gibbs sampler. Derive the required full conditional distributions.

[25]

5. Suppose we have observations $X_1, \ldots, X_n$. Under model M_0, X_i are i.i.d. $N(0,1)$ and under model M_1, X_i are i.i.d. $N(\theta,1)$, $\theta \in \mathrm{R}$. As there are difficulties with improper noninformative prior, one may like to use a uniform prior over $[-K, \mathrm{K}]$ for a large K under M_1.

(a) Explain why it is not a good solution to the problem.

(b) Will there be a conflict between P-value and the posterior probability of M_0 ? Justify your answer.

[8+7=15]

*****

INDIAN STATISTICAL INSTITUTE

First Semestral Examination: 2012-13

M.STAT 2nd year. Advanced Design of Experiments. Date- 21.11.12

Total marks 65. Maximum you can score is 60 Duration: Three hours

Answer all questions.

Keep your answers brief and to the point.

1. State whether each of the following statements is True or False and give **brief** justifications in each case: [$5 \times 3 = 15$]

   a) A Hadamard matrix of order 2^t exists for all $t \geq 2$.

   b) An orthogonal array $OA(49, 5, 7, 2)$ does not exist.

   c) For a 2^5 experiment, if interactions involving 4 or more factors are absent then a design given by an $OA(32, 5, 2, 4)$, with its 32 rows interpreted as the treatment combinations, will necessarily allow the estimability of all main effects and two-factor and three-factor interactions.

   d) A balanced uniform crossover design with 8 treatments, 8 subjects and 8 periods does not exist.

   e) If a design d is A-optimal for inferring on a parameter θ in a class $\mathcal{D}$, then it need not remain A-optimal in $\mathcal{D}$ for inferring on a non-singular transformation of θ.

2. a) Indicate the construction of a Hadamard matrix of order 12 after clearly stating the result you use to construct this. (proof of result not required). Write down the first two rows of this Hadamard matrix explicitly.

   Hence, or otherwise, show that $OA(12, 11, 2, 2)$ and $OA(24, 12, 2, 3)$ both exist.

   b) Prove that the existence of a Hadamard matrix of order 12 implies the existence of a symmetric BIB design. Write down the parameters of this BIB design. [(5+2)+3=10]

3. a) Define E-optimality and explain its statistical significance.

   b) Prove that a BIB design with parameters v, b, k will be E-optimal for treatment effects in the class of all block designs with parameters v, b, k.

   c) Prove that a regular generalized Youden Square design is universally optimal in a suitable class of row-column designs. Where is the 'regularity' needed in your proof?

   [(1+2)+3+4=10]

4. a) Define a strongly balanced crossover design with t treatments, n subjects and p periods. What restrictions must n, p, t obey in order that the design may exist?

   b) Construct a strongly balanced design in 5 treatments and 10 subjects.

   c) Show that under the usual model for crossover designs a strongly balanced uniform design has a completely symmetric information matrix for direct effects. (You may begin by assuming the form of the information matrix jointly for direct and carryover effects.) [(1+1)+3+5=10]

5. a) Consider a $2 \times 2 \times 4$ factorial arrangement, involving factors $F_1 F_2, F_3$. Explicitly write down any one contrast belonging to the interaction $F_1F_2F_3$ and any contrast belonging to F_2F_3. (you must specify the coefficients numerically). Check if these 2 contrasts are mutually orthogonal or not.

   b) Suppose a fractional factorial of the experiment in a) above is to be run as a completely randomized design and suppose that all main effects and all interactions are present in the model. Will the two contrasts considered in a) be estimable with such a design? Justify your answer.

   c) Consider a fraction of a $3 \times 4 \times 5$ factorial consisting of the treatment combinations $000, 100, 200, 010, 020, 030, 001, 002, 003, 004$. Under the absence of all two-factor and three-factor interactions, is it possible to estimate the general mean and the main effect contrasts from this fraction? justify your answer. [4+3+3=10]

INDIAN STATISTICAL INSTITUTE

Semestral Examination

M. Stat. - II Year (Semester - I)

Graph Theory and Combinatorics

Date : 20·11·12 Maximum Marks : 100 Duration : 3:00 Hours

Note : The question is of 130 marks. You may answer any part of any question, but maximum you can score is 100.

1. *2-factor of a graph* G is a 2-regular subgraph span all vertices of G. 1-factor implies a perfect matching.

   (i) Prove that every regular graph of even degree has a 2-factor.

   (ii) Prove that a loopless graph G has a loopless $\Delta(G)$-regular supergraph.

   (iii) Use (i) and (ii) to prove that for any loopless graph G with $\Delta(G)$ even, $\chi'(G) \leq 3\Delta(G)/2$.

   [15+10+15=40]

2. Let D be a list of $n > 1$ nonnegetive integers. Prove that D is graphic if and only if D' is graphic, where D' is the list of size $n-1$ obtained from D by deleting its largest element Δ and subtracting 1 from its Δ next largest elements. [15]

3. Let A be a set of $d+2$ points in $\Re^d$. Prove that there exist two disjoint subsets A_1, A_2 of A such that $Conv(A_1) \cap Conv(A_2) \neq \phi$. Using this result, prove Helly's theorem. Note that, $Conv(A)$ is the minimum area convex polygon enclosing point set of A.

   [10+15=25]

4. Consider a set of points $A \subset \Re^n$. a and b are two constants. Construct a set A with $\binom{n}{2}$ points where all the pairwise distances between points in A are equal to either a or b.

   Prove that any two-distance set in $\Re^n$ has at most $\frac{1}{2}(n+1)(n+4)$ points. [15+15=30]

5. Let $R(s,t)$ is the minimum number n such that any graph on n vertices contains either an independent set of size s or a clique of size t. Prove that for any $s, t \geq 1$, $R(s,t) \leq \binom{s+t-2}{s-1}$.

   [20]

INDIAN STATISTICAL INSTITUTE

Semestral Examination: 2012-13 (First Semester)
Master of Statistics (M. Stat.) II Year
Advanced Probability I

Teacher: Parthanil Roy

Total Points: 55 | **Date:** 22·11·12 | **Duration: 3 hours**

Note:

- **Please write your name and roll number on top of your answer booklet(s).**
- **There are five problems in this exam. Problem 1 is worth 5 points and will count towards your Assignments score. Problems 2 - 5 (worth 50 points in total) will count towards your Semestral Exam score.**
- **Show all your works and write explanations when needed.**
- **This is an open note examination. You are allowed to use your own hand-written notes (such as class notes, exercise solutions, list of theorems, formulas etc.). Please note that no printed or photocopied materials are allowed. In particular, you are not allowed to use books, photocopied class notes etc. If you are caught using any, you will get a zero in the semestral examination.**

1. (5 points) Show that there exists a discrete-parameter Gaussian process $\{B_n\}_{n\geq 1}$ satisfying $E(B_n) = 0$ for all $n \geq 1$ and $\text{Cov}(B_m, B_n) = \min\{m, n\}$ for all $m, n \geq 1$

2. Let $\{B_n\}_{n\geq 1}$ be as in Problem 1. Define $B_0 \equiv 0$.
   (a) (4 points) Show that $\{B_n\}_{n\geq 0}$ is a martingale, and $\{B_n^2\}_{n\geq 0}$, $\{(B_n^2 - 25)_+\}_{n\geq 0}$ are submartingales with respect to the natural filtration of $\{B_n\}_{n\geq 0}$.
   (b) (6 points) Compute the Doob decomposition of $\{B_n^2\}_{n\geq 0}$.

3. Fix $a_0 \in (0, 1)$. Let $\{X_n\}_{n\geq 0}$ be a (time-inhomogeneous) Markov Chain with $X_0 \equiv a_0$ and transition probabilities given by

$$P\left(X_{n+1} = \frac{a_n}{2} \,\middle|\, X_n = a_n\right) = 1 - a_n, \quad P\left(X_{n+1} = \frac{1 + a_n}{2} \,\middle|\, X_n = a_n\right) = a_n$$

   for all $n \geq 0$.
   (a) (3 points) Show that $E[(X_{n+1} - X_n)^2] = E[(X_n(1 - X_n)]/4$ for all $n \geq 0$.
   (b) (5 points) Prove that $\{X_n\}_{n\geq 0}$ is a martingale (with respect to its natural filtration) that converges almost surely to a random variable Y.
   (c) (7 points) Determine the distribution of Y.

[P.T.O.]

4. (a) (8 points) Let S_0 be an integrable random variable and $\{X_i\}_{i\geq 1}$ be a sequence of i.i.d. random variables with finite mean μ. Define

$$S_n := S_0 + X_1 + X_2 + \cdots + X_n, \ n \geq 1.$$

Let τ be a stopping time with respect to the natural filtration of $\{S_n\}_{n\geq 0}$ satisfying $E(\tau) < \infty$. Using martingale techniques or otherwise, show that

$$E(S_\tau) = E(S_0) + \mu E(\tau).$$

(b) (7 points) Consider the Gambler's Ruin Problem with a biased coin as described in class. Compute the expected number of tosses needed in the game. Justify <u>all</u> your steps.

5. State whether the following statements are *true* or *false* and provide detailed reasons supporting your answers.

(a) (5 points) If ν is a σ-finite measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ which is absolutely continuous with respect to the Lebesgue measure, then $\nu(C) < \infty$ for any compact subset C of $\mathbb{R}$.

(b) (5 points) If $\{Y_n\}_{n\geq 0}$ is an L^2-bounded martingale, then $\sum_{i=0}^{\infty} E[(Y_{i+1} - Y_i)^2] < \infty$.

Wish you all the best

Indian Statistical Institute
Semester Examination: 2012-2013
MS(QE) I/ M.Stat.II: 2012-2013

Game Theory I

Date: 23/11/2012 Maximum Marks: 50 Duration: 3 Hours

Answer any FOUR questions. Your total score cannot exceed 50.

1. (a) Prove that every finite extensive form game with perfect information has a subgame perfect Nash equilibrium.

   (b) Consider a Prisoners' dilemma game which is played infinitely repeatedly. Show that if the players are sufficiently patient, then playing the trigger strategy is a subgame perfect equilibrium in which the cooperative outcome is sustained.

   [6+7=13].

2. Consider a two-player three-period standard bargaining game of alternating offers. Let the utility function of player i be $U_i(x_i, t) = \delta^t x_i$, $0 \leq \delta \leq 1$; $t = 1,2,3$. Find the equilibrium outcome of the game. If the utility function be $U_i(x_i, t) = x_i - c_i t$, where $c_i > 0$ is the cost of delay for player i, find the corresponding equilibrium outcome.

   [6+7=13]

3. Consider a game in which the following simultaneous-move game is played twice.

2 1	L	M	R
T	10, 10	2, 12	0, 13
M	12, 2	5, 5	0, 0
B	13, 0	0, 0	1, 1

The players observe the actions chosen in the first play of the game prior to the second play. Specify and characterize the possible pure strategy subgame perfect Nash equilibria of this game.

[13]

4. Consider the following simultaneous move game played between a potential entrant (E) and an incumbent monopolist (M).

E\M	a	f
e	1, 1	-1, k
n	0, 3	0, 3

Here k can take two values: either $k = -1$ or $k = 2$. But the true value of k is known only to the monopolist; the entrant knows that $k = -1$ occurs with probability p and $k = 2$ with probability $1 - p$. All these are common knowledge. So this is a game of asymmetric information. Clarify the concept of Bayesian Nash equilibrium and solve the game for pure strategy Bayesian Nash equilibrium.

[3+10=13]

5. There are two firms, 1 and 2, competing in prices in a homogeneous good market; prices can be only <u>integer</u> quantities. The market demand for the product is $D(p) = \max[0, 20 - p]$, where p is the price of the product. Consumers always buy from the low-price seller. In case the firms charge the same price, each firm gets one-half of the customers. The unit costs of firm 1 and firm 2 are 3 and 5 respectively. (a) Suppose the firms charge prices simultaneously. What will be the equilibrium prices? (b) If firm 1 decides its price first, then firm 2 decides, find the corresponding equilibrium prices.

[10+3=13]

INDIAN STATISTICAL INSTITUTE

Semestral Examination: (2012–2013)

M. Stat 2nd Year

Statistical Computing

Date: 26.11.12 Marks: 100. Duration: 3 hours.

Attempt all questions

1. (a) The standard linear regression model can be written in matrix notation as $\mathbf{X} = \mathbf{A}\boldsymbol{\beta} + \mathbf{U}$. Here $\mathbf{X}$ is the $r \times 1$ vector of dependent variables, $\mathbf{A}$ is the $r \times s$ design matrix, $\boldsymbol{\beta}$ is the $s \times 1$ vector of regression coefficients, and $\mathbf{U}$ is the $r \times 1$ normally distributed error vector with mean $\mathbf{0}$ and variance $\sigma^2\mathbf{I}$. The dependent variables are right censored if for each i there is a constant c_i such that $Y_i = \min\{c_i, X_i\}$ is observed. Derive an EM algorithm for estimating the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma^2)'$ in the presence of right censoring.

(b) Multidimensional scaling attempts to represent q objects as faithfully as possible in p-dimensional space giving a weight $w_{ij} > 0$ and a dissimilarity measure y_{ij} for each pair of objects i and j. If $\boldsymbol{\theta}_i \in \mathbb{R}^p$ is the position of object i, then the $p \times q$ parameter matrix Θ with i-th column $\boldsymbol{\theta}_i$ is estimated by minimizing the stress

$$\sigma^2(\Theta) = \sum_{1 \le i < j \le q} w_{ij}(y_{ij} - \| \boldsymbol{\theta}_i - \boldsymbol{\theta}_j \|)^2,$$

where $\| \boldsymbol{\theta}_i - \boldsymbol{\theta}_j \|$ is the Euclidean distance between $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_j$. Construct an MM algorithm to estimate Θ. With reasons state any assumptions that you make.

Marks: 10+15=25

2. (a) Using the properties of Fourier transform, solve the following partial differential equation

$$\frac{\partial^2}{\partial t^2}u(t,x) - \frac{\partial^2}{\partial x^2}u(t,x) = 0,$$

with initial conditions

$$\begin{aligned} u(0,x) &= f(x), \\ \frac{\partial}{\partial t}u(0,x) &= g(x). \end{aligned}$$

(b) Denote by V the collection of functions f with $f'' \in L^2[0,1]$ and consider the subspace

$W_2^0 = \{f(x) \in V : f, f' \text{ absolutely continuous and } f(0) = f'(0) = 0\}$.

Define an inner product on W_2^0 as

$$< f, g >= \int_0^1 f''(t)g''(t)dt.$$

(i) Show that for $f \in W_2^0$, and for any s, $f(s)$ can be written as

$$f(s) = \int_0^1 (s-u)_+ f''(u)du,$$

where $(a)_+$ is a for $a > 0$ and 0 for $a \leq 0$.

(ii) Hence, obtain the reproducing kernel of W_2^0.

Marks: 10+5+10=25

3. For some sequence $\{\theta_n; n = 1, 2, \ldots\}$, consider the following generalized accept-reject method:

At iteration n $(n \geq 1)$

(i) Generate $X_n \sim g_n$ and $U_n \sim Uniform(0,1)$, independently.

(ii) If $U_n \leq \theta_n f(X_n)/g_n(X_n)$, accept $X_n \sim f$;

(iii) Otherwise, move to iteration $n+1$.

Let Z be the random variable denoting the output of this algorithm.

(a) Show that Z has the cdf

$$P(Z \leq z) = \sum_{n=1}^{\infty} p_n \int_{-\infty}^{z} f(x)dx$$

where $p_1 = \theta_1$ and $p_n = \theta_n \prod_{m=1}^{n-1}(1-\theta_m)$ for $n = 2, 3, \ldots$.

(b) Show that

$$\sum_{n=1}^{\infty} p_n = 1 \text{ if and only if } \sum_{n=1}^{\infty} \log(1 - p_n) \text{ diverges.}$$

(c) Give one example of a sequence $\{\theta_n\}$ that satisfies $\sum_{n=1}^{\infty} p_n = 1$ and one example of a sequence $\{\theta_n\}$ that does not satisfy it.

Marks: 5+10+10=25

4. (i) Let P be the transition probability matrix of an irreducible, aperiodic, finite-state Markov chain. Then prove that there is an integer m such that for $n \geq m$, the matrix P^n has strictly positive entries.

(ii) In the above, suppose that π is the invariant distribution. Then prove that there exist $r \in (0, 1)$ and $c > 0$ such that

$$\| P^n(x_0, \cdot) - \pi(\cdot) \| \leq cr^n,$$

where $\| \cdot \|$ denotes the total variation distance and x_0 is an arbitrary point in the state-space.

Marks: 15+10=25

INDIAN STATISTICAL INSTITUTE

Semestral Examination, First Semester: 2012-13

M.Stat. II Year (AS)

Life Contingencies

Date: November 27, 2012 Maximum marks: 100 Duration: 4 . hours

Students are permitted to use non-programmable calculators and Actuarial Formulae and Tables.

1. If $l_x = \max\{100 - x, 0\}$ and the interest rate is 5% per annum, determine

   (a) ${}_{5|}q_{60}$, [1]

   (b) P_{25}, [3]

   (c) the appropriate interpolation rule between integer ages. [2]

   [Total 6]

2. If $q_{70} = 0.017$ and $q_{71} = 0.02$, calculate $p_{70.3}$ by assuming constant force of mortality. [3]

3. Indicate, with formal justification, whether the following pairs of random variables are positively correlated, negatively correlated or uncorrelated. (Assume a constant force of interest.)

   (a) The joint life status and the last survivor status of a pair of independent lives. [2]

   (b) The curtate future life and the fractional part of future life (assuming that 'uniform distribution of death' assumption prevails). [3]

   (c) The present value of the aggregate premium income and the present value of benefit payout in a deferred term insurance arrangement with level benefit payable at the end of the year of death and premium payable annually in advance during the term. [4]

   [Total 9]

4. A MStat student aged 21 exact has passed the first year and is about to enter the second year. The probability of his death during the second year is equal to $1 - e^{-.0005}$. He decides to appear for professional examinations of an actuarial society during this year. The stress of preparing for the additional examinations produces an additional force of decrement (for failure in the second year only), which increases linearly over the year from 0 to 0.002, and is independent of his usual force of decrement.

   (a) Calculate the probability of the student surviving the second year of MStat. [3]

   (b) Further assume that, in the absence of the 'actuarial decrement', the probability of the student's survival to age $21 + t$ (for $0 \leq t < 1$) is $e^{-.0005(3t-2t^2)}$. Calculate, in the double decrement situation, the probability that the student dies due to the 'actuarial decrement' before reaching the age 22. [4]

   (c) Under the assumption of part (b), compare numerically the probabilities of death in the second year of MStat due to the 'actuarial decrement' in the single and double decrement situations. Explain why one is smaller than the other. [2+2]

   [Total 11]

5. (a) Explain what role a select life table can play in dealing with the heterogeneity of a population. [3]

(b) A life insurer looking for business in a tribal area suffers from shortage of data. A limited study produces the following levels of exposure and mortality of male lives.

Age	Initial exposed to risk	Observed number of deaths
40	1750	3
50	1420	5
60	1100	12

If the insurer uses the AM92 life table, which value of standardized mortality ratio should be used? [3]

[Total 6]

6. An actuarial student of age x wins a prize in the form of a whole life insurance with benefit S payable at the end of the year of death. The student immediately sells off the insurance, and uses the money to buy a pension in the form of whole life annuity due at an appropriate rate. Further, she plans to use every instalment of the pension to purchase a whole life insurance with appropriate benefit payable at the end of the year of death. There is a constant rate of interest, and no expenses.

   (a) Give an expression for the amount of pension received by the student each year. [1]

   (b) Give an expression for the benefit amount of the whole life insurance to be purchased in the year k. [1]

   (c) Derive an expression for the expected present value of the cumulative benefits of all the whole life insurances to be purchased. [2]

   (d) Show that the student indeed stands to gain through this plan, i.e., the expected present value calculated in part (c) is larger than S. [4]

[Total 8]

7. Arrange the following in increasing order of magnitude, with justification.

   (a) $\bar{A}^{1}_{x:\overline{n}|}$, $\bar{A}_{x:\overline{n}|}$, $\bar{A}_x$. [2]

   (b) $\ddot{a}_x$, a_x, $\ddot{a}_x^{(2)}$, $a_x^{(2)}$, $\ddot{a}_x^{(4)}$, $a_x^{(4)}$, $\bar{a}_x$. [6]

[Total 8]

8. A 10-year endowment insurance policy for a life aged 55 consists of a basic sum assured of Rs. 1 lakh. Compound reversionary bonus at the rate of 1.9231% is added at the beginning of every year. Level premiums are payable annually in advance throughout the term of the policy. The initial expense is Rs. 1,000 plus 50% of premium. The renewal expenses, incurred at the time of premium payment, are Rs. 200 plus 5% of premium. Calculate the gross annual premium from the equivalence, assuming that 6% rate of interest and mortality as per AM92 Ultimate prevail. [6]

9. A pension scheme provides for an retirement age benefit of $n/60$ths of the final pensionable salary, where n is the number of years of service, with fractions counting proportionately. Final pensionable salary is defined as the average salary over the three years prior to retirement. A member who has joined the scheme at age 30 exact, wants to calculate his expected retirement benefits after reaching age 52 exact. If the member has earned a salary of Rs. 1 lakh in each of twelve months from age 51 exact to age 52 exact, calculate the expected present value of his total pension, using the functions defined in the pension scheme table of the Actuarial Tables. [4]

10. Prove from first principles the result

$$\left[{}_tV^1_{x:\overline{n}|} + P^1_{x:\overline{n}|}\right](1+i) = q_{x+t} + p_{x+t} \times {}_{t+1}V^1_{x:\overline{n}|}.$$ [5]

11. A three-year unit-linked endowment assurance contract, described below, is issued to a 50-year old male.

Premiums	: Rs. 1,100 paid annually in advance
Part of premium allocated to units	: 50% in the first year
	: 95% in the renewal years
Bid-offer spread	: 3%
Annual management charge	: 1% of bid value of units, deducted at the end of the year
Death benefit	: Bid value of units, subject to minimum assured benefit of Rs. 1,000, payable at the end of the year of death
Maturity benefit	: Bid value of units
Surrender benefit	: None
Unit reserves	: Bid value of units
Non-unit reserves	: None

(a) Determine the yearly status of the unit fund. [4]

(b) Determine the projected profit vector. [4]

(c) Calculate the profit signature vector. [2]

(d) Calculate the expected present value of the profit. [1]

(e) Calculate the profit margin in percentage. [1]

Basis:

Mortality rate	: AM92 Select
Surrender rate	: nil
Interest rate	: 4% per annum
Unit fund growth rate	: 8% per annum
Initial expense	: Rs. 200
Renewal expenses	: Rs. 50 at the beginning of renewal year
Initial commission	: 10% of premium
Renewal commission	: 3% of premium at the beginning of renewal year
Risk discount rate	: 10% per annum

[Total 12]

12. A life insurance contract for a male life age 65 and a female life age 60 provides for a life annuity at the rate of Rs. 50,000 to be paid annually in arrears to the spouse of the person dying first. The annuity commences from the end of the year of death. The couple have to pay level premium (annually in advance) at a suitable rate till the joint life status continues. If the expenses are 10% of premium (incurred at the time of premium payment), calculate the gross premium from the equivalence principle. [5]

Basis:

Mortality: PMA92C20 and PFA92C20

Annual rate of interest: 4%

13. A 20 year term insurance is issued to a group of 100 male policyholders and 100 female policyholders, each having age 55. The assured sum is Rs. 10 lakhs payable at the end of the year of death. The premium is to be paid annually in advance for the duration of the policy. The lives follow PMA92C20 and PFA92C20 mortality tables. The interest rate is 4% per annum. Expenses are ignored.

   (a) Using the equivalence principle, calculate the net premium per policy separately for male and female policyholders. [4]

   (b) Using the equivalence principle, calculate the net premium reserve per policy (separately for male and female policyholders) that is required to be held at the end of the tenth year of the benefit period. [3]

   (c) Calculate the death strain at risk during the tenth year of the policy, per policy in force at the beginning of the tenth year, separately for male and female policyholders. [1]

   (d) At the end of nine years, policies for 98 males and 98 females were in force. There were two deaths in the following year, both of female policyholders. Calculate the mortality profit for the tenth year, aggregated over all the 196 policies. [3]

   (e) Compare and comment on the components of the mortality profit or loss attributable to the male and female policyholders. [1]

[Total 12]

14. A five-year unit-linked policy is issued to a life aged 75. When no non-unit reserve is held, the year-end cash-flows turn out to be as follows.

$$557.01, \; 491.91, \; 372.86, \; -797.37, \; 101.64.$$

The company wants to hold minimum non-unit reserves at the end of the policy years (as late during the policy period as possible), so that the cash-flows at end of each policy year is non-negative. Using the mortality rates given in AM92 Ultimate table and an interest rate of 3% per annum, calculate the sequence of reserves needed. [5]

Date: 28.11.12
Time: 3 hours

Statistical Methods in Genetics – I
M-Stat (2nd Year)
First Semester Examination 2012-13

This paper carries 60 marks. Answer all questions.

Question 1

Consider a disorder controlled by two unlinked autosomal biallelic loci with alleles (A_1,a_1) and (A_2,a_2), respectively. Suppose an individual is affected if and only if he/she has genotype $A_1A_1a_2a_2$.

(a) If p_1 is the conditional probability of an offspring being affected given that exactly one of the parents is affected and p_2 is the conditional probability that an offspring is affected given both parents are unaffected, show that $p_2=p_1^2$.

(b) In a study on trios (two parents and an offspring), 30 families are selected with exactly one affected parent and 50 families are selected with both parents unaffected. In the first set of families, 6 offspring are affected, while in the second set of families, 2 offspring are affected. Obtain the maximum likelihood estimate of the prevalence of the disease. [5 + 4]

Question 2

Suppose genotype data are available on a random set of individuals at an autosomal biallelic locus in an inbred population. If an analysis ignores the fact that the population is inbred, what is the impact on the maximum likelihood estimators of the allele frequencies at the locus? Show that the variances of the above estimators are the minimum in the absence of inbreeding in the population. [4 + 5]

P.T.O.

Question 3

(a) Consider a disease controlled by an autosomal biallelic locus. If an individual is affected, show that irrespective of the disease model, it is equally likely for the aunt and the grandson of the individual to be affected.

(b) Consider a recessive disorder controlled by an autosomal biallelic locus. Obtain the disease allele frequency for which the absolute difference in the allele frequencies between cases and controls at a biallelic marker locus in linkage disequilibrium with the disease locus is the minimum. [6 + 6]

Question 4

Consider two autosomal biallelic loci with alleles (*D*,*d*) and (*M*,*m*) respectively. The following are the genotype data on two nuclear families:

Family 1: father is *DM*/*dm*, mother is *Ddmm*, offspring are *DDMm*, *DdMm* and *DDmm*

Family 2: both parents are *Dm*/*dM*, offspring are *DdMM*, *DDmm* and *ddMm*

Using the LOD score approach, test whether the two loci are linked. [10]

Question 5

(a) Explain why linkage disequilibrium exists over much smaller distances on the genome compared to linkage.

(b) Given genotype data on a random set of individuals at two autosomal biallelic loci, develop an EM algorithm to obtain the maximum likelihood estimator of the coefficient of linkage disequilibrium between the two loci. [3 + 7]

Class Presentation carries 10 marks.

INDIAN STATISTICAL INSTITUTE

Semestral Examination : Semester I (2012-13)

M. Stat. II Year

Actuarial Methods

Date: 30.11.2012 Maximum marks: 100 Time: $3\frac{1}{2}$ hours

Calculator and Actuarial table can be used.

1. In an NCD system with three discount categories (0%, 25% and 50%), the rules are as follows: (*i*) following a claim-free year, the discount increases by one level, or remains at category 3, (*ii*) following a year in which exactly one claim has been made, the discount decreases by one level, or remains at category 1, (*iii*) following a year in which more than one claim has been made, the discount returns to category 1.

   The number of claims made in each year under each policy is assumed to follow a *Poisson* distribution with mean λ. Write down the transition matrix of the probabilities p_{ij} that a policyholder in category i in one year will move to category j in the following year. Obtain the proportions of policyholders in different categories under equilibrium. Calculate the average premium amount in equilibrium assuming the basic premium to be c. [5+5+2=12]

2. (a) The Bayesian approach using quadratic loss always produces an estimate which can be readily expressed in the form of a credibility estimate. Prove or disprove the statement.

   (b) Ten independent and identically distributed observations on number of claims following a $Poisson(\lambda)$ distribution are 3, 4, 3, 1, 5, 5, 2, 3, 3, 2. Assuming an $Exp(0.2)$ prior distribution for λ, find the Bayes estimator of λ under squared error loss. Express the estimate in the form of a credibility estimate. [6+(5+1)=12]

3. (a) Describe domination of strategies in the context of zero-sum two-players game.

   (b) An entrepreneur is trying to decide whether to expand his business. He can either go ahead and *expand*, or *wait* until a more suitable time. He believes that there are two possible economic scenarios for the current year: *good* and *poor*. If he waits, he expects to make a profit of Rs 70,000 in the coming year if the economy is good, and Rs 30,000 if the economy is poor. If he expands, he expects to make a profit of Rs 150,000 if the economy is good, and a loss of Rs 6,000 if the economy is poor. The entrepreneur wishes to minimize his expected loss in the coming year. Let p be the probability of a recession (poor economy) in the current year. What values of p will make him wait?

   [2+6=8]

4. (a) Based on recent experience, an insurance company believes that individual claims in a particular category follows a *Log-normal* distribution with mean Rs 5,000 and standard deviation Rs 7,500. Estimate the proportion of claims that will exceed Rs 25,000.

   (b) The random variable X has a *Burr* distribution with density function given by

   $$f(x) = \frac{\alpha\gamma\lambda^{\alpha}x^{\gamma-1}}{(\lambda + x^{\gamma})^{\alpha+1}}$$

   with parameters $\gamma = 2$ and $\lambda = 500$. Find maximum likelihood estimate of α based on a random sample $x_1, \cdots, x_n$.

(c) Claims occur as a *Pareto* distribution with parameters $\alpha = 6$ and $\lambda = 200$. A proportional reinsurance agreement is in force with a retained proportion of 80%. Find the mean of the amount paid by the reinsurer on an individual claim. What this mean would be if, instead, there is an excess of loss reinsurance agreement with retention level 50?

[3+2+(2+5)=12]

5. (a) Explain product liability in the context of general insurance.

   (b) List four perils for household and commercial buildings insurance.

   (c) Specify two objectives of a No Claim Discount (NCD) system.

   (d) Define Loss Ratio and Ultimate Loss Ratio.

   (e) Assume the claims to follow *Exponential* distribution with mean $1/\lambda$ and an XOL reinsurance with retention limit Rs 1000. In a sample of 100 claims, the mean of 90 claims that do not exceed Rs 1000 is Rs 82.9. Find MLEs of λ and the probability that the re-insurer needs to pay on a particular claim.

[2+2+2+3+3=12]

6. The cumulative claims paid each year under a certain cohort of insurance policies are recorded in the table below for accident years 1998-2001.

Accident	Development Year			
Year	0	1	2	3
1998	2457	4196	4969	5010
1999	2648	4715	5561	
2000	3084	5315		
2001	3341			

The rates of claim inflation over these years 1999-2001 are, respectively, 2.1%, 10.5% and 3.2%. Assume a constant 5% claim inflation rate for the future years. State the assumptions underlying the inflation-adjusted basic chain-ladder method. Using the method, estimate the inflation-adjusted outstanding claims reserve as on 31 December 2001. [3+14=17]

7. (a) Define the surplus process, in the context of ruin theory, and, then, the probability of ultimate ruin as a function of initial surplus. Prove that this ruin probability is a decreasing function of the initial surplus.

   (b) An insurer plans to issue 5000 one year policies at the start of a year. The annual aggregate claims for each policy has a compound Poisson distribution with Poisson parameter $\lambda = 0.5$ and the individual claim amounts have a lognormal distribution with $\mu = 5.04$ and $\sigma = 1.15$. The premium for each policy is Rs 160. Claims are assumed to be paid at the end of the year. Calculate the minimum interest rate the insurer must earn if the profit at the end of the year is to exceed Rs 52,500 with probability 0.9. Use normal approximation.

[(1+1+1)+10=13]

8. (a) Write the *Binomial* regression model as a GLM and give the corresponding canonical link function. Give expression for the corresponding deviance residuals and explain their role in model selection.

(b) The following time series model is used for monthly inflation rate (Y_t) in a particular country:

$$Y_t = 0.4Y_{t-1} + 0.2Y_{t-2} + e_t + 0.025.$$

If this is considered as an ARIMA(p, d, q) model, derive the values of p, d and q. Determine whether Y_t is stationary.

(c) Consider

$$f(x) = \begin{cases} kx^{-0.5}, & \text{if } 0 \leq x \leq 1 \\ ke^{-x}, & \text{f } x > 1, \end{cases}$$

with $k > 0$ being the normalizing constant. Describe a method for simulating an observation from this distribution.

[(2+2+2)+(3+2)+3=14]

INDIAN STATISTICAL INSTITUTE

Semestral exam. (Semester I: 2012-2013)

Course Name: M. Stat. 2nd year

Subject Name: Analysis of discrete data

Date: 30.11.12 , Maximum Marks: 55. Duration: 3 hr.

1. Geometrically interpret *Risk Ratio* = 1 for a 2×2 contingency table. [10]

2. Suppose there are three urns, labelled as urn A, urn B and urn C, each having 4 yellow, 5 blue and 6 orange balls initially. Let Y_1, Y_2, Y_3 be three discrete random variables. Draw a ball from the urn A, Y_1 represents the indicator variable corresponding to the colour of the drawn ball. Consequently we add an additional ball of the same colour of the drawn ball to the urn B. Now draw one ball from this urn B, and Y_2 represents the indicator corresponding to the colour of the drawn ball. Consequently add one ball of the same colour as the drawn ball (from urn B) to urn C. Now draw one ball from urn C, and Y_3 corresponds to the colour of this drawn ball. Find the marginal distributions of Y_1, Y_2 and Y_3. Find $\tau_{12}, \tau_{13}, \tau_{23}$, where τ_{ij} is the Goodman and Kruskal's τ for Y_j on Y_i. Find τ_{13} as a function of τ_{12} and τ_{23}. [3+9+1]

3. Each week *Variety* magazine summarizes reviews of new movies by critics in several cities. Each review is categorized as pro, con, or mixed, according to whether the overall evaluation is positive, negative, or a mixture of the two. The following Table summarizes the ratings from April 1995 through September 1996 for two Chicago film critics Gene Siskel and Roger Ebert. Discuss kappa and generalized kappa as measures of agreement in this context. Find the values (setting appropriate weights for generalized kappa) in this case, and interpret the results.

	Ebert		
Siskel	Con	Mixed	Pro
Con	24	8	13
Mixed	8	13	11
Pro	10	9	64

[4+(4+2)]

4. Five groups of animals were exposed to a dangerous substance in varying concentrations. The following table gives the number of animals exposed (n_i) and the number that died (y_i) corresponding to the ith concentration level (dose), $i = 1, 2, \cdots, 5$.

Concentration level (i)	n_i	y_i
1×10^{-5}	6	0
1×10^{-4}	6	1
1×10^{-3}	6	4
1×10^{-2}	6	6
1×10^{-1}	6	6

Describe the fit of an appropriate logit model for π_i (probability of death corresponding to ith dose) as a function of $\log_{10}$(concentration level (i)). [Discuss the applicability of Newton-Raphson procedure in this context.] How can you test for LD_{50} in this context? [5+4]

5. Discuss the latent variable approach to model ordinal categorical variable with possible covariates. Write down the likelihood assuming normality of the latent variable. How can you model a bivariate response vector using latent variable approach where one variable is ordinal categorical and the other is continuous? Write down the likelihood in this context. [3+2+4+4]

INDIAN STATISTICAL INSTITUTE

First Semestral Back Paper Examination: 2012-13

M.STAT 2nd year. Advanced Design of Experiments. 26/12/12

Total marks 100. Duration: Three hours

Answer all questions.

Keep your answers brief and to the point.

1. State whether each of the following statements is True or False and give brief justification in each case: [$5 \times 4 = 20$]

   a) An orthogonal array OA$(64, 10, 8, 2)$ does not exist.

   b) A $3 \times 4 \times 2$ factorial experiment will lead to 24 independent contrasts belonging to the 3-factor interaction effect.

   c) The conditions $vr = bk$ and $r(k-1) = \lambda(v-1)$ are necessary for the existence of a BIB design with parameters v, b, r, k, λ.

   d) A BIB design with parameters v, b, k is universally optimal for estimating a complete set of orthonormal treatment contrasts in the class of all block designs with v treatments arranged in b blocks, each of size k.

2. a) Define a Hadamard matrix.

   b) Can there exist a Hadamard matrix of order 18? Justify your answer with a proof.

   c)Construct a Hadamard matrix of order 8.

   d) Prove that the existence of an orthogonal array OA$(N, k, 2, 2s)$ implies the existence of an OA($2N$, $k+1$, 2, 2s+1). [2+6+6+6=20]

3. a) Define A-optimality and explain its statistical significance.

   b) Prove that a randomized block design with parameters $v, b, k(= v)$ is A-optimal for a full set of orthonormal treatment contrasts in the class of all block designs with parameters $v, b, k(= v)$.

   c) Prove that a Latin square design with t treatments is universally optimal for a full set of orthonormal treatment contrasts in the class of all row-column designs with t treatments and t rows and t columns. [5+7+8=20]

4. a) Define a balanced uniform crossover design.

   b) What are the necessary conditions that the parameters of the design in (a) must satisfy in order that such a design may exist?

   c) Construct a strongly balanced uniform design with 3 treatments and the minimum number of subjects and periods required for such a design to exist.

   c) Let $\mathcal{D}$ be the class of all crossover designs with t treatments, n subjects and p periods. Let d_0 be a strongly balanced uniform design in $\mathcal{D}$. Show that d_0 is universally optimal in $\mathcal{D}$ for carryover effects under the usual model. [2+3+5+10=20]

5. a) Consider a factorial experiment in 10 factors, $F_1, \ldots, F_{10}$, the factor F_i having s_i levels, $i = 1, \ldots, 10$. Define a contrast belonging the interaction effect $F_3F_4F_5$.

b) Write down the expression for a full set of orthonormal contrasts belonging to the interaction $F_3F_4F_5$ and show that this indeed satisfies your definition in (a) above.

c) Either construct an OA(18, 8, 3, 2) or show that such an OA is non-existent. [4+8+8=20

INDIAN STATISTICAL INSTITUTE

Semestral exam. (Semester I: 2012-2013) - Back Paper

Course Name: M. Stat. 2nd year

Subject Name: Analysis of discrete data

Date: 28 12.12 , Maximum Marks: 100. Duration: 3 hrs.

Note: Total marks: 110. Answer as much as you can.

1. Suppose there are three urns, each having $\sum_{j=0}^{k} a_j$ balls with a_j balls of lebel j, $j = 0, 1, \cdots, k$. Let Y_1, Y_2, Y_3 be three discrete random variables. Draw a ball from the urn '1', if the drawn ball is of lebel 'j_1', then $Y_1 = j_1$. Consequently add addtional b balls of lebel j_1 to the urn '2'. Now draw a ball from the urn '2', if it is of lebel 'j_2', then $Y_2 = j_2$. Add additional b balls of lebel j_2 to the urn '3'. Draw a ball from the urn '3', if it is of lebel 'j_3', set $Y_3 = j_3$. Show that Y_i's are identically distributed. Find $corr(Y_1, Y_2)$, $corr(Y_1, Y_3)$ and $corr(Y_2, Y_3)$. [8+12]

2. Geometrically discuss the difference in interpretation of odds ratio $= 0.25, 1$ and 4. [15]

3. Derive the joint asymptotic distribution of log odds ratios in a 2×3 contingency table. Consequently find 95% confidence interval for log odds ratio for the following table.

	Myocardial Infraction	
	Yes	No
Placebo	28	656
Aspirin	18	658

[8+4]

4. Define Theil's entropy measure of association for nominal responses. Show that it reduces to the form

$$-\frac{\sum_i \sum_j \pi_{ij} \log(\pi_{ij}/\pi_{i+}\pi_{+j})}{\sum_j \pi_{+j} \log \pi_{+j}},$$

where π_{ij} is the cell probability of the (i,j)th cell and $\pi_{i+} = \sum_j \pi_{ij}$, $\pi_{+j} = \sum_i \pi_{ij}$. [3+4]

5. Discuss kappa and generalized kappa as measures of agreement. Calculate kappa for a 4×4 table having $n_{ii} = 5$ for all i, $n_{i,i+1} = 15$, $i = 1, 2, 3$, $n_{41} = 15$, and $n_{ij} = 0$ otherwise. Explain why strong association does not imply strong agreement. Also find the value of generalized kappa by considering suitable weights and interpret the results. [5+4+3+3]

6. Describe Cochran-Mantel-Haenszel test for conditional independence. Carry out the test procedure for the following data of a multicenter clinical trial, and interpret your result.

Center	Treatment	Success	Failure
1	Drug	11	25
	Control	10	27
2	Drug	16	4
	Control	22	10
3	Drug	14	5
	Control	7	12
4	Drug	2	14
	Control	1	16
5	Drug	6	11
	Control	0	12
6	Drug	1	10
	Control	0	10
7	Drug	1	4
	Control	1	8
8	Drug	4	2
	Control	6	1

[5+9]

7. Discuss Fisher scoring for a regression using general link function. Discuss the special case of logistic regression. [7+3]

8. Data on pre- and post-operative conditions (classified as bad, moderate, good) of 100 patients are given along with their age, sex and another important covariate related to the initial condition of the disease. Give a latent variable based model and illustrate an approach to test whether there is significant improvement after the operation or not. Discuss any computational problem that might be encountered in the analysis. [6+7+4]

INDIAN STATISTICAL INSTITUTE

Back-paper Examination : Semester I (2012-13)

M. Stat. II Year

Actuarial Methods

Date: 26·12·12 Maximum marks: 100 Time: $3\frac{1}{2}$ hours

Calculator and Actuarial table can be used.

1. The profit per client-hour made by a privately owned health centre depends on the variable cost involved. Variable cost, over which the owner of the health centre has no control, takes one of the three levels θ_1 = high, θ_2 = medium and θ_3 = low. The owner has to decide at what level to set the number of client-hours that can be either $d_1 = 16,000$, $d_2 = 13,400$ or $d_3 = 10,000$. The profit (in Rs.) per client-hour is as follows:

	θ_1	θ_2	θ_3
d_1	85	95	110
d_2	105	115	130
d_3	125	135	150

Determine the minimax solution. Given the probability distribution $p(\theta_1) = 0.1$, $p(\theta_2) = 0.6$, $p(\theta_3) = 0.3$, determine the solution based on the Bayes criterion. [3+5=8]

2. Consider n independent and identically distributed observations on claims following a $N(\mu, 100)$ distribution. Assuming that μ is equally likely to be either 50 or 60, find the Bayes estimator of μ under squared error loss. Find this Bayes estimator of μ, when it is only given that the mean of the sample exceeds 57. [4+4=8]

3. Claim amounts from a portfolio have the distribution with pdf

$$f(x) = 2cxe^{-cx^2},\ x \geq 0,\ c > 0.$$

An XOL reinsurance arrangement is in force with retention limit $M = 3$. A sample of reinsurer's payment amounts gives the following values: $n = 10$, $\sum y_i = 8.7$ and $\sum y_i^2 = 92.3$. Find maximum likelihood estimate of c. [8]

4. (a) Explain the concept of benefits and perils in the context of general insurance by means of examples.

   (b) Describe a credibility estimate, specifying its characteristics.
   The Bayesian approach using quadratic loss always produces an estimate which can be readily expressed in the form of a credibility estimate. Prove or disprove the statement.

[(2+2)+(3+5)=12]

5. (a) Describe a compound Poisson distribution in the context of aggregate claims. Derive expressions for its mean and variance in terms of those of number of claims and claim size distributions.

   (b) Consider a proportional reinsurance arrangement wherein the direct writer retains a proportion k. Find the mgf of Y, the net individual claim amount paid by the direct writer, in terms of that of X. Hence find expressions for the mgf's of the aggregate claim amounts paid by the direct writer and the reinsurer separately, if the number of claims has a $Poisson(\lambda)$ distribution.

[(2+1+2)+(2+2+1)=10]

6. (a) Describe the difference between collective risk model and individual risk model.

   (b) A portfolio of policies consists of one-year term assurances on 100 lives aged exactly 30 and 200 lives aged exactly 40. The probability of a claim during the year on any one of the lives is 0.0004 for the 30 year olds and 0.001 for the 40 year olds. If the sum assured on a life aged x is uniformly distributed between $1000(x-10)$ and $1000(x+10)$, calculate mean and variance of the aggregate claims from this portfolio during the year. Which of the two risk models, mentioned in part (a), has been used here?

[3+(2+4+1)=10]

7. The aggregate claim process for a particular risk is a compound Poisson process with Poisson parameter $\lambda = 20$ per year. Individual claim amounts are Rs 100 w.p. 1/4, Rs 200 w.p. 1/2 and Rs 250 w.p. 1/4. The initial surplus is Rs 1000. Using a normal approximation, calculate the smallest premium loading factor such that the probability of ruin at year 3 is at most 0.05. [10]

8. (a) Describe a typical No Claim Discount (NCD) system specifying its objective(s).

   (b) An NCD system with three discount categories (0%, 25% and 50%), the rules are as follows: (*i*) following a claim-free year, the discount increases by one level, or remains at category 3, (*ii*) following a year in which exactly one claim has been made, the discount decreases by one level, or remains at category 1, (*iii*) following a year in which more than one claim has been made, the discount returns to category 1.

   The number of claims made in each year under each policy is assumed to follow a *Poisson* distribution with mean λ. Write down the transition matrix of the probabilities p_{ij} that a policyholder in category i in one year will move to category j in the following year. Obtain the proportions of policyholders in different categories under equilibrium. Calculate the average premium amount in equilibrium assuming the basic premium to be c.

[3+(5+4+2)=14]

9. The table below shows the cumulative claims (in Rs '000s) incurred on a particular class of insurance policies, divided by accident year and development year.

Accident Year	Development Year 0	1	2	3
1997	502	556	589	600
1998	487	565	593	
1999	608	640		
2000	551			

State the assumptions underlying the basic chain-ladder method. Using the method, estimate the outstanding claims reserve as on 31 December 2000. [2+6=8]

10. (a) Discuss Generalized Linear Models (GLM) in contrast with Linear Models specifying the assumptions clearly. What is canonical link function?

    (b) Identify the process $2X_t = 7X_{t-1} - 9X_{t-2} + 5X_{t-3} - X_{t-4} + e_t - e_{t-2}$ as an ARIMA model.

    (c) Using the Acceptance-Rejection method, generate a discrete random variable from $p(1) = 1/3$ and $p(2) = 2/3$ with an unbiased coin only.

[(3+1)+4+4=12]

INDIAN STATISTICAL INSTITUTE

Backpaper Examination: 2012-13 (First Semester)
Master of Statistics (M. Stat.) II Year
Advanced Probability I

Teacher: Parthanil Roy

Total Points: 100 **Date:** 31·12·12 **Duration: 3 hours**

Note:

- **Please write your name and roll number on top of your answer booklet(s).**
- **Show all your works and write explanations when needed.**
- **This is an open note examination. You are allowed to use your own hand-written notes (such as class notes, exercise solutions, list of theorems, formulas etc.). Please note that no printed or photocopied materials are allowed. In particular, you are not allowed to use books, photocopied class notes etc. If you are caught using any, you will get a zero in the backpaper examination.**

1. (20 points) Consider the probability space $\left([0,1]^2, \mathcal{B}_{[0,1]^2}, \lambda_2\right)$ where λ_2 is the Lebesgue measure on $[0,1]^2$. Let $M : [0,1]^2 \to [0,1]$ be the function $M(x,y) = \max(x,y)$. Find a regular conditional probability on $\mathcal{B}_{[0,1]^2}$ given the random variable M.

2. (20 points) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $(X_n)_{n\geq 1}$ be i.i.d. Bernoulli $\left(\frac{1}{2}\right)$ random variables defined on it. Let $Y_n := I_{[X_n=1 \text{ and } X_{n+14}=1]}$ for $n \geq 1$. Let $\mathbf{P}$ and $\mathbf{Q}$ be the laws of the two process $(X_n)_{n\geq 1}$ and $(Y_n)_{n\geq 1}$ respectively on $(\mathbb{R}^\infty, \mathcal{B}^\infty)$. Show that $\mathbf{P}$ and $\mathbf{Q}$ are mutually singular.

3. (20 points) Suppose $\{(X_n, \mathcal{F}_n)\}_{n\geq 0}$ is a martingale and τ is a finite stopping time such that X_τ is integrable and $\liminf_{n\to\infty} E\left(|X_n| I_{\tau>n}\right) = 0$. Show that the stopped process $\{X_{\tau\wedge n}\}_{n\geq 0}$ is uniformly integrable and converges in L^1 to X_τ as $n \to \infty$.

4. Let $\{X_n\}_{n\geq 1}$ be a sequence of i.i.d. continuous random variables. We call X_k a *record value* for the sequence if $X_k > X_l$ for $1 \leq l < k$. Let I_k be the indicator function for the event that X_k is a record value. Let R_n is the number of record values in the first n observations $X_1, X_2, \ldots X_n$.

   (a) (10 points) Show that
   $$\sum_{k=1}^{\infty} \frac{I_k - k^{-1}}{\log k} < \infty$$
   almost surely.

(b) (10 points) Using (a) or otherwise, show that

$$\frac{R_n}{\log n} \to 1$$

almost surely as $n \to \infty$.

5. Fix a positive integer n and a positive real number λ. Let $\{(S_k, \mathcal{F}_k)\}_{0\le k\le n}$ be an L^2-martingale with $S_0 \equiv 0$. Define

$$\tau := \min\{0 \le k \le n : |S_k| > \lambda\}$$

with the convention that $\min \emptyset = n$. Define $X_k := S_k - S_{k-1}$ for all $1 \le k \le n$.

(a) (5 points) Show that for all $2 \le k \le n$,

$$S_{k-1}^2 + \sum_{i=1}^{k-1} X_i^2 = 2S_{k-1}S_k - 2\sum_{i=1}^{k} S_{i-1}X_i.$$

(b) (15 points) Prove that

$$E\left(S_{\tau-1}^2 + \sum_{i=1}^{\tau-1} X_i^2\right) \le 2\lambda E(|S_n|).$$

Wish you all the best

INDIAN STATISTICAL INSTITUTE

Back-paper Examination, First Semester: 2012-13

M.Stat. II Year (AS)

Life Contingencies

Date: 27·12·12 Maximum marks: 100 Duration: 3 hours and 30 minutes

Students are permitted to use non-programmable calculators and Actuarial Formulae and Tables.

1. If the force of interest is twice the force of mortality, and both rates are constant, calculate the standard deviation of the present value of a continuous whole life annuity, assuming that its expected value is 12.5. [4]

2. A select table (select period 5 years) is based on the following rates of mortality: $q_{[x]+t} = \frac{0.02}{1.02}$ for all values of x and values of $t < 5$; $q_x = \frac{0.03}{1.03}$. If $l_{[10]} = 100,000$, calculate (a) l_{45}, and (b) $l_{[40]+1}$. [2+1=3]

3. In a situation where the force of interest is a constant (δ) per year, it is claimed that one rupee today can be used to pay a continuous whole life annuity to (x) at the rate Rs. δ per year, as well as to provide a whole life assurance to x with benefit of Re. 1. Express this statement in standard actuarial notation, and prove or disprove it from first principles. [5]

4. One thousand independent lives aged 25 are offered a 20-year deferred 15-year endowment assurance of Rs. 10 lakhs payable at the end of the year of death or at maturity. The premium is to be paid annually in advance for 20 years. The lives follow AM92 Select mortality and the interest rate is 6% per annum.

   (a) Using the equivalence principle, calculate the net premium provision per policy that is required at the end of the premium-paying period. [2]

   (b) Using the equivalence principle, calculate the net premium provision per policy that is required at the end of the first year of the benefit period. [2]

   (c) Calculate the death strain at risk, per policy in force at the beginning of the benefit period, during the first year of benefit period. [1]

   (d) At the end of 20 years, 976 policies were in force. There were 6 deaths in the following year. Calculate the mortality profit. [2]

   (e) Comment on the mortality profit or loss. [2]

   [Total 9]

5. An annuity-due makes quarterly payments to a life aged 60 exact where each payment is 1.0094797 times greater than the one immediately preceding. The first quarterly amount is Rs. 1,000. Calculate the actuarial present value of the annuity.

   Basis: Mortality: AM92 ultimate with UDD; Interest: 8% per annum. [5]

6. If $P_x = .017$, $i = 0.03$, $q_x = 0.02$, $q_{x+1} = 0.022$ and $q_{x+2} = 0.024$, calculate ${}_3V_x$. [4]

7. The future lifetimes of two individuals aged x and y are independent, and are subject to constant forces of mortality of 0.02 and 0.03 respectively. Assuming interest rate $i = 3\%$, find the actuarial present value of a whole life annuity-due of Rs. 100,000 per year, payable monthly to the last survivor, commencing at the end of the month of termination of the joint life. [7]

P.T.O

8. A male aged x is due to retire in n years. A benefit scheme provides for a sum of Rs. 100,000 payable to his wife aged y at the end of the year of his death, provided that she is alive at that time. Further, if he dies in service and his wife is alive at that time, she will receive a reversionary whole life annuity of Rs. 30,000 per year payable annually in arrear. Derive a concise expression for the net present value of this benefit, using appropriate actuarial notations. [6]

9. An impaired life aged 35 experiences 5 times the force of mortality of a life of the same age subject to standard mortality. A two year term assurance policy is sold to this impaired life. The policy has a sum assured of Rs. 2,50,000 payable at the end of the year of death.

   (a) Show that the probability of survival of an impaired life aged x for a further year is p_x^5, where p_x is the corresponding probability of a non-impaired life. [2]

   (b) Calculate the expected present value of the benefits payable to each life under the above policy, assuming that standard mortality is AM92 Ultimate and interest is 4% pa. [4]

[Total 6]

10. Assume that m single-decrement tables are given, each representing a different cause of decrement.

   (a) Explain in detail how to obtain a multiple-decrement table embodying these m causes of decrement by proving the result $\mu_x^{(T)} = \sum_{k=1}^{m} \mu_x^{(k)}$. [7]

   (b) Consider a special case of two decrements to prove the following:

$$d_x^{(k)} = \frac{q_x^{(k)}}{q_x^{(1)} + q_x^{(2)}} . d_x^{(T)}, \quad k = 1, 2.$$ [3]

[Total 10]

11. A life insurance company sells 5-year-term, single-premium, unit-linked bonds, each for a single premium of Rs. 15,000. There is no bid/offer spread and the allocation percentage is 100%.

   (a) Assuming that the only charge is a 3% annual management charge and assuming unit growth of 9% pa, calculate the unit provision at the start and the end of each year and the management charge each year. [3]

   (b) Calculate the net present value of the contract assuming:

   - Commission of 5% of the premium,
   - Initial expenses of Rs. 150,
   - Annual renewal expenses of Rs. 50 in the 1st year, inflating at 5% pa,
   - Independent rate of mortality is a constant 0.5%,
   - Independent rate of surrender is 5% pa,
   - Non-unit fund interest rate is 9% pa,
   - Risk discount rate 12% pa.

   The company holds unit provisions equal to the full value of the units and zero non-unit provisions. You may assume that expenses are incurred at the start of the year and that death and surrender payments are made at the end of the year. [4]

[Total 7]

P·

12. The death in service benefit of a pension scheme is four times the member's salary on the date of death. Normal Pension Age is 60. Give an expression for the present value of this benefit to a life aged 35 exact with salary of Rs. 25,000, who has just received a salary increase. Define all the symbols used, and do not use commutation functions.

[5]

13. A 22-year endowment policy has to be issued to a life aged 33 exact, with mortality following the AM92 Ultimate table. The benefit amount B is to be paid at the end of the year of death or at maturity. Level premiums of amount G are paid annually in advance during the 22 years of the contract. The interest rate is 4% per annum.

The insurer incurs the following fixed and variable expenses.

Time of expense	Expense
Policy issue	0.2% of B plus 40% of G
Payment of subsequent premium	0.02% of B plus 2% of G
Payment of death benefit	0.03% of B
Payment of maturity benefit	0.02% of B

Using the equivalence principle, calculate the gross premium G as a multiple of the benefit amount B. [7]

14. Consider the three-state continuous-time Illness-Death Markov model, with states *Healthy*, *Critically ill* and *Dead*, with the following constant forces of transition:

(i) σ for transition from *Healthy* to *Critically ill*,
(ii) μ for transition from *Healthy* to *Dead* and
(iii) ν for transition from *Critically ill* to *Dead*.

No transition from *Critically ill* to *Healthy* is possible. A life insurance company uses this model to price its stand-alone critical illness policies. Under these policies, a lump sum benefit is payable on the occasion that a life becomes critically ill during a specified policy term. No other benefits are payable.

A 20-year policy with sum assured Rs. 200,000 is issued to a healthy life aged 40 exact.

(a) Write down a formula, in integral terms, for the expected present value of benefits under this policy. [2]

(b) Calculate the expected present value at the outset for this policy.
Basis: $\mu = 0.01$, $\sigma = 0.02$, $\nu = 3\mu$, Interest is 8% pa. [3]

[Total 5]

15. A ten-year unit-linked policy is issued to a life aged 18. Assuming no non-unit reserves, the year-end cash-flows turn out to be as follows.

705, −293, 433, 497, 101, 297, −350, 127, 256, 100.

The company wants to hold the minimum non-unit reserves at the end of policy years 1,2,. . . ,9, which are needed to make the cash-flows in the following years non-negative. Using the mortality law ${}_tp_x = \frac{x^2}{(x+t)^2}$ and interest rate 5% per annum, calculate the sequence of reserves needed. [8]

16. A 20-year endowment policy for a life aged 40 has a sum assured of Rs. 5 lakhs. There is a bonus of 10% of original sum assured and an additional bonus of 1.92308% on all previously announced bonuses, both of which vest at the end of the year. Net premium is collected annually in advance for the duration of the policy.

(a) Calculate the net premium. [6]

(b) At the end of 10 years, and after bonus have vested as per the above plan, the insurer wishes to convert the with-profit contract into that of a fixed benefit contract, without changing the rate of premium. Calculate the benefit amount that would be equivalent to the benefit committed under the existing contract.

[3]

Basis: Mortality: AM92 Ultimate; Interest: 6%; Expenses: Ignore.

[Total 9]

Date: 31.12.12
Time: 3 hours

Statistical Methods in Genetics – I
M-Stat (2nd Year)
Backpaper Examination 1st Semester 2012-13

This paper carries 100 marks. Answer all questions.

1. Suppose you are given data on *A*-*B*-*O* blood groups for a random set of individuals in an inbred population. Describe an EM algorithm to obtain the maximum likelihood estimators of the frequencies of the alleles *A*, *B*, *O* and the inbreeding coefficient. [15]

2(a) Suppose that the mutation rate from allele *A* to allele *a* is 0.002 per generation and that from *a* to *A* is 0.004 per generation. In how many generations will the frequency of allele *a* increase from 0.1 to 0.2?

(b) Consider a dominant disorder controlled by an autosomal biallelic locus. In one family, both parents are affected and three of the four offspring are affected. In another family, exactly one parent is affected and all three offspring are affected. Compute the joint likelihood of the two pedigrees.
[10 + 10]

3(a) Suppose data on affection status (i.e., affected or unaffected) are collected on sibships ascertained through an affected proband and both parents unaffected. If the ascertainment probability of a proband is small, show that the maximum likelihood estimator (m.l.e.) of the segregation ratio is approximately the proportion of affected sibs among all sibs after eliminating all the probands in the study. How does this m.l.e. compare with the proportion of affected sibs among all sibs? Based on your m.l.e., is it possible to infer whether the disease is recessive or not?

(b) What is the expected identity-by-descent score of a pair of sibs, both heterozygous at an autosomal biallelic locus? [(6+2+4) + 8]

4. Consider two biallelic loci with alleles (A,a) and (B,b), respectively and the recombination fraction between them is θ. Consider a nuclear family where both parents are heterozygous at both loci and there are four offspring with genotypes Aabb, AABB, AaBb and aaBb. Obtain the posterior

probabilities of the phase combinations of the parents given these data. Suppose it is known that the father is in coupling phase while the mother is in repulsion phase. Compute the LOD score for evaluating linkage for this family. [20]

(b) Consider the following data from an affected sib-pair study (both sibs are affected and affection status of parents are unknown) on Coronary Artery Disease (CAD). *MTHFR* (methylenetetrahydrofolate reductase) on Chromosome 1 is believed to be a candidate gene for CAD. Twelve sib-pairs comprising all affected siblings (along with their parents) were genotyped at a triallelic marker locus *D1S1012* near this candidate gene. Do these data provide evidence of linkage between *D1S1012* and a locus controlling CAD?

Sib-pair	Parental genotypes	Genotypes of affected sibs
1	AA,AB	AA,AB
2	AB,BC	AB,BC
3	AB,CC	AC,BC
4	BB,BC	BB,BB
5	AB,AB	AB,AB
6	AC,*	AA,AA
7	AB, BB	AB,BB
8	AA,AB	AA,AB
9	BB ,*	BC,BC
10	CC,CC	CC,CC
11	AB,AC	AC,BC
12	AC,AC	AA,AC

* denotes missing genotype [25]

INDIAN STATISTICAL INSTITUTE
Mid-semester Examination: 2012-2013 (Second Semester)

M. Stat II Year

Survival Analysis

Date: February 18, 2012 Maximum Marks: 50 Duration: 2 hours 30 minutes.

Note: Total Mark is 55

1. (a) What is the difference between censoring and truncation?

   (b) Suppose that the time to death T, follows an exponential model with hazard rate λ, and that the random-right censoring time C is exponential with hazard rate θ. Let $X = \min\{T, C\}$ and $\delta = I(T \leq C)$. Assume that T and C are independent. Show that X and δ are independently distributed.

   (c) If independent lifetimes T_1 and T_2 have proportional hazards

$$\lambda_i(t) = \lambda_0(t)\eta_i,$$

   for $i = 1, 2$, respectively, then compute $P(T_1 < T_2)$.

[2 + 8 + 5 = 15]

2. Consider two survival functions $S_0(t)$ and $S_1(t)$ with the corresponding mean residual lifetime $r_0(t)$ and $r_1(t)$ satisfying

$$r_1(t) = \theta r_0(t), \quad \text{for all } t > 0, \theta > 0.$$

   Then show that $S_1(t) = S_0(t)\left[\int_t^\infty \frac{S_0(y)dy}{r_0(0)}\right]^{\frac{1}{\theta}-1}$. [10]

3. (a) Suppose that a survival variable T with covariate z has survival function $S_z^T(t)$ and hazard function $\lambda_z^T(t)$ satisfying

$$\lambda_z^T(t) = \lambda_0^T(t)g(z),$$

   where λ_0^T is the baseline hazard and g is a positive function satisfying $g(0) = 1$. Then show that the hazard function of $Y = \log T$, $\lambda_z^Y(t)$ satisfies

$$\lambda_z^Y(t) = \lambda_0^Y(t)g(z)$$

   (b) Obtain the log-linear representation of Weibull distribution.

[6 + 4 = 10]

4. The data below show remission times, in weeks, for 24 leukemia patients randomly assigned to two treatments A and B. Asterisks denote censoring times.

Treatment A: 1, 2, 3, 6*, 7, 10, 12*, 14, 15*, 18, 20*, 22.
Treatment B: 1, 1, 2, 2*, 3, 4, 5, 8, 8, 9*, 11, 12.

(a) Compute Kaplan-Meier estimate of survival probabilities in Group A.

(b) Suppose that failure times from two groups of individuals are exponential with failure rates λ and λe^{β}, respectively. Derive a test for the homogeneity of the two groups giving full details. What is your conclusion about the homogeneity of the two groups A and B?

$[8 + 12 = 20]$

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination
Semester II : 2012-2013
M.Stat. II Year
Advanced Probability II

Date : 18.02.13 | **Maximum Score : 40** | Time : $2\frac{1}{2}$ Hours

Note : This paper carries questions worth a **total of 52 marks**. Answer as much as you can. The **maximum** you can score is 40.

1. Let $\{X_t, t \in [0,\infty)\}$ be a non-negative submartingale with a.s. continuous paths. Using results from the discrete-parameter case, show that for any $0 \le S < T < \infty$ and any $\lambda > 0$, $P\left(\sup_{S \le t \le T} X_t > \lambda\right) \le \frac{EX_T}{\lambda}$.

   [6]

2. Let $\{B_t, t \in [0,\infty)\}$ denote a standard Brownian motion.
   Denoting $X_n = \sup_{3^n \le t \le 3^{n+1}} B_t^2/t^2$ for $n \ge 1$, show that $\sum_n P(X_n > 2^{-n}) < \infty$.
   Deduce that $B_t/t \longrightarrow 0$ with probability 1, as $t \to \infty$.
   Using the above (or otherwise), show that the process $\{X_t, t \in [0,\infty)\}$ defined as $X_t = tB_{\frac{1}{t}}$ for $t > 0$ and $X_0 \equiv 0$, is a standard Brownian motion.

   (6+4+6)=**[16]**

3. With $(S, \mathcal{S})$ denoting a Polish space equipped with its Borel σ-field, state clearly the definition of a continuous-path markov process with state space S.
   For a markov process with state space S, define what is meant by its resolvent operators R_λ on $C_b(S)$.
   State and prove the resolvent identity and hence show that (i) the R_λ for all λ have the same range and trivial null space; and (ii) the operator $\lambda I - R_\lambda^{-1}$ defined on the common range of the R_λ does not depend on λ.

   (2+2+6+6+6)=**[22]**

4. Let $\{B_t\}$ be an $\{\mathcal{F}_t\}$-Brownian motion on some probability space $(\Omega, \mathcal{F}, P)$. Show that for any $f \in L_2([0,\infty))$, if $\{M_t, t \in [0,\infty)\}$ denotes the continuous-path Weiner integral $M_t = \int_0^t f_s dB_s$, then the process $X_t = M_t^2 - \int_0^t f_s^2 ds$, $t \in [0,\infty)$ is a continuous $\{\mathcal{F}_t\}$-martingale.

   [8]

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination: 2012-13 (Second Semester)
Master of Statistics (M. Stat.) II Year
Stochastic Processes I

Teacher: Parthanil Roy

Maximum Marks: 40 **Date: 21/02/2013** **Duration: 2:30 pm - 5:00 pm**

Note:

- **Please write your name and roll number on top of your answer booklet(s).**
- **There are four problems with a total of 40 points. Show all your works and write explanations when needed.**
- **Please justify all your steps. If you are using a result proved in class, please quote that result. Note that your answer should be consistent with the class-notes.**
- **This is an open note examination. You are allowed to use your own hand-written notes (such as class notes, exercise solutions, list of theorems, formulas etc.). Please note that no printed or photocopied materials are allowed. In particular, you are not allowed to use books, photocopied class notes etc. If you are caught using any, you will get a zero in the mid-semestral examination.**

Suppose (X, d) is a Polish space with Borel σ-field $\mathcal{B}(X)$ and $\mathcal{P}(X)$ denotes the set of all probability measures on $(X, \mathcal{B}(X))$ equipped with the topology of weak convergence. Based on this notation, answer the following questions.

1. Suppose $\{\mu\} \cup \{\mu_n : n \geq 1\} \subseteq \mathcal{P}(X)$ such that $\mu_n(A) \to \mu(A)$ for all Borel set $A \subseteq X$ satisfying $\mu(\partial A) = 0$. Show that $\{\mu_n\}_{n\geq 1}$ is a tight family. [10]

2. Suppose $\{\mu_n : n \geq 1\} \subseteq \mathcal{P}(X)$ is a tight family. State whether the following statements are *true* or *false*. Justify your answer in both cases.

   (a) If any weakly convergent subsequence of $\{\mu_n : n \geq 1\}$ converges to $\mu \in \mathcal{P}(X)$, then $\mu_n \Rightarrow \mu$. [5]

   (b) If $\{\mu_n : n \geq 1\}$ has a subsequence $\{\mu_{n_k} : k \geq 1\}$ such that any weakly convergent subsequence of $\{\mu_{n_k} : k \geq 1\}$ converges to $\mu \in \mathcal{P}(X)$, then $\mu_n \Rightarrow \mu$. [5]

3. Let $\{Y_n\}_{n\geq 1}$ and $\{Z_n\}_{n\geq 1}$ be two sequences of X-valued random variables defined on the same probability space. Suppose $\{Y_n\}_{n\geq 1}$ is a tight family and $d(Y_n, Z_n) \to 0$ in probability as $n \to \infty$. Show that $\{Z_n\}_{n\geq 1}$ is also a tight family. [10]

4. Let $\mu_n \Rightarrow \mu$ in $\mathcal{P}(X)$ and $\phi : X \to \mathbb{R}$ be a measurable map with $D := \{x \in X : \phi \text{ is discontinuous at } x\}$.

   (a) Show that $D \in \mathcal{B}(X)$. [3]

   (b) If $\mu(D) = 0$, then show that $\mu_n \circ \phi^{-1} \Rightarrow \mu \circ \phi^{-1}$ in $\mathcal{P}(\mathbb{R})$. [7]

Wish you all the best

INDIAN STATISTICAL INSTITUTE
Mid-Semestral Examination: (2012-2013)
MS (Q.E.) I Year
Macroeconomics I

Date: 22.02.13 | Maximum Marks 40 | Duration 3 hours

Use separate booklets for group A & B

Group A
Answer *any two* questions

1. Discuss the possibility of an *inflation-unemployment trade-off* in the context of the initial version of the Phillips Curve. What sort of implications does *expectation-augmented* Phillips Curve have on this trade-off? Explain.
(For first part it is enough to draw the Phillips curve and carry out discussion in terms of that curve only without bringing in AF curve. For the second part it is enough to consider only *static* or *naïve* expectation).

[4 + 6] = [10]

2. The following aggregate demand-aggregate supply model seeks to capture the effects of real (supply) shocks on an economy with wage indexation..
   (i) (demand) $y = m - p$ (m, nominal money supply, is a policy variable)
   (ii) (supply) $y = y^* + p - w - u$ (u is a white noise with a given variance)
   (iii) (wage indexation) $w = \theta p$ ($0 \leq \theta < 1$),
   where all variables are in logarithms (and hence, negative or zero value of a variable is admissible); y is output, p is price, w is money wage rate and θ (a parameter) measures the *degree of wage indexation* with respect to the current price, p.

(*a*) Take the initial money supply to be given at $\overline{m}$, initial value of θ to be a *positive fraction and* initial value of $u = 0$. Also using (ii), draw the *demand and supply* curves on a p-y plane. Find equilibrium values of p and y.

(*b*) Suppose now that there is an adverse supply shock ($u > 0$). Draw the demand and supply curves and find equilibrium values of p and y. Show that an appropriate monetary policy can restore output to its level before the shock.

(c) Suppose $u = 0$, $m = \overline{m}$, *but* there is full indexation ($\theta = 1$). Draw the demand and supply curves and find equilibrium values of p and y. Suppose, now that there is an adverse supply shock ($u > 0$). Draw the demand and supply curves and find equilibrium values of p and y. Can any monetary policy restore output to its level before the shock? Explain.

[3 + 3 + 4] = [10]

3. An economy (with no public institution viz. government) has only **three** domestic producing units - a Wheat Farm (**WF**), a Hydroelectric Plant (**HEP**), and a Hand Loom Factory (**HLF**). HEP and WF use freely available rain water and WF uses wheat seed (from its last year's stock). They do not use any other intermediate goods. Figures (in Rs. crores) on the production activities of these units as well as final expenditures by the different agents during a year are given below.

Good	*Purchased from*	*For the Purpose of*					
		Intermediate Use	**Final** Use				
		by the HLF	Consumption (C) *by* Household	Investment: GFI *by* HEP	Investment: Δ S *by* WF	Export: EX *by* WF	Export: EX *by* HLF
Domestic Producers							
Wheat	WF		100		10	40	
Electri--city	HEP	20	90				
cloth	HLF		100				40
Foreign	**Producers**						
Raw Cotton	Abroad	60					
Turbines	Abroad			90			

Here C = consumption, GFI = gross fixed investment, Δ S = change in stock, EX= export.

(a) Find gross value of output of each domestic sector. Compute the value of GDP of this economy (Y) by both the *Value Added* method and the *Final Expenditure* method.

(b) If investment in turbines (all imported from abroad) by the HEP were higher by Rs.10 crores in this year, would the GDP in this year have been higher by Rs 10 crores also, or by more or by less etc.? Argue.

[6 + 4 = 10]

Group B

Answer all

1. Show the condition that perceived demand curve be more elastic than the market share demand curve is sufficient to guarantee that the market equilibrium in the Dixit- Stiglitz model is unique. [10]

2. Show that beginning from a monopolistic equilibrium in the Blanchard- Kiyotaki model if all agents (labourers and producers) could coordinate on an equi-proportionate cut in wages and prices then that would increase both the real profits and utility [10]

INDIAN STATISTICAL INSTITUTE

Mid-semester Examination : Semester II (2012-2013)

Course Name : BSDA (M. Stat. 2nd year)

Subject Name : Statistical Methods in Biomedical Research

Date : 22·02·13 , Maximum Marks : 30. Duration : 1 hr. 30 min.

1. Suppose 60 patients are to be treated in a clinical trial for comparing two treatments A and B. Out of the first 18 patients, 10 are treated by A and 8 are treated by B. What will be the allocation probability for the 19th patient to treatment A for

   (i) Complete randomization;

   (ii) Random allocation rule (RAR);

   (iii) Truncated binomial design (TBD);

   (iv) Permuted block design ($B = 6$, $b = 10$);

   (v) Permuted block design ($B = 10$, $b = 6$);

   (vi) Eforn's biased coin design (mention p);

   (vii) Big stick rule ($c = 2$);

   (viii) Biased coin design with imbalance tolerance (BCDWIT(p,$c = 3$));

   (ix) Friedman-Wei's urn design ($\alpha = \beta$);

   (x) Ehrenfest urn design ($w = 5$);

   (xi) Generalized biased coin design (GBCD($\rho = 5$))?

   (xii) If, in addition, it is known that there are 7 successes by A and 3 successes by B out of the first 18 patients, what will be this probability for RPW($\alpha = \beta$) rule? Here B, b, p, c, α, β w and ρ are design parameters of the respective designs in standard notations.

   [12]

2. If there are only three groups in a group sequential study, give one form of type I error spending function which will spend the total type I error in a 1:7:19 fashion in the three groups (assuming equal time interval for each group). How the type I error will be distributed if there are four groups assuming equal time interval for each group? [2+2]

3. In the Zelen's play-the-winner rule, if the two treatments under consideration have success probabilities 0.7 and 0.4 respectively, find the unconditional probability that the 10-th patient will result in a success. Modify this probability

if it is known that the 7-th patient is treated by the first treatment. How the probability changes if, in addition, it is known that the first patient was treated by the second treatment? What will be the expected number of successes if 100 patients are treated? [4+3+2+5=14]

February 22, 2013

INDIAN STATISTICAL INSTITUTE
M.STAT SECOND SEMESTER, 2012-13
MID-SEMESTRAL EXAMINATION

Full Marks: 60 Applied Multivariate Analysis Time: $2\frac{1}{2}$ Hours

1. Find a value of α that minimizes $g(\alpha) = \sum_{t=1}^{20} |t^2 - \alpha| - 10\alpha$. Is this choice of α unique? Justify your answer. [4+2]

2. Consider a linear regression model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim N(0, 0.25\mathbf{I})$, and the design matrix

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

If $\widehat{\boldsymbol{\beta}}$ is the usual least square estimator of $\boldsymbol{\beta}$, find $\mathbf{t}$ with $\|\mathbf{t}\| = 1$ such that the variance of $\mathbf{t}'\widehat{\boldsymbol{\beta}}$ is minimum. Calculate this minimum variance. Is this choice of $\mathbf{t}$ unique? Justify your answer. [5+2+2]

3. Show that half-space depth of an observation $\mathbf{x}$ w.r.t. the multivariate normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is given by $1 - \Phi\left\{\sqrt{(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}\right\}$, where $\Phi(\cdot)$ is the c.d.f. of the standard normal distribution. [5]

4. If F is spherically symmetric, show that for any $\mathbf{x} \neq \mathbf{0}$, show that $r(\mathbf{x}) = sign(\mathbf{x})\ g(\|\mathbf{x}\|)$, where $r(\mathbf{x})$ denotes the spatial rank of $\mathbf{x}$ w.r.t. the distribution F, $sign(\mathbf{x}) = \mathbf{x}/\|\mathbf{x}\|$ and g is a monotonically decreasing function. [8]

5. If $(\mathbf{X} - \boldsymbol{\theta})$ and $H(\mathbf{X} - \boldsymbol{\theta})$ has the same distribution for all orthogonal matrix H, and the distribution has finite second order moments, show that $E(\mathbf{X}) = \boldsymbol{\theta}$ and $Disp(\mathbf{x}) = \sigma^2\mathbf{I}$ for some $\sigma^2 > 0$. [2+3]

6. Define projection pursuit regression model. Give an example where projection pursuit regression is expected to perform better than additive regression. Give proper justification to your answer. [2+2+2]

7. Suppose that $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are independent observations from a multivariate normal distribution, and we want to test whether this distribution have spherical probability contours. Find the likelihood ratio test statistic and show that it be expressed as a function of the eigenvalues of the moment based sample dispersion matrix. Describe how you will perform the test when (i) n is large (ii) n is small. [6+3+3]

8. Suppose that $\binom{\mathbf{x}_1}{\mathbf{y}_1}, \binom{\mathbf{x}_2}{\mathbf{y}_2}, \ldots, \binom{\mathbf{x}_n}{\mathbf{y}_n}$ are independent realizations of a $(p+q)$ dimensional normal variate $\binom{\mathbf{x}}{\mathbf{y}}$, where p and q are the dimensions of $\mathbf{x}$ and $\mathbf{y}$, respectively. Construct an appropriate test statistic for testing the independence between $\mathbf{x}$ and $\mathbf{y}$. Check whether this test statistic can expressed as a function of canonical correlation coefficients. [4+5]

Statistical Methods in Public Health
Mid-Semestral Examination
M.Stat. II Year, 2012-13
Full Marks - 70
Time - 2 hrs. 30 mins.

Indian Statistical Institute
Kolkata 700 108, INDIA

DATE- 25.02.13

Attempt all questions:

1. Provide two empirical estimates of Relative Growth Rate (RGR) based on the size data available for two time points. Under suitable assumptions find out expressions for bias and MSE of the RGR metric under measurement errors. Also suggest estimates for bias and MSE based on the size data of n individuals available for two specific time points.

[2 + 8 + 4 = 14]

2. Let us define $X(t)$ = size, $R(t) = \frac{1}{X(t)}\frac{dX(t)}{dt}$ = relative growth rate (RGR) at time point t. We assume $(R(1), \ldots, R(q))' \sim N_q(\theta, \Sigma)$, where $E(R(t)) = \theta(t) = f(\phi, t)$, a suitable rate profile, $t = 1(1)q$. Suppose we are interested in testing the hypothesis of Gompertz growth curve model(GGCM), i.e., to test

$$H_0 : \theta(t) = ae^{-bt}ag : H_1 : \text{not } H_0.$$

Using the approximate expression for expectation and variance of the logarithm of ratio of RGR for two consecutive time points, describe two testing procedures and critical regions in testing the null hypothesis of GGCM. Also suggest required modifications of test statistics when the errors are non-normal.

[15 + 5 = 20]

3. (a) Define Lyapunov stability and asymptotic stability for a single species model based on an autonomous differential equation.

(b) Consider the following dynamical system

$$\frac{dx}{dt} = rx(1 - \frac{x}{K}) - \alpha xy$$
$$\frac{dy}{dt} = \beta xy - \gamma y,$$

where x, y be the prey and predator populations respectively. All the parameters have their usual meanings.

(*i*) Find out the biologically feasible equilibria.

(*ii*) Find out the conditions for which both the species will co-exist.

(*iii*) Interpret your findings in biological sense.

(*iv*) When predator population is absent and the prey population grows in a theta-logistic fashion, find out the conditions for stability (Prove necessary theorem if needed).

[3 + 2 + 5 + 2 + 6 = 18]

4. Consider two growth equations for single species dynamics as follows

$$(i)\frac{dx}{dt} = rx\left(1 - \frac{x}{K}\right)^{\theta}$$

$$(ii)\frac{dx}{dt} = rx^2\left[1 - \left(\frac{x}{K}\right)^{\theta}\right],$$

where parameters have their usual interpretations.

(*a*) Express (*i*) and (*ii*) in difference of birth and death rates (For (*i*) you can consider the approximate expression of right hand side).

(*b*) Derive the quasi equilibrium probabilities in each case. Also determine the expression for the approximate mean and variance for model (*ii*) using stochastic perturbation.

[4 + 6 + 8 = 18]

Indian Statistical Institute
Midterm Examination
Asymptotic Theory of Inference
M.Stat. Second Year
Second Semester (2012-2013)

Date:- 25.02.13 Maximum Marks: 30 Duration :- 2 hours

Answer as many questions as you want. The maximum you can score is 30.

1. (a) Consider $X_1, \ldots, X_n$ i.i.d. P_θ, $0 < \theta < 1$ where

$$P_\theta = \begin{cases} \text{Bin}(1, \theta) & \text{if } \theta \text{ is rational} \\ \text{Bin}(1, 1-\theta) & \text{if } \theta \text{ is irrational.} \end{cases}$$

Is θ consistently estimable based on the observations ? Prove your answer. Is there a polynomial in θ which is consistently estimable based on the observations? Prove your assertion. [4+2=6]

(b) Let $X_1, \ldots, X_n$ have some joint distribution indexed by a parameter θ which belongs to a parameter space Θ, a metric space with a certain metric d. Consider an estimator of θ defined as a zero of a random (depending on $X_1, \ldots, X_n$) real-valued criterion function $\theta \to \psi_n(\theta)$. Suppose, for a fixed real-valued function ψ and every $\epsilon > 0$,

$$\sup_{\theta \in \Theta} |\psi_n(\theta) - \psi(\theta)| \xrightarrow{P} 0 \text{ and}$$

$$\inf_{\theta : d(\theta, \theta_0) \geq \epsilon} |\psi(\theta)| > 0 = \psi(\theta_0), \text{ where } \theta_0 \in \Theta.$$

Then show that any sequence of zeros of the criterion function, i.e. estimators $\hat{\theta}_n$ satisfying $\psi_n(\hat{\theta}_n) = 0$ converges in probability to θ_0, i.e $d(\hat{\theta}_n, \theta_0)$ converges to zero in probability. Can we say the same thing about any sequence of *approximate* zeros i.e estimators $\hat{\theta}_n$ satisfying $\psi_n(\hat{\theta}_n) = o_p(1)$? [4+1=5]

2. Is there any similarity in the basic architechture of proofs of consistency of m.l.e by Wald and Ibragimov-Hasminskii ? Justify your answer briefly. [4]

3. (a) Give an example of two sequences of mutually contiguous probability measures on $(-\infty, \infty)$ with the usual Borel sigma field, such that none of them is tight. Prove your assertion. Recall that a sequence of probability measures $\{P_n\}$ on a metric space is tight if given any $\epsilon > 0$, there exists a compact subset of K_ϵ of the metric space such that $P_n(K_\epsilon) > 1 - \epsilon$ for each $n \geq 1$. [4]

(b) Suppose $\{P_{1n}\}_{n \geq 1}$ and $\{P_{2n}\}_{n \geq 1}$ are two sequences of probability measurs on $(R, \mathcal{B})$ such that P_{1n} is absolutely continuous with respect to P_{2n} for each n. Does this mean that $\{P_{1n}\}_{n \geq 1}$ has to be contiguous with respect to $\{P_{2n}\}_{n \geq 1}$? Prove the result or give a counterexample. [4]

(c) Give an example of a sequence of measurable spaces $(\mathcal{X}_n, \mathcal{A}_n)$ and two sequences $\{P_{1n}\}_{n\geq 1}$ and $\{P_{2n}\}_{n\geq 1}$ of probability measures on these measurable spaces such that
(i) $\{P_{1n}\}_{n\geq 1}$ is contiguous with respect to $\{P_{2n}\}_{n\geq 1}$ and
(ii) for some sequence of real valued random variables T_n defined on $(\mathcal{X}_n, \mathcal{A}_n)$, the measure $P_{2n}T_n^{-1}$ converges weakly while $P_{1n}T_n^{-1}$ does not. [5]

(d) State and prove LeCam's Third Lemma. Describe in a few sentences how this lemma might be useful in asymptotic inference problems. [5+2]

INDIAN STATISTICAL INSTITUTE
Mid-Semester Examination: 2012-2013, Second Semester
M-Stat II (MSP)
Ergodic Theory

Date: 26.02.13 Max. Marks 30 Duration: 2 Hours

Note: Answer all questions.
All Measures Are Probability Measures Unless Stated Otherwise

1. Construct a probability space and a measure preserving transformation T on it so that T and T^2 are ergodic but T^3 is not. [5]

2. a) A measure preserving transformation T is weak mixing on $(X, \mathcal{B}_X, \mu)$, if and only if $T \times S$ is ergodic for all ergodic measure preserving transformation S on any probability space $(Y, \mathcal{B}_Y, \nu)$—Prove.

   b) Are both the parts of the statement in (a) correct if 'weak mixing T is replaced with 'ergodic T'? Prove or give counterexample whichever appropriate. [7+8]

3. Let T be an invertible measure preserving transformation on $(X, \mathcal{B}_X, \mu)$. Let $B \in \mathcal{B}_X$ be such that $\mu(B) > 0$. Let $T_B : (B, \mathcal{B} \cap B, \mu|_B) \to (B, \mathcal{B} \cap B, \mu|_B)$ be the transformation on B induced by T. Prove that T_B is invertible measure preserving. Show that T_B is ergodic if T is ergodic. [8]

4. Let T be an invertible measure preserving transformation on a probability space $(X, \mathcal{B}, \mu)$ and let $B \in \mathcal{B}$ with $\mu(B) > 0$. Let $Y = X \times \{0\} \cup B \times \{1\}$. Y is made into a probability space in a natural way i.e. $\mathcal{B}_Y = \mathcal{B} \times \{0\} \cup \mathcal{B} \cap B \times \{1\}$ and μ_Y is normalized $\mu_0 + \mu_1$, where μ_0 is μ and μ_1 is μ restricted to B. Define a transformation S on Y by:

$$S(y) = \begin{cases} (Tx, 0) & \text{if } y = (x,0) \text{ with } x \in B^c \\ (x, 1) & \text{if } y = (x,0) \text{ with } x \in B \\ (Tx, 0) & \text{if } y = (x,1) \text{ with } x \in B \end{cases}$$

Verify that S is a measure preserving transformation on Y and that S is ergodic if T is ergodic. [7]

Indian Statistical Institute
Statistical Methods II
B-I, Midsem

Date: Feb 27, 2013 Duration: 2hrs.

Attempt all questions. The maximum you can score is 40. Justify all your steps. This is a closed book, closed notes examination. You may use your own calculator. No need to perform more than three steps of any iterative method.

If copying is detected in the solution for any problem, all the students involved in the copying will get 0 for that problem. Also an additional penalty of 5 will be subtracted from the overall aggregate of each of these students.

1. A box contains 5 white balls and 2 black balls. A coin with unknown $P(Head) = p$ is tossed. A white ball is added to the box if the outcome is head; otherwise a black ball is added. Then a ball is drawn at random from the box, and the colour is recorded. If the colour is "white", then find the mle of p. [10]

2. Consider the following random experiment. A coin with unknown $P(Head) = p \in (0, 1)$. is tossed. If it shows head, then a random variable is generated with distribution $Uniform(0, p)$, otherwise X is generated from $Uniform(p, 1)$. This random experiment is performed 10 times independently (each time starting with a fresh coin toss) to produce data $X_1, ..., X_{10}$. Find a suitable estimator (mle, mme or something else) of p based on this data. Comment on the your choice of the estimator. [10]

3. Find mle and mme of θ based on a random sample $X_1, ..., X_{10}$ from the distribution $Uniform(-\theta, \theta)$. Suggest how you can compare the performances of the two estimators. [15]

4. Consider the discrete uniform distribution over the set

$$\{\theta - 2, \ \ \theta - 1, \ \ \theta, \ \ \theta + 1, \ \ \theta + 2\},$$

where $\theta \in \mathbb{R}$ is an unknown parameter. Assume that the following is a random sample from this distribution:

$$8, 6, 8, 10, 9.$$

Find mle of θ based on this data. Justify your steps. [5]

5. Based on two iid observations 2.3 and 4.1 from the Cauchy distribution with density

$$f_\theta(x) = \frac{1}{\pi \cdot (1 + (x - \theta)^2)}, \quad x \in (-\infty, \infty),$$

find mle of θ. [10]

Indian Statistical Institute

Second Midsemestral examination

M. Stst. II (MSP)

Statistical Inference II

Date: 28 February, 2013 Maximum marks: 100 Duration: 3 hrs.

Answer all Questions. *Paper carries 110 points.*

1 (a) Briefly describe the notion of statistical experiments.

(b) Let $\mathcal{E} = (\mathcal{X}, \mathcal{A}, \{P_\theta\})$ and $\mathcal{F} = (\mathcal{Y}, \mathcal{B}, \{Q_\theta\})$ be two statistical experiments. Define the deficiency of $\mathcal{E}$ with respect to $\mathcal{F}$. (you have to define all the mathematical concepts and notations required for the definition)

[10 + 15 =25]

2 (a) Suppose $\mathcal{E}, \mathcal{F}$ and the parameter space Θ are finite. Show that there is a Markov kernel which minimizes the deficiency defined above.

(b) Let $\mathcal{E}$ and $\mathcal{F}$ denote Bin(k, θ) and Bin(n, θ) with $0 \leq \theta \leq 1$, respectively, where $k < n$. Find the value of the deficiency $\delta(\mathcal{E}, \mathcal{F})$ along with the the optimizing kernel explicitly. What can you say about $\delta(\mathcal{F}, \mathcal{E})$?

[15 + 20 = 35]

3. Let P_0 and P_1 be two probability measures on some measurable space $(\mathcal{X}, \mathcal{A})$. Show that there is a non-randomized most powerful test for testing P_0 against the alternative P_1 at level α if and only if there is a measurable subset A satisfying (i) $P_0(A) = \alpha$ and (ii) $A = \arg\max_{C \in \mathcal{A}} \left[P_1(C) - \lambda P_0(C)\right]$ for some $\lambda \geq 0$. [25]

4. Describe the three basic principles for statistical evidence, namely, sufficiency (S), conditionality (C) and the likelihood (L) principles with one example each. Prove that evidentially (C) + (S) imply (L) (the basic formulation and assumptions must be clearly stated). [25]

INDIAN STATISTICAL INSTITUTE
Mid-Semestral Examination, Second Semester: 2012-13
M.Stat. II Year (AS)
Actuarial Models

Date: March 1, 2013 Maximum marks: 60 Duration: 2 hours

Answer any combination of questions worth 60 marks. The entire paper is worth 65 marks.

1. A No-Claims Discount system operated by a motor insurer has four levels: 0% discount (Level 1), 25% discount (Level 2), 40% discount (Level 3), 50% discount (Level 4).
The rules for moving between these levels are as follows:
Following a claim-free year, move to the next higher level, or remain at level 4.
Following a year with one or more claims:
move back one level, or remain at level 1, if, in the year before the most recent year, there were no claims;
move back two levels, or move to level 1 or remain at level 1 if, in the year before the most recent year, there was one or more claims.
For a given policyholder, the probability of no claims in a year is 0.8.

   (i) Let $X(t)$ denote the state, either 1, 2, 3 or 4, of the policyholder in year t. Explain why $\{X_t\}_{t=1}^{\infty}$ is *not* a Markov chain. [2]

   (ii) (a) By increasing the number of states, define a new stochastic process $\{Y_t\}_{t=1}^{\infty}$ which *is* Markov and is such that $Y(t)$ indicates the discount level for the policyholder in year t.

   (b) Write down the transition matrix for the Markov chain $\{Y_t\}_{t=1}^{\infty}$.

   (c) Calculate the long-run probability that the motorist is in discount level 3.

   [10]

   [Total 12]

2. An investigation was conducted into the effect marriage has on mortality and a model was constructed with three states: Single, Married and Dead. It is assumed that transition rates between states are constant.

   (i) Sketch a diagram showing the possible transitions between states. [2]

   (ii) Write down an expression for the likelihood of the data in terms of transition rates and waiting times, defining all the terms you use. [3]

   The following data were collected from males and females in their thirties.

Years spent in Married state	: 40,062
Years spent in Single state	: 10,298
Number of transitions from Married to Single	: 1,382
Number of transitions from Single to Dead	: 12
Number of transitions from Married to Dead	: 9

   (iii) Derive the maximum likelihood estimator (MLE) of the transition rate from Single to Dead. [4]

   (iv) Estimate the constant transition rate from Single to Dead and its variance. [2]

   (v) Derive the MLE of the probability that a married person will die before becoming single again, and evaluate this MLE for the above data. [5]

   [Total 16]

3. A time-inhomogeneous Markov jump process has state space $\{A, B\}$ and the transition rate for switching between states equals $2t$, regardless of the state currently occupied, where t is time.

The process starts in state A at time $t = 0$.

(i) Calculate the probability that the process remains in state A until at least time s. [2]

(ii) Show that the probability that the process is in state B at time t, and that it is in the first visit to state B, is given by $t^2 \times e^{-t^2}$. [3]

(iii) (a) Sketch the probability function given in (ii).

(b) Give an explanation of the shape of the probability function.

(c) Calculate the time at which it is most likely that the process is in its first visit to state B. [7]

[Total 12]

4. There is a population of ten cats in a certain neighbourhood. Whenever a cat with fleas meets a cat without fleas, there is a 50% probability that some of the fleas transfer to the other cat such that both cats harbour fleas thereafter. Contacts between two of the neighbourhood cats occur according to a Poisson process with rate μ, and these meetings are equally likely to involve any of the possible pairs of individuals. Assume that once infected a cat continues to have fleas, and that none of the cats owners has taken any preventative measures.

(i) If the number of cats currently infected is x, explain why the number of possible pairings of cats which could result in a new flea infection is $x(10 - x)$. [2]

(ii) Show how the number of infected cats at any time, $X(t)$, can be formulated as a Markov jump process, specifying the state space and the Kolmogorov differential equations in matrix form. [4]

(iii) State the distribution of the holding times of the Markov jump process. [2]

(iv) Calculate the expected time until all the cats have fleas, starting from a single flea-infected cat. [3]

[Total 11]

5. A street-side vendor makes a certain number of egg-rolls, K, before the stall opens each afternoon, and then continues to make c egg-rolls per hour. The j^{th} customer of the evening arrives with a demand for N_j egg-rolls.

(i) Describe a model which would allow you to estimate the probability that the vendor will run out of rolls. State any assumption you make. [4]

(ii) Define a stochastic process appropriate for this study, clearly indicating whether it operates in continuous or discrete time and whether it has a continuous or discrete state space. [3]

(iii) Determine the relevant expression for the probability that the vendor never runs out of rolls in a particular evening, in terms of K, c, and N_j. [3]

(iv) Identify the data one needs to record in order to be able to estimate the above probability. [4]

[Total 14]

INDIAN STATISTICAL INSTITUTE
Mid-Semester Examination: 2012-13
M. Stat. II Year
Sample Survey

Date: 01 · 03 · 13 **Maximum Marks: 40** **Duration: 3 Hours**

Answer any four questions, each carrying 10 marks.

Assignments to be submitted on the day of the examination carry 10 marks.

1. Explain why you look for an unbiased estimator with a low variance in estimating a finite population total.

2. Show why you may never find an unbiased estimator for a finite population total with the uniformly least possible variance.

3. Believing that an enemy's arsenal releases tanks serially numbered 1,2,... through K, make a reasonable guess of the value of K on noting your own army has captured three tanks bearing labels 210, 38 and 155. What & how can you argue about the level of the accuracy in this exercise?

4. Show that Horvitz & Thompson's estimator for a finite population total is admissible among all unbiased estimators.

   Suggest one more such admissible estimator as an alternative giving justifications.

   [8+2]

5. Given the following in usual notations

   U = (1, 2, 3, 4, 5, 6, 7)

   Y = (12, 18, 0, 7, 4, 18, 17)

   s= (3, 6, 5, 2, 3)

   Survey data d = ?
   Sufficient statistic d* = ?

   Answer the question. Give a complete reasoning

   [5+5]

6. Find Hartley & Ross's unbiased estimator for a finite population total, supplying an unbiased estimator for its variance. Give a theory for checking if this variance estimator is uniformly non-negative.

   [4+2+4]

7. Illustrate with valid arguments how a sufficient statistic may be useful in estimating a finite population parameter.

*****

INDIAN STATISTICAL INSTITUTE

M.Stat. 2[nd] year, 2012-13 (Mid-semestral Examination)

Subject: Theory of Games and Statistical Decisions

Date: 04/03/2013 Full Marks: 40 Time: 1 hr 15 mins

The paper contains questions of 45 marks. Attempt all. You can score a maximum of 40.

1. a) Consider a non-cooperative game Γ with more than two players.

   Discuss how a cooperative game is naturally derived from it on the same set of players.

   Show that the characteristic function thus derived is super-additive.

   b) Define essential games and inessential games.

   Show that each essential game is strategically equivalent to a unique game in 0-1 reduced form.

   [11+ 7]

2. a) Show that if every 2x2 submatrix of a matrix A has a saddle point, then A has a saddle point.

   b) Consider a two-player zero-sum non-cooperative game with set of strategies

   $S_I = S_{II} = \{ X : X \text{ is a random variable and } 0 \leq X \leq 1 \}$

   and pay-off of the situation (X,Y) for the first player is given by $E(|X-Y|)$ where X,Y are independent.

   Find the value of the game and the set of all optimal strategies.

   c) Let A be a 3x3 diagonal matrix with diagonal elements a_1, a_2, a_3 respectively. Considering the mixed extension of the matrix game A, find the value and the set of all optimal strategies.

   [10+ 10+ 7]

****

INDIAN STATISTICAL INSTITUTE

Semestral Examination: 2012-13 (Second Semester)
Master of Statistics (M. Stat.) II Year
Stochastic Processes I

Instructor: Parthanil Roy

Total Points: 60 | **Date: 26/04/2013** | **Duration: 3 hours**

- **Please write your name and roll number on top of your answer booklet(s).**
- **Please justify all your steps. If you are using a result proved in class, please quote that result. Note that your answer should be consistent with the class-notes.**
- **This is an open note examination. You are allowed to use your own hand-written notes (such as class notes, exercise solutions, list of theorems, formulas etc.). Please note that no printed or photocopied materials are allowed. In particular, you are not allowed to use books, photocopied class notes etc. If you are caught using any, you will get a zero in the semestral examination.**

1. This problem sketches an alternative proof of reflection principle for Brownian motion without using its strong Markov property. Suppose $X_1, X_2, \dots$ are independent, identically distributed and centred random variables. Set $S_0 \equiv 0$ and for all $n \geq 1$, define $S_n := X_1 + X_2 + \cdots + X_n$ and $Y_n = \max\{S_0, S_1, \dots, S_n\}$. Solve (a) - (c) without using *strong Markov property* or *reflection principle for Brownian motion.*

   (a) (10 points) If further
   $$P(X_1 = 1) = P(X_1 = -1) = \frac{1}{2}, \tag{1}$$
   then show that $n^{-1/2} Y_n$ converges weakly and find its weak limit.

   (b) (6 points) Let W be the Wiener measure on $C[0,1]$. Find the law of
   $$Z(\omega) = \sup_{0 \leq s \leq 1} \omega(s), \quad \omega \in C[0,1]$$
   under W.

   (c) (2 points) Will the sequence $n^{-1/2} Y_n$ converge weakly if X_i's have finite second moment but they are not necessarily of the form (1)? If yes, find the weak limit. If not, justify your answer.

[P. T. O.]

(d) (2 points) Give a proof of reflection principle for Brownian motion without using its strong Markov property.

2. Find whether the following statements are true or false. If a statement is true, give a detailed proof. If it is false, disprove it by producing a counter-example.

   (a) (6 points) If $(s,\omega) \mapsto Y_s(\omega)$ is a bounded real-valued jointly measurable map defined on $[0,\infty) \times C[0,\infty)$, then $(x,s) \mapsto E_x(Y_s)$ is also a jointly measurable map.

   (b) (6 points) If $\omega \mapsto Y(\omega)$ is a bounded real-valued measurable map defined on $C[0,\infty)$, then $x \mapsto E_x(Y)$ is a continuous map.

   (c) (6 points) If X is a Polish space and $\mu_n \Rightarrow \mu$ in $\mathcal{P}(X)$, then

$$\liminf_{n\to\infty} \int_X f(x)\,\mu_n(dx) \geq \int_X f(x)\,\mu(dx)$$

   for any bounded measurable function $f : X \to \mathbb{R}$ satisfying $\liminf_{x\to x_0} f(x) \geq f(x_0)$ for all $x_0 \in X$.

3. (12 points) Suppose $\{B_t\}_{t\geq 0}$ is a standard Brownian motion starting from 0. Fix $u > 0$. Compute the joint probability density function of $M_u := \sup_{0\leq s\leq u} B_s$ and B_u.

4. (10 points) Assignments.

APRIL 30, 2013

INDIAN STATISTICAL INSTITUTE
SECOND SEMESTRAL EXAMINATION : 2012-2013
SECOND SEMESTER, M.STAT - II YEAR

FULL MARKS: 100 APPLIED MULTIVARIATE ANALYSIS TIME: $3\frac{1}{2}$ HOURS

1. Suppose that a random vector $\mathbf{X} = (X_1, X_2 \ldots, X_{10})'$ follows a 10-dimensional normal distribution with the dispersion matrix $\mathbf{\Sigma} = (1-\rho)\mathbf{I} + \rho\mathbf{1}\mathbf{1}'$, where $\mathbf{I}$ is the 10×10 identity matrix and $\mathbf{1}$ is the 10×1 vector with all elements unity.

   (a) Define $S = \{\alpha \in R^{10} : \|\alpha\| = 1\}$ as the set of 10-dimensional vectors with unit norm.
   (i) Find an α_0 that maximizes the variance of $\alpha'\mathbf{X}$ with respect to α in S.
   (ii) Find an α_0 that minimizes the variance of $\alpha'\mathbf{X}$ with respect to α in S. [5+5]

   (b) Find the correlation co-efficient between the two linear combinations obtained in (i) and (ii) of (a) when $\rho = 0.5$. [2]

2. Suppose that $\mathbf{Z} = (Z_1, Z_2, Z_3)'$ follows a distribution with the mean vector $(10.5, 12.3, 13.9)'$ and the dispersion matrix

$$\mathbf{\Sigma} = \begin{bmatrix} 1.00 & 0.63 & 0.45 \\ 0.63 & 1.00 & 0.35 \\ 0.45 & 0.35 & 1.00 \end{bmatrix}.$$

   (a) Is it possible to generate $\mathbf{Z}$ using a valid $k = 1$ factor model ? If you think that it is possible, find the communalities and the specific variances. If you think that it is not possible, give justification to your answer. [6]

   (b) The eigen values and the eigen vectors of $\mathbf{\Sigma}$ are $\lambda_1 = 1.96$, $\mathbf{e}_1 = (0.625, 0.593, 0.507)'$, $\lambda_2 = 0.68$, $\mathbf{e}_2 = (-0.219, -0.491, 0.843)'$, $\lambda_3 = 0.36$ and $\mathbf{e}_3 = (0.749, -0.638, -0.177)'$. Assuming $k = 2$ factor model, calculate the factor loading matrix and the specific variances. [6]

   (c) Use a suitable two dimensional plot and check whether factor rotation helps you in interpreting the result obtained in (b). [3]

3. Suppose that the distribution of a four dimensional random vector $\mathbf{X} = (X_1, X_2, X_3, X_4)'$ is an equal mixture of two independent normal distributions $N(-\mathbf{1}, (1-\rho)\mathbf{I} + \rho\mathbf{1}\mathbf{1}')$ and $N(\mathbf{1}, (1+\rho)\mathbf{I} - \rho\mathbf{1}\mathbf{1}')$, where $0 < \rho < 1/3$, $\mathbf{I}$ is the 4×4 identity matrix, and $\mathbf{1}$ is the four dimensional vector with all elements unity.

   (a) Find the dispersion matrix of this distribution. [4]

   (b) Define $Y_1 = a_1X_1 + a_2X_2$ and $Y_2 = a_3X_3 + a_4X_4$. Show that the correlation coefficient between Y_1 and Y_2 cannot exceed 2/3. [8]

4. Assume that $\mathbf{X} = (X_1, X_2, X_3)'$ follows an elliptically symmetric unimodal distribution with the mean vector $(0, 0, 0)'$ and the dispersion matrix $\mathbf{\Sigma} = 3\mathbf{I} + \mathbf{1}\mathbf{1}'$, where $\mathbf{I}$ is the 3×3 identity matrix and $\mathbf{1}$ is the three dimensional vector with all elements unity.

   (a) Show that for any given α $(0 < \alpha < 1)$, the smallest region (in terms of volume) that has probability α is an ellipsoid. [4]

   (b) Derive the equation of the smallest region that has probability at least 0.75. [5]

   (c) Find the volume of this region. [3]

5. (a) Consider a two-class classification problem, where the density functions of the two-classes are $f_1(x) = 1,\ 0 \le x \le 1$ and $f_2(x) = 1 + \beta sin(2k\pi x),\ 0 \le x \le 1$, where k is a positive integer and $\beta \in (0, 1)$. Assuming the prior probabilities of the two classes are equal, find the classifier with the smallest total misclassification probability and show that the total misclassification probability of this classifier cannot be smaller than 1/3. Is this classifier minimax ? Justify your answer. [2+5+2]

   (b) Give an example of a two-class classification problem, where the competing classes have different density functions, but the Bayes risk is equal to smaller of the two prior probabilities. [3]

6. (a) Show that logistic discriminant analysis can be viewed as a generalization of Fisher's linear discriminant analysis. [4]

   (b) Show that a kernel density estimator with the triangular kernel function can be viewed as a limiting case of an average shifted histogram density estimator. [5]

   (c) Consider the following data set consisting of 10 observations.

X_1	17	22	18	25	14	17	24	14	21	25
X_2	21	16	18	15	23	19	24	12	18	17
Class	A	A	A	A	A	B	B	B	B	B

   Sketch the class boundary of the 1-nearest neighbor classifier constructed based on this data set. [6]

7. Assume that we have n independent observations from a mixture of three unimodal distributions differing in their locations with mixing proportions p^2, $2pq$ and q^2, where $p > 1/2$.

   (a) Describe how you will use the idea of cluster analysis to estimate p. [6]

   (b) If it is known that all three unimodal distributions are normal, how will you modify your estimation procedure ? [6]

8. Computer assignments [10]

Indian Statistical Institute

Statistical Methods in Public Health
Semestral Examination
M.Stat. II Year, 2012-13
Full Marks - 100
Time - 3 hrs. 30 mins.

Attempt all questions

Group A

1. (a) State and prove the Bendixon Negative criterion to show that there are no periodic orbits of any 2-D autonomous differential equation. Illustrate the above criterion in a simple epidemic model.

(b) Write down the stochastic version of the simple epidemic model. Show that the outcome of this model is not the same as the result obtained from the analogous deterministic model.

[(6 + 4) + (3 + 7) = 20]

2. Write down the Kermack- Mckendrick model for general epidemics and hence find out the basic reproduction number. How does this model differ from simple epidemics? Show that in the case of general Epidemics the disease dies out from lack of infectives but not for lack of susceptibles.

[6 + 2 + 7 = 15]

3. State clearly the basic assumptions of the gonorrhea model. How does this model differ from other infectious disease model? Discuss the nature of the dynamics of the gonorrhea model.

[4 + 2 + 9 = 15]

Group B

4. Initially an experiment is set up with one toxic (*Dinophysis sp.*) and two non-toxic phytoplankton (*Chaetocerous gracilis* and *Biddulphia regia*) species. All the three phytoplankton growth profiles are monitored with regular recorded biomass of 16 samples on each of the 8 experimental days. The phytoplankton are collected from the deltaic region of river Subarnarekha ($87^o31^{'}$E and $21^o37^{'}$N) and the isolation is done in the laboratory. Species culture are maintained at optimal conditions in the laboratory in 16 separate bikers although the species might exhibits a negligible amount of genetic variation.

(a) Assuming multivariate normality with suitably chosen covariance structure among the recorded biomass of 8 time points, suggest two estimates of Relative Growth Rate (RGR). Find the asymptotic mean and the variance of any one of these two estimates. Suggest a suitable test procedure for testing the equality of expected RGRs for three species at the final time period.

[2 + 5 + 5 = 12]

(b) The above experiment (as described in Question (1)) may involve measurement errors due to initial irregularities of the laboratory setup. Stating suitable assumptions, find out expressions for the bias and the MSE of the Fisher's RGR metric under these measurement errors for the initial time interval(Second order of approximation). Assuming logistic growth for the biomass profile, suggest an improved metric for RGR in comparison with the usual Fisher's RGR based on three consecutive time points.

[7 + 4 + 4 = 15]

5.(a) Discuss "Propagation of Errors" in the context of measurement errors as described in Question 4(b). Based on the available 16 samples provide two estimates of relative error for initial time point. Comment on its asymptotic behaviour.

[3 + 2 + 7 = 12]

or

Let us consider the model

$$d(x(t)) = \mu(x(t)) + \sigma(x(t)) * dW(t),$$

where, $d(x(t))$ is the approximate population size change in time interval $d(t)$ and $d(W(t))$ is the differential of Wiener process having mean zero and variance $d(t)$. The infinitesimal mean $\mu(x(t))$ specifies underlying deterministic growth curve, while the infinitesimal variance $\sigma(x(t))$ corresponds to stochastic fluctuations.

(a) Assuming exponential growth and homoscadastic variance structure, find the expression for expected time to extinction with initial population size as $x(0)$.

(b) Derive the limiting growth model when

$$\mu(x(t)) = r_m \left(\frac{x(t)}{a} - 1\right)\left(1 - \left(\frac{x(t)}{k}\right)^{\theta}\right)$$

and comment on expected time to extinction.

[5 + 4 + 3 = 12]

6. Let us assume that the biomass profile (as described in Question (4)) of one non-toxic phytoplankton species used to follow θ-logistic growth law and genetic variations are not in negligible amount. Further we assume that due to this variation, the instantaneous growth rate parameter of θ-logistic law follows a normal distribution with suitable mean and variance. Using this information construct a profile likelihood to estimate θ. You can assume carrying capacity as a nuisance parameter. Discuss two other methods of estimation in this context. Suggest a suitable testing procedure for

$$H_0 : \theta = 1, \quad vs. \quad H_1 : \theta > 1,$$

using the score function.

[4 + 3 + 4 = 11]

or

Consider the growth equation for single species dynamics as follows

$$\frac{dx}{dt} = rx^{\gamma}\left[1 - \left(\frac{x}{K}\right)^{\theta}\right],$$

where parameters have their usual interpretations.

Derive the quasi equilibrium probability (write necessary steps only) and determine the expression for the approximate mean and variance incorporating stochastic perturbation in the given model.

[5 + 6 = 11]

INDIAN STATISTICAL INSTITUTE

Second Semestral Examination (2012-2013)

M.Stat. 2nd Year

ASYMPTOTIC THEORY OF INFERENCE

Date: 30 April, 2013 Max. Marks: 100 Duration: $3\frac{1}{2}$ Hours

Answer as many questions as you can. The Maximum you can score is 100.

1. Let $\{(\mathcal{X}^n, \mathcal{A}^n), P_\theta^n, \theta \in \Theta\}$, $n \geq 1$, be a sequence of statistical experiments where Θ is an open subset of R. Assume that for some fixed $\theta_0 \in \Theta$ and for all $u \geq 0$,

$$\frac{dP^n_{\theta_0+u/n}}{dP^n_{\theta_0}} = \begin{cases} \exp\left(u\Delta_n(\theta_0) + \epsilon_n(u, \theta_0)\right), & \text{if } Z_n > u \\ 0, & \text{if } Z_n < u \end{cases}$$

where $\Delta_n(\theta_0)$ is a sequence of random variables converging in $P^n_{\theta_0}$ probability to some positive constant $c(\theta_0)$, $\epsilon_n(u, \theta_0)$ converges in $P^n_{\theta_0}$ probability to zero and Z_n is a sequence of random variables converging weakly to some distribution. Show that $\{P^n_{\theta_0+u/n}\}$ is contiguous to $\{P^n_{\theta_0}\}$ for all $u \geq 0$ if and only if $P^n_{\theta_0}(Z_n > t) \to \exp(-tc(\theta_0))$ for all $t \geq 0$ (in this case the family $\{P^n_\theta, \theta \in \Theta\}$ is said to be locally asymptotically exponential at θ_0).

[12]

2. Let $X_1, X_2, \ldots, X_n$ be i.i.d. with a common density $f(x, \theta) = e^{-(x-\theta)}$, $x > \theta$, $\theta \in R$ and let P^n_θ denote the joint distribution of $X_1, X_2, \ldots, X_n$ under θ.

(a) Show that the family $\{f(\cdot, \theta), \theta \in R\}$ is not quadratic mean differentiable but the corresponding family $\{P^n_\theta, \theta \in R\}$ is locally asymptotically exponential at all $\theta_0 \in R$.

(b) Verify whether for fixed $\theta_0 \in R$ and $u \in R$, $\{P^n_{\theta_0}\}$ is contiguous to $\{P^n_{\theta_0+u/n}\}$.

[(10+6)+6 = 22]

3. Let $X_1, X_2, \ldots, X_n$ be a random sample from a $N(\theta, 1)$ population. Consider the problem of testing $H_0 : \theta = 0$ vs $H_1 : \theta = \delta/\sqrt{n}$, $\delta > 0$ (fixed). Find the limiting power of the corresponding sign test using Le Cam's lemma.

Also find Pitman's asymptotic relative efficiency of the sign test relative to the test based on the sample mean.

[20]

4. Describe Fisher's notion of asymptotic efficiency of estimators and comment on this in the light of Hodges' example. Write a short note on the Hajek-Le Cam theory of efficient estimation.

[14]

5. (a) Define a regular sequence of estimators. State the Hajek-Inagaki convolution theorem.

Show that the Hodges estimator is not regular (assume that the observations are i.i.d. $N(\theta, 1)$, $\theta \in R$).

(b) Use the convolution theorem to find, under LAN condition, a (non-trivial) lower bound to the limiting risk of a regular estimator for a subconvex loss.

[15+12=27]

6. Let $X_1, X_2, \ldots, X_n$ be i.i.d. with a common distribution function $F \in \mathcal{F}$ where $\mathcal{F}$ is the class of all distribution functions on R. Consider a real valued functional $T(\cdot)$ on $\mathcal{F}$.

(a) Define the k-th order Gateaux differential of $T(\cdot)$ at F in the direction of some fixed $G \in \mathcal{F}$.

(b) Consider the problem of estimation of $T(F) = E_F h(X_1, \ldots, X_m)$ for a kernel $h(x_1, \ldots, x_m)$, $m \leq n$. Find the k-th order Gateaux differential of the functional $T(\cdot)$ and hence express the first order Gateaux differential of $T(\cdot)$ at F in the direction of the sample distribution function F_n as an average of i.i.d. random variables.

Show, under suitable assumptions, that $\sqrt{n}(T(F_n) - T(F))$ is asymptotically normal. Use the asymptotic equivalence of a U-statistic and its projection and that of a U-statistic and a V-statistic to prove this result.

[2+12=14]

Semester 2, 2012-2013 FINAL EXAM Date: 03 May 2013

MStatII DIRECTIONAL DATA ANALYSIS Time: 3 hrs.

Show all your work. Marks are indicated in the margin. Total marks: 100.

Q. 1. [25=4+8+7+6] Consider the circular distribution with p.d.f. given by $f(\theta) = K.[1 + 2\rho cos(\theta - \mu) + 2\rho^2 cos2(\theta - \mu)]$

(a) Obtain K. (b) Obtain the mean direction and the circular variance. (c) Derive the characteristic function for $f(\theta)$. (d) Verify whether $f(\theta)$ is unimodal.

Q. 2. [25=7+7+11] (a) Let $\Psi = \Theta - \Phi$, where Θ has distribution $f(\theta)$ given in Q. 1 above and Φ has a Circular Uniform distribution, and Θ and Φ are independently distributed. Obtain the distribution of Ψ.

(b) Identify the wrapped Cauchy distribution as a member of the wrapped stable family of distributions. Show that its corresponding Fourier series form has an equivalent single term representation.

(c) Derive the Locally Most Powerful Invariant test for Isotropy against the distribution $f(\theta)$ defined in Q. 1 above.

Q. 3. [25=6+(10+3+6)(a) Establish the entries of the table for Analysis of Mean Directions (ANOMED) for k independent von Mises distributions, stating your assumptions explicitly. Give a real-life example of the application of the corresponding test.

(b) (i) Starting with linear and circular probability distributions as members of exponential families, show how in general you can derive the distribution of the points in a unit disc. State precisely, without proof, the basic theorem you may need for this derivation. (ii) Give a specific example of a distribution in (i). (iii) Show explicitly how you can obtain consistent estimators of the parameters of the distribution in (ii).

Q. 4. [25] TAKE HOME. Given the directional data set in the class, suggest for it a probability distribution defined on the hyperdisc. Obtain estimates of the parameters of this distribution based on the technique of obtaining consistent estimators for such distributions.

Semester 2, 2012-2013 FINAL EXAM Date: 03 May 2013

MStatII DIRECTIONAL DATA ANALYSIS Time: 3 hrs.

Show all your work. Marks are indicated in the margin. Total marks: 100.

Q. 1. [25=4+8+7+6] Consider the circular distribution with p.d.f. given by $f(\theta) = K.[1 + 2\rho cos(\theta - \mu) + 2\rho^2 cos2(\theta - \mu)]$

(a) Obtain K. (b) Obtain the mean direction and the circular variance. (c) Derive the characteristic function for $f(\theta)$. (d) Verify whether $f(\theta)$ is unimodal.

Q. 2. [25=7+7+11] (a) Let $\Psi = \Theta + \Phi$, where Θ has distribution $f(\theta)$ given in Q. 1 above and Φ has a Circular Uniform distribution, and Θ and Φ are independently distributed. Obtain the distribution of Ψ.

(b) Identify the wrapped Cauchy distribution as a member of the wrapped stable family of distributions. Show that its corresponding Fourier series form has an equivalent single term representation.

(c) Derive the Locally Most Powerful Invariant test for Isotropy against the distribution $f(\theta)$ defined in Q. 1 above.

Q. 3. [25=6+(10+3+6)(a) Establish the entries of the table for Analysis of Mean Directions (ANOMED) for k independent von Mises distributions, stating your assumptions explicitly. Give a real-life example of the application of the corresponding test.

(b) (i) Starting with linear and circular probability distributions as members of exponential families, show how in general you can derive the distribution of the points in a unit disc. State precisely, without proof, the basic theorem you may need for this derivation. (ii) Give a specific example of a distribution in (i). (iii) Show explicitly how you can obtain consistent estimators of the parameters of the distribution in (ii).

Q. 4. [25] TAKE HOME. Given the directional data set in the class, suggest for it a probability distribution defined on the hyperdisc. Obtain estimates of the parameters of this distribution based on the technique of obtaining consistent estimators for such distributions.

INDIAN STATISTICAL INSTITUTE

Semester Examination: 2012-2013, Second Semester

M-Stat II

Ergodic Theory

Date: 3.5'13 Max. Marks 60 Duration: 3 Hours

Note: 1. Answer all questions.

2. All the measures considered are probability measures unless otherwise stated.

3. Total Marks: 65. Maximum you can score: 60.

1. a) Let T_1 and T_2 be two measure-preserving transformations on $(X_1, \mathcal{B}_1, m_1)$ and $(X_2, \mathcal{B}_2, m_2)$ respectively. Show that if $T_1 \times T_2$ on $(X_1 \times X_2, \mathcal{B}_1 \otimes \mathcal{B}_2, m_1 \otimes m_2)$ is ergodic then $E(T_1) \cap E(T_2) = \{1\}$, where $E(T_1)$ (respectively $E(T_2)$) denotes the set of eigenvalues of T_1 (respectively T_2).

   b) Let $K = \{z \in \mathbb{C} : |z| = 1\}$ with Borel σ-field and Lebesgue measure. Let $\alpha \in K$ and T_α denote the rotation by α acting on K.
   i) Show that T_α is not weak mixing.
   ii) If α, $\beta \in K$, then find necessary conditions α and β must satisfy if $T_\alpha \times T_\beta$ is ergodic.

   [7 + (7+7)]

2. Suppose T is a measure preserving transformation on a probability space $(X, \mathcal{B}, m)$. An eigenvalue λ is called rational eigenvalue if there exists a positive integer k such that $\lambda^k = 1$.

   a) Show that if T^n is ergodic for all integers $n \geq 1$, then T does not admit any rational eigenvalue except 1.

   b) If T is ergodic but T^2 is not, then show that -1 is an eigenvalue of T.

   [6+6]

3. a) Let X be a Polish space and $\mathcal{B}$ its Borel σ-field and m a probability measure. Let $(\tilde{\mathcal{B}}, \tilde{m})$ denote the corresponding measure algebra. If T

is a measure-preserving transformation on $(X, \mathcal{B}, m)$ then show that T is invertible mod 0 if and only if $T^{-1}(\tilde{\mathcal{B}}) = \tilde{\mathcal{B}}$. (You must state clearly if you use any theorem on Polish spaces.)

b) Let $(X_i, \mathcal{B}_i, m_i, T_i)$, $i = 1, 2$, be two measure-preserving dynamical systems where X_i are Polish spaces and $\mathcal{B}_i$ are Borel σ-fields. If T_1 and T_2 are conjugate then show that T_1 is invertible mod 0 if and only if T_2 is so.

[6+6]

4. a) Let $(X, \mathcal{B}, m, T)$ be a dynamical system and let $\mathcal{A}$ be a finite sub-field of $\mathcal{B}$. Show that $h(T, \mathcal{A}) = H(\mathcal{A})$ if and only if $\{T^{-n}\mathcal{A}, n \geq 0\}$ is an independent sequence.

b) Let $(X, \mathcal{B})$ be a measurable space and m_1 and m_2 two probability measures on $\mathcal{B}$. Let T be measure preserving with respect to both m_1 and m_2. Then show that, if $\alpha \in [0, 1]$,

$$h_{\alpha m_1 + (1-\alpha) m_2}(T) = \alpha h_{m_1}(T) + (1 - \alpha) h_{m_2}(T),$$

where $h_m(T)$ denotes the entropy of the dynamical system $(X, \mathcal{B}, m, T)$.

c) Show that if $p_i \leq q_j$ for all i and j, then the entropy of the Bernoulli shift $\mathcal{B}(p_1, p_2, \ldots, p_n)$ is greater than or equal to the entropy of $\mathcal{B}(q_1, q_2, \ldots, q_m)$.

[6+7+7]

INDIAN STATISTICAL INSTITUTE

Second-semesteral Examination : (2012-2013)

M. Stat 2nd Year

Pattern Recognition and Image Processing

Date: May 3, 2013 Maximum marks: 100 Time: 3 hours.

Note: Attempt all questions. Maximum you can score is 100. Answer Group A and Group B questions in separate answerscripts.

Group A

1. For a two-class problem with equal prior probabilities and misclassification costs, let the class densities be $p_i(\mathbf{x}) = N(\mu_i, \sigma^2 I)$, for $i = 1, 2$.

   (a) Express the optimal error rate P^* in terms of the c.d.f. of the standard normal density and $\frac{\|\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1\|}{2\sigma}$. [8]

   (b) Let, $\boldsymbol{\mu}_1 = \mathbf{0}$ and $\boldsymbol{\mu}_2 = \mu \mathbf{1}_M$, where $\mathbf{1}_M$ is the M-dimensional vector with all entries equal to one. Show that in this case, P^* goes to 0 as M goes to infinity. Give an interpretation of this result. [8]

2. Assuming equal prior probabilities and equal misclassification costs, find the overall Bayes error and the overall asymptotic probability of error for the 1-NN rule in a 2-class problem based on a single feature x where the class densities $p_1(x)$ and $p_2(x)$ are as given below: [6 + 8]

$$p_1(x) = \begin{cases} x & \text{for } 0 \le x \le 1, \\ 2 - x & \text{for } 1 \le x \le 2 \end{cases} \quad \text{and} \quad p_2(x) = \begin{cases} x - 1 & \text{for } 1 \le x \le 2, \\ 3 - x & \text{for } 2 \le x \le 3. \end{cases}$$

3. Show the required architecture (with appropriate weights) of a multilayer perceptron which can solve the two-dimensional two-class classification problem where the first class consists of all points in the first and the third quadrants of the measurement space and the second class is the complement of the first class. [10]

4. (a) For any binary tree T, let $| T |$ be the number of nodes and $| \widetilde{T} |$ be the number of terminal nodes of T respectively. Show that $| T | = 2 | \widetilde{T} | - 1$. (Hint: Use induction) [5]

   (b) Describe briefly the minimal cost-complexity pruning algorithm for generating the α-sequence and the corresponding smallest minimizing subtree sequence in CART. [10]

(Please Turn Over)

Group B

1. (a) Describe briefly the complete-linkage algorithm for clustering a set of n observations on a vector random variable $\boldsymbol{X}$.

   (b) Use this algorithm to split a set of 6 data points into three clusters, if the pairwise distances between them are as given in the following matrix:

$$\begin{bmatrix} 0 & 1 & 2 & 6 & 11 & 5 \\ 1 & 0 & 2 & 5 & 10 & 4 \\ 2 & 2 & 0 & 4 & 9 & 4 \\ 6 & 5 & 4 & 0 & 5 & 1 \\ 11 & 10 & 9 & 5 & 0 & 6 \\ 5 & 4 & 4 & 1 & 6 & 0 \end{bmatrix}$$

   (c) Sketch the dendrogram for the clustering obtained.

   (d) Describe briefly any method known to you for assessing the validity of clusters obtained by any (clustering) algorithm.

[4+7+4+5=20]

2. (a) Explain the general principle behind frequency domain methods for enhancement of a graylevel image. Give an example of a filter that implements smoothing in the frequency domain.

   (b) If a graylevel image is corrupted by "salt" noise, describe any filter that can be used to restore it.

   (c) Discuss how the Laplacian of a graylevel image can be used to sharpen it.

[(5+2)+4+4=15]

3. (a) Explain how a graylevel image can be segmented into regions having homogeneous gray levels, through region splitting and merging based on quadtrees, stating clearly the underlying assumptions.

   (b) Describe how the Hough transform is used to detect subsets of dark pixels which lie on straight line segments in a binary image.

   (c) Discuss, with examples if necessary, how redundancy can be reduced in a graylevel image by run-length coding. What compression ratio is achieved by this coding for an $M \times N$ image in which all pixels have the same intensity value?

[6+7+(4+3)=20]

INDIAN STATISTICAL INSTITUTE

Final Examination : Semester II (2012-2013)

Course Name : BSDA (M. Stat. 2nd year)

Subject Name : Statistical Methods in Biomedical Research

Date : 06·05·13 , Maximum Marks : 60. Duration : 3 hrs.

[Separate booklets are to be used for the two groups.]

Group A

1. In a trial of a new treatment for stroke, a thrombolytic (blood-thinning) agent was compared with an existing treatment. The outcome was whether or not there was a favourable outcome (i.e. a binary variable) as assessed with the aid of a scale widely used by physicians in the field. The trial was designed to have 80% power to find an improvement of 0.1 in the mean proportion with favourable outcome at the 5% significance level. How many patients, n, would be needed in each group if the proportion responding favourably in the existing treatment group was 0.3? How would this change if the proportion responding in the existing treatment group was thought to be 0.2? What would be the number if the proportion responding to the existing treatment were 0.6? Compare the results and comment. [4+2+2+2=10]

2. Discuss the randomized version of the play-the-winner rule. Suppose you start with one ball of each kind. If the difference in the success probabilities by the two treatments is 0.4, how many balls need to be added for every entering patient so as to achieve a probability of 0.55 of allocating the second patient to the better treatment? For that choice of design parameter if, in addition, it is known that the probability that the third incoming patient is treated by the better treatment is 0.582, what are the probabilities of success of the two treatments?

[2+4+4=10]

Statistical Methods in Biomedical Research

M. Stat-II (2012-2013)

Semester Examination

Group B

Note: The following questions carry a total of 50 marks. Answer as many as you can. The maximum you can score is 40 in Group B. Use a separate Answer script .

1.Consider the following data on two groups of eight rabbits allotted at random to the two preparations of flaxedill. One μg per cc solution was injected at the rate of 4 cc per minute to each of them. Injection was stopped as soon as there was a head drop and the corresponding dose was observed for each rabbit. The head drop doses are presented in the table below.

Standard preparation μg	Test preparation μg
840	580
350	500
420	570
400	700
540	800
600	600
560	550
550	470

Identify the type of assay. Estimate the relative potency of the test preparation of flaxedill relative to the standard preparation along with the standard error of the estimate.

1+2+4=7

2.a) Define the "Preparation contrast", "Combined slope contrast" and "Parallelism contrast" justifying the associated names in a 2k-point symmetrical Parallel Line assay. Using the above contrasts derive an estimator of the relative potency ρ of the test preparation.

b) Suggest a suitable block design different from RBD and BIBD useful for a Parallel Line assay. Under the suggested design, derive the sums of squares due to the "Parallelism contrast."

(3x3+5)+ (4+5)=23

3.What is the linearizing transformation of dose and response in Slope-ratio assay? Suggest a suitable block design d_0 superior to BIBD with respect to the estimation of the regression coefficients associated with the relative potency ρ in a Slope ratio assay. Derive the form of "Blank contrast" required for validity test and obtain its estimate under the design d_0 you have suggested.

2+4+7+7=20

INDIAN STATISTICAL INSTITUTE
Second Semester Examination: 2012-13
M. Stat. II Year
Advanced Sample Survey

Date: 6 May 2013 **Maximum Marks: 50** **Duration: 3 Hours**

Answer any four questions, each carrying 10 marks.

Records of Assignments to be submitted on the day of the examination carry 10 marks.

1 Describe how you may draw a sample circular systematically with probabilities proportional to given positive measures of sizes of the units. Explain how to use such a sample to unbiasedly estimate a finite population total. Discuss how a modification of this sampling scheme may serve a useful purpose.

2 Explain why and how a generalized regression predictor of a population total may be useful in the context of survey sampling. How do you measure its accuracy?

3 Discuss one device to generate randomized responses and its use each for unbiasedly estimating the (i) population proportion bearing a stigmatizing trait in a dichotomized finite population and also the (ii) population total amount of money earned through dubious ways by people in a given community.

Indicate how in each case you need to sample to get an unbiased estimate along with an unbiased estimate of the variance of the unbiased estimate working out detailed formulae in the context.

4 Explain with illustrations uses of replicated sampling and derive Murthy's almost unbiased estimator for the ratio of totals of two variables.

5 Given a sample by a design with positive inclusion probabilities varying with the population units, how will you estimate the correlation coefficient between two variables defined on a finite population? How will you make sure that your estimated correlation coefficient will never assume value outside the interval (-1,+1)?

INDIAN STATISTICAL INSTITUTE
Semestral Examination – Semester II : 2012-2013
M.Stat. II Year (MSP)
Advanced Probability II

Date : 06.05.13 **Maximum Score : 60** Time : $3\frac{1}{2}$ Hours

Note : **Qn.6** carries **20** marks and is **compulsory**. **Qns.1-5** carry a total of **60** marks. Answer as much as you can. The **maximum** you can score in **Qns.1-5** is **40**. Whenever you use any result proved in class, state it clearly.

In each problem, there is a probability space $(\Omega, \mathcal{F}, P)$, equipped with a filtration $(\mathcal{F}_t)$. Assume, whenever necessary, that $(\Omega, \mathcal{F}, P), (\mathcal{F}_t)$ satisfy the "usual hypotheses".

1. Let $\{B_t\}$ be a standard Brownian Motion. We define $\{Z_n(t), t \geq 0\}$ recursively as follows: $Z_0(t) \equiv 1$; and for each $n \geq 1$, $Z_n(t) = \int_0^t Z_{n-1}(s)dB(s), t \geq 0$.
(a) Show, by induction, that $E[Z_n^2(t)] \leq t^n/(n!)$ for all t and all $n \geq 0$. Conclude that, for each $n \geq 0$, $Z_n \in \mathcal{L}_{2,T}$ for all $T > 0$ and hence each $\{Z_n(t)\}$ is a continuous square-integrable martingale.
(b) Deduce that, with probability 1, the series $\sum_{n\geq 0} Z_n(t)$ converges uniformly on every compact interval and hence $X(t) = \sum_{n\geq 0} Z_n(t)$ defines a continuous adapted process.
(c) Show that $X \in \mathcal{L}_{2,T}$ for every $T > 0$ and satisfies the Stochastic Differential Equation $dX(t) = X(t)dB(t)$, $X(0) = 1$.
(d) Conclude that $\sum_{n\geq 0} Z_n(t) = \exp[B(t) - (t/2)]$, $t \geq 0$. (5+5+5+5)=[20]

2. Let $\{M_t\}$ be a continuous square-integrable martingale and let $h \in \mathcal{L}^{2,\langle M \rangle}$.
Show that, if $\{\widetilde{M}_t\}$ is a continuous square-integrable martingale with the property that, for any continuous square-integrable martingale $\{N_t\}$, one has $\langle \widetilde{M}, N\rangle_t = \int_0^t h_s\, d\langle M, N\rangle_s$ for all t, then $\widetilde{M} = \int h\, dM$. [10]

3. (a) Let $\{M_t\}$ be a bounded continuous martingale with $\{\langle M\rangle_t\}$ also bounded. Fix $T > 0$ and define, for each $n \geq 1$,
$t_0^{(n)} = 0$, and, $t_{i+1}^{(n)} = T \wedge \inf\{t > t_i^{(n)} : |t - t_i^{(n)}| \vee |M_t - M_{t_i^{(n)}}| \vee |\langle M\rangle_t - \langle M\rangle_{t_i^{(n)}}| > 2^{-n}\}$, $i \geq 0$.
(a) Denoting $\Delta_i^n = (M_{t_{i+1}^{(n)}} - M_{t_i^{(n)}})^2 - (\langle M\rangle_{t_{i+1}^{(n)}} - \langle M\rangle_{t_{i+1}^{(n)}})$, show that $\lim_{n\to\infty} \sum_{i\geq 0} E[(\Delta_i^n)^2] = 0$.
(b) Conclude that $\sum_{i\geq 0}(M_{t_{i+1}^{(n)}} - M_{t_i^{(n)}})^2 \longrightarrow \langle M\rangle_T$ in probability, as $n \to \infty$. (6+4)=[10]

4. Let $\{X_t\}$ be a continuous adapted process such that (i) $\{M_t = X_t - \int_0^t \beta(X_s)ds\}$ is a square-integrable martingale, and (ii) $\langle M\rangle_t = \int_0^t \alpha(X_s)ds$, $t \geq 0$, where $\alpha(\cdot)$, $\beta(\cdot)$ are bounded continuous functions on $\mathbb{R}$ with $\alpha(\cdot) > 0$. Define $\varphi(x) = \int_0^x \exp\{-\int_0^y [2\beta(z)/\alpha(z)]\, dz\}\, dy$, for $x \in \mathbb{R}$. Show that $\{\varphi(X_t)\}$ is a square-integrable martingale. [10]

5. Let $\{M_t\}$ be a continuous square-integrable martingale.
(a) Let τ be a finite stopping time and let $\sigma = \inf\{t > \tau : \langle M\rangle_t > \langle M\rangle_\tau\}$. Show that, with probability 1, $M_t(\omega) = M_\tau(\omega)$ for $\tau(\omega) \leq t \leq \sigma(\omega)$.
(b) Deduce that, there is a P-null set N, such that, for all $\omega \notin N$, any interval of constancy of $\langle M\rangle_t(\omega)$ is also an interval of constancy of $M_t(\omega)$. (6+4)=[10]

6. **Self-study and presentation project.** [20]

INDIAN STATISTICAL INSTITUTE
Semester Examination: 2012-2013 (Second Semester)

M. Stat II Year

Survival Analysis

Date: 26 April, 2013 Maximum Marks: 80 Duration: 3 hours.

Note: Answer all the questions. This paper contains 83 marks.

1. (a) Show that if the survival time T follows a Weibull distribution, then the proportional hazards (PH) model and the accelerated failure time (AFT) model coincide.

   (b) Discuss how you will test for proportional hazard assumption using a time-dependent covariate.

   (c) Consider the linear regression model for the survival random variable T

   $$E(T_i) = \alpha + \beta z_i, \text{ for } i = 1, \ldots, n$$

   Suppose you have random censored observations (X_i, δ_i, z_i), $i = 1, \ldots, n$, where $X_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \leq C_i)$ under independent censoring. Define $X_i^* = X_i\delta_i + E(T_i|T_i > X_i)(1 - \delta_i)$. Show that $E(X_i^*) = \alpha + \beta z_i$.

   [5 + 3 + 7 = 15]

2. Suppose that the survival prospects for two groups of patients are to be compared using Cox proportional hazard model based on random right censored data. Write down the PH model by defining a suitable covariate. Derive a score test based on an appropriate partial likelihood for testing the equality of survival functions.

   [3 + 12 = 15]

3. Suppose the continuous lifetimes T given z is from a PH model $\lambda(t; z) = \lambda_0(t)\exp(\beta' z)$ are grouped into $k + 1$ intervals $I_i = [a_{i-1}, a_i)$, $i = 1, \ldots, k + 1$, with $a_0 = 0$ and $a_{k+1} = \infty$. Define $p_i = P[T \geq a_i | T \geq a_{i-1}]$ and $p_i(z) = P[T \geq a_i | T \geq a_{i-1}; z]$.

   (a) Find the expression for $p_i(z)$ in terms of p_i and β.

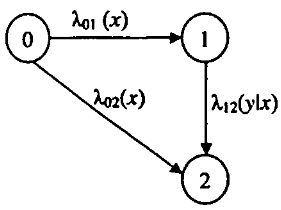
   (b) Write down the likelihood function to obtain maximum likelihood estimate of p_i's and β. Assume that all censoring in I_i takes place just prior to time a_i, after all deaths in I_i have occurred.

   (c) Show that when $\beta = 0$, the mle $\hat{p}_i$ of p_i is the life table estimate, when censoring takes place at the end of the intervals.

   [3 + 3 + 9 = 15]

4. Consider the proportional hazard model $\lambda(t; z) = \lambda_0(t)\exp(\beta' z)$, where $\lambda_0(t)$ is unspecified and arbitrary. Derive Breslow's estimator for $\Lambda_0(t)$, the cumulative base line hazard function. Show that Breslow's method also gives partial likelihood for estimating β. [7 + 3 = 10]

5. Consider the illness-death model shown in the following Figure, where 0, 1 and 2 represents healthy, illness and death states, respectively. Let T_0 be the transition time from state 0 to state 1 or 2, whichever is earlier and T_{12}, the sojourn time spent in state 1 before death. Assume that $\lambda_{01}(x) = \lambda_{01}$, $\lambda_{01}(x) = \lambda_{02}$ and $\lambda_{12}(y|x) = \lambda_{12} \times y$. Let C be the censoring random variable independent of both T_0 and T_{12}. Suppose there are n starting individuals at time 0 at state 0. Describe the data with appropriate notation. Find the maximum likelihood estimate of the parameters.



[4 + 8 = 12]

6. You have competing risks data of the form (T_i, J_i), $i = 1, 2, \ldots, n$, where T_i is the observed duration for the ith case, and J_i takes the value 0 when T_i is a censored observation, and the value k when T_i is the time of failure from cause k. There are K possible causes of failure (i.e., $1 \le k \le K$), thought to be governed by K cause-specific hazard rates.

   (a) Give an expression for the Nelson-Aalen estimator of the integral (from time 0 to t) of the kth cause-specific hazard rate. The expression should be in terms of the given data.

   (b) Give an expression of the estimated asymptotic variance of this estimator (based on the counting process theory), also in terms of the given data.

   (c) Suggest a 95% confidence band for the integrated cause-specific hazard for the kth risk, and specify whether its width is fixed or changes with t.

   (d) Can the estimator of the integrated cause-specific hazard be used to obtain an estimator of the cause-specific hazard itself? Explain.

   (e) It is suspected that the cause-specific hazard rate for the Kth risk is bigger than that for the first risk. Describe an asymptotic test for the equality of the two rate functions.

[3+3+3+3+4=16]

INDIAN STATISTICAL INSTITUTE

Semestral Examination, Second Semester: 2012-13

M.Stat. II Year (AS)

Actuarial Models

Date: May 6 , 2013 Maximum marks: 100 Duration: 3 hours

Answer all questions. Standard actuarial notations are followed.

1. The placement office of a college arranges exactly one campus interview per week for the final year students of a college, with no restriction on participation. For a particular placement season consisting of 40 weeks, one has data on the entry level salary for each job offer, name of the student and time (week number) of offer.

   (a) Let X_t denote the maximum salary offer received by a student till week t of the placement season. If X_t is modelled as a stochastic process, describe whether the time set and the state space are continuous or discrete. [1]

   (b) Does X_t have the Markov property? Explain. [2]

   (c) Can the above processes for different students be regarded as independent? Explain. [2]

   (d) Let, for $n > 1$, Y_n be the indicator of whether the nth job offer received by a student involves a higher starting salary than the $(n-1)$th offer received by him/her. Can Y_n be regarded as a Markov chain? Explain. [2]

   (e) Describe a test for the Markov property of Y_n, and indicate why it may be difficult to implement for the available data. [3]

   [Total 10]

2. (a) Express the values of ${}_tp_x$ in terms of t and p_x according to the uniform distribution of death (UDD), the constant force of mortality (CFM) and the Balducci interpolation rules, where x is an integer and $0 < t < 1$. [2]

   (b) Indicate, for each interpolation, whether the plot of ${}_tp_x$ vs. t is convex, concave or linear. [3]

   (c) For any given value of t and p_x, arrange the three interpolated values of ${}_tp_x$ in the increasing order. [3]

   [Total 8]

3. The number of claims settled by an insurer is modelled by a non-homogeneous Poisson process. The process starts from time 0 with no claims. At any point of time t, the rate of occurrence of a new claim is λt. You are given the data on exact times of arrival of all the claims till time t_{max}.

   (a) Write down the likelihood of λ on the basis of the available data. [2]

   (b) Derive an expression for the maximum likelihood estimator of λ. [1]

   (c) Give an expression for the asymptotic variance of λ. [2]

   (d) You suspect that the rate of occurrence of a new claim at time t may be λt^{α}, where α may not be equal to 1. Indicate how you would test the hypothesis $\alpha = 1$, giving as much detail as you can. [5]

   [Total 10]

4. The crude mortality rates estimated for a population of tribal females were graduated by using the parametric formula $\mathring{q}_x = a + bc^x$. This was done by adjusting the parameters a, b and c through trial and error to bring the graduated rates in the vicinity of the crude rates. The initial exposed to risk, crude and graduated mortality rates and the standardized deviation for different years of age are summarized in the following table.

Age (years)	Initial exposed to risk (person years)	mortality rate crude	mortality rate graduated	standardized deviation
x	E_x	$\hat{q}_x$	$\mathring{q}_x$	z_x
45	13244	0.00290	0.00348	−1.133
46	14412	0.00333	0.00358	−0.503
47	13985	0.00300	0.00368	−1.328
48	11590	0.00300	0.00379	−1.384
49	10148	0.00319	0.00391	−1.162
50	12649	0.00427	0.00402	0.444
51	8892	0.00472	0.00415	0.836
52	14450	0.00399	0.00428	−0.534
53	16846	0.00406	0.00441	−0.686
54	13775	0.00375	0.00455	−1.395
55	14976	0.00409	0.00469	−1.075
56	12680	0.00407	0.00485	−1.264
57	11024	0.00512	0.00500	0.179
58	16832	0.00456	0.00517	−1.104
59	13897	0.00466	0.00534	−1.100

(a) Determine, by using the chi-squared test, whether the graduated rates adhere to the data. [3]

(b) Plot the deviations and identify a possible inadequacy in graduation. [4]

(c) Explain why this aspect is not detectable through the chi-square test. [1]

(d) Carry out a formal test for this apparent deficiency. [2]

(e) Suggest how the graduation may be adjusted to correct this deficiency. [2]

[Total 12]

5. Consider a continuous time sickness-death model with constant rates of transition from 'Well' state to 'Sick' state (σ), from 'Sick' state to 'Well' state (ρ), from 'Well' state to 'Dead' state (μ), and from 'Sick' state to 'Dead' state (ν).

(a) Indicate how a sample path of this Markov jump process can be recorded. [1]

(b) Assume that $\sigma = 10$, $\rho = 20$, $\mu = 2$ and $\nu = 5$, and that the process starts from the 'Well' state at time 0. Using the pseudo-random samples from the uniform distribution over the unit interval: 0.3158, 0.9279, 0.4949, 0.7776, 0.5895, 0.7316, 0.7946, 0.2089, 0.0284, 0.5517, 0.3813, 0.4757, 0.8109, 0.1396, 0.0036, 0.3483, 0.5870, 0.8369, 0.0613, 0.9747, generate as many sample paths from the process as possible. [6]

(c) Using the data you have generated, evaluate the MLE of σ and its estimated standard error. (Use standard formulae; no need to derive any expression.) [3]

[Total 10]

6. A sample of 50 fresh graduates, who started working for various organizations on 1st July 2012, were tracked for a period of 200 days to determine how long they take to hop to another job. A few of the subjects were lost to follow up after being transferred elsewhere or fired by their employer. The following is a summary of the data collected.

Days since joining	*Number continuing to work*	*Event*
31	49	One subject fired
46	48	One subject transferred
50	47	One hopped to another job
61	46	One subject transferred
73	45	One hopped to another job
91	44	One hopped to another job
108	43	One subject fired
145	42	One subject fired
186	41	One hopped to another job

(a) Calculate the Nelson-Aalen estimator of the cumulative hazard. [5]

(b) If an exponential distribution is to be fitted to the data, calculate the MLE of the mean of that distribution. [2]

(c) Plot the estimate of part (a) along with the cumulative hazard of the exponential distribution fitted in part (b). [5]

(d) Specify a 95% asymptotic confidence interval around the Nelson-Aalen estimator at 180 days. [2]

(e) It had been found in an earlier study that there is a 30% chance of a new recruit hopping to another job within 180 days. Use the computations of part (d) to test the hypothesis that this probability has not changed. [1]

[Total 15]

7. An association of pedestrians has set up a fund for victims of traffic accidents. Every member contributes a fixed amount c to the fund every month. If a member has a traffic accident in any month (which happens with probability 0.05), s/he receives a fixed 'treatment subsidy' amount b. This arrangement is for at most one accident per member per month. If treatment of the victim continues into the following month (which happens with probability .1), the same benefit continues. In case the victim needs more than two months of treatment, the flow of benefit is interrupted every third month. The event of death is ignored in this simple model.

(a) Construct a discrete time Markov chain to model this scheme. [2]

(b) Draw the transition diagram. [2]

(c) Write down the transition matrix. [2]

(d) Explain whether this Markov chain is irreducible and/or aperiodic. [2]

(e) Calculate the stationary probability distribution of the chain. [4]

(f) Calculate the average fund outgo per month per member in the steady state. [2]

[Total 14]

8. Consider the regression model for hazard rate

$$h(t; z_i) = h_0(t)e^{\beta z_i},$$

where z_i is the value of a single covariate for the ith individual.

(a) Identify the regression model specifically when $h_0(t)$ is (i) specified as c^t with an unspecified parameter c, (ii) completely unspecified. [2]

(b) Given censored lifetime data with covariate values data (t_i, δ_i, z_i), $i = 1, 2, \ldots, n$, describe the likelihood you would use to infer β when $h_0(t)$ is (i) specified as c^t with an unspecified parameter c, (ii) completely unspecified. [2]

[Total 4]

9. In a mortality investigation, the following data have been recorded for ten independent lives ($i = 1, \ldots, 10$) observed between exact ages 70 and 71.

a_i = the time in months after exact age 70 when the ith life came under observation;

b_i = the time in months after exact age 70 when ith life was censored;

d_i = 1 if the ith life died before $70 + b_i$,
= 0 if the ith life survived to $70 + b_i$.

Life i	a_i	b_i	d_i
1	0	6	0
2	1	12	0
3	1	3	1
4	2	12	0
5	3	9	1
6	4	12	0
7	5	11	1
8	7	12	0
9	8	10	1
10	9	12	0

(a) Using the Binomial Model of Mortality write down the likelihood of these observations. [3]

(b) Using the Balducci assumption over the interval (70, 71), express this likelihood as a function of q_{70}. [4]

(c) Estimate q_{70} from the given data, by using an explicit estimator. [2]

[Total 9]

10. The deaths recorded during the period of a mortality investigation, conducted from 1-1-2007 to 31-12-2009, were classified by age at the date of death, defined as $x = y_d - y_b$, y_d and y_b being the calendar years of death and birth, respectively. The number of alive persons (in the relevant group) as of 1st January of 2007, 2008, 2009 and 2010 are available and classified by the age last birthday on the day of each census.

(a) What is the rate year implied by the classification of deaths? [1]

(b) What is the range of ages of the lives at the beginning of the rate year? [1]

(c) If one wishes to estimate from the above data the mortality rate μ_{x+f} for different integer ages x and a fixed fraction f, which is the appropriate value of f? [2]

(d) Give an expression for the central exposed to risk in terms of the population counts described above, in accordance with the principle of correspondence, which can be used to estimate μ_{x+f}. [4]

[Total 8]

Indian Statistical Institute
Second Semestral Examination
M. Stst. II (MSP)
Statistical Inference II

Date: 08. 05. 2013 Maximum marks: 100 Duration: 3 hrs.

Answer all Questions. This is an open notes examination as per instructions.
Paper carries 110 points.

1. A statistic T is said to be *partially sufficient* for θ in presence of a nuisance parameter η if the parameter space is cartesian product of possible $\theta-$ and $\eta-$ values and if the following two conditions hold: (a) the conditional distribution of the data X given $T = t$ depends only on η; (b) the marginal distribution of T depends only on θ. Show that under the above conditions there is a UMP test based on T for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$ (η unknown).
[20]

2. Let $X_1, X_2, \ldots, X_n$ be a sample from the uniform distribution over the integers $1, 2, \ldots, \theta$ and let a be a positive integer. Show that
(i) The sufficient statistic $X_{(n)}$ is complete if the parameter spacc is $\Theta = \{\theta : \theta \leq a\}$.
(ii) Show that $X_{(n)}$ is not complete is the parameter space is $\Theta = \{\theta : \theta \geq a\}$ and $a \geq 2$. Find a complete sufficient statistic in this case.
(iii) Is there a UMP unbiased test for $H_0 : \theta \leq a$ versus $H_1 : \theta > a$ in this case?
[5+ 10 + 5 = 20]

3 (i) Let x be an $n \times s$ matrix ($s \leq n$) of rank s, and G be the group of transformation $gx = xB$ where B is any $s \times s$ non-singular transformation matrix. Find a suitable maximal invariant function for this group action on x.
(ii) Let $X = (X_1, X_2, \ldots, X_n)$, and suppose that the density of X is $f_i(x_1 - \theta, x_2 - \theta, \ldots, x_n - \theta)$ under H_i ($i = 0, 1$), where θ ranges over $\mathbb{R}$. By considering the group $gx \mapsto (x_1 + c, x_2 + c, \ldots, x_n + c)$ find the UMP invariant test for $H_0 : f = f_0$ versus $H_1 : f = f_1$ (θ unknown).
[10 + 10 =20]

P. T. O.

4. Define the notion of maximin testing procedure (the null, the indifference zone, and, the alternative). Let $(\mathcal{X}, \mathcal{A})$ be an Euclidean measurable space, and let the distributions $P_\theta, \theta \in \Theta$ be dominated by appropriate Lebesgue measure. Show that for any mutually exclusive subsets Θ_0 and Θ_1 there is a level- α test statistic ϕ (that is, $E_\theta \phi \leq \alpha$ for every $\theta \in \Theta_0$) which maximizes $\inf_{\theta \in \Theta_1} E_\theta \phi$.

[20]

5. Consider testing a simple versus simple hypothesis between two finitely supported pmf's p and q. Formulate the NP optimization as a linear programming problem. Derive its Lagrangian and the dual. Hence describe the saddle points of the Lagrangian and their interpretation in the derivation of the most powerful tests. How would you define P-value in terms of the dual of the problem?

[30]

INDIAN STATISTICAL INSTITUTE

M.Stat 2nd year

2012 - 13 (Semestral Examination)

Subject: Theory of Games and Statistical Decisions

Date of Exam.: 10/05/2013 Time: 3 hrs Full Marks: 60

Note: ***The paper contains questions of 66 marks. Maximum marks you can score is 60.***

1. **a)** Show that a function $f: A \times B \to R$ possesses a saddle point if and only if $\max_{a\in A} \inf_{b\in B} f(a,b)$ and $\min_{b\in B} \sup_{a\in A} f(a,b)$ both exist and are equal.

   b) In continuation to 1.a), show that $\exists$ A_1 and B_1, with $A \supset A_1$ and $B \supset B_1$ such that the set of all saddle points is given by $A_1 \times B_1$.

   [12 + 4]

2. **a)** Let there be n players each with set of strategies {1,2}. Let H_i be the pay-off function for the i^{th} player. Considering mixed extension describe a method to find algebraically the set of all admissible situations for player 1.

   b) Let players 1, 2, 3 be three industrial enterprises who use the water of a lake and each has the following two strategies.

   Strategy 1: To put back used water to the lake after purification at the cost of 1 unit.

   Strategy 2: To put back used water to this lake without purification.

   If at most one player takes strategy 2, the water of the lake does not get contaminated.
   If at least two players take strategy 2, the water of the lake gets contaminated causing a loss of 3 units to each.

   Describe the problem in game theoretic framework in mixed extension. Find sets of admissible situations for the players and hence derive the set of all equilibrium situations.

   [8 + 12]

3. **a)** Consider the 2x4 matrix game with matrix A having first row (4 3 3 3) and the 2^{nd} row (1 3 2 4). Find the value of the game in mixed extension and sets of all optimal strategies for each player.

P.T.O.

b) Let in a two-player antagonistic game player 1 has a set of 3 strategies $\{\theta_1, \theta_2, \theta_3\}$. Let pay-off of player 1 be $r_1(d)$, $r_2(d)$ and $r_3(d)$ with strategies θ_1, θ_2, θ_3 respectively when the strategy of the 2nd player is d.

Let $A = \{(r_1(d), r_2(d), r_3(d))$: $d \in$ set of all possible strategies of player 2$\}$
When $A = \{(x_1, x_2, x_3) : (x_1 - 3)^2 + (x_2 - 4)^2 + (x_3 - 5)^2\} \leq 8\}$
& player 1 uses mixed strategies on $\{\theta_1, \theta_2, \theta_3\}$, find the value of the game and set of all optimal strategies of player 1.

c) Let there be 3 players I, II III each with set of strategies $\{1, 2, 3, 4, 5\}$. Let each player has pay-off function H given by

$$H(i_1, i_2, i_3) = 1 \text{ if } i_1 = i_2 = i_3$$
$$= 0 \text{ otherwise.}$$

Describe the set of all equilibrium situations of the game in mixed extension.
Hence show that there can exist two equilibrium situations such that pay-off of the same player differs at the two situations.

[8+10+(10+2)]

XXXXXXXXXXXXXXXXXXXX

INDIAN STATISTICAL INSTITUTE
Second Semester Back Paper Examination: 2012-13
M. Stat. II Year
Advanced Sample Survey

Date: 28.06.13 **Maximum Marks: 100** **Duration: 3 Hours**

Answer any four questions each carrying an equal score

1 Derive Godambe & Thompson's bound on the model-expected design-variance of a design unbiased estimator of a finite population total.

2 Show that the Brewer's predictor and the ratio predictor are both special cases of a generalized regression predictor.

3 Discuss the problem of small area estimation and illustrate the use of "Borrowing Strength" in tackling it.

4 Explain and illustrate the similarities and dissimilarities in "Randomized Response Approach" and "Multi-Stage Sampling".

5 Show how to find an Empirical Bayes estimator for a Domain Total in Survey Sampling.

*****

INDIAN STATISTICAL INSTITUTE
Back Paper Examination: 2012-2013 (Second Semester)

M. Stat II Year

Survival Analysis

Date: 02.07.13 Maximum Marks: 100 Duration: 3 hours.

Note: Answer all the questions

1. (a) Let $X \sim$ Weibull (α, λ) and $T = [X]$, the largest integer less than or equal to X. Find the p.m.f. and discrete hazard of T and check whether it is IFR or DFR distribution.

   (b) Consider type-II censored data from exponential (λ) life distribution. Prove that the maximum likelihood estimate of expected lifetime is unbiased. Give a method of generating type-II censored observations from exponential (λ) life distribution. There is credit for using the fewer random numbers.

$[(2 + 3 + 3) + (7 + 5) = 20]$

2. A model used in the construction of life tables is a piecewise, constant hazard rate model. the time axis is divided into k intervals, $[\tau_{i-1}, \tau_i)$, $i = 1, \ldots, k$, with $0 = \tau_0 < \tau_1 < \ldots < \tau_k = \infty$. The hazard rate is given by

$$\lambda(t) = \begin{cases} \lambda_1, & \text{if } 0 \leq t < \tau_1 \\ \lambda_2, & \text{if } \tau_1 \leq t < \tau_2 \\ \vdots & \\ \lambda_{k-1}, & \text{if } \tau_{k-2} \leq t < \tau_{k-1} \\ \lambda_k, & \text{if } t > \tau_{k-1}. \end{cases}$$

   Find the mean residual lifetime.

[12]

3. Suppose a continuous lifetime random variable T given z satisfying proportional hazard model

$$\lambda(t; z) = \lambda_0(t) \exp(\beta' z)$$

   is grouped into intervals $I_1, I_2, \ldots, I_{k+1}$, where $I_i = [a_{i-1}, a_i]$ and $0 = a_0 < a_1 < \ldots < a_k < a_{k+1} = \infty$. Define $T_d = i$, if $a_{i-1} \leq T \leq a_i$. Find the discrete hazard of T_d for given z.

[6]

4. Derive the Kaplan-Meier estimate of a survival function based on random right censored data using the nonparametric maximum likelihood argument. Then derive the corresponding variance estimate.

[8 + 5 = 13]

5. Suppose the survival prospect of two groups of patients are to be compared using a covariate z. Group 0, indicated by $z = 0$, has survival function $S_0(t) = \exp(-\lambda t)$. Group 1, indicated by $z = 1$, has survival function $S_1(t) = S_0(t)^{\psi}$.

   (a) Show that the two groups follow the proportional hazards model.

   (b) Consider random right censored data (x_i, δ_i, z_i), $i = 1, \ldots, n$. Find the maximum likelihood estimate of λ and ψ.

[4 + 8 = 12]

6. Consider the following right censored data with covariate (x_i, δ_i, z_i), $i = 1, \ldots, n$: x = 54, 127, 297, 389, 1536; δ = 1, 0, 1, 0, 0 and z = 2.09, 0.36, 0.60, 1.44, 0.91. Consider Cox proportional hazard model to study the effect of covariate.

   (a) Construct Cox partial likelihood for this data set.

   (b) Derive a score test for testing the hypothesis $H_0 : \beta = 1$.

[5 + 8 = 13]

7. Consider a competing risks problem with survival time T and mode of failure J. Suppose there are m causes of failure and jth cause-specific hazard function $\lambda_j(t)$ is of the form

$$\lambda_j(t) = w_j \lambda(t), \quad \text{for } j = 1, \ldots, m,$$

where $0 < w_j < 1$ and $\sum w_j = 1$. Find the subdistribution function $F_j(t)$. Show that T and J are independently distributed. What is w_j here?

[4 + 4 + 2 = 10]

8. Consider the bivariate lifetime distribution with survivor function

$$S(t_1, t_2) = \exp[-(\lambda_1 t_1 + \lambda_2 t_2 + \lambda_1 \lambda_2 \theta t_1 t_2)],$$

where $\lambda_1 > 0$, $\lambda_2 > 0$ and $0 \leq \theta \leq 1$. Suppose that data available in the form (T, J), where $T = \min(T_1, T_2)$ and $J = j$ such that $T = T_j$ is observed.

   (a) Obtain the cause-specific hazard functions and also the marginal hazard functions of T_1 and T_2.

   (b) Discuss why it would not be possible to assess the adequacy of the joint model for (T_1, T_2) on the basis of data on (T, J)

[8 + 6 = 14]

INDIAN STATISTICAL INSTITUTE

SECOND SEMESTRAL (BACKPAPER) EXAMINATION : 2012-2013

SECOND SEMESTER, M.STAT - II YEAR

DATE - 01·07·13

FULL MARKS: 100 APPLIED MULTIVARIATE ANALYSIS TIME: 3 HOURS

1. Prove or disprove the following statements.

(a) Consider a multivariate distribution, which is symmetric about the origin. If all principal components of this distribution have the same variance, the distribution is spherically symmetric. [5]

(b) Consider two random vectors $\mathbf{X} = (X_1, X_2, \ldots, X_p)'$ and $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_p)'$ each with the variance-covariance matrix $\mathbf{I}_p$, the $p \times p$ identity matrix. If $\rho(X_i, Y_j) = r$ for all $i = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, q$, the value of r cannot exceed $1/p$. [5]

(c) Consider two random vectors $\mathbf{X} = (X_1, X_2, X_3)'$ and $\mathbf{Y} = (Y_1, Y_2, Y_3, Y_4)'$. If $(\alpha_1'\mathbf{X}, \beta_1'\mathbf{Y})$ and $(\alpha_2'\mathbf{X}, \beta_2'\mathbf{Y})$ are two pairs of canonical variables, then $\rho(\alpha_1'\mathbf{X}, \beta_2'\mathbf{Y}) = 0$. [5]

(d) Consider two random vectors $\mathbf{X} = (X_1, X_2, X_3)'$ and $\mathbf{Y} = (Y_1, Y_2, Y_3)'$. If $\mathbf{X}$ and $\mathbf{Y}$ have the same marginal distribution and the first canonical correlation coefficient between $\mathbf{X}$ and $\mathbf{Y}$ is 0, then $\mathbf{X}$ and $\mathbf{Y}$ are independent. [5]

(e) If the dispersion matrix of $\mathbf{Z} = (Z_1, Z_2, Z_3)'$ is given by

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & 0.65 & 0.43 \\ 0.65 & 1 & 0.78 \\ 0.43 & 0.78 & 1 \end{bmatrix},$$

$\mathbf{Z} = (Z_1, Z_2, Z_3)'$ cannot be generated using a valid $k = 1$ factor model. [5]

(f) Consider a histogram with a fixed common bin width h. For any given $\mathbf{x}$, the variance of the histogram density estimator of $f(x)$ converges to 0 as the sample size tends to infinity. [5]

(g) Consider a two-class classification problem in dimension 2, where both of the competing classes have continuous probability density functions. If these two classes have the same mean vector and the same prior probability, the total misclassification probability of a classifier cannot be zero. [5]

(h) Consider a classification problem between the two discrete distributions given below.

Distribution-1				
x	0	1	2	3
$f(x)$	$\frac{\alpha}{2}$	$\frac{\alpha}{2}$	$\frac{1-\alpha}{2}$	$\frac{1-\alpha}{2}$

Distribution-2				
x	0	1	2	3
$f(x)$	$\frac{1-\alpha}{2}$	$\frac{1-\alpha}{2}$	$\frac{\alpha}{2}$	$\frac{\alpha}{2}$

For this classification problem, the total misclassification probability of the Bayes classifier is $\min\{\alpha, 1-\alpha\}$. [5]

(i) Consider a classification problem between a standard normal and a standard Cauchy distribution. Also consider a classifier that classifies an observation x to the normal population if $|x| \nless C$, where C is a positive constant. Irrespective of the choice of C, this classifier is admissible. [5]

(j) If all measurement variables are continuous, the k medoids clustering algorithm and the k means clustering algorithm are identical. [5]

2. Assume that the distribution of a random vector $\mathbf{X} = (X_1, X_2, X_3, X_4)'$ is an equal mixture of two normal distributions $N(-\mathbf{1}, 0.5\ \mathbf{I})$ and $N(\mathbf{1}, 1.5\ \mathbf{I})$, where $\mathbf{I}$ is the 4×4 identity matrix and $\mathbf{1}$ is the four dimensional vector with all elements unity. Define $\mathbf{Y}_1 = (X_1, X_2)'$ and $\mathbf{Y}_2 = (X_3, X_4)'$. Find the largest canonical correlation coefficient between $\mathbf{Y}_1$ and $\mathbf{Y}_2$. [10]

3. (a) Suppose that you have 200 independent observations on a 10 dimensional random vector $\mathbf{X}$, which is assumed to be normally distributed. Describe how you will test whether a $k = 4$ factor model is appropriate for the data. [6]

   (b) What is the varimax criterion for factor rotation ? How can that be helpful in factor analysis ? [2+2]

4. Consider a two-class classification problem, where Class-1 is an equal mixture of two bivariate normal distributions $N((1,1)', \mathbf{I})$ and $N((-1,-1)', \mathbf{I})$, and Class-2 is an equal mixture of $N((1,-1)', \mathbf{I})$ and $N((-1,1)', \mathbf{I})$. Assuming that the prior probabilities of these two classes are equal, find the Bayes classifier for this classification problem and show that the total misclassification probability of this Bayes classifier is $2\Phi(1)\Phi(-1)$, where $\Phi(\cdot)$ denotes the distribution function of the univariate standard normal distribution. [4+4]

5. Suppose that you have two sets of independent observations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{10}$ and $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{10}$ from two bivariate normal distributions. Construct a suitable classifier based on these labeled observations. If another set of 100 unlabeled observations $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{100}$ from these two classes is available to you, describe how you will modify your classifier. [2+6]

6. The following matrix shows all pairwise distances for 6 observations. Use the average linkage method for clustering and construct the corresponding dendogram.

0	44	42	17	39	9
44	0	2	27	5	35
42	2	0	25	3	33
17	27	25	0	22	8
39	5	3	22	0	30
9	35	33	8	30	0

Use Dunn index or any other suitable cluster validation index to estimate the number of clusters present in the data. [8+6]

INDIAN STATISTICAL INSTITUTE

Final Examination : Semester II (2012-2013) - Back Paper

Course Name : BSDA (M. Stat. 2nd year)

Subject Name : Statistical Methods in Biomedical Research

Date : 10.07.31 , Maximum Marks : 100. Duration : 3 hrs.

[Use separate answerscripts for each group]

Group A

1. Describe Efron's biased coin design for allocation with two treatments, say A and B. Find the expectation of the proportion of allocation by A. [3+4]

2. Suppose three treatments A, B and C are being compared in a clinical trial. The first patient is treated randomly by choosing any treatment with same probability. For any subsequent patient i, if the response of the $(i-1)$th patient is a success, we treat the ith patient by the same treatment as the $(i-1)$th patient. If the response of the $(i-1)$th patient is a failure, we treat the ith patient by any of the remaining two treatments by tossing a fair coin. If the success probabilities of the three treatments are 0.7, 0.4 and 0.3 respectively, find the probabilities that the 10th patient is treated by A and the 10th patient results in a success. [12]

3. Suppose patients are entering in a system in a sequential way. If the response time of any patient is constant ($= c$), and interarrival times are exponential with expectation b, find the probability that the response of the 4-th patient will be obtained prior to the entrance of the 9-th patient. If, instead, interarrival times are normally distributed with mean μ and variance σ^2, find the required probability. [4+3]

4. Discuss the concept of type I error spending functions in group sequential analysis. Discuss how type I error spending functions can be constructed by accumulating boundary crossing probabilities. If there are only two groups, give one form of type I error spending function which will spend the total type I error in 1:3 way in the two groups (assuming equal time interval for each group). [3+7+4]

5. Coeliac disease is a condition that impairs the ability of the gut to absorb nutrients. A useful measure of nutritional status is the bicep's skinfold thickness, which has standard deviation 2.3 mm in this population. A new nutritional programme is proposed and is to be compared with the present programme.

(a) If two groups of equal size are compared at the 5% significance level, how large should each group be if there is to be 90% power to detect a change in mean skinfold thickness of 0.5 mm? How many would I need if the power were 80%? [2]

(b) Suppose I can recruit 300 patients, what difference can I detect with 80% power? [2]

(c) Suppose I decide that a change of 1 mm in mean skinfold thickness is of interest after all. How many patients do I need to achieve a power of 80%? [2]

(d) What would be the effect on this value if 2.3 mm underestimates σ by 20% and if it overestimates σ by 20%? [2]

(e) Assuming that 2.3 mm is a satisfactory estimate of σ, what sample sizes would we need to achieve 80% power to detect a mean difference of 1 mm if we have opted to allocate patients to the new and the control treatments in the ratio 2:1? [2]

Statistical Methods in Biomedical Research
M. Stat-II (2012-2013)
Back paper Examination
Part B

Note: Use separate Answer script for Part B.

1.State and prove the Fieller's Theorem. Use this theorem to obtain the fiducial limits of the relative potency of the test preparation in comparison to the standard preparation in a Direct assay.

2+4+6=12

2.What is the metameric transformation for dose and response in a Parallel line assay? Obtain the form of the relative potency of the test preparation. Suggest a 6 point symmetric Parallel line assay design in 18 blocks, 12 replications under which the "Preparation contrast" and the "Combined slope contrast" can be estimated with full efficiency. Under this design write down the ANOVA table , explicitly deriving the different components . How can you perform a validity test to check the underlying assumptions of a Parallel line assay?

2+3+5+12+3=25

3. Derive the form of " Intersection contrast " in a Slope ratio assay. Write down another useful Orthogonal contrast justifying its importance in a Slope ratio assay. Construct a 9 point symmetrical Slope ratio assay design in 12 blocks of size 7 each.

6+3+4=13

INDIAN STATISTICAL INSTITUTE

Backpaper Examination: 2012-13 (Second Semester)
Master of Statistics (M. Stat.) II Year
Stochastic Processes I

Instructor: Parthanil Roy

Total Points: 100 | **Date:** 10.07.13 | **Duration: 3 hours**

- **Please write your name and roll number on top of your answer booklet(s).**
- **Please justify all your steps. If you are using a result proved in class, please quote that result. Note that your answer should be consistent with the class-notes.**
- **This is an open note examination. You are allowed to use your own hand-written notes (such as class notes, exercise solutions, list of theorems, formulas etc.). Please note that no printed or photocopied materials are allowed. In particular, you are not allowed to use books, photocopied class notes etc. If you are caught using any, you will get a zero in the backpaper examination.**

1. Find whether the following statements are true or false. If a statement is true, give a detailed proof. If it is false, disprove (in details) with a counter-example.

   (a) (15 points) If X is a Polish space, $P_n \Rightarrow P$ in $\mathcal{P}(X)$ and $f : X \to \mathbb{R}$ is a bounded function such that it is continuous except for one point $x_0 \in X$, then $\int f dP_n \to \int f dP$.

   (b) (10 points) If X is a Polish space, $P_n \Rightarrow P$ in $\mathcal{P}(X)$ and F is a uniformly bounded and equicontinuous family of real valued functions defined on X, then $\int f dP_n \to \int f dP$ uniformly in $f \in F$.

   (c) (10 points) If $\{B_t\}_{t\geq 0}$ is a standard Brownian motion starting at 0, then $\inf\{t > 0 : B_t = 0\} = 0$ almost surely.

   (d) (15 points) If $(X_n)_{n\geq 1}$ is a sequence of i.i.d. Random variables (defined on the same probability space) taking values in a Polish space with a common law P and for each n and ω, $P_{n,\omega}$ is the empirical distribution of the first n variables corresponding to the sample point ω, then for almost all ω, $P_{n,\omega} \Rightarrow P$.

2. (15 points) Consider the probability space $(C[0,1], \mathcal{C}, W)$, where $\mathcal{C}$ is the Borel σ-field on $C[0,1]$ and W is the Wiener measure. Define a random variable ξ on this space as

$$\xi(\omega) = \omega(1), \quad \omega \in C[0,1].$$

   Find a regular conditional probability given the random variable ξ. **[P. T. O.]**

3. Let $\{B_t\}_{0\leq t\leq 1}$ be a Wiener process defined on a sample space $(\Omega, \mathcal{F}, P)$. For each sample point $\omega \in \Omega$, define a number

$$V(\omega) = \int_0^1 B_s(\omega)ds.$$

(a) (3 points) Show that V is well-defined and it is a random variable.

(b) (7 points) Calculate the characteristic function of V.

4. (25 points) Suppose $\{B_t\}_{t\geq 0}$ is a standard Brownian motion starting from 0. Compute the cumulative distribution function of $L = \sup\{t < 1 : B_t = 0\}$.

INDIAN STATISTICAL INSTITUTE

Back-paper Examination – Semester II : 2012-2013

M.Stat. II Year (MSP)

Advanced Probability II

ate: 12·07·13 | **Maximum Score : 45** | Time : 3 Hours

ote : This paper contains five questions each with **20** marks. **Maximum** you can score is **45**.
henever you use any result proved in class, state it clearly.
nless otherwise specified, $(\Omega, \mathcal{F}, P)$ is to be assumed to be a probability space equipped with a tration $\{\mathcal{F}_t\}$ and satisfying the "usual hypotheses".

1. Let $\{B_t\}$ be a standard Brownian Motion on a probability space $(\overline{\Omega}, \overline{\mathcal{F}}, \overline{P})$ and let $Y_t = \alpha t + \beta B_t$, where α, β are real numbers with $\beta > 0$. Denote Ω to be $C[0,\infty)$, that is, the set of all real continuous functions on $[0,\infty)$ and $\mathcal{F}$ the σ-field on Ω generated by the coordinate maps.
   (a) Show that for every real x, the map on $\overline{\Omega}$ into Ω defined by $\overline{\omega} \mapsto x + Y_{(\cdot)}(\overline{\omega})$ is measurable.
   (b) Denoting P^x to be the probability on $(\Omega, \mathcal{F})$ induced by the measurable map in (a), show that, with the usual definition of filtration $\{\mathcal{F}_t\}$ and shift operartors $\{\theta_t\}$ on Ω, the family $\{P^x, x \in \mathbb{R}\}$ is a Markov process and derive its generator. (6+14)=**[20]**

2. Recall that $\mathcal{L}_2^{loc}$ denotes the set of all real functions f on $[0,\infty) \times \Omega$ such that $f(t,\omega)$ is jointly measurable in (t,ω), $f(t,\cdot)$ is $\mathcal{F}_t$-measurable for each t, and $\int_0^t f^2(s,\omega)\,ds < \infty$, P-almost surely, for each t.
   (a) Show that for $f \in \mathcal{L}_2^{loc}$, the map $\omega \mapsto \int_0^t f^2(s,\omega)\,ds$ is $\mathcal{F}_t$-measurable for each t.
   (b) Let $M_t = \int_0^t f_s dB_s$, $N_t = \int_0^t g_s dB_s$, $t \geq 0$ for f, $g \in \mathcal{L}_2^{loc}$. Prove the following:
   (i) If τ is a finite stopping time such that $f(t,\omega) = g(t,\omega)$ for $t \leq \tau(\omega)$, then $M_{t\wedge\tau} = N_{t\wedge\tau}\ \forall\, t$.
   (ii) $\{M_t N_t - \int_0^t f_s g_s ds\}$ is a continuous local martingale. (6+(7+7))=**[20]**

3. (a) Define what are meant by (i) the predictable σ-field on $[0,\infty) \times \Omega$ and (ii) a predictable process on $(\Omega, \mathcal{F}, P)$.
   (b) Show that for any two $\{\mathcal{F}_t\}$-stopping times τ and η, the stochastic interval $[\tau, \eta)$, defined to be the set $\{(t,\omega) : \tau(\omega) \leq t < \eta(\omega)\}$, belongs to the predictable σ-field.
   (c) Let $\mathcal{L}$ be a linear space of real-valued bounded measurable functions on $[0,\infty) \times \Omega$. Suppose $\mathcal{L}$ contains all bounded $\{\mathcal{F}_t\}$-adapted processes with left-continuous trajectories. Suppose also that $\mathcal{L}$ satisfies the property: if $X_n \in \mathcal{L}$, $n \geq 1$, $X_n \uparrow X$ and X is bounded, then $X \in \mathcal{L}$. Show that $\mathcal{L}$ contains all bounded predictable processes. ((2+2)+6+10)=**[20]**

4. Let $\{A_t\}$ be a predictable process with non-decreasing trajectories and satisfying $A_0 \equiv 0$ and $E(A_t) < \infty\ \forall\, t$. Denote $\mathcal{L}_2(A)$ to be the class of all predictable processes $\{h_t\}$ such that $\int_\Omega \int_0^t h^2(s,\omega)\, A(ds,\omega) P(d\omega) < \infty$ for all $t \in [0,\infty)$.
   (a) Show that $\mathcal{L}_2(A)$ is a Hilbert space with the norm $\|h\|_{\mathcal{L}_2(A)} = \sum_{n\geq 1} 2^{-n}(d_n(h) \wedge 1)$, where $d_n(h) = \left[\int_\Omega \int_0^n h^2(s,\omega)\, A(ds,\omega) P(d\omega)\right]^{1/2}$.
   (b) Let $\mathcal{L}_0$ denote the class of all processes $\{h_t\}$ for which there is a sequence $0 = t_0 < t_1 < t_2 < \cdots \uparrow \infty$, such that $h_t = \xi_i$ for $t \in (t_i, t_{i+1}]$ where ξ_i, $i \geq 0$ are $\mathcal{F}_{t_i}$-measurable random variables that are uniformly bounded. Show that $\mathcal{L}_0 \subset \mathcal{L}_2(A)$. Show also that $\mathcal{L}_0$ is dense in $\mathcal{L}_2(A)$. (10+10)=**[20]**

5. Let $M \in \mathcal{M}_2$ and $h \in \mathcal{L}_{2,\langle M\rangle}$. Denote $N = \int h dM$.
   (a) Show that if $g \in \mathcal{L}_{2,\langle N\rangle}$, then $hg \in \mathcal{L}_{2,\langle M\rangle}$ and $\int g dN = \int hg dM$.
   (b) Assuming that M is continuous, show that the process $X_t = \exp\{N_t - \frac{1}{2}\int_0^t h_s^2 d\langle M\rangle_s\}$, $t \geq 0$ is a continuous local martingale. (10+10)=**[20]**

INDIAN STATISTICAL INSTITUTE

Backpaper Examination, Second Semester: 2012-13
M.Stat. II Year (AS)
Actuarial Models

Date: 12th.7.2013 Maximum marks: 100 Duration: 3 hours

Answer all questions. Standard actuarial notations are followed.

1. (a) Define, in the context of stochastic processes, (i) a stationary process, (ii) a counting process. [2]
   (b) Give an example of application of each type of process. [2]

   [Total 4]

2. The price of a stock can either take a value above a certain point (state A), or take a value below that point (state B). Assume that the evolution of the stock price in time can be modelled by a two-state Markov jump process with homogeneous transition rates $\sigma_{AB} = \sigma$, $\sigma_{BA} = \rho$. The process starts in state A at $t = 0$ and time is measured in weeks.
   (a) Write down the generator matrix of the Markov jump process. [1]
   (b) State the distribution of the holding time in each of states A and B. [1]
   (c) If $\sigma = 3$, find the value of t such that the probability that no transition to state B has occurred until time t is 0.2. [2]
   (d) Assuming all the information about the price of the stock is available for a time interval $[0, T]$, explain how the model parameters σ and ρ can be estimated from the available data. [3]
   (e) State what you would test to determine whether the data support the assumption of a two-state Markov jump process model for the stock price. [1]

   [Total 8]

3. (a) Explain the difference between a time-homogeneous and a time inhomogeneous Poisson process. [1]
   (b) An insurance company assumes that the arrival of motor insurance claims follows an inhomogeneous Poisson process. Data on claim arrival times are available for several consecutive years. Describe one statistical test that can be used to test the validity of this assumption. [3]
   (c) The above mentioned company concludes that an inhomogeneous Poisson process with rate $\lambda(t) = 3\cos(2\pi t)$ is a suitable fit to the claim data (where t is measured in years).
      i. Comment on the suitability of this transition rate for motor insurance claims. [2]
      ii. Write down the Kolmogorov forward equations for $P_{0j}(s,t)$. [2]
      iii. Verify that these equations are satisfied by:

$$P_{0j}(s,t) = \frac{(f(s,t))^j \exp(-f(s,t))}{j!}$$

      for some $f(s,t)$ which you should identify. [2]

iv. Comment on the form of the solution compared with the case where λ is constant. [2]

[Total 12]

4. A motor insurance company wishes to estimate the proportion of policyholders who make at least one claim within a year. From historical data, the company believes that the probability a policyholder makes a claim in any given year depends on the number of claims the policyholder made in the previous two years. In particular:

- the probability that a policyholder who had claims in both previous years will make a claim in the current year is 0.25;
- the probability that a policyholder who had claims in one of the previous two years will make a claim in the current year is 0.15; and
- the probability that a policyholder who had no claims in the previous two years will make a claim in the current year is 0.1.

(a) Construct this as a Markov chain model, identifying clearly the states of the chain. [2]

(b) Write down the transition matrix of the chain. [1]

(c) Explain why this Markov chain will converge to a stationary distribution. [2]

(d) Calculate the proportion of policyholders who, in the long run, make at least one claim in a given year. [4]

[Total 9]

5. A manufacturer uses a test rig to estimate the failure rate in a batch of electronic components. The rig holds 100 components and is designed to detect when a component fails, at which point it immediately replaces the component with another from the same batch. The following are recorded for each of the n components used in the test $(i = 1, 2, \ldots, n)$:

$$\begin{aligned} s_i &= \text{time at which component } i \text{ is placed on the rig,} \\ t_i &= \text{time at which component } i \text{ is removed from rig,} \\ f_i &= \begin{cases} 1 & \text{if component is removed due to failure,} \\ 0 & \text{if component is working at end of test period.} \end{cases} \end{aligned}$$

The test rig was fully loaded and was run for two years continuously. You should assume that the force of failure, μ, of a component is constant and component failures are independent.

(a) Show that the contribution to the likelihood from component i is:

$$\exp\left(-\mu\left(t_i - s_i\right)\right) \cdot \mu^{f_i}.$$ [2]

(b) Derive the maximum likelihood estimator for μ. [4]

[Total 6]

6. (a) Assume that the force of mortality between consecutive integer ages, y and $y+1$, is constant and takes the value μ_y. Let T_x be the future lifetime after age x $(x \leq y)$ and $S_x(t)$ be the survival function of T_x. Show that

$$\mu_y = \log[S_x(y-x)] - \log[S_x(y+1-x)].$$ [4]

(b) An investigation was carried out into the mortality of male life office policyholders. Each life was observed from his 50th birthday until the first of three possible events occurred: his 55th birthday, his death, or the lapsing of his policy. For those policyholders who died or allowed their policies to lapse, the exact age at exit was recorded. Using the result from part (a) or otherwise, describe how the data arising from this investigation could be used to estimate μ_{50} and ${}_5q_{50}$. [4]

[Total 8]

7. A national mortality investigation is carried out over the calendar years 2002, 2003 and 2004. Data are collected from a number of insurance companies. Deaths during the period of the investigation, θ_x, are classified by age nearest at death. Each insurance company provides details of the number of in-force policies on 1 January 2002, 2003, 2004 and 2005, where policyholders are classified by age nearest birthday, $P_x(t)$.

   (a) State the rate year implied by the classification of deaths. [1]

   (b) State the ages of the lives at the start of the rate interval. [2]

   (c) Derive an expression for the exposed to risk, in terms of $P_x(t)$, which may be used to estimate the force of mortality in year t at each age. State any assumptions you make. [3]

   (d) Describe how your answer to part (c) would change if the census information provided by some companies was $P_x^*(t)$, the number of in-force policies on 1 January each year, where policyholders are classified by age last birthday. [3]

[Total 9]

8. An investigation took place into the mortality of persons between exact ages 60 and 61 years. The table below gives an extract from the results. For each person it gives the age at which they were first observed, the age at which they ceased to be observed and the reason for their departure from observation.

Person	*Age at entry*		*Age at exit*		*Reason for exit*
	years	*months*	*years*	*months*	
1	60	0	60	6	withdrew
2	60	1	61	0	survived to 61
3	60	1	60	3	died
4	60	2	61	0	survived to 61
5	60	3	60	9	died
6	60	4	61	0	survived to 61
7	60	5	60	11	died
8	60	7	61	0	survived to 61
9	60	8	60	10	died
10	60	9	61	0	survived to 61

   (a) Estimate q_{60} using the Binomial model. [5]

   (b) List the strengths and weaknesses of the Binomial model for the estimation of empirical mortality rates, compared with the Poisson and two-state models. [3]

[Total 8]

9. An investigation was undertaken of the mortality of persons aged between 40 and 75 years who are known to be suffering from a degenerative disease. It is suggested that

the crude estimates be graduated using the formula:

$$\overset{\circ}{\mu}_{x+1/2} = \exp\left[b_0 + b_1\left(x+\frac{1}{2}\right) + b_2\left(x+\frac{1}{2}\right)^2\right].$$

(a) Explain why this might be a sensible formula to choose for this class of lives. [1]

(b) Suggest two techniques which can be used to perform the graduation. [2]

(c) The table below shows the crude and graduated mortality rates for part of the relevant age range, together with the exposed to risk at each age and the standardized deviation at each age.

Age last birthday	*Graduated force of mortality*	*Crude force of mortality*	*Exposed to risk*	*Standardized deviation*
x	$\overset{\circ}{\mu}_{x+1/2}$	$\hat{\mu}_{x+1/2}$	E_x^c	$z_x = \dfrac{E_x^c(\overset{\circ}{\mu}_{x+1/2} - \hat{\mu}_{x+1/2})}{(E_x^c \overset{\circ}{\mu}_{x+1/2})^{1/2}}$
50	0.08127	0.07941	340	–0.12031
51	0.08770	0.08438	320	–0.20055
52	0.09439	0.09000	300	–0.24749
53	0.10133	0.10345	290	0.11341
54	0.10853	0.09200	250	–0.79336
55	0.11600	0.10000	200	–0.66436
56	0.12373	0.11176	170	–0.44369
57	0.13175	0.12222	180	–0.35225

Test this graduation for (i) overall goodness-of-fit, (ii) bias, and (iii) the existence of individual ages at which the graduated rates depart to a substantial degree from the observed rates. [9]

[Total 12]

). A life insurance company has carried out a mortality investigation. It followed a sample of independent policyholders aged between 50 and 55 years. Policyholders were followed from their 50th birthday until they died, they withdrew from the investigation while still alive, or they celebrated their 55th birthday (whichever of these events occurred first).

(a) Describe the censoring that is present in this investigation. [2]

An extract from the data for 12 policyholders is shown in the table below.

Policyholder	*Last age at which policyholder was observed (years and months)*	*Outcome*
1	50 years 3 months	Died
2	50 years 6 months	Withdrew
3	51 years 0 months	Died
4	51 years 0 months	Withdrew
5	52 years 3 months	Withdrew
6	52 years 9 months	Died
7	53 years 0 months	Withdrew
8	53 years 6 months	Withdrew
9	54 years 3 months	Withdrew
10	54 years 3 months	Died
11	55 years 0 months	Still alive
12	55 years 0 months	Still alive

(b) Calculate the Nelson-Aalen estimate of the survival function. [6]

(c) Sketch on a suitably labelled graph the Nelson-Aalen estimate of the survival function. [3]

[Total 11]

11. An investigation was undertaken into the effect of a new treatment on the survival times of cancer patients. Two groups of patients were identified. One group was given the new treatment and the other an existing treatment. The following model was considered:

$$h_i(t) = h_0(t) \exp\left(\beta^T z\right),$$

where $h_i(t)$ is the hazard at time t, where t is the time since the start of treatment

$h + 0(t)$ is the baseline hazard at time t

z is a vector of covariates such that:

$z_1 =$ sex (a categorical variable with $0 =$ female, $1 =$ male)

$z_2 =$ treatment (a categorical variable with $0 =$ existing treatment, $1 =$ new treatment), and

β is a vector of parameters, (β_1, β_2).

The results of the investigation showed that, if the model is correct, then (i) the risk of death for a male patient is 1.02 times that of a female patient, and (ii) the risk of death for a patient given the existing treatment is 1.05 times that for a patient given the new treatment.

(a) Estimate the value of the parameters β_1 and β_2. [4]

(b) Estimate the ratio by which the risk of death for a male patient who has been given the new treatment is greater or less than that for a female patient given the existing treatment. [2]

(c) Determine, in terms of the baseline hazard only, the probability that a male patient will die within 3 years of receiving the new treatment. [2]

[Total 8]

INDIAN STATISTICAL INSTITUTE
Second Semester Backpaper Examination: 2012-13

M. Stat. II Year
Asymptotic Theory of Statistical Inference

Date: 19 .07.2013 **Maximum Marks : 100** **Duration: 3 Hours**

Answer all questions

1 Describe Fisher's notion of asymptotic efficiency of estimators and comment on this in the light of Hodges' example. Write a short note on the Hajek-Le-Cam theory of efficient estimation.

[16]

2 Let P_{on} and P_{1n} be probabilities on (χ_n, A_n), $n \geq 1$ such that P_{1n} is contiguous to P_{on}. Let $\wedge_n = \log \frac{dP_{1n}}{dP_{on}}$ be the corresponding log-likelihood ratio and T_n be a sequence of random variables such that $\mathscr{L}\{(T_n, \wedge_n) | P_{on}\}$ converges to a distribution G_0. Show that $\mathscr{L}\{(T_n, \wedge_N) | P_{1n}\}$ converges to a distribution G_1 where $\frac{dG_1(t,\lambda)}{dG_0} = e^{\lambda}$.

[16]

3. Let $X_1, X_2, \ldots, X_n$ be i.i.d. with a common density $f(x,\theta), \theta \in \mathbb{R}$. *Fix* $\theta_0 \in \mathbb{R}$ and define $Z_n(u) = \prod_{i=1}^{n} f(X_i, \theta_0 + u) \Big/ \prod_{i=1}^{n} f(X_i, \theta_0)$. Suppose for any closed set G not containing zero, $\sup_{u \in G} Z_n(u) \to 0$ almost surely under θ_0. Show that the MLE, if it exists, converges to θ_0 almost surely under θ_0.

[9]

4. Let $\{P_\theta^n, \theta \in R\}$, $n \geq 1$, be a sequence of statistical experiments such that for a fixed $\theta \in \mathbb{R}$ and for all $u \in \mathbb{R}$,

$$\log \frac{dP^n_{\theta + u/\sqrt{n}}}{dP^n_\theta} = u\Delta_n - \frac{1}{2}u^2 I(\theta) + o_p(1)$$

where $I(\theta)$ is a finite positive number and Δ_n is a sequence of random variables converging in distribution to some random variable Δ (under θ). Show that for all $u \in \mathbb{R}$, $\{P^n_{\theta+u/\sqrt{n}}\}$ is contiguous to $\{P^n_\theta\}$ if and only if Δ follows $N(0, I(\theta))$.

[15]

P.T.O.