INSTITUTE OF ACTUARIES OF INDIA

EXAMINATIONS

16th May 2008

Subject CT6 – Statistical Methods

Time allowed: Three Hours (10.00 – 13.00 pm)

Total Marks: 100

INSTRUCTIONS TO THE CANDIDATES

1. ***Do not write your name anywhere on the answer scripts. You have only to write your Candidate's Number on each answer script.***

2. ***Mark allocations are shown in brackets.***

3. ***Attempt all questions, beginning your answer to each question on a separate sheet. However, answers to objective type questions could be written on the same sheet.***

4. ***Fasten your answer sheets together in numerical order of questions. This, you may complete immediately after expiry of the examination time.***

5. ***In addition to this paper you should have available graph paper, Actuarial Tables and an electronic calculator.***

Professional Conduct:

"It is brought to your notice that in accordance with provisions contained in the Professional Conduct Standards, If any candidate is found copying or involved in any other form of malpractice, during or in connection with the examination, Disciplinary action will be taken against the candidate which may include expulsion or suspension from the membership of IAI."

Candidates are advised that a reasonable standard of handwriting legibility is expected by the examiners and that candidates may be penalized if undue effort is required by the examiners to interpret scripts.

AT THE END OF THE EXAMINATION
Please return your answer scripts and this question paper to the supervisor separately.

Q1)

(i) What is a saddle point? (1)

(ii) Which strategy can be used in a game in which there is no saddle point? (2)

(iii) For what values of x and y does the matrix given below have a saddle point? (5)

$$\begin{bmatrix} 4 & 7 & 6 \\ 3 & 2 & y \\ x & 5 & 8 \end{bmatrix}$$

[8]

Q2)

Time plot of a time series shows a linear trend.

(i) Explain how a linear trend can arise from the following models

a) a linear function of time plus a stationary process, (1)

b) cumulative sum of a stationary process with non-zero mean. (1)

(ii) Describe two ways to remove linear trend from a time series. (2)

[4]

Q3) The total number of claims arising in a year have the Poisson distribution with mean 100. The claim sizes are independent and follow the log-normal distribution with mean Rs, 2,000 and standard deviation Rs. 1,000. There is an excess-of-loss reinsurance arrangement with retention limit Rs. 2,500.

(i) Determine the percentage of claims that involve the re-insurer. (3)

(ii) Determine the mean of the claim amounts paid by the re-insurer. (3)

(iii) If the number of claims paid by the direct insurer in a year happens to be n, give an expression for the probability that exactly m of these claims involve the re-insurer. (1)

(iv) If the re-insurer pays m claims in a given year, determine the range of possible values of the total number of claims paid by the direct insurer. (1)

(v) Calculate the unconditional probability that exactly m claims involve the re-insurer, to show that the number of claims experienced by the re-insurer has a Poisson distribution. (3)

(vi) Calculate the expected aggregate claim amount paid by the re-insurer in one year. (1)

[12]

Q4) It is presumed that during the initial years of a company, the events of posting net annual profit in successive years occur independently with probability p. A watchdog agency gives a profitability score (X) to start-up companies, and keeps a record of the number of years (Y) taken by the company to post net annual profit for the first time (Y=1,2,3,…). An analyst wishes to fit a Generalized Linear Model to the paired data (X,Y) for several start-up companies.

(i) What is the distribution of Y? (1)

(ii) Calculate the mean of Y. (1)

(iii) Show that the distribution of Y belongs to an exponential family, and identify the natural parameter, the mean and the variance function. (4)

(iv) Using the canonical link function, state a model that relates X to the mean of Y. (2)

(v) The analyst intends to fit the above model to the data $(X_1,Y_1),\ldots,(X_n,Y_n)$, for n start-up companies. Write down the likelihood function explicitly in terms of the model parameters and the data. (2)

[10]

Q5) Aggregate annual claims on a portfolio of insurance policies have a compound Poisson distribution with parameter λ. Individual claim amounts have the exponential distribution:

$$f(x) = \exp(-x), \qquad x > 0 .$$

The insurer calculates premium using a loading of α and the initial surplus U.

(i) If the first claim occurs at time t, what is the probability of that claim causing ruin? (3)

(ii) Show that the probability of ruin at the first claim is $\exp(-U) / (2 + \alpha)$ (5)

(iii) Determine the minimum loading α if the insurer wishes to make sure that the probability of ruin at the first claim is less than 5%. (2)

[10]

Q6) Imagine you are a juror for a murder trial. The victim had been stabbed to death at her home. The defendant is a neighbour of the victim. After questioning a number of the victim's other neighbours, the police framed charges against this defendant. During the trial, after you hear most of the prosecution's evidence against the defendant, you believe the evidence is very weak. You estimate there is only about a 10% chance that the defendant is guilty. This is the prior probability of guilt. However, at the last minute, the prosecution introduces evidence they recently obtained from a laboratory test of blood found at the crime scene. They state that DNA tests performed on the blood indicate that it has a genetic profile that matches the defendant's profile. The prosecution states that this DNA test has only 1% chance of false match, and 1% chance of false mismatch. Thus, there is only a 1% chance that a randomly selected, innocent person's DNA would match the DNA found at the crime scene, and only 1% chance that the actual murderer's DNA would not match it. You are also told that the DNA test itself is error-free. What is the probability that the defendant is guilty, given that there was a DNA match? **[5]**

Q7) Consider the following probability mass function of a discrete random variable X.

$P_x(X = 2) = 0.15$

$P_x(X = 3) = 0.20$

$P_x(X = 5) = 0.25$

$P_x(X = 7) = 0.20$

$P_x(X = 11) = 0.20$

Using the following pseudo-random numbers from the Uniform (0,1) distribution, generate nine samples from P_x: 0.011, 0.757, 0.438, 0.258, 0.981, 0.518, 0.400, 0.351, 0.672.

[4]

Q8) An insurance company charges an annual premium of Rs 500 and operates a No Claims Discount system as follows:

Level 1	:	0% discount
Level 2	:	25% discount
Level 3	:	50% discount

The rules for moving between levels are as follows:

If the policyholder does not make a claim during the year, they move up one level or are eligible to stay at level 3.

If the policyholder makes 1 claim during the year, they move down one level or stay at level 1.

If the policyholder makes 2 or more claims during the year, they move straight down to, or remain at, level 1.

The insurance company has recently introduced a "protection" system where on reaching, or remaining eligible to remain at, level 3, policyholders are immediately offered the opportunity to "protect" their discount for an additional annual premium of Rs 50. If they make no claims or 1 claim during the year they can remain at level 3. However, if they make 2 or more claims during the year, they move straight t level 1.

Out of the policyholders who had "protected" their discount level at the beginning of the year and are still eligible to stay at level 3 at the end of the year, 25% chose to "protect" their discount again the following year. From all the policyholders eligible for level 3 at the end of the year, 10% chose to take the "protection" option.

Policyholders at different levels are found to have different rates of claiming. The number of claims made per year follows a Poisson distribution with parameter λ as follows:

Level	1 and 2	3 and "protected"
λ	0.6	0.4

(i) Derive the transition matrix. (7)

(ii) Calculate the proportions at each of the levels 1, 2, 3 and the "protected" level, when the system reaches a steady state. (6)

(iii) Determine the average premium per policy. (2)

[15]

Q9) Claims paid to date on a motor insurance account are as follows:

Figures in Rs 000's

	Development Year			
Policy Year	0	1	2	3
2004	1500	1300	900	350
2005	2050	1575	1200	
2006	2400	1900		
2007	3000			

Inflation for the 12-month period to the middle of each year was as follows:

2005	7%
2006	5%
2007	5.5%

You are given the following further information:

- Annual premiums written in 2007 were Rs 55,00,000
- Future inflation from mid 2007 is estimated to be 5% p.a.
- The ultimate loss ratio (based on mid 2007 prices) has been estimated at 83%
- Claims are assumed to be fully run off by the end of development year 3

Estimate the outstanding claims arising from policies written in 2007 only, taking explicit account of the inflation statistics, using the Bornhuetter-Ferguson method.

[13]

Q10)

(i) Explain what a conjugate prior distribution is. (2)

(ii) The random variables $X_1, X_2, \ldots, X_n$ are independent and have a density function

$$f(x) = \lambda * \exp(-\lambda x), \quad x > 0.$$

Show that the Gamma distribution can be a conjugate prior for λ. (3)

(iii) a) If the density function of λ is

$$f(\lambda) = \frac{s^{\alpha}\lambda^{\alpha-1}e^{-s\lambda}}{\Gamma(\alpha)}, \quad \lambda > 0,$$

show that $E[1/\lambda] = s/(\alpha-1)$

b) If $X_1, X_2, \ldots, X_n$ is an independent random sample from the exponential distribution with parameter λ, show that the posterior mean of 1/λ (for a Gamma prior) can be expressed as a credibility estimate, i.e., a weighted average of the prior mean of 1/λ and the sample average.

(6)

[11]

Q11)

(i) Define a Markov Process. (2)

(ii) For what value of p an AR(p) process is a (one-step) Markov Process? (1)

(iii) Rearrange terms of the AR(2) process $X_t = 0.5\, X_{t-1} + X_{t-2} + e_t$ such that the resultant vector AR process becomes a (one-step) Markov Process. (2)

[5]

Q12) Residuals are plotted after fitting a time series model based on a sample of 100 observations. The graph has 43 turning points. Test independence of the residuals at the 5% level of significance. **[3]**

****