INSTITUTE OF ACTUARIES OF INDIA

EXAMINATIONS

7th November 2008

Subject CT6 – Statistical Methods

Time allowed: Three Hours (10.00 – 13.00 Hrs)

Total Marks: 100

INSTRUCTIONS TO THE CANDIDATES

1. *Please read the instructions on the front page of answer booklet and instructions to examinees sent along with hall ticket carefully and follow without exception*
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. *In addition to this paper you will be provided with graph paper, if required.*

AT THE END OF THE EXAMINATION

Please return your answer book and this question paper to the supervisor separately.

Q1) a) Explain what is meant by 'prior distribution' and 'posterior distribution'. (2)

b) Random variable Y has density function $f(y) = \theta y \exp(-\frac{\theta y^2}{2}),\ y > 0, \theta > 0$, where θ is an unknown parameter. A random sample of 12 values of Y is given below:

2.58, 2.65, 2.89, 2.25, 3.06, 2.57, 2.41, 2.68, 2.93, 2.30, 2.43, 2.93.

Assuming that the prior distribution of θ is Gamma (12, 21)
(i) Determine the posterior distribution of θ; (5)
(ii) Calculate the Bayesian estimate of θ using the squared error loss. (1)

[8]

Q2) a) In the theory of games, what do you understand by:
i) Pure strategy, (1)
ii) Minimax criterion. (2)

b) Consider a zero sum game with the following loss matrix for player A.

		Player A	
		Strategy I	Strategy II
Player B	Strategy 1	4	–1
	Strategy 2	–3	2

Determine the optimal strategy for each player based on the minimax criterion. (4)

c) A random variable has density function $f(x) = 2x/\theta^2$, $0<x<\theta$. A statistician is trying to decide between the two possibilities, θ = 3 and θ = 5, on the basis of a single observed value from the distribution. It will cost him Rs 20 if he chooses θ = 3 (incorrectly) and Rs 50 if he chooses θ = 5 (incorrectly). A correct decision incurs no loss. He has decided that he would choose θ = 3 if the observed value from the distribution is less than a constant k and θ = 5 if the observed value is greater than k. What value of the constant k should be chosen to minimize the maximum risk? (6)

[13]

Q3) a) What are the advantages that an insurer can have by offering a No Claims Discount (NCD) to its policyholders? (2)

b) The NCD system operated by a motor insurer has three levels of discount: 0%, *X%* and 2*X%*. The system operates according to the following rules:

- A driver who makes no claim in a year moves up one level of discount or remains at the 2*X%* discount level,
- A driver who makes one or more claims in a particular year moves down one level of discount or remains at the 0% discount level.

The basic premium payable by drivers at the 0% discount level is *P*. The company has insured 9940 drivers for many years. The company classifies its policyholders as 'good' drivers and 'bad' drivers. There are 7735 good drivers and 2205 bad drivers. For a good driver, the probability of having a claim free year is 0.9, for a bad driver this probability is 0.8.

i.) Calculate the expected number of good drivers and bad drivers at each discount level, assuming steady state. (7)

ii.) Derive expressions, in terms of P and X, for the average premium paid by the good drivers and the average premium paid by the bad drivers. (2)

iii.) Determine the value of X required to ensure that the average premium paid by the good drivers is equal to half the average premium paid by the bad drivers. Comment on your computed discount levels. (2)

[13]

Q4) Consider a portfolio of insurance policies with aggregate claims $S = X_1 + X_2 + \cdots + X_N$, where the claim sizes (X_i's) are independent and identically distributed random variables, and the number of claims N is independent of the X_i's.

i.) Write down a formula for the variance of S. (1)

ii.) If the X_i's have the exponential distribution with mean $1/\lambda$ and N has the geometric distribution with probability density function

$$P(N = n) = q^n p,\ n = 0, 1, 2, \ldots,\ \text{show that}\ V(S) = \frac{q(p+1)}{\lambda^2 p^2}.$$ (3)

iii.) Suppose that the number of claims (N) till time t has the Poisson distribution with parameter θt, and the individual claim size distribution is exponential with mean $1/\lambda$. The insurer collects premium continuously at a constant rate that makes the amount of premium collected over a period of time equal to the expected aggregate claim size over the same period. What is the annual rate of the premium? If the insurer's initial surplus is U, give an expression for the probability of ruin at the end of one year. (4)

[8]

Q5) a. In the context of Generalized Linear Models, define the terms:

i. Covariate, (1)

ii. Link function. (2)

b. When is a probability distribution said to belong to an exponential family? (1)

c. Justify whether or not the following distributions belong to the exponential family:

i. Exponential distribution with parameter λ for $\lambda > 0$. (2)

ii. Weibull distribution with density function $f(x) = crx^{r-1}\exp(-cx^r),\ x > 0, r \neq 1, c > 0$. (2)

iii. Geometric distribution with density function $f(x) = (1-q)q^x$ for $x = 0, 1, 2, \ldots,\ 0<q<1$. (2)

[10]

Q6) The following data are taken from the records of a non-life insurance company.

Claim amounts paid in the year of accident and incremental amounts paid in subsequent years are as follows:

(Amounts in ‘000s)

		Year of payment				
		1997	**1998**	**1999**	**2000**	**2001**
Accident year	**1997**	28,791	22,063	2,805	378	78
	1998		27,620	2,310	17,725	8,256
	1999			26,935	11,925	9,872
	2000				36,661	9,222
	2001					18,619

Annual rate of claims inflation

Year	Rate
1997-98	5%
1998-99	7%
1999-00	6%
2000-01	9%

Expected future rate of claims inflation

Year	Rate
2001-02	6%
2002-03	6%
2003-04	5%
2004-05	5%

Reserves held in deposit on or after 31/12/2001 are expected to earn interest @8% per annum.

Assuming that 1997 claims would have “run-off” fully by the end of 2001, you have to estimate the reserves needed in respect of claims outstanding at that time.

a)
i) Use the Basic Chain Ladder method without taking into account the given data regarding inflation and interest earnings. (5)

ii) What is the underlying assumption regarding inflation? (1)

b) Use inflation adjusted chain ladder method, taking into account the inflation, both past and future, and interest earned on deposits from 31/12/2001 onwards (Hint: interest earned on deposits is to be used in calculating discounted reserves). (10)

[16]

Q7) The aggregate claims from a risk have the compound Poisson distribution with parameter μ, where μ is the expected number of claims. Individual claim amounts have the Pareto distribution with parameters $\alpha = 3$, $\lambda = 1000$. The insurer of this risk calculates the premium using a premium loading factor of 0.2 (i.e., it charges 20% in excess of the risk premium). The insurer is considering effecting excess of loss reinsurance with retention limit of Rs 1000 for each claim. The reinsurance premium would be calculated using a premium loading factor of 0.3. The insurer’s expected profit is defined to be the premium charged by the insurer, less the reinsurance premium, and less the claims paid by the insurer net of reinsurance.

a) Show that the insurer’s expected profit before reinsurance is 100μ. (2)

b) Calculate the insurer’s expected profit after effecting reinsurance, and hence the percentage reduction in the insurer’s expected profit. (6)

[8]

Q8) The density function f is defined by

$$f(x) = k\phi(x), \text{ for } -3 < x < 3$$
$$= 0, \text{ otherwise}$$

where $\phi(x)$ is the density function of standard normal.

a) Find the value of k rounded off at 4 decimal points. (2)

b) What linear transformation of a Uniform (0,1) random variate will give a random variate from Uniform(–3,3)? (2)

c) Give a step-wise algorithm to generate random variates from this truncated normal distribution using Uniform(0,1) random variates (4)

[8]

Q9) a) Define Partial autocorrelation function (PACF). Also briefly mention how it can be derived by minimizing the expected value of "square of difference between X_t and a linear functions of $X_{t-1}, X_{t-2},\ldots, X_{t-k}$". (2)

b) Let ϕ_k denote the PACF between X_t and X_{t-k}. Derive and explain why is PACF $\phi_k = 0$ for k>2 for an AR(2) process. (4)

[6]

Q10) The time series X_t is assumed to follow ARIMA(p,d,q) process defined by

$X_t = 1 + \frac{11}{6}X_{t-1} - X_{t-2} + \frac{1}{6}X_{t-3} + \varepsilon_t$, where ε_t are N(0,1) random variables. One of the roots of the characteristic equation is 1.

a. Derive the other roots of the characteristic equation. (3)
b. State whether the process is stationary. (1)
c. Determine p, d and q. (1)
d. What is the mean of $X_t - X_{t-1}$? (2)
e. What is the mean of X_t? (1)
f. If X_{98}=2.4, X_{99}=1.6 and X_{100}=1.2, forecast X_{101}. (2)

[10]

****