INSTITUTE OF ACTUARIES OF INDIA

EXAMINATIONS

12th November 2013

Subject CT6 – Statistical Models

Time allowed: Three Hours (10.30 – 13.30 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.***

4. ***Please check if you have received complete Question Paper and no page is missing. If so, kindly get new set of Question Paper from the Invigilator.***

AT THE END OF THE EXAMINATION

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

Q. 1) There is a group of 100 yearly premium paying mode, term insurance policies, just issued, each with term 6 years and yearly premium of Rs. 1000. After issuance, each policy may take their entire premium back within 15 days which has a probability of 10% .Under each policy, if the premium is not taken back within 15 days, the number of total premium installment received over it's lifetime is denoted by the random variable N, where P(N = i) = 1/6 for i = 1,2,3,..,6.

i) Find the mean and variance of total premium received out of a single policy just issued over its lifetime. (3)

ii) Derive an expression to express P(X = m), where the random variable X denotes the total amount of premiums received in Rs. over their lifetime from this group of 100 policies, where each of the polices has crossed 15 days from issue date and no refund of premium has been taken.

(You may assume premium payments from each of the policies are independent) (6)

iii) Check that your derived expression gives correct result for m = 100000. (1)

[10]

Q. 2) Consider the time series model defined by $X_t = (4/3)\ X_{t-1} - (7/12)\ X_{t-2} + (1/12)\ X_{t-3} + \varepsilon_t$, where $\{\varepsilon_t\}$ is the white noise.

i) Derive the autocorrelation coefficient with lag 1 and lag 2 i,e, ρ_1 and ρ_2 for this process. (4)

ii) Comment on the stationarity of this model. (3)

iii) Calculate the partial autocorrelation coefficient Φ_1 and Φ_2 from this model. (2)

[9]

Q. 3) A student has written the following algorithm to generate random variates for an unknown distribution. RAND() is the syntax to generate a random number from U (0, 1) distribution & LN denotes logarithm to the base e.

Step 1 : U = 0, X = 0;

Step 2 : Z= RAND() ;

Step 3 : U = LN(1 – Z);

Step 4 : U = - U / 3;

Step 5 : U = $U^{(1/4)}$;

Step 6 : X = U;

Step 7 : RETURN X ;

Study the algorithm and derive the density function of the unknown distribution. Also, state the name of the distribution with parameters.

[6]

Q. 4) An insurance company designs a group health insurance product suitable for employers wishing to reimburse minor medical expenses of the employee and his / her family members. Under this scheme the employee or any family member can claim reimbursement for medical consultation fee paid and the cost of the medicines prescribed. The aggregate yearly medical consultation expenses of an individual employee's family (including the employee) has a compound Poisson distribution with Poisson parameter 0.5 and the individual consultation expense has a Gamma (α,λ) distribution. The prescribed medicine cost is independent of the consultation fee and is a uniformly distributed random variable between Rs. 100 to Rs. C ($C > 100$). Every month, the employer has to provide Rs. 40 for each employee to fund the scheme. Any deficit or surplus from the scheme is funded or taken back by the insurer at the year end to bring the net fund position at zero. The random variable S denotes the total yearly claimed consultation & medicine expenses from the scheme. You may assume S has an approximately normal distribution. You may also ignore any interest earned by the scheme. The insurer is accessing the target employers to whom it can market the product.

i) If $\alpha = 5.5$, $\lambda = 0.01$ and $C = 400$, derive the minimum number of employees an employer should have to be 99% sure that the insurer will not have to fund any deficit at the year end. (7)

ii) The values of α, λ & C are not known with certainty but are estimated as $5 \leq \alpha \leq 6.5$, $0.008 \leq \lambda \leq 0.011$ and $100 \leq C \leq 500$. Considering the worst possible combination of values of α, λ & C from the insurer's perspective, derive the minimum number of employees the employer should have to be 99% sure that the insurer will not have to fund any deficit at the year end (6)

[13]

Q. 5) **i)** A non life insurer insures all cattle of the farmers of a small district. On death of any domestic animal, the insurer reimburses the monetary value of the cattle to its owner. The yearly premium is determined using a security loading factor of 25%. The premium is shared among the farmer's co-operative and the local government at the ratio 50:50.

The historical record shows that the number of yearly death of the cattle in the district has a Poisson distribution with mean λ. The monetary value is INR of the cattle have an exponential distribution with parameter 0.0001.

Ignoring interest and expenses calculate the insurer's adjustment coefficient and derive an upper bound of the insurer's probability of ruin if it has an initial surplus of INR 1000000. (6)

ii) Explain from the first principle as to why the upper bound as calculated above in part a, is not dependent upon the Poisson parameter λ. Explain in the process which element is affected by the Poisson parameter λ. (4)

[10]

Q. 6) The term insurance business of a small life insurer is spread over three geographical regions of a country. It has collected 35 weeks' number of claims data of its term insurance business from the three geographical regions. In the data, only the Region I have 4 weekly claims data where number of claims in the week is 2. All the other weekly claims data have values either 0 or 1. The first 10 data is for Region I, the next 5 data is for Region II and the last 20 data is for Region III. The total number of claims in the data is 11, 3 and 4 respectively for Region I, Region II and Region III respectively. It wishes to use a Poisson model to analyze the data.

i) Show that the Poisson distribution is a member of the exponential family of distribution. (2)

ii) The insurer decides to use a model (model I) for which

$$\log \mu_i = a \quad \text{for Region I,}$$
$$= b \quad \text{for Region II,}$$
$$= c \quad \text{for Region III,}$$

Here μ_i is the mean of the relevant Poisson distribution.

Derive the likelihood functions for the model and hence the maximum likelihood estimates for a, b & c. (4)

iii) The insurer also wishes to analyze the easier model, $\log \mu_i = a$ for all the regions (model II). Find the maximum likelihood estimate for a under this model. (2)

iv) Derive the scaled deviance for model I and model II. (You may assume that $f(y) = y \log y$ is equal to 0 for $y = 0$) (6)

v) Compare model I directly with model II by using an appropriate test statistics. (3)

[17]

Q. 7) There is a group of m independent 'mediclaim' policies which are in the book of an insurer since long time. Under each policy, at the most one claim is possible in any month as per the contract. The probability of a claim in a month for each policy is p ($0 < p < 1$). The total monthly number of claims from the group of m policies are $x_1, x_2, \ldots, x_n$ in the past n months. The prior distribution of p is given by the density function $f(p) \propto \{p(1-p)\}^a$ where $a > -1$.

i) Derive the posterior distribution of p given $x_1, x_2, \ldots x_n$. (4)

ii) Derive maximum likelihood estimate of p ($\underline{p}$) . (3)

iii) Derive the Bayesian estimate of p under quadratic loss and show it takes the form of a credibility estimate $Z\underline{p} + (1 - Z)k$, where k is a scalar (which you should specify) in terms of prior distribution of p. (3)

iv) Explain what happens to Z when n increases gradually. (1)

v) Calculate the Bayesian estimates of p and Z if $m = 100$, $n = 12$ and $x_1 + x_2 + \ldots + x_{12} = 15$ when $a = 0$ and $a = 3$. (2)

vi) Considering the prior variance, comment on effect on Z of increasing a and also relate this effect to the quality of prior information of p in each case. (2)

[15]

Q. 8) The random variable X has Weibull distribution function $F(x) = 1 - exp\,(-cx^{1/4})$, x > 0, where c is an unknown parameter.

i) **a)** Show that, for any positive integer m:

$$E\,(X^m) = c \int_0^\infty y^{4m}\, e^{-cy} dy$$

where y is a function of x. (4)

b) By comparing the integrand in (a) with a gamma density, or otherwise, show that:
E (X^m) = c^{-4m} (4m)! , where! denotes factorial function. (3)

ii) A sample of 100 claims from an insurance portfolio gave $\sum_{i=1}^{100} x_i^{1/4} = 1430$, where x_i denotes the amount of the i^{th} claim amount.

Use the information given above to fit the distribution function F(x) to this set of data by calculating the maximum likelihood estimate of c. (4)

iii) The table below shows the number of the claim amounts in part (ii) in each of the ranges indicated:

Range	**Number of claims**
0 < x < 100	12
100 ≤ x < 1,000	15
1,000 ≤ x < 10,000	18
10,000 ≤ x < 100,000	18
x ≥ 100,000	37

For the distribution fitted in part (ii):

a) Calculate the expected number of claim amounts in each of the ranges in the above table, (4)

b) Calculate the χ^2 goodness of fit statistic, and (2)

c) Comment briefly on the quality of the fit. (3)

[20]

****