INSTITUTE OF ACTUARIES OF INDIA

EXAMINATIONS

15^{th} March 2022

Subject CS1A – Actuarial Statistics (Paper A)

Time allowed: 3 Hours 30 Minutes (9.30 - 13.00 Hours)

Total Marks: 100

Q. 1) Assume for a health insurance policy the number of claims follows a Poisson process with a rate of 0.3 per year.

i) Determine the probability that no claim arises in the policy in one year. (1)

ii) For the policy determine the probability that, out of four consecutive years, there are one or more claim(s) in two of the years and no claim in the remaining two years. (2)

iii) Assuming that a claim has just occurred, determine the probability that more than three year will elapse before occurrence of the next claim. (2)

iv) For the health insurance policy, provide the steps to be followed in order to simulate an observation of N number of claims, occurring in one year. (1)

v) Using the following two random numbers (uniformly distributed between 0 and 1) simulate the number of claims in one year for the above health insurance policy.

a) 0.7521
b) 0.6512 (2)

[8]

Q. 2) Following measures are provided for 2 independent random samples.

Sample	Sample size	$\sum x_i$	$\sum x_i^2$
1	15	297	5970
2	15	297	6200

Sample Mean for both the samples = 19.8

i) Calculate sample variance for both the samples. (1)

ii) Assuming that the samples are drawn from a normally distributed population

a) Carry out t-tests of the hypothesis H_0: $\mu = 18$ vs H_1: $\mu \neq 18$ at 98% level of confidence for sample 1. (2)

b) Calculate confidence interval for sample 2 to test the null hypothesis stated in sub-part ii(a). (2)

c) Discuss the reasons for differences and similarities while comparing the test results of Sample 1 and Sample 2. (2)

d) A new 100 sized sample 3 is collected from the same population.
Explain how the width of the confidence interval of sample 3 will differ from confidence interval of sample 2. (2)

e) It was observed that one of the claims in the sample 2 has an extremely large value and can be considered as an outlier. The value is now replaced with a new randomly selected one, which is not an outlier anymore.
For the updated sample, explain how the confidence intervals will differ from that of sample 2 calculated in sub-part ii(b). (2)

[11]

Q. 3) The claim amounts X and Y (in units of INR 1000) for two different types of insurance policy are modelled using a gamma distribution with parameters α = 5, λ = 1/8 and α = 3, λ = 1/4 respectively i.e. X ~ Gamma(5,1/8), Y ~ Gamma(3,1/4). Assume that X and Y are independent of each other.

i) Identify which one of the following options describes the moment generating function of X

A. (1-8t)^(-4)

B. (1-t/8)^(-4)

C. (1-8t)^(-5)

D. None of the above (1)

ii) Use moment generating function to show $\frac{1}{4}X \sim \chi^2_{10}$ (2)

iii) Calculate the probability that the claim amount, X, exceeds INR 40,000. (2)

iv) An analyst argues that sum of X and Y must follow Gamma(8,3/8) i.e. X + Y ~ Gamma(8,3/8)

Comment on the analyst's argument using moment generating function. (2)
[7]

Q. 4) An analyst is working on renewals of health insurance policies. He has estimated following aggregate claim statistics for four policies over last 3 years:

Policy Number	$\sum_{j=1}^{3} X_j$	$\sum_{j=1}^{3} (X_j - \bar{X})$
1	12,183	12,504
2	13,098	12,718
3	12,822	12,432
4	13,453	12,242

i) Using EBCT Model 1, compute the credibility premium of Policy Number 4 for the upcoming renewal. Students may refer actuarial table for the formula. (4)

ii) The medical inflation in last few years is very high and increasing at a rate of 20% per year on average. Describe the adjustment to be made in the above estimation (as determined in sub part i) in order to incorporate the medical inflation. Also, state any additional data that will be required. (3)

iii) Explain, without performing any further calculation, what change is expected to the credibility factor (determined in sub-part i), if Policy Number 1 is excluded. (2)
[9]

Q. 5) The joint probability density function of random variables X and Y is

$$f(x,y) = \begin{cases} ke^{-\left(3x+\frac{y}{5}\right)}, & x > 0, y > 0 \\ 0, & Otherwise \end{cases}$$

i) Determine the value of k. (3)

ii) Identify the right marginal density function of X and Y from the following expressions

A. $f_X(x) = 1/3e^{-3x}$, $f_Y(y) = 5*e^{-\frac{y}{5}}$

B. $f_X(x) = 3e^{-3x}$, $f_Y(y) = 1/5*e^{-\frac{y}{5}}$

C. $f_X(x) = 1/3e^{-\frac{x}{3}}$, $f_Y(y) = 1/5*e^{-\frac{y}{5}}$

D. $f_X(x) = 3e^{-3x}$, $f_Y(y) = 5*e^{-5y}$

E. None of the above (2)

iii) State whether X and Y are independent based on your answer in part (ii). (1)

iv) Determine the conditional density function f(x|X>5). (3)

v) Identify which one of the following expressions is equal to the conditional expectation of E[x|X>5]

A. $\int_0^\infty 15t\, e^{-3t}dt + \int_0^\infty 3\, e^{-3t}dt$

B. $\int_0^\infty 3\, e^{-3t}dt + \int_0^\infty 15t\, e^{-3t}dt$

C. $\int_0^\infty 15(t\,\hat{}2)e^{-3t}dt + \int_0^\infty 3\, e^{-3t}dt$

D. $\int_0^\infty 3t\, e^{-3t}dt + \int_0^\infty 15\, e^{-3t}dt$

E. None of the above (1)

[10]

Q. 6) Aggregate claim amounts of a health insurance portfolio follows a normal distribution with mean μ where prior distribution of μ is $N(\mu_0, \sigma^2)$. For this distribution a model has been built using n sized sample data with mean $\bar{x}$, variance s^2 and credibility factor Z.

State the relationship (increasing, decreasing or nothing) of the credibility factor with

μ_0,
σ^2
and s^2

and provide the reason for the same. **[5]**

Q. 7) Housing prices, X_i (in units of 1000) in region X follow a normal distribution with mean 500 and standard deviation 10 i.e. $X_i \sim N(500, 10\hat{}2)$ for i=1,……..,10.
For another region Y the housing prices Y_j (in units of 1000) are normally distributed with mean 510 and standard deviation 5 i.e. $Y_j \sim N(510, 5\hat{}2)$ for j=1,……..,5. Assume that two samples are independent of each other. Let $\bar{X}$ and $\bar{Y}$ denotes the means of the two samples and let S_X^2 and S_Y^2 be the sample variance.

i) Calculate the probability of "the region X sample mean is greater than the region Y sample mean".

A. 0.00561
B. 0.00489
C. 0.00494
D. 0.00572
E. None of the above (3)

ii) Calculate the probability of sample variance for region X is greater than 100. (2)

iii) Calculate the approximate probability for "sample variance of region X is less than the sample variance of region Y".

A. 2.5%
B. 4.2%
C. 6%
D. 10%
E. None of the above (3)

iv) Calculate the probability of "the region X sample standard deviation is more than 4 times greater than the region Y sample standard deviation". (2)

[10]

Q. 8) **i)** Define conjugate prior distribution. (2)

ii) Assume that $X_1, X_2, \ldots\ldots\ldots, X_n$ are random samples from an exponential distribution with parameter λ, where λ is a random variable.
Prove that the conjugate prior distribution for λ is a Gamma distribution. (3)

iii) If λ ~ Gamma (α, β),

a) Deduce that $E(\frac{n}{\lambda}) = \frac{n\beta}{(\alpha-1)}$ where n is the sample size. (2)

b) Considering the distribution in part (ii), show that the posterior mean of n/λ can be expressed as
$Z * \sum_i^n x_i + (1- Z) *$ prior mean of n/λ (3)

[10]

Q. 9) **i)**

a) Poisson distribution $(e^{-mu} * mu^{y}) / (y!)$ can be written as exp(y*log(mu)-mu-log y!), in the form of a member of the exponential family.

Identify correct option indicating Variance function i.e. V(mu)

A. V(mu) = log(mu)
B. V(mu) = (1/mu) -1
C. V(mu) = (y/mu) -1
D. V(mu) = mu
E. V(mu) = 1/mu (2)

b) Identify correct option indicating mean (mu) and variance as a function of 'mu' for a particular distribution when written in the form of a member of the exponential family having

- b(theta) = -log(-theta)
- theta = -1/mu
- a(phi) = 1

A. mean = mu and variance = mu
B. mean = mu^2 and variance = mu^2
C. mean = mu and variance = mu^2
D. mean = mu^2 and variance = mu
E. None of the above (2)

ii)
a) A student modelling accidental hospitalisation claims found that interaction between 'age band' and 'gender' is significant. Explain what this means (for the model predicting accidental hospitalisation claims) with reference to beta parameter for gender.

(You are told that males are more prone to accidents than females and hence males require more accidental hospitalisation compared to female across all age groups.) (4)

b) A student is fitting GLM model to predict policy renewal rate for a group of policies. Comment on the choice of link function that can be used with suitable expression. (4)

iii) Answer the following questions using ANOVA Table as provided

Source of variation	**d.f.**	**Sum of Squares**	**Mean Sum of Squares**
Regression	1	**A**	SS_{REG}
Residual	12	**B**	**C**
Total	13	5.38	

a) Identify A, B and C to complete the above table and calculate F value when

SS_{REG} =1.38 (for Model 1) and
SS_{REG} =2.38 (for Model 2) (4)

b) Comment on the results in part (a) above using tabulated values provided below, clearly stating the hypothesis

$F_{1,12}$ table values:

1%	**2.50%**	**5%**	**10%**
9.33	6.554	4.747	3.177

(3)

c) Comment on the suitability of Model 1 and Model 2 using R^2.
Note - Students are required to provide calculation of R^2 separately for Model 1 and Model 2 clearly stating the formula. (3)

iv)

a) A statistician is comparing models using R. He observes following results using R software

Result 1

Analysis of Deviance Table

Model 1: y ~ 1
Model 2: y ~ x

	Resid. Df	Resid. Dev	Df	Deviance	F	Pr(>F)	
1	11	1888					
2	10	118.71	1	1769.3	149.05	2.48E-07	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Result 2

Analysis of Deviance Table

Model 1: y ~ x
Model 2: y ~ x * region

	Resid. Df	Resid. Dev	Df	Deviance	F	Pr(>F)	
1	10	118.706					
2	8	31.289	2	87.417	11.175	0.00483	**

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Comment on the results and model you would prefer with reasons. (4)

b) You then further observe following result in R

Result 3

Model 1: y ~ x + region
Model 2: y ~ x * region

	Resid. Df	Resid. Dev	Df	Deviance	F	Pr(>F)
1	9	31.327				
2	8	31.289	1	0.0374	0.0096	0.9245

Comment on the new result and model that you would prefer with reasons based on results in part a and part b. (4)

[30]

****