INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

17 September 2019 (am)

Subject CS1A – Actuarial Statistics
Core Principles

Time allowed: Three hours and fifteen minutes

INSTRUCTIONS TO THE CANDIDATE

1. *Enter all the candidate and examination details as requested on the front of your answer booklet.*

2. *You must not start writing your answers in the booklet until instructed to do so by the supervisor.*

3. *Mark allocations are shown in brackets.*

4. *Attempt all questions, begin your answer to each question on a new page.*

5. *Candidates should show calculations where this is appropriate.*

Graph paper is NOT required for this paper.

AT THE END OF THE EXAMINATION

Hand in BOTH your answer booklet, with any additional sheets firmly attached, and this question paper.

In addition to this paper you should have available the 2002 edition of the Formulae and Tables and your own electronic calculator from the approved list.

1 A survey showed that 40% of investors invest in at least two companies in order to diversify their risk.

Calculate an approximate probability that more than 100 investors have invested in at least two companies in a random sample of 300 investors. [3]

2 Let $X_1, X_2, \ldots, X_n$ be a random sample consisting of independent random variables with mean μ and variance σ^2. Consider the sample mean:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

(i) Derive the expected value of $\bar{X}$. [1]

(ii) Derive the variance of $\bar{X}$. [2]

(iii) Comment on the variance of variable $\bar{X}$ as compared to the variance of X_i. [1]

An actuary is interested in exploring the difference in the size of claim losses from two insurance portfolios, and can take samples of claims from these portfolios.

(iv) Explain how the answer to part (iii) can affect the precision of the actuary's comparison. [2]

[Total 6]

3 The table below shows the annual aggregate claim statistics for three risks over four years. The annual aggregate claim for risk i, in year j, is denoted by X_{ij}.

Risk i	$\bar{X}_i = \frac{1}{4}\sum_{j=1}^{4} X_{ij}$	$s_i^2 = \frac{1}{3}\sum_{j=1}^{4}\left(X_{ij} - \bar{X}_i\right)^2$
1	2,109	3,959,980
2	6,152	7,543,626
3	3,016	3,151,286

(i) Calculate the value of the credibility factor for Empirical Bayes Model 1. [4]

(ii) Comment on how each of the following features of the data affects the value of the credibility factor calculated in part (i):

(a) the number of years of data
(b) the variance of the claim amounts.

[2]

[Total 6]

4 X and Y are discrete random variables with joint distribution as follows:

	$X = 0$	$X = 1$	$X = 3$
$Y = -1$	0.08	0.03	0.00
$Y = 0$	0.03	0.12	0.20
$Y = 3$	0.11	0.11	0.06
$Y = 4.5$	0.04	0.20	0.02

(i) Calculate:

(a) $E(Y \mid X = 1)$
(b) $\mathrm{Var}(X \mid Y = 3)$.

[5]

(ii) Calculate the probability functions of the marginal distributions for X and Y. [2]

(iii) Determine whether X and Y are independent. [2]

[Total 9]

5 An insurance portfolio has a set of n policies ($i = 1, 2, \ldots, n$), for which the company has recorded the number of claims per month, Y_{ij}, for m months ($j = 1, 2, \ldots, m$). It is assumed that the number of claims for each policy, for each month, are independent Poisson random variables with $E[Y_{ij}] = \mu_{ij}$. These random variables are modelled using a simple generalised linear model, with $\log(\mu_{ij}) = \beta_i$, for ($i = 1, 2, \ldots, n$).

(i) Derive the maximum likelihood estimator of β_i. [4]

(ii) Show that the deviance for this model is:

$$D = 2\sum_{i=1}^{n}\sum_{j=1}^{m}\left\{y_{ij}\log\frac{y_{ij}}{\bar{y}_i} - \left(y_{ij} - \bar{y}_i\right)\right\}$$

where $\bar{y}_i$ is the average number of claims per month for policy i:

$$\bar{y}_i = \sum_{j=1}^{m}\frac{y_{ij}}{m}$$

[4]

The company has data for each month over a three-year period. For one policy, the average number of claims per month was 18.95. In the most recent month for this policy, there were seven claims.

(iii) Determine the part of the total deviance that comes from this single observation. [2]

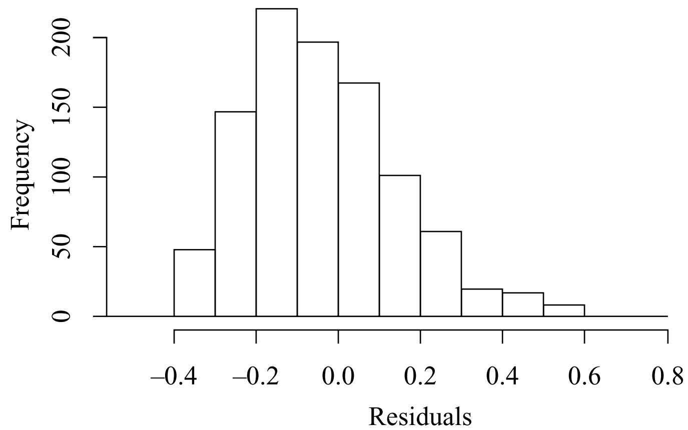
[Total 10]

6 An actuary is asked to check a linear regression calculation performed by a trainee.

The trainee reports a least squares slope parameter estimate of $\hat{b} = 13.7$ and a sample correlation coefficient $r = -0.89$.

(i) Justify why this suggests that the trainee has made an error. [2]

In a different simple linear regression model, a histogram of the residuals is shown below.



(ii) Comment on the validity of the assumptions of the linear model. [2]

The following pairs of data are available:

x	0	1	2	3	4	5	6	7	8	9
y	−1.35	−4.96	− 9.20	−13.15	−16.70	−21.23	−25.14	−28.44	−33.68	−37.39

for which

$$\bar{y} = -19.124, \quad \sum_{i=1}^{10}(y_i - \bar{y})^2 = 1{,}329.523, \quad \sum_{i=1}^{10}(x_i - \bar{x})^2 = 82.5,$$

$$\sum_{i=1}^{10}(x_i - \bar{x})(y_i - \bar{y}) = -331.05$$

A linear model of the form $y = \alpha + \beta x + \varepsilon$ is fitted to the data, where the error terms (ε) independently follow a $N(0, \sigma^2)$ distribution, and where α, β and σ^2 are unknown parameters.

(iii) Determine the fitted line of the regression model. [3]

(iv) Calculate a 95% confidence interval for the predicted mean response if $x = 11$. [5]

(v) Comment on the width of a 95% confidence interval for the predicted mean response if $x = 3.5$, as compared to the width of the interval in part (iv), without calculating the new interval. [2]

[Total 14]

7 An actuary has designed a new product to insure luxury apartments. If there is a claim, her insurance company pays a fixed sum of £1 million per claim. The probability of a claim on a policy in a given year is θ and the probability of more than one claim on a policy in any given year is zero. The actuary's prior beliefs about θ are given by a Beta distribution with parameters $a = 3$ and $b = 5$.

In the first year, the company insured 300 apartments and in the second year it insured $300 + x$ apartments, where x is an integer. In year 1 the total amount of claims was £39 million, while in year 2 it was £60 million.

(i) Show that the posterior distribution of θ is Beta with parameters 102 and $506 + x$. [7]

(ii) Derive the Bayesian estimate of θ in terms of x, under quadratic loss. [2]

(iii) Derive the Bayesian estimate of θ in terms of x, under all-or-nothing loss. [4]

(iv) Justify that, in this case, the Bayesian estimate of θ cannot be the same under quadratic and all-or-nothing loss. [2]

[Total 15]

8 A city is experiencing a high crime rate, particularly burglaries. A sample of 100 streets is taken, and in each of the sampled streets, a sample of six similar houses is taken. The table below shows the number of sampled houses, x, which have had burglaries during the last six months, and the corresponding frequency, f, in terms of number of streets.

Number of houses burgled, x	0	1	2	3	4	5	6
Number of streets, f	39	36	19	4	1	1	0

It is assumed that the number of sampled houses per street that have been burgled during the last six months follows a Binomial distribution, i.e. $X \sim Bin(6, p)$.

(i) State any assumptions needed to justify the use of a Binomial distribution for X. [2]

(ii) Show that the maximum likelihood estimate of p, the probability that a sample house has been burgled during the last six months, is $\hat{p} = 0.1583$. [5]

(iii) Determine the fitted values of the Binomial model using the estimate of p from part (ii). [2]

(iv) Comment on the fit of the model in part (iii), without doing any formal tests. [1]

An insurance company works on the basis that the probability of a house being burgled over a six-month period is 0.13.

(v) Perform a test to investigate whether the Binomial model with this value of p provides a good fit for the data. [7]
[Total 17]

9 An actuary wants to model a particular type of claim size and has been advised to use a Gamma distribution with probability distribution function:

$$f(x, \alpha, \theta) = \frac{x^{\alpha - 1}}{\Gamma(\alpha)\theta^{\alpha}} e^{-\frac{x}{\theta}}, \quad 0 < x < \infty, \quad \alpha > 0, \quad \theta > 0.$$

(i) Show, using moment generating functions, that:

(a) $E(X) = \alpha\theta$

(b) $E(X^2) = \alpha(\alpha + 1)\theta^2$

(c) $E(X^3) = \alpha(\alpha + 1)(\alpha + 2)\theta^3$.

[3]

The shape parameter alpha is assumed to be $\alpha = 4$.

(ii) (a) Determine the variance of the claim size distribution in terms of θ.

(b) Calculate the coefficient of skewness of the claim size distribution, which is defined as:

$$\frac{E\left[\left(X - E(X)\right)^3\right]}{\left\{E\left[\left(X - E(X)\right)^2\right]\right\}^{1.5}}.$$

[4]

Let $X_1, X_2, \ldots, X_n$ be a random sample of n claim sizes for such claims.

(iii) Show that the maximum likelihood estimator (MLE) of θ is given by:

$\hat{\theta} = \frac{\bar{X}}{4}$. [3]

(iv) Show that $\hat{\theta}$ is an unbiased estimator of θ. [1]

A sample of $n = 100$ claim sizes yields $\sum x_i = 796.2$ and $\sum x_i^2 = 8{,}189.4$.

(v) Calculate the MLE of θ. [1]

(vi) (a) Calculate the sample variance.

(b) Compare the result in part (vi)(a) with the variance of the distribution evaluated at $\hat{\theta}$.

[2]

The sample coefficient of skewness is given as 1.12.

(vii) Comment on its comparison with the coefficient of skewness of the distribution, calculated in part (ii)(b). [1]

(viii) Calculate an appropriate 95% confidence interval for θ by using an approximate 95% confidence interval for the mean of the distribution of the claim size. [3]

(ix) (a) Determine the variance of the distribution of θ at both lower and upper limits of the confidence interval calculated in part (viii).

(b) Comment on the result in part (ix)(a) with reference to your answer in part (vi)(a) above.

[2]

[Total 20]

END OF PAPER