

INSTITUTE AND FACULTY OF ACTUARIES



EXAMINATION

3 April 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** The amount of money customers spend in a single trip to the supermarket is modelled using an exponential distribution with mean €15.
- (i) Calculate the probability that a randomly selected customer spends more than €20. [2]
- (ii) Calculate the probability that a randomly selected customer spends more than €20, given that it is known that she spends more than €15. [3]
- [Total 5]
- 2** Consider an estimator of a parameter θ , denoted as $\hat{\theta}$.
- (i) State the definition of the bias of $\hat{\theta}$ by giving an appropriate formula. [1]
- (ii) State the definition of the mean square error of $\hat{\theta}$, denoted as $MSE(\hat{\theta})$, by giving an appropriate formula. [1]
- (iii) Derive an expression for $MSE(\hat{\theta})$ in terms of the variance and the bias of $\hat{\theta}$. [3]
- [Total 5]
- 3** The number of claims on a certain type of policy follows a Poisson distribution with claim rate λ per year. For a group of 200 independent policies of this type, the total number of claims during the last calendar year was 82.
- Determine an approximate 95% confidence interval for the true annual claim rate for this type of policy based on last year's claims. [4]
- 4** Alice and Bob are playing a game of dice. Two fair six-sided dice are rolled. Consider the following events:
- A = 'sum of two dice equals 3'
 B = 'sum of two dice equals 7'
 C = 'at least one of the dice shows a 1'.
- (i) Show that $P(C) = 11/36$. [1]
- (ii) Calculate $P(A|C)$. [2]
- (iii) Calculate $P(B|C)$. [2]
- (iv) Determine whether A and C are independent. [1]
- (v) Determine whether B and C are independent. [1]
- [Total 7]

- 5** (i) State the central limit theorem for independent identically distributed random variables X_1, X_2, \dots, X_n with finite mean μ and finite (non-zero) variance σ^2 . [2]
- (ii) Show that if the random variable B has the binomial distribution with parameters (n, p) , then $\frac{B - np}{\sqrt{np(1-p)}}$ approximately follows a standard normal distribution for large n , using the central limit theorem. [4]

Two players have played a large number of independent games. In a sample of 100 of these games, one player has won 57 games and the other player has won 43.

- (iii) Derive a 95% confidence interval for the probability p that the first player wins a given game, using the normal approximation in part (ii). [4]
[Total 10]

6 Let X and Y be two continuous random variables.

- (i) State the definition of independence of the random variables X and Y in terms of their joint probability density function. [2]

The joint probability density function of X and Y is given by:

$$f_{X,Y}(x,y) = \begin{cases} 8xy & \text{for } 0 < x < y < 1, \\ 0 & \text{otherwise.} \end{cases}$$

- (ii) (a) Determine the marginal density functions of X and Y . [2]
- (b) State whether or not X and Y are independent based on your answer in part (ii)(a). [1]
- (iii) Derive the conditional expectation $E[X | Y = y]$. [3]
[Total 8]

7 Consider a random sample X_1, \dots, X_9 of size 9 from a Normal distribution with expectation 3 and variance 4, that is, $X_i \sim N(3,4)$ for $i = 1, \dots, 9$, and a sample Y_1, \dots, Y_{18} of size 18 from a $N(4,10)$ distribution. Assume that the two samples are independent. Let \bar{X} and \bar{Y} denote the means of the two samples and let S_X^2 and S_Y^2 be the sample variances.

- (i) Calculate the probability for the event that $\bar{X} > 4$. [2]
- (ii) Derive the probability for the event that $\bar{X} > \bar{Y}$. [3]
- (iii) Calculate the probability for the event that $S_X^2 > 4$. [2]
- (iv) Show that the probability for the event $S_X^2 > S_Y^2$ is approximately 0.05. [3]
- (v) Explain whether the exact probability in part (iv) is greater or less than 0.05. [2]

[Total 12]

- 8 The Poisson distribution with mean and variance μ has the following density function:

$$f(y) = \frac{\mu^y e^{-\mu}}{y!}$$

- (i) (a) Show that this probability function can be written in the standard form of the exponential family of distributions, stating the natural and scale parameters, θ and ϕ , and the associated functions of these parameters. [4]
- (b) Verify the mean and variance of the Poisson distribution, using the expression from part (a) together with the properties of the exponential family of distributions. [2]

A wildlife researcher is investigating the national population of a particular species of bear. The researcher believes that the number of bears detected over one year, at each of $i = 1, 2, \dots, 30$ observation points across the country, may follow a Poisson distribution with parameter μ_i . She also believes that μ_i depends on the density of trees, t_i , at the observation point i . The tree density t is measured in ‘trees per square kilometre’.

The researcher specifies the following linear predictor, where α and β are parameters to be estimated:

$$\eta(\mu_i) = \alpha + \beta t_i$$

The researcher then runs a computer model that fits a generalised linear model (using the Poisson canonical link function) to bear-count and tree density data collected from the 30 observation points.

Parameters:

	<i>Estimate</i>	<i>Standard Error</i>
Intercept, α	-1.54520	0.29190
Tree density, β	0.42408	0.09352

- (ii) (a) Explain, using the model output shown above, whether the variable ‘tree density’ is significant or not. [3]
- (b) Estimate, using the fitted model, the expected number of bears that would be detected over a one-year period in a woodland area with a tree density of 12. [3]
- [Total 12]

- 9 Consider a sample of 1,000 motor insurance policies. We assume that the annual total claim amounts per policy are independent and identically distributed. We denote by X the number of policies with a total amount of over £5,000 claimed in a calendar year, and assume that X has a Binomial distribution, $X \sim \text{Bin}(p, 1,000)$, with expectation $E[X] = 1,000p$.

An analyst wishes to estimate the unknown proportion p of claims with amount greater than £5,000 per year.

- (i) Derive the maximum likelihood estimator for p . [3]

Suppose now that the analyst has some prior knowledge about p and assumes a Beta

prior distribution with density function $f(p) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}$.

- (ii) Derive the density of the Bayesian posterior distribution of p in terms of n , X , α and β . [2]
- (iii) State the type of the posterior distribution of p with its parameters. [2]
- (iv) Comment on the relationship between the prior distribution and the posterior distribution of p in this context. [2]

Assume that 50 policies out of 1,000 policies in an actual sample have a total claim amount of over £5,000.

- (v) Estimate p using the MLE in (i). [1]
- (vi) Estimate p using the Bayesian estimator under quadratic loss, based on the posterior distribution derived in parts (ii) and (iii). Assume that the parameters of the prior distribution are $\alpha = 2$ and $\beta = 2$. [3]
- (vii) Comment on the difference between the values estimated in (v) and (vi). [1]
- (viii) State the Bayesian estimator from part (vi) in the form of a credibility interval, determining the credibility factor. [3]

[Total 17]

- 10** A professional body wishes to analyse the performance of its students on a particular two-part examination. It records the scores obtained by a sample of 12 students on the first part of the exam, and the scores obtained by the same students on the second part of the exam. The results are as follows:

Student	A	B	C	D	E	F	G	H	I	J	K	L
First-part exam score x (%)	82	49	73	60	61	77	65	85	91	53	59	73
Second-part exam score y (%)	76	58	75	66	70	71	76	92	87	59	63	71

$$\sum x = 828 \quad \sum y = 864 \quad \sum x^2 = 59,054 \quad \sum y^2 = 63,362 \quad \sum (x - \bar{x})(y - \bar{y}) = 1,334$$

- (i) Calculate the fitted linear regression equation of y on x . [3]
- (ii) Assuming the full Normal model:
- (a) Calculate an estimate for the error variance σ^2 . [2]
- (b) Determine the 90% confidence interval for σ^2 . [2]
- (iii) Test whether the data are positively correlated, by considering the slope parameter. [4]
- (iv) Calculate a 95% confidence interval for the mean second-part exam score corresponding to an individual first-part exam score of 57. [3]
- (v) Test whether these data could come from a population with a correlation coefficient equal to 0.75. [4]
- (vi) Calculate the proportion of variation explained by the model. [1]
- (vii) Comment on the fit of the model, using your answer in part (vi). [1]
- [Total 20]

END OF PAPER