INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

15 April 2021 (am)

Subject CS1 – Actuarial Statistics
Core Principles

Paper A

Time allowed: Three hours and fifteen minutes

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

If you encounter any issues during the examination please contact the Assessment Team on T. 0044 (0) 1865 268 873.

1 A random variable, X, is modelled using a gamma distribution with parameters $\alpha = 50$ and $\lambda = 0.25$.

(i) Calculate an approximate value for $P(X > 270)$ using the chi-square distribution. [2]

(ii) Calculate an approximate value for $P(X > 270)$ using the central limit theorem. [4]

(iii) Comment on the difference between your answers to parts (i) and (ii). [2]

[Total 8]

2 Consider two random variables, X and Y. The conditional expectation and conditional variance of Y given X are denoted by the two random variables U and V, respectively; that is, $U = E[Y|X]$ and $V = \text{Var}[Y|X]$.

Assume that Y is Normally distributed with expectation 5 and variance 4. Also assume that the expectation of V is 2.

(i) Calculate the expected value of U. [1]

(ii) Calculate the variance of U. [2]

[Total 3]

3 Consider two random variables, X and Y, with a uniform distribution on the interval [0,1]; that is, $X \sim U(0,1)$ and $Y \sim U(0,1)$. Assume that X and Y are independent.

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

A $\frac{1}{t}(e^{-t} - 1)$ for $t \neq 0$

B $\frac{1}{t}(e^{t} - 1)$ for $t \neq 0$

C $\frac{1}{t}(1 - e^{-t})$ for $t \neq 0$

D $\frac{1}{t}(1 - e^{t})$ for $t \neq 0$

[2]

(ii) Derive the value of the moment generating function $M_X(t)$ of X at $t = 0$. [1]

An analyst argues that the sum of X and Y must have a uniform distribution on the interval [0,2] since both X and Y are uniformly distributed on [0,1].

(iii) Derive the moment generating function for the random variable Z with a $U(0,2)$ distribution. [2]

(iv) Comment on the analyst's argument by determining if the random variable $Z = X + Y$ has a uniform distribution on [0,2] using moment generating functions. [3]

[Total 8]

4 Consider a random sample of size $n = 25$ from a Normal distribution with mean 10, variance 4 and sample variance S^2.

(i) Write down the sampling distribution of S^2. [2]

(ii) Calculate, using your answer in part (i), the expected value of S^2. [1]

(iii) Calculate, using your answer in part (i), the variance of S^2. [1]

[Total 4]

5 The joint probability density function of random variables X and Y is:

$$f(x,y) = \begin{cases} ke^{-(x+2y)}, & x > 0, y > 0 \\ 0, & \text{otherwise} \end{cases}$$

[**Hint:** You may find it helpful to define the functions $g_X(x) = e^{-x}$ and $g_Y(y) = e^{-2y}$, using this notation in your answers.]

(i) Demonstrate that X and Y are independent. [1]

(ii) Verify that $k = 2$. [3]

(iii) Demonstrate that $f_Y(y)$, the marginal density function of Y, is:

$$2e^{-2y} \text{ for } y > 0.$$

[1]

(iv) Demonstrate that the conditional density function $f(y|Y > 3)$ is:

$$f(y|Y > 3) = 2e^{6-2y} \text{ for } y > 3.$$

[**Hint:** Consider $P(Y \leq y|Y > 3)$.] [3]

(v) Identify which **one** of the following expressions is equal to the conditional expectation $E[Y|Y > 3]$:

A $\int_0^\infty te^{-2t}dt + \int_0^\infty 3e^{-2t}dy$

B $\int_0^\infty te^{-2t}dt + \int_0^\infty 6e^{-2t}dy$

C $\int_0^\infty 2te^{-2t}dt + \int_0^\infty 3e^{-2t}dy$

D $\int_0^\infty 2te^{-2t}dt + \int_0^\infty 6e^{-2t}dy$

[1]

(vi) Determine the value of the conditional expectation $E[Y|Y > 3]$. [2]

(vii) Identify which **one** of the following options is the conditional expectation $E[Y^2|Y > 3]$:

A 12.5
B 13.5
C 14.5
D 15.5.

[2]

(viii) Determine the conditional variance $\text{Var}[Y|Y > 3]$. [1]

[Total 14]

6 A tutor believes that the number of exams passed by students sitting three different exams follows a binomial distribution with parameters $n = 3$ and p. A random sample of 120 students showed the following results:

Number of exams passed	0	1	2	3
Number of students	40	60	15	5

(i) (a) Identify which **one** of the following corresponds to the log likelihood function of p given the observed data:

A $\log L \propto 255 \log(1-p) + 105\log(p)$
B $\log L \propto 115 \log(1-p) + 80\log(p)$
C $\log L \propto 265 \log(1-p) + 115\log(p)$
D $\log L \propto 175 \log(1-p) + 85\log(p)$

[2]

(b) Show, using your answer to part (i)(a), that the maximum likelihood estimate for p is $\hat{p} = 0.2917$. You are **not** required to check that it is a maximum. [3]

(ii) Perform a goodness of fit test for the binomial model Bin(3, p) at a significance level of 5%. [8]

[Total 13]

7 A telecommunications company has performed a small empirical study comparing phone usage in rural and urban areas, collecting data from a total of 35 people who use their phones independently. The average number of hours that each person spent using their phone during a week is denoted by Y.

In the following table, $\overline{Y}$, denotes the sample mean of Y in rural and urban areas, and S_Y denotes the sample standard deviations; that is, $S_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$.

	Sample size n	$\overline{Y}$	S_Y
Rural areas	15	3.7	2.1
Urban areas	20	4.4	1.9

A statistical test is to be performed, at the 5% significance level, to determine whether the null hypothesis that mean phone usage in rural areas is the same as mean phone usage in urban areas, i.e. for:

H_0: phone usage is equal versus H_1: phone usage is not equal.

(i) State a suitable distribution for the test statistic with its parameter(s). [1]

(ii) Justify any assumption(s) required to perform this test. [2]

(iii) Identify which **one** of the following options gives the correct value of the test statistic for this test:

A –1.031
B –0.519
C –3.019
D –1.455.

[2]

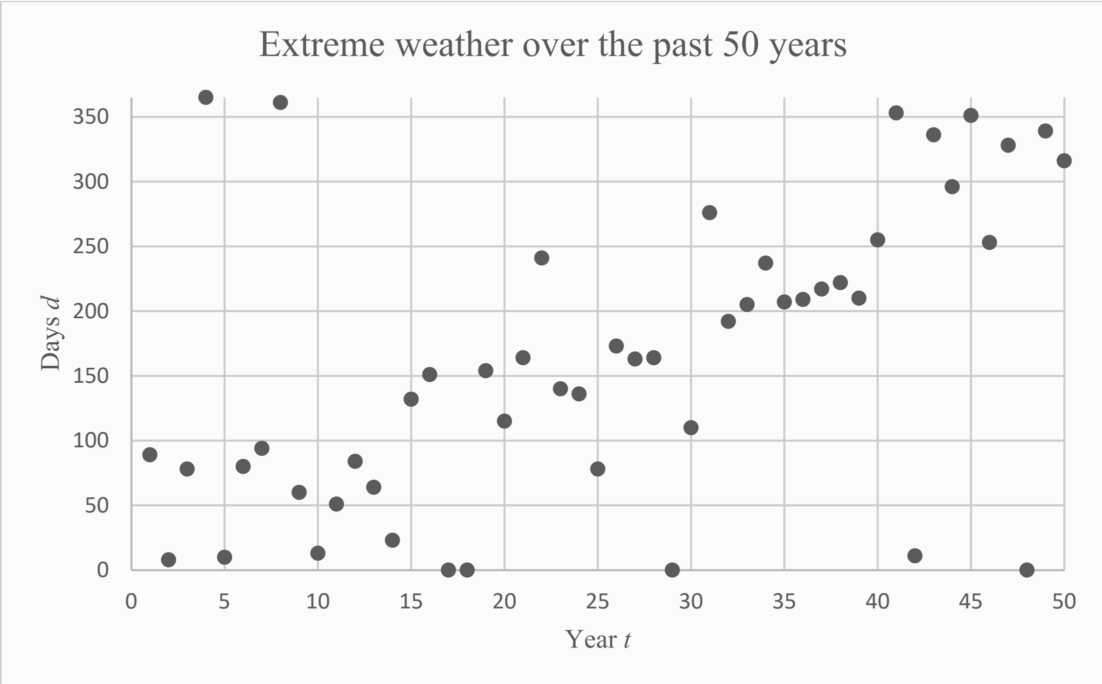
(iv) Write down the conclusion of the test including the relevant critical value(s) from the Actuarial Formulae and Tables. [3]

(v) Determine a 95% confidence interval for the mean phone usage (hours per week) for rural areas, stating any assumption(s) you make. [4]

[Total 12]

8 An initial investigation into climate change has been conducted using climate change data from the past 50 years, collected by the International Meteorological Society. For each year, t, the number of consecutive days, d, of extreme weather was recorded. The total number of days in any year is 365 and extreme weather is defined as a rainless day with temperatures in excess of 28 degrees Celsius.

An Actuary has performed a preliminary statistical analysis on the data. Below is a scatter plot of the Actuary's findings:



The Actuary also fitted a least squares regression line for extreme weather days on year, giving:

$$\hat{d} = 147.39 - 5.82601t,$$

and calculated the coefficient of determination for this regression line as:

$$R^2 = 91.5\% .$$

(i) Comment on the plot and the Actuary's analysis. [2]

A separate analysis, on the same data, is undertaken independently by a statistician. Below are the key summaries of their analysis:

$$\sum t = 1{,}275 \quad \sum t^2 = 42{,}925 \quad \sum d = 8{,}502 \quad \sum d^2 = 1{,}911{,}378 \quad \sum td = 282{,}724$$

(ii) Verify that the equation of the statistician's least squares fitted regression line of extreme weather days on year is given by:

$$\hat{d} = 8.59592 + 6.33114t.$$

[3]

(iii) (a) Determine the standard error of the estimated slope coefficient in part (ii).

(b) Test the null hypothesis of ‘no linear relationship’ at the 1% confidence level, using the equation in part (ii).

(c) Determine a 99% confidence interval for the underlying slope coefficient for the linear model, using the equation in part (ii). [7]

Further climate change data are collected from an alternative independent data source, also covering the past 50 years. These data were analysed and resulted in an estimated slope coefficient of:

$$\hat{\beta} = 5.21456 \text{ with standard error } 1.98276$$

(iv) (a) Test the ‘no linear relationship’ hypothesis at the 1% confidence level based on the further climate change data. [2]

(b) Determine a 99% confidence interval for the underlying slope coefficient β based on the alternative climate change data. [2]

(v) Comment on whether or not the underlying slope coefficients, for the statistician’s data in part (ii) and the independent data in part (iv), can be regarded as being equal. [3]

(vi) Discuss why the results of the tests in parts (iii)(b) and (iv)(a) seem to contradict the conclusion in part (v). [4]

[Total 23]

9 The number of claims received by a motor insurance company on any given day follows a Poisson distribution with mean u. Prior beliefs about u are expressed through a gamma distribution with parameters a and b. Over a period of n days the observed number of claims received per day are $x_1, x_2, \ldots, x_n$.

(i) Identify which **one** of the following is the posterior density of u:

A $f(u|x) \propto u^{b + \Sigma x_i - 1} e^{-(a+n)u}$

B $f(u|x) \propto u^{a + \Sigma x_i} e^{-(b+n+1)u}$

C $f(u|x) \propto u^{a + \Sigma x_i - 1} e^{-(b+n)u}$

D $f(u|x) \propto u^{b + \Sigma x_i + 1} e^{-(a+n-1)u}$

[3]

(ii) Write down the posterior density of the parameter u and specify its parameters. [2]

(iii) (a) Determine the Bayesian estimate of u under quadratic loss. [2]

(b) Write down the Bayesian estimate of u under quadratic loss as a credibility estimate and state the credibility factor. [2]

Suppose that $a = 9$, $b = 3$ and that the company receives 320 claims in total during a 6-day period.

(iv) Calculate the Bayesian estimate of u under quadratic loss. [2]

(v) Calculate the variance of the posterior distribution of u. [2]

An industry expert suggests that prior beliefs about u are better expressed through a gamma distribution with parameters $a = 18$ and $b = 6$.

(vi) Explain how these prior beliefs would affect the variance of the posterior distribution of u, without explicitly calculating the variance of the posterior distribution. [2]

[Total 15]

END OF PAPER