INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

17 September 2021 (am)

Subject CS1 – Actuarial Statistics Core Principles

Paper A

Time allowed: Three hours and twenty 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 sample of size 15 is taken from a Normal distribution with mean 19 and variance 2.

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

(ii) Explain why your answer in part (i) is valid for this random sample. [1]

[Total 3]

2 A statistician wants to simulate values from certain distributions and has available only a random number generator that yields independent samples from a $N(0,1)$ distribution.

(i) Describe an algorithm to simulate random numbers from a t distribution with 1 degree of freedom. [3]

(ii) Describe an algorithm to simulate random numbers from a gamma distribution with parameters $\frac{3}{2}$ and $\frac{1}{2}$. [3]

(iii) Describe an algorithm to simulate random numbers from an F distribution with (1,1) degrees of freedom. [2]

[Total 8]

3 The random variable X follows a distribution with mean $E[X] = \frac{b}{a-1}$ and variance $\text{Var}[X] = \frac{ab^2}{(a-1)^2(a-2)}$ where $a = 4$ and $b = 6$ are the parameters of the distribution.

Y is a random variable such that

$E(Y \mid X = x) = 3x + 6$

and

$\text{Var}(Y \mid X = x) = x^2 + 4$

Calculate the unconditional standard deviation of Y.

[6]

4 The number of pizzas ordered in a restaurant each day follows a Poisson distribution with unknown mean m. The prior distribution for m follows a gamma distribution with mean 35 and standard deviation 5. The restaurant receives 135 pizza orders over 7 days.

(i) Write down an expression of the prior probability density function for m leaving out any coefficient of proportionality. [3]

(ii) Identify which **one** of the following expressions gives the correct posterior probability density function for m.

A $f_{\text{posterior}}(m) \propto m^{135} e^{-7m}$

B $f_{\text{posterior}}(m) \propto m^{183} e^{-7.7m}$

C $f_{\text{posterior}}(m) \propto m^{184} e^{-8.4m}$

D $f_{\text{posterior}}(m) \propto m^{183} e^{-8.4m}$

[3]

(iii) Calculate a point estimate for the number of pizzas ordered each day, using Bayesian estimation under all-or-nothing loss. [4]

(iv) Calculate a point estimate for the number of pizzas ordered each day, using Bayesian estimation under squared-error loss. [2]

[Total 12]

5 The probability that a claim is made on a car insurance policy in a particular year is 0.06. The policies are assumed to be independent among them. 500 of these policies are selected at random.

(i) Calculate the probability that no more than 40 of these policies will result in a claim during the year, stating any approximations you make. [5]

Past data from the insurer indicate that the standard deviation of claim amounts is £75. The insurer wishes to construct a 95% confidence interval for the mean claim amount, with an interval width of £10.

(ii) Calculate the sample size needed to achieve this level of accuracy for a 95% confidence interval. [4]

[Total 9]

6 Consider independent observations $y_1, y_2, \ldots, y_n$ of a random variable Y with probability density function

$$f(y) = 2cy\exp(-cy^2), \quad y > 0,$$

where $c > 0$ is an unknown parameter. Let $F(y)$ denote the cumulative distribution function (CDF) of Y.

(i) Identify which **one** of the following expressions gives the inverse of the CDF of Y:

A $\quad y = \left\{-\frac{1}{c}\log(1 - F(y))\right\}$

B $\quad y = 1 - \left\{-\frac{1}{c}\log(1 - F(y))\right\}$

C $\quad y = \left\{-\frac{1}{c}\log(1 - F(y))\right\}^{1/2}$

D $\quad y = 1 - \left\{-\frac{1}{c}\log(1 - F(y))\right\}^{1/2}$

[2]

(ii) Determine how values of this random variable can be generated using the inverse transform method. [2]

A gamma prior distribution is assumed for c with parameters a and b.

(iii) Identify which **one** of the following expressions is correct for the posterior density of parameter c:

A $\quad p(c \mid y) \propto c^{n+a-1}\exp\left\{-\left(ab + \sum_{i=1}^{n} y_i^2\right)c\right\}$

B $\quad p(c \mid y) \propto c^{n+a-1}\exp\left\{-\left(b + \sum_{i=1}^{n} y_i^2\right)c\right\}$

C $\quad p(c \mid y) \propto c^{n+a}\exp\left\{-\left(ab + \sum_{i=1}^{n} y_i^2\right)c\right\}$

D $\quad p(c \mid y) \propto c^{n+a}\exp\left\{-\left(b + \sum_{i=1}^{n} y_i^2\right)c\right\}$

[2]

(iv) Determine the posterior distribution of parameter c with all relevant parameters. [2]

[Total 8]

7 Let X_i, $i = 1, 2, \ldots, n$ be independent random variables, each following an exponential distribution with parameter b. We consider the random variable $Y = \sum_{i=1}^{n} X_i$.

(i) Justify why $M_Y(t)$, the moment generating function (MGF) of variable Y, is given by

$$M_Y(t) = \left(1 - {}^{t}\!/_{b}\right)^{-n} \qquad [2]$$

Let Z be a random variable such that the MGF of Z is $M_z(t) = \sqrt{M_Y(t)}$.

(ii) Determine the value of b for which Z follows a chi-square distribution, specifying the degrees of freedom of the chi-square distribution. [3]

[Total 5]

8 The number of hospital admissions for respiratory conditions in a big city was recorded over 150 days. The level of the concentration of a certain pollutant was also recorded ('low', 'medium', 'high'), together with the mean temperature (in degrees Celsius) on the day. Part of the data is shown below.

Day	*Temperature* (X_1)	*Pollutant concentration* (X_2)	*Hospital admissions* (Y)
1	10	Low	26
2	8	Low	37
.	.	.	.
.	.	.	.
.	.	.	.
50	12	Low	32
51	7	Medium	31
.	.	.	.
.	.	.	.
.	.	.	.
120	3	Medium	28
121	5	High	35
.	.	.	.
.	.	.	.
.	.	.	.
150	6	High	31

A generalised linear model is to be fitted to investigate the dependence of the number of hospital admissions on mean temperature and pollutant concentration.

(i) Write down a suitable model for the number of hospital admissions. [3]

(ii) Justify the inclusion of the terms that you have used in the linear predictor in part (i). [2]

A statistician fitted a GLM, and obtained the following summary:

Coefficients:	*Estimate*	*Std. error*	*z value*	*Pr(>|z|)*
(Intercept)	–0.372	0.053	–6.916	4.66e-12 ***
X_1	0.090	0.015	5.676	1.38e-08 ***
X_2 Medium	–0.100	0.080	–1.244	0.213570
X_2 High	0.298	0.082	3.614	0.000301 ***
$X_1 : X_2$ Medium	0.036	0.023	1.551	0.120933
$X_1 : X_2$ High	–0.076	0.028	–2.705	0.006825 **

Suppose that, on a different day, the pollutant concentration is High and the mean temperature is 19 degrees Celsius.

(iii) Write down the linear function of the parameters the statistician should use in constructing a predictor of the number of hospital admissions on that day. [1]

(iv) Explain why estimates for X_2 Low and $X_1 : X_2$ Low are not shown in the summary of the results above. [1]

(v) Comment on the impact of the pollutant concentration on the number of hospital admissions, based on the summary of results above. [2]

[Total 9]

9 An actuarial analyst working in an investment bank believes that a firm's first year percentage return (y) depends on its revenues (x). The table below provides a summary of x, y and the natural logarithmic revenue (z) for 110 firms.

	Mean	*Median*	*Sample standard deviation*	*Minimum*	*Maximum*
y	0.106	–0.130	0.824	–0.938	4.333
x (£ million)	134.487	39.971	261.881	0.099	1455.761
$z = \log(x)$	3.686	3.688	1.698	–2.316	7.283

The analyst determined that the correlation between y and x is −0.0175 and that the linear regression line of the return on the revenue is

$$\hat{y} = \hat{a} + \hat{b}x.$$

(i) (a) Identify which **one** of the following options gives the correct values of the coefficient estimates $\hat{a}$ and $\hat{b}$:

A $\hat{a} = 0.113$ and $\hat{b} = -5.506 \times 10^{-5}$

B $\hat{a} = -5.506 \times 10^{-5}$ and $\hat{b} = 0.113$

C $\hat{a} = 748.1227$ and $\hat{b} = -5.562$

D $\hat{a} = -5.562$ and $\hat{b} = 748.1227$

(b) Calculate the fitted return for a firm with revenue 95.55.

[3]

The analyst estimated the regression using the logarithm revenues (z) and y as

$$\hat{y} = 0.438 - 0.090z$$

(ii) (a) Calculate the fitted return for the firm with revenue 95.55 (£ million) using the regression model with the logarithmic revenues.

(b) Comment on the result in parts (ii)(a) and (i)(b).

(c) Calculate the value of the sum S_{zy}.

[3]

(iii) Perform a statistical test at the 10% significance level to determine if the logarithmic revenues significantly affect the percentage returns. [5]

The analyst speculated that, other things being equal, firms with greater revenues will be more stable and thus enjoy a larger return. They considered the null hypothesis of no relation between z and y.

(iv) Perform a statistical test at the 10% significance level to determine whether the analyst's speculation is correct. Your answer should include the hypotheses of the test. [3]

(v) Calculate Pearson's correlation coefficient between z and y. [1]

A client is considering investing in a firm that has $z = 2$.

(vi) (a) Calculate the client's predicted first year percentage return.

(b) Calculate an approximate 95% confidence interval corresponding to the predicted percentage return in part (vi)(a).

[4]

A firm in the data has logarithmic revenue $z = 1.76$ and the highest first year percentage return $y = 4.333$.

(vii) (a) Calculate the residual for this observation.

(b) Comment on the observed data for this firm using part (vii)(a).

[3]

[Total 22]

10 Total yearly aggregate claims in a particular company are modelled as a random variable X, where X is assumed to follow a Normal distribution with unknown mean μ and variance $\sigma^2 = 12{,}000^2$. Aggregate claims from the last 5 years are as follows:

146,000	142,000	153,000	127,000	132,000

An analyst wishes to estimate the unknown parameter μ.

(i) Identify which **one** of the following gives the correct expression of the derivative of the log-likelihood function:

A $\frac{dl(\mu)}{d\mu} = -\sum_{i=1}^{n}(x_i - \mu)$

B $\frac{dl(\mu)}{d\mu} = \sum_{i=1}^{n}(x_i - \mu)$

C $\frac{dl(\mu)}{d\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)$

D $\frac{dl(\mu)}{d\mu} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)$

[2]

(ii) Calculate the maximum likelihood estimate for μ, using your answer to part (i). [1]

(iii) Calculate a 95% confidence interval for μ. [4]

The analyst assumes a Normal prior distribution for μ with density function

$$f(\mu) \propto e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}}, \quad \mu_0 > 0 \text{ and } \sigma_0 > 0.$$

For such a prior, the analyst derives the posterior distribution for μ as

$$p(\mu|\underline{x}) \propto \exp\left(-\frac{1}{2}(n\tau + \tau_0)\left(\mu - \frac{n\tau\overline{x} + \tau_0\mu_0}{n\tau + \tau_0}\right)^2\right)$$

where $\tau = \frac{1}{\sigma^2}$ and $\tau_0 = \frac{1}{\sigma_0^2}$.

Prior information about μ suggests that $\mu_0 = 150{,}000$ and $\sigma_0^2 = 10{,}204.08^2$.

(iv) Write down the distribution corresponding to the density $p(\mu|\underline{x})$ above, with all its parameters values. [2]

(v) Comment on the relationship between the prior distribution and the posterior distribution of μ. [1]

(vi) Calculate the value of the Bayesian credibility estimate for μ under quadratic loss. [2]

(vii) Calculate an approximate 95% Bayesian interval for μ, based on its posterior distribution. [2]

(viii) Comment on the intervals estimated in parts (iii) and (vii). [1]

Another analyst assumes a Uniform prior distribution for μ with mean $\mu_0 = 150{,}000$ and variance $\sigma_0^2 = 10{,}204.08^2$.

(ix) Identify which **one** of the following gives the correct expression of the posterior distribution for μ:

A $\quad p(\mu|\underline{x}) \propto \left(\frac{\mu-\mu_0}{\sigma_0^2}\right)\exp\left(-\frac{n}{2\sigma^2}(\mu-\bar{x})^2\right)$

B $\quad p(\mu|\underline{x}) \propto \exp\left(-\frac{n}{2\sigma^2}(\mu-\bar{x})^2\right)$

C $\quad p(\mu|\underline{x}) \propto \exp\left(-\frac{1}{2}\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right)\left(\mu-\frac{\frac{n\bar{x}}{\sigma^2}+\frac{\mu_0}{\sigma_0^2}}{\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}}\right)^2\right)$

D $\quad p(\mu|\underline{x}) \propto (\mu-\mu_0)^2\exp\left(-\frac{n}{2\sigma^2}(\mu-\bar{x})^2\right)$

[3]
[Total 18]

END OF PAPER