INSTITUTE AND FACULTY OF ACTUARIES

EXAMINERS' REPORT

September 2018

Subject CT6 – Statistical Methods
Core Technical

Introduction

The Examiners' Report is written by the Principal Examiner with the aim of helping candidates, both those who are sitting the examination for the first time and using past papers as a revision aid and also those who have previously failed the subject.

The Examiners are charged by Council with examining the published syllabus. The Examiners have access to the Core Reading, which is designed to interpret the syllabus, and will generally base questions around it but are not required to examine the content of Core Reading specifically or exclusively.

For numerical questions the Examiners' preferred approach to the solution is reproduced in this report; other valid approaches are given appropriate credit. For essay-style questions, particularly the open-ended questions in the later subjects, the report may contain more points than the Examiners will expect from a solution that scores full marks.

The report is written based on the legislative and regulatory context pertaining to the date that the examination was set. Candidates should take into account the possibility that circumstances may have changed if using these reports for revision.

Mike Hammer
Chair of the Board of Examiners
December 2018

A. General comments on the *aims of this subject and how it is marked*

1. The aim of the Statistical Methods subject is to provide a further grounding in mathematical and statistical techniques of particular relevance to financial work.

2. Errors carried over normally only lose credit the first time they appear.

3. Generally arithmetic errors are not treated as harshly as method errors.

4. Markers exercise judgement when answers are partly correct and can award partial marks if appropriate. In particular, where a candidate has not used the method in the marking schedule, but has shown some understanding by their working, some credit is given.

5. Errors just due to rounding do not lose marks unless the rounding is excessive (e.g. rounding an interim step to just 2 sig fig, say) and significantly compromises accuracy.

B. General comments on *student performance in this diet of the examination*

Student performance was strong in this diet, with many candidates able to score well even on the harder questions. Some of the candidates who performed poorly may have improved their scores by taking more care in reading the question and planning their answers.

C. Pass Mark

The Pass Mark for this exam was 60.

Solutions

Q1

For a given λ, we know that one can sample from *X* as

$$X = -\frac{1}{\lambda}\log U \qquad U \sim U(0,1)$$ [2]

Algorithm:

1. Generate $U_1 \sim U(0,1)$ [½]
2. $$\lambda = \begin{cases} 1 & U_1 \in (0, 0.3] \\ 2 & U_1 \in (0.3, 0.6] \\ 3 & U_1 \in (0.6, 1] \end{cases}$$ [1]

3. Generate $U_2 \sim U(0,1)$

[1]

4. $X = -\frac{1}{\lambda}\log U_2$ [½]

[Total 5]

This question was generally well answered, but a disappointing number of candidates used the probability density function, rather than the cumulative density function.

Q2

(i) The distribution of the response variable. [1]
A linear predictor as a function of the covariates. [1]
A link function between the response variable and the linear predictor. [1]

(ii)
$$f(y,\mu) = \exp\left[n\left(y\ln\mu + (1-y)\ln(1-\mu)\right) + \ln\binom{n}{ny}\right]$$
[1]
$$= \exp\left[n\left(y\ln\left(\frac{\mu}{1-\mu}\right) + \ln(1-\mu)\right) + \ln\binom{n}{ny}\right]$$

Hence

$$\theta = \ln\left(\frac{\mu}{1-\mu}\right)$$ [1]

$\varphi = n$ [½]

$$a(\varphi) = \frac{1}{\varphi}$$ [½]

$$b(\theta) = \ln\left(1+e^{\theta}\right)$$ [½]

$$c(y,\varphi) = \ln\binom{\varphi}{\varphi y}$$ [½]

(iii) Link function: $g(\mu) = \ln\left(\frac{\mu}{1-\mu}\right)$ [1]

[Total 8]

Well prepared candidates were able to score highly here, although many candidates gave insufficiently detailed answers in part (i), and didn't give expressions in part (ii) as a function of the correct variable e.g. leaving b(theta) as a function of mu.

Q3

(i) Under Bayesian estimation the parameter being estimated is itself considered to be a random variable. [2]

(ii) Quadratic loss function – the mean of the posterior distribution [1]
Absolute error function – the median of the posterior distribution [1]
All-or-nothing loss function – the mode of the posterior distribution [1]

(iii) A suitable prior is $U(0,1)$, so $f_\theta(\theta) = 1$

Distribution of $X|\ \theta$ is Bin(500, θ), so the likelihood function is:

$L(\theta) = {}^{500}C_{326}\theta^{326}(1-\theta)^{174}$ [2]

Apply Bayes, PDF of posterior is proportional to $\theta^{326}(1-\theta)^{174}$ [1]

Hence the distribution of $\theta|X$ is beta(327,175), and the estimate of θ under quadratic loss is 327/(327+175) = 65.1%. [1]

[4]

[Total 9]

Again many candidates were able to score well here. A few candidates tripped up by assuming the posterior was binomial rather than beta.

Q4

(i) $P(X < M) = 1 - \left(\frac{100}{100+M}\right)^3 = 0.9$ [1½]

$\left(\frac{100}{100+M}\right) = 0.1^{1/3} \Rightarrow M = \frac{100 - 100 * 0.1^{1/3}}{0.1^{1/3}} = 115.4$ [2½]

(ii) Let Y be the claim amount paid by the reinsurer, so that

$$Y = \begin{cases} 0 & X <= M \\ X - M & X > M \end{cases}$$

$E(Y \mid X > M) = \frac{E(Z)}{P(X > M)}$ [1]

$E(Z) = \int_M^\infty (x - M) f(x) dx = \int_M^\infty (x - M) \frac{3 * 100^3}{(100 + x)^4} dx$ [1]

$u = (x - M); \frac{dv}{dx} = \frac{3 * 100^3}{(100 + x)^4}$ [1]

$$E(Z)=\left[-(x-M)\frac{100^3}{(100+x)^3}\right]_M^{\infty}+\int_M^{\infty}\frac{100^3}{(100+x)^3}dx$$ [1]

$$=0+\left[\frac{-100^3}{2(100+x)^2}\right]_M^{\infty}=\left(\frac{100^3}{2(100+M)^2}\right)=10.772$$ [1]

$$E(Y\mid X>M)=\frac{10.772}{0.1}=107.7$$ [1]

[Total 10]

Many candidates got full marks on part (i), but only the strongest candidates were able to score well on part (ii).

Q5

(i) df1 = 534 / 528 = 1.011 364 … [1]
df2 = (528+541)/(469+525) = 1.075 453 … [½]
df3 = (469 + 525 + 558) / (435 + 485 + 509) = 1.086 074 … [½]

Accident Year	Development Year			
	0	1	2	3
2014	435	469	528	534
2015	485	525	541	547.15
2016	509	558		606.92
2017	544			642.62

[2]

(ii) Cost per claim:

Accident Year	Development Year			
	0	1	2	3
2014	8.982	11.953	11.479	11.507
2015	9.961	13.941	13.808	
2016	11.870	14.842		
2017	12.980			

[2]

Grossing up table:

Accident Year	Development Year			
	0	1	2	3
2014	78.050%	103.872%	99.754%	100%
2015	71.962%	100.716%	99.754%	
2016	81.811%	102.294%		
2017	77.274%			

[2]

Ultimate average cost per claim:

Accident Year	Development Year			
	0	1	2	3
2014	8.982	11.953	11.479	11.507
2015	9.961	13.941	13.808	13.842
2016	11.870	14.842		14.509
2017	12.980			16.797

[1]

Ultimate total claim amounts:

Accident Year	Development Year			
	0	1	2	3
2014	3,907	5,606	6,061	6,145
2015	4,831	7,319	7,470	7,574
2016	6,042	8,282		8,806
2017	7,061			10,794

[1]

Outstanding = (6145 +7574 + 8806 + 10794) – 19544 = 13,775 [1]

[Total 11]

This question posed few problems for well-prepared candidates.

Q6

(i) For Player 1, Strategy VI dominates strategies IV & V [1]
For Player 2, Strategy C dominates Strategy F … [½]
… and Strategies D & E with IV & V eliminated [1]
With D, E & F eliminated, Strategy III dominates Strategy II [½]

(ii) There are no saddle points, since the maximum in any column is never equal to the minimum of any row. [2]

(iii) 'b' is always dominated for α. [½]

For β, we require $18p + 31(1 - p) < 22$ [1]
For γ, we require $23p + 19(1 - p) < 22$ [½]
So $9/13 < p < 3/4$ [1]

(iv) Now α is also dominated for Player 1, so we are left with

	a	c
β	18	31
γ	23	19

[1]

Hence Player 2 should choose a randomised strategy with
$18p + 31(1 - p) = 23p + 19(1 - p)$ [1]
$p = 12/17$ (i.e. choose strategy a with probability 12/17) [½]
with value 21.8 [½]

[Total 11]

Performance in this question was mixed, with many candidates dropping marks in part (i) for insufficient explanation of their answer; and fewer candidates were able to score well in parts (iii) & (iv).

Q7

(i) The adjustment coefficient is the unique positive solution to

$\lambda M_X(R) - \lambda - \lambda(1 + \theta)\, E(X)\, R = 0$ [1]

(ii) Cancelling the λ terms we have

$M_X(R) = E(e^{RX}) = 1 + (1 + \theta)\, E(X)R$ [½]

$$E\left(1 + RX + \frac{R^2X^2}{2} + \ldots\right) = 1 + (1 + \theta)\, E(X)R$$ [½]

And truncating the expression we get

$E(1 + RX + R^2X^2/2) = 1 + (1 + \theta)\, E(X)R$ [½]

i.e. $1 + Rm_1 + R^2m_2/2 = 1 + (1 + \theta)\, m_1R$ [½]

i.e. $R^2m_2 = 2\theta m_1R$ [½]

i.e. $R = \dfrac{2\theta m_1}{m_2}$ [½]

(iii) $1 + (1 + \theta)\, E(X)R = M_X(R)$

$$1 + (1+\theta)R/\gamma = \left(1 - R/\gamma\right)^{-1}$$

$$\left[1 + (1+\theta)R/\gamma\right]\left(1 - R/\gamma\right) = 1$$

$$\frac{1+\theta}{\gamma^2}R^2 - \theta R/\gamma = 0$$

$$R/\gamma\left((1+\theta)R/\gamma - \theta\right) = 0$$

[2]

Rejecting $R = 0$; we therefore have

$$R = \frac{\theta\gamma}{(1+\theta)}$$ [1]

(iv) m_1 is 200, $m_2 = Var(X) + [E(X)]^2 = 200^2 + 200^2 = 80000$ [1]

Approximation is therefore 2 * 12% * 200 / 80,000 = 0.000 6 [½]

True value is (12% * 1/200) / 1.12 = 0.000 536 [½]

As expected, the approximation is an upper bound for R. [1]

(v) Use Lundberg, $e^{-RU} = e^{-0.000536*5000} = 6.9\%$ [1]

(vi) Can reduce by holding higher initial surplus / reinsurance / higher premium loading etc. [2]

[Total 13]

This question was generally well answered, although candidates lost marks when not clearly showing their steps in parts (ii) & (iii).

Q8

(i) Mean = $\alpha/\beta \Rightarrow \overline{\beta} = \alpha/\overline{X}$ [2]

(ii) The MLE is the estimate that maximises the likelihood of having observed the sample data. [2]

(iii)
$$L(\beta) = \prod_{i=1}^{n} \frac{\beta^{\alpha}}{\Gamma(\alpha)} x_i^{\alpha-1} e^{-\beta x_i}$$
$$\log L(\beta) \propto n\alpha \ln \beta - \beta \sum_{i=1}^{n} x_i$$
[1½]

Then differentiate

$$\frac{n\alpha}{\overline{\beta}} - \sum_{i=1}^{n} x_i = 0$$

$$\Rightarrow \beta = \alpha n / \sum x_i = \alpha / \overline{X} = 6 / \overline{X}$$ [1½]

Check for maximum

$$\frac{d^2 \ln L}{d\beta^2} = -6n/\beta^2 < 0$$ [1]

In this case the MLE and method of moments lead to the same result. [1]

(iv) $\left(1-\dfrac{t}{\beta}\right)^{-\alpha}$ [1]

(v) $M_Y(t) = E(e^{tY}) = E(e^{2n\beta t\overline{X}}) = E(e^{2\beta t\sum X_i}) = \prod_{i=1}^{n} E(e^{2\beta tX_i})$ [1½]

by independence [½]

$M_Y(t) = \prod_i \left(1-\dfrac{2\beta t}{\beta}\right)^{-\alpha} = (1-2t)^{-n\alpha}$ [1]

By the uniqueness property of MGFs [1]

This is a Chi Squared distribution with parameter $2n\alpha$ [1]

[Total 15]

Many candidates were able to score well on this question, although again marks were dropped when not showing all the steps in parts (iii) & (v).

Q9

(i)

$r_1 = \dfrac{C}{B} = \dfrac{6}{29} = 0.2068966$ [1½]

$r_2 = \dfrac{D}{B} = -\dfrac{24}{29} = -0.8275862$ [1½]

(ii) $\phi_{11} = \rho_1 = 0.2068966$ [1]

$\phi_{22} = \dfrac{\rho_2 - \rho_1^2}{1-\rho_1^2} = -0.9093168$ [1]

(iii) Model X:

$a_1 = \rho_1 = 0.2068966$ [1]

$\hat{\sigma}^2 = \gamma_0 - a_1\gamma_1 = \gamma_0 - a_1\rho_1\gamma_0 = 2.9*(1-0.2068966^2) = 2.775862$ [1½]

$a_0 = \overline{y}(1-a_1) = 10.5*(1-0.2068966) = 8.327586$ [1]

Model Y:

$\gamma_1 = b_1\gamma_0 + b_2\gamma_1, \qquad \rho_1 = b_1 + b_2\rho_1$ [1]

$\gamma_2 = b_1\gamma_1 + b_2\gamma_0, \qquad \rho_2 = b_1\rho_1 + b_2$ [1]

$\hat{b}_2 = \hat{\Psi}_{22} = -0.9093168$ [1]

$\hat{b}_1 = \rho_1(1-b_2) = 0.2068966*(1-(-0.9093168)) = 0.3950312$ [1]

$\sigma^2 = \gamma_0 - b_1\gamma_1 - b_2\gamma_2 = \gamma_0\left(1 - b_1\rho_1 - b_2\rho_2\right)$

$= 2.9*(1-0.2068966*0.3950312-0.9093168*0.8275862) = 0.480621$ [1½]

$b_0 = \bar{y}(1 - b_1 - b_2) = 10.5*(1 - 0.3950312 + 0.9093168) = 15.9$ [1]

(iv) Markov property requires that the process depends only on the previous observation. [1]
This holds for model X, which is AR1 … [1]
… but not for Y, which is AR2 [1]
[Total 18]

Most candidates were able to score well on parts (i), (ii) & (iv) but only the strongest candidates were able to gain most or all of the marks in part (iii).

END OF EXAMINERS' REPORT