INSTITUTE AND FACULTY OF ACTUARIES

EXAMINERS' REPORT

April 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.

Luke Hatter
Chair of the Board of Examiners
June 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 generally better in this diet than in the recent past, with many candidates able to score well even on the harder questions. As has been noted previously, candidates struggled on topics where, although the questions were relatively straightforward, the topic had been examined infrequently or not been examined for a number of years.

C. Pass Mark

The Pass Mark for this exam was 60.

Solutions

Q1 (a) Pareto so clear $\lambda = 1$ and $\alpha = 5$ [1]

From tables $E\left(X^2\right) = \dfrac{\Gamma(\alpha-2)\Gamma(1+2)}{\Gamma(\alpha)} = \dfrac{2*2}{24} = \dfrac{1}{6}$ [½]

$E\left(X^4\right) = \dfrac{\Gamma(\alpha-4)\Gamma(1+4)}{\Gamma(\alpha)} = 1$ [½]

So $E\left(X^4 - 4d^2X^2 + d^4\right) = 1 - \dfrac{2d^2}{3} + d^4$ as required [½]

(b) Differentiating with respect to *d* and setting equal to 0

$-\dfrac{4d}{3} + 4d^3 = 0$ [1]

So $4d^2 = \dfrac{4}{3}, d = \dfrac{1}{\sqrt{3}}$ [1]

Check for minimum: second derivative is $12d^2 - \dfrac{4}{3}$, > 0 so this is indeed a minimum. [½]

[Total 5]

Many candidates struggled on part (a), as this has not been examined for some time, although the majority of candidates were able to score well on part (b), including checking for a minimum.

Q2 (i) The occurrence of the claim and the amount of the claim can be modelled separately, they are independent. [1]

(ii) The maximum likelihood estimate yields the highest probability of observing what has been observed, [2]

(iii) Method of moments [1]
Method of percentiles [1]
Bayesian estimation [1]
[Max 1]

(iv) Insufficient claims [1]

Large claims not recorded (reinsurance) [1]
Small claims not recorded (policy excess) [1]
Change in nature of policy over time [1]
Complications with modelling future inflation [1]
[Max 2]
[Total 6]

> *Candidates with a strong knowledge and understanding of the bookwork were able to score well here, although many struggled to articulate their points sufficiently well in parts (i) and (ii); and to give clearly differentiated examples in part (iv).*

Q3 (i) $m(\theta)$ is the average claim amount for each risk for a given value of θ_i [1]

$s^2(\theta)$ is the variance of the claim amount for each risk given a value of θ_i [1]

$E(s^2(\theta))$ is the average variability of data values from year to year for a single risk, I [1]

$\text{var}(m(\theta))$ is the variability of the average data values for different risks [1]

Z is the credibility factor for the EBCT 1 model / weight placed on the sample mean [1]

(ii) (a) *Z* increases as n increases, since we place more weight on the data for that risk [1]

(b) As $E\left(s^2(\theta)\right)$ increases, *Z* decreases since the variance of the data from the individual risk is high and so we place more weight on the collective data. [2]

(c) As $\text{var}\left(m(\theta)\right)$ increases, *Z* increases since it implies that the means of the individual risks are very different, so we place more weight on the individual risk data compared to the collective. [2]
[Total 10]

> *Most candidates scored well here, although some lost marks through a combination of only using formulae in part (i) and giving insufficient explanations in part (ii).*

Q4 Ultimate loss ratio = 4/4.32 = 92.5926% [1]

DF3 = 4/3.85 = 1.03896 [1]

DF2 = (3.85+4.15)/(3.38+3.67) = 1.134752 [1]

DF1 = (3.38+3.67+3.86)/(3.01+3.3+3.32) = 1.132918 [1]

Adjusted expected ultimate claim for
AY2 = 4.15+0.925926*4.41*(1 – 1/1.03896) = 4.3031 [1½]

Adjusted expected ultimate claim for AY3
= 3.86+0.925926*4.55*(1 – 1/(1.03896*1.134751)) = 4.4995 [1½]

Adjusted expected ultimate claim for AY4
= 3.74+0.925926*4.68*(1–1/(1.03896*1.134751*1.132918)) = 4.8290 [1½]

So reserve = 4 + 4.3031 + 4.4995 + 4.8290 – 13.5 = 4.13m [½]
[Total 9]

Most candidates scored very well on this straightforward chain-ladder question, although some candidates did not appear to know the method required.

Q5 (i) For risk one, let x_i, y_i, z_i be the numbers of claims in month i for risks one, two & three respectively. Let $\mu_I, \mu_{II}, \mu_{III}$ be the monthly rate for these three risks then the likelihood function is:

$$\log L(\mu_I, \mu_{II}, \mu_{III}) = \log L(\mu_I) + \log L(\mu_{II}) + \log L(\mu_{III})$$ [½]

where

$$\log L(\mu_I) = \sum_{i=1}^{36} x_i \log \mu_I - \sum_{i=1}^{36} \mu_I - \sum_{i=1}^{36} \log x_i! = 20 \log \mu_I - 36\mu_I - \sum_{i=1}^{36} \log x_i!$$ [1]

$$\log L(\mu_{II}) = \sum_{i=1}^{30} y_i \log \mu_{II} - \sum_{i=1}^{30} \mu_{II} - \sum_{i=1}^{30} \log y_i! = 18 \log \mu_{II} - 30\mu_{II} - \sum_{i=1}^{30} \log y_i!$$

$$\log L(\mu_{III}) = \sum_{i=1}^{24} z_i \log \mu_{III} - \sum_{i=1}^{24} \mu_{III} - \sum_{i=1}^{24} \log z_i! = 16 \log \mu_{III} - 24\mu_{III} - \sum_{i=1}^{24} \log z_i!$$

After differentiating and equating to zero we have

$$\frac{\partial \log L(\mu_I, \mu_{II}, \mu_{III})}{\partial \mu_I} = -0.36 + \frac{20}{\mu_I}$$

[1]

Second derivative is $-20/\widehat{\mu_I^{\,2}} < 0$ [½]

so $\widehat{\mu_I} = \frac{20}{36} = \frac{5}{9}$

[1]

Similarly we can see that $\widehat{\mu_{II}} = \frac{18}{30} = \frac{3}{5}$, $\widehat{\mu_{III}} = \frac{16}{24} = \frac{2}{3}$ [1]

[Total 5]

(ii) For testing whether the three models are the same we carry out the likelihood ratio test.

We fit the same rate to the three risks using this log likelihood function

$$\log L(\mu) = \left(\sum x_i + \sum y_i + \sum z_i\right)\log\mu - 90\mu - \sum_{i=1}^{36}\log x_i! - \sum_{i=1}^{30}\log y_i! - \sum_{i=1}^{24}\log z_i!$$

[1]

and similar to the above the corresponding MLE is $\hat{\mu} = \frac{54}{90} = \frac{27}{45}$ [½]

$$2\left(\log L(\mu_I, \mu_{II}, \mu_{III}) - \log L(\mu)\right)$$
$$= 2\left(20\log\frac{20}{36} - 20 + 18\log\frac{18}{30} - 18 + 16\log\frac{16}{24} - 16 - 54\log\frac{54}{90} + 54\right)$$
$$= 2*\left(20*log\left(\frac{20}{36}\right) + 18*\log\left(\frac{18}{30}\right) + 16*\log\left(\frac{16}{24}\right) - 54*\log\left(\frac{54}{90}\right)\right)$$
$$= 0.2930949$$

[1½]

The difference in the parameters between the models is 3 – 1 = 2, therefore we compare this test statistics against the χ_2^2 which at the 5% upper level has critical value 5.991>. 0.2931. Therefore there is insufficient evidence to reject the hypothesis that the three risks have a common rate. [2]

[5]

[Total 10]

Variants of this question have been seen many times before and so candidates were able to score very well here, although only the strongest candidates were able to pick up full marks.

Q6 (i) Guess 1, then guess 2 if "Low", then guess 3 if "Low" again. (A)
Guess 1, then guess 3 if "Low", then guess 2 if "High" (B) [1]
Guess 2, then guess 1 if "High", or guess 3 if "Low" (C) [1]
Guess 3, then guess 2 if "High", then guess 1if "High" again (D) [1]
Guess 3, then guess 1 if "High", then guess 2 if "Low" (E) [1]

(ii)

Tarik \ Liam	*A*	*B*	*C*	*D*	*E*
1	1	1	2	3	2
2	2	3	1	2	3
3	3	2	2	1	1

[3]

(iii) There is no saddle point. [1]
This is because there is no element in the matrix which is both the highest in the row and lowest in the column, and vice versa. [1]
[Total 9]

Candidates with a good understanding of the relevant theory were able to score very well here, although some candidates struggled to formulate the strategies required in part (i).

Q7 (i) The characteristic polynomial has a triple root of $B = 1/\alpha$ [1]
and therefore for $|\alpha| < 1$ ensures stationarity [1]

(ii) Expanding the cubic sum in the initial equation we have

$$\left(1 - 3\alpha B + 3\alpha^2 B^2 - \alpha^3 B^3\right) X_t = \varepsilon_t$$

$$X_t - 3\alpha X_{t-1} + 3\alpha^2 X_{t-2} - \alpha^3 X_{t-3} = \varepsilon_t \quad [1]$$

We have an AR(3) process with parameters $a_1 = 3\alpha$, $\alpha_2 = -3\alpha^2$ and $a_3 = \alpha^3$
. [2]

The Yule walker equations for this process are for ρ_1:

$$\rho_1 = a_1 + a_2\rho_1 + a_3\rho_2$$
$$\rho_1(1-a_2) = a_1 + a_3\rho_2 \qquad (1)$$ [1]

And for ρ_2:

$$\rho_2 = a_1\rho_1 + a_2 + a_3\rho_1$$
$$\rho_2 = (a_1 + a_3)\rho_1 + a_2 \qquad (2)$$ [1]

And therefore rearranging (1) and (2)

$$\rho_1 = \frac{a_1 + a_3 a_2}{1 - a_2 - a_1 a_3 - a_3^2} = \frac{3\alpha - 3\alpha^5}{1 + 3\alpha^2 - 3\alpha^4 - \alpha^6}$$ [1]

And

$$\rho_2 = (a_1 + a_3)\rho_1 + a_2 = \left(3\alpha + \alpha^3\right)\frac{3\alpha - 3\alpha^5}{1 + 3\alpha^2 - 3\alpha^4 - \alpha^6} - 3\alpha^2$$ [1]

(iii) Using the definition (Tables have these too)

$$\phi_1 = \rho_1 \; and \; \phi_2 = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$ [1]

(iv) For $\alpha = 1$, the process is not stationary but $\nabla^3 X_t$ is as it becomes a white noise. [1]

In this case we need to fit the white noise to the $T-3$ third differenced observations $\nabla^3 X_1, \nabla^3 X_2 \ldots \nabla^3 X_{T-3}$. [1]

[Total 12]

Many candidates scored well on this relatively straightforward time-series question, although a good number struggled with the algebra in part (ii) and only the strongest candidates were able to answer part (iv) correctly.

Q8 (i) r is the unique positive root of the equation:
$\lambda + cr = \lambda M_X(r)$ [½]

Here $c = 1.2\lambda E(X)$, so the equation simplifies to
$1 + 1.2E(X)r = M_X(r)$ [½]

$M_X(r) = e^{\mu r + \frac{1}{2}\sigma^2 r^2}$ (from tables) [½]

$E(X) = 500$ (from question) [½]

So $1 + 600r - e^{500r + \frac{1}{2}200r^2} = 0$ [1]

At r = 0.000 7075, LHS is +0.000 03, and at r = 0.000 708 5, LHS is –0.000 08 [1]

So the adjustment coefficient must be 0.000 708 to 3sf

(ii) By Lundberg's inequality, upper bound given by $e^{-RU} = e^{-0.000\,708*5000} =$ 0.029 [1]

(iii) $\lambda = 0.002$ [½]

MGF is $\dfrac{\lambda}{\lambda - r}$ [½]

$1 + 600r = \dfrac{0.002}{0.002 - r}$ [1]

$(1 + 600r)(0.002 - r) = 0.002 \Rightarrow 1.2r - r - 600r^2 = 0$ [1]

$r(0.2 - 600r) = 0 \Rightarrow r = 1/3000$ [1]

$e^{-RU} = 0.189$ [1]

(iv) Normal distributions allow the possibility of negative claim amounts [1]

Normal distributions do not have "fat tails", commonly observed in insurance claims [1]

Normal distributions are not positively skewed, unlike typical claim amounts [1]

[Max 1]

Any sensible alternative (gamma, Pareto, Weibull etc.) [1]

[Total 12]

Most candidates are now familiar with the method required in part (i), and were able to score well throughout this question.

Q9 (i) This is a heterogeneous group, … [1]
… since the parameters vary by patient, not at the overall level. [1]

(ii) (a) $E(\lambda_i) = 0.5*0.1 + \frac{1}{3}*0.3 + \frac{1}{6}*0.9 = 0.3$ [½]

$$E(\lambda_i^2) = 0.5*0.1^2 + \frac{1}{3}*0.3^2 + \frac{1}{6}*0.9^2 = 0.17$$

So $\text{Var}(\lambda_i) = 0.17 - 0.3^2 = 0.08$ [1]

(b) $E[S_i \mid \lambda_i] = [\lambda_i m_1]$, $[\text{var}[S_i \mid \lambda_i]] = [\lambda_i m_2]$ [1]

(c) $E(S_i) = E[E[S_i \mid \lambda_i]] = E[\lambda_i m_1] = 0.3m_1 = 0.3*250 = 75$ [½]

[1½]

(d) For the whole portfolio, since the variables are independent and identically distributed, the mean and variance are just 100*75 and 100*23,810 i.e. 7,500 and 2,381,000 respectively. [½]

(iii) Now homogeneous

So $\text{Var}(\lambda_i) = 0.5*(0.2^2 + 0.4^2) - 0.3^2 = 0.01$ [1]

$$E\left[\sum_{i=1}^{n} S_i\right] = nE(S_1) = 0.3*100*250 = 7500$$ [1]

$$\text{Var}\left[\sum_{i=1}^{n} S_i\right] = E\left[\text{Var}\left(\sum_{i=1}^{n} S_i \mid \lambda\right)\right] + \text{Var}\left[E\left(\sum_{i=1}^{n} S_i \mid \lambda\right)\right]$$
$$= E[n\lambda m_2] + \text{var}[n\lambda m_1]$$
$$= 0.3*100*200 + 0.01*100^2*250^2$$
$$= 6256000$$

[2]

(iv) The mean is the same, but the variance is much higher. This makes sense since the parameter uncertainty is at the overall level rather than at the individual level, so affects the aggregate claim variance much more. [2]

[Total 13]

This question was the poorest answered in the entire paper, despite being closely related to the example given in the Core Reading. Since this topic has not come up for some time, it is likely that many candidates were under prepared in this area.

Q10 (i) $F(x) = 1 - e^{-\lambda x}$ [1]

First, sample u from $U \sim U(0,1)$ [1]

Then $X = -\frac{1}{\lambda}\log(1-U)$ [1]

(ii) $M = \sup \frac{f}{h} = \sup_{x>0} \frac{\sqrt{2}}{\lambda\sigma\sqrt{\pi}} e^{\frac{-x^2}{2\sigma^2}+\lambda x}$ [½]

the maximum being achieved at $x = \sigma^2\lambda$, [1]

and $M = \frac{\sqrt{2}}{\lambda\sigma\sqrt{\pi}} e^{\frac{\lambda^2\sigma^2}{2}}$ [½]

(iii) therefore $g(x) = \frac{f(x)}{Mh(x)} = e^{\frac{-x^2}{2\sigma^2}+\lambda x-\frac{\lambda^2\sigma^2}{2}}$ [1]

The rejection algorithm is then:

Sample *x* from $f(x)$ as in (i) [Step 1] [½]

Sample u from $U(0,1)$ [Step 2] [½]

If $u < e^{\frac{-x^2}{2\sigma^2}+\lambda x-\frac{\lambda^2\sigma^2}{2}}$ then set $y = x$, otherwise go to step 1. [1]

(iv) The optimal λ is the one minimizing M, [1]

this can be found by taking $\log M$ and differentiating

$$\log M' = \left(-\log\lambda + \log\frac{\sqrt{2}}{\sigma\sqrt{\pi}} + \frac{\lambda^2\sigma^2}{2}\right)' = -\frac{1}{\lambda} + \lambda\sigma^2$$ [1½]

which becomes zero for $\lambda = \dfrac{1}{\sigma}$ [½]

(v)

Y is a sample as in (iii)

Sample U from $U(0,1)$ [1]

$X = \mu + Y$, if $U > 0.5$ [1]

$X = \mu - Y$, if $U <= 0.5$ [1]

[Total 14]

Most candidates scored well in parts (i) to (iii), although a number failed to use the cumulative distribution in part (i).
Only the best prepared candidates were able to score well in parts (iv) and (v).

END OF EXAMINERS' REPORT