INSTITUTE AND FACULTY OF ACTUARIES

EXAMINERS' REPORT

April 2017

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
July 2017

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*

1. The general performance of students in this examination was slightly weaker than in the recent past. Well prepared candidates were able to score highly.

2. In general, candidates did not score as well as hoped on the elements of bookwork that were tested, which may be because they had not been examined in the recent past and hence did not feature heavily in past papers.

C. Pass Mark

The Pass Mark for this exam was 60.

Solutions

Q1 (i) Strategy II is dominated since strategy I is better under all the opponent's strategies. [2]

(ii) e.g. $d = 2, 0, 0$ [1]

[Total 3]

> This straightforward question was very well answered by the vast majority of candidates.

Q2 (i)

$$f_X(x) = \int_0^\infty f_{X,\lambda}(x,\lambda)\,d\lambda = \int_0^\infty f_\lambda(\lambda)\, f_{X|\lambda}(x|\lambda)\,d\lambda$$ [½]

$$\int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1}\exp(-\beta\lambda)\frac{\lambda^k}{\Gamma(k)} x^{k-1} exp(-\lambda x)\,d\lambda$$

$$= \frac{\beta^\alpha x^{k-1}}{\Gamma(\alpha)\Gamma(k)} \frac{\Gamma(\alpha+k)}{(\beta+x)^{\alpha+k}} \int_0^\infty \frac{(\beta+x)^{(\alpha+k)}}{\Gamma(\alpha+k)} \lambda^{\alpha+k-1}\exp\left(-(\beta+x)\lambda\right)d\lambda\,.$$ [2]

The integral sums to 1 so we are left with

$$\frac{\Gamma(\alpha+k)\beta^\alpha}{\Gamma(\alpha)\Gamma(k)} \frac{x^{k-1}}{(\beta+x)^{\alpha+k}}, x > 0$$ [1]

Which is the PDF of the Generalised Pareto distribution. [½]

(ii) The Exponential is a special case of the Gamma distribution with $k = 1$. [½]

So the mixture distribution is $\frac{\Gamma(\alpha+1)\beta^\alpha}{\Gamma(\alpha)\Gamma(1)} \frac{x^{1-1}}{(\beta+x)^{\alpha+1}} = \frac{\alpha\beta^\alpha}{(x+\beta)^{\alpha+1}}$ [1]

Which is the PDF of a Pareto distribution with parameters α and β. [½]

[Total 6]

> Candidates who were familiar with the necessary bookwork were able to score very well here, although a disappointing number were not. Most candidates spotted the first step in part (ii) – setting k equal to 1.

Q3 (i) Insurance claims are often very positively skewed, with large claims often being several multiples of smaller claims. This suits distributions with heavy tails. [2]

(ii) $1-e^{-c15^{\gamma}}=0.25$ so $c15^{\gamma}=-\ln 0.75$ [½]

Similarly $c80^{\gamma}=-\ln 0.25$ [½]

So $\left(\frac{15}{80}\right)^{\gamma}=\frac{\ln 0.75}{\ln 0.25}=0.207519\ldots$ [1]

So $\gamma=\frac{\ln 0.207519}{\ln\left(\frac{15}{80}\right)}=0.9394$ [½]

And $c=-\frac{\ln 0.75}{15^{0.9394}}=0.0226$ [½]

(iii) Since γ is less than 1, the Weibull distribution in this case has a heavier tail than the exponential. [2]

[Total 7]

Candidates generally scored well on parts (i) and (ii), but only the better candidates were familiar with the bookwork for part (iii).

Q4 (i) $\lambda \sim \Gamma(\alpha,\ \beta)$ where $\frac{\alpha}{\beta}=60$ and $\frac{\alpha}{\beta^2}=360,$ so $\beta=\frac{1}{6}$ and $\alpha=10$. [1½]

Hence $f(\lambda)\propto\ \lambda^9 e^{-\left(\frac{\lambda}{6}\right)}$; [1]

the likelihood $L(\lambda)\propto\ \lambda^{200}e^{-\left(\frac{10\lambda}{3}\right)}$; [1]

So the posterior $f(\lambda|x)\propto\ \lambda^9 e^{-\left(\frac{\lambda}{6}\right)} * \lambda^{200}e^{-\left(\frac{10\lambda}{3}\right)} = \lambda^{209}e^{-\left(\frac{7\lambda}{2}\right)}$ [1]

This is gamma with parameters 210 and 3.5. [1]

$$\ln f\left(\hat{\lambda}|x\right)=\text{const}+209\ln\hat{\lambda}-3.5\,\hat{\lambda}$$

Differentiating $\frac{209}{\hat{\lambda}}=3.5 \Rightarrow \hat{\lambda}=59.7$ [1½]

Differentiating again gives $-\frac{209}{\hat{\lambda}^2}$ which is clearly negative and so we have a maximum.

(ii) The Bayesian estimate is very similar to the mean of the prior distribution, [½]

which is unsurprising since the average number of realised claims has been in line with this. [½]

[Total 8]

Many candidates were able to score well here, especially on part (i).

Q5 (i) From the definition

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

$$\exp\left[n\left(y\log\left(\frac{\mu}{1-\mu}\right) + \log(1-\mu)\right) + \log\binom{n}{ny}\right] \quad [1]$$

Hence

$$\theta = \log\left(\frac{\mu}{1-\mu}\right) \quad [½]$$

$$\varphi = n \quad [½]$$

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

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

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

(ii) From the theory we know that

$$E(y) = b'(\theta) = \log\left(1+e^{\theta}\right)' = \frac{e^{\theta}}{1+e^{\theta}} = \mu \quad [2]$$

$$\text{Var}(y) = a(\varphi)b''(\theta) = \frac{1}{n}\left(\frac{e^{\theta}}{1+e^{\theta}}\right)' = \frac{1}{n}\frac{e^{\theta}}{1+e^{\theta}}\left(1-\frac{e^{\theta}}{1+e^{\theta}}\right) = \frac{\mu(1-\mu)}{n} .$$ [2]

[Total 8]

Most candidates were able to make some progress, but only the better candidates were able to obtain the precise specification. Most candidates knew the standard results from the theory for part (ii), but then struggled to apply it in this particular case.

Q6 (i) The method of moments (or method of least squares) and maximum likelihood estimation. [2]

(ii) The method of moments (or method of least squares) does not make any assumptions about the distribution of ε_t. [2]

(iii) The parameter estimates are obtained from Y-W equations for the AR(1) processes. [1]

Namely, $\rho_1 = \alpha$ and $\sigma^2 = \gamma_0 - \alpha\gamma_1$ where ρ_1, γ_0 and γ_1 are estimated from the sample quantities. [2]

(iv) Both models above are identical as the observations from one also satisfy the difference equation of the other. [2]

(v) Answer in (iv) implies that the models are equivalent so the process is stationary regardless of the value of *c*. [2]

[Total 11]

Candidates familiar with the theory were able to score well on parts (i) to (iii). Unfortunately there was a typo in the introduction to part (iv), since the mu was excluded, so marks were awarded generously where there was any evidence this had caused confusion in the answers to parts (iv) and (v).

Q7 (i) $\overline{P_A} = 350,\ \overline{P_B} = 300,\ \overline{P_C} = 550,\ \bar{P} = 1200$

$$P^* = \frac{1}{11}\left[350^*\left(1-\frac{350}{1200}\right)+300^*\left(1-\frac{300}{1200}\right)+550^*\left(1-\frac{550}{1200}\right)\right] = 70.1$$ [1]

$$\overline{X_A} = \frac{4.8}{350} = 0.0137,\ \overline{X_B} = \frac{3.7}{300} = 0.0123,\ \overline{X_C} = \frac{6.7}{550} = 0.0122,\ \bar{X} = \frac{15.2}{1200} = 0.0127$$

[1]

Now need to calc $\sum_{j=1}^{4} P_{ij}\left(X_{ij} - \overline{X_i}\right)^2$

$$i = 1, 85*\left(\frac{1.16}{85} - 0.0137\right)^2 + 88*\left(\frac{1.18}{88} - 0.0137\right)^2 + 85*\left(\frac{1.14}{85} - 0.0137\right)^2 +$$

$$92*\left(\frac{1.32}{92} - 0.0137\right)^2 = 0.000053286$$ [1]

$$i = 2, 68*\left(\frac{0.85}{68} - 0.0123\right)^2 + 82*\left(\frac{1.02}{82} - 0.0123\right)^2 + 70*\left(\frac{0.96}{70} - 0.0123\right)^2 +$$

$$80*\left(\frac{0.87}{80} - 0.0123\right)^2 = 0.000306436$$ [½]

$$i = 3, 110*\left(\frac{1.48}{110} - 0.0122\right)^2 + 132*\left(\frac{1.52}{132} - 0.0122\right)^2 + 143$$

$$*\left(\frac{1.78}{143} - 0.0122\right)^2 + 165*\left(\frac{1.92}{165} - 0.0122\right)^2 = 0.000296037$$ [½]

$$E\left[s^2(\theta)\right] = \frac{1}{3}\sum_{i=1}^{3}\left\{\frac{1}{3}\sum_{j=1}^{4} P_{ij}\left(X_{ij} - \overline{X_i}\right)^2\right\}$$

$$= \frac{1}{3}\left\{\frac{1}{3}\left(0.000005329 + 0.000030644 + 0.000029604\right)\right\}$$

$$= 0.000072862$$ [1]

Now need to calc $\sum_{j=1}^{4} P_{ij}\left(X_{ij} - \overline{X}\right)^2$

$$i = 1, 85*\left(\frac{1.16}{85} - 0.0127\right)^2 + 88*\left(\frac{1.18}{88} - 0.0127\right)^2 + 85*\left(\frac{1.14}{85} - 0.0127\right)^2$$

$$+92*\left(\frac{1.32}{92} - 0.0127\right)^2 = 0.00043741$$ [1]

$$i = 2, 68*\left(\frac{0.85}{68} - 0.0127\right)^2 + 82*\left(\frac{1.02}{82} - 0.0127\right)^2 + 70*\left(\frac{0.96}{70} - 0.0127\right)^2$$

$$+80*\left(\frac{0.87}{80} - 0.0127\right)^2 = 0.000339769$$ [½]

$$i = 3, 110*\left(\frac{1.48}{110} - 0.0127\right)^2 + 132*\left(\frac{1.52}{132} - 0.0127\right)^2 + 143$$

$$*\left(\frac{1.78}{143} - 0.0127\right)^2 + 165*\left(\frac{1.92}{165} - 0.0127\right)^2 = 0.000425330$$ [½]

Now

$$\frac{1}{Nn-1}\sum_{i=1}^{3}\sum_{j=1}^{4}P_{ij}\left(X_{ij} - \bar{X}\right)^2$$

$$= \frac{1}{11}\left(0.00043741 + 0.000339769 + 0.000425330\right)$$

$$= 0.000109319$$ [1]

So $\text{var}\left[m(\theta)\right] = \frac{1}{70.1}\left(0.000109319 - 0.000072862\right) = 0.000000520$ [½]

So the credibility factor $Z = \dfrac{\sum P_j}{\sum P_j + \dfrac{E\left[s^2(\theta)\right]}{\text{var}\left[m(\theta)\right]}} = \dfrac{550}{550 + \dfrac{0.000072862}{0.000000520}}$

$$= 0.7970$$ [1]

So the expected premium per policy for company C is

$$0.7970*0.0122 + (1 - 0.7970)*0.0127 = 0.012\ 3$$ [1]

So total expected claims in 2016 is 180 * 0.012 3 = $2.21m [½]

(ii) EBCT Model 2 allows for the increasing number of policies that Company C has seen year on year so is arguably a better estimate. [2]

[Total 13]

> Some candidates were clearly unfamiliar with the techniques needed to apply EBCT II from scratch, although many candidates were able to score well.

Q8 (i) The general form can be expressed as follows:

$$C_{ij} = r_j s_i x_{i+j} + e_{ij}$$ [1]

where

C_{ij} is the cumulative or incremental entry in the run-off triangle; [½]
r_j is the development factor for Development Year j; [1]
s_i represents the exposure (or number of claims / policies); [1]
x_{i+j} is a parameter varying by calendar year; [1]
e_{ij} is an error term. [½]

(ii) DF from year 2 to year 3 is 10078 / 6847 = 1.471885 [1]
DF from year 1 to year 2 is (7123 + 6847) / (3215 + 2986) = 2.252862 [1]
For AY 2015, expected ultimate loss is 0.91 * 12012 = 10931 [1]
Expected loss to date is 10931 / 1.471885 = 7,426 [1]
So the adjusted ultimate loss is 10931 – (7426 – 7123) = 10627 [1]
For AY 2016, expected ultimate loss is 0.91 * 12867 = 11709 [1]
Expected loss to date is 11709 / (1.471885 * 2.252862)= 3531 [1]
So the adjusted ultimate loss is 11709 – (3531 – 4167) = 12345 [1]
So the reserve is 10078 + 10627 + 12345 – 21186 = 11864 [1]

[Total 14]

> Many candidates were unfamiliar with the bookwork required to answer part (i), although most scored well on part (ii).

Q9 (i) The function $\dfrac{f(x)}{h(x)} = \dfrac{1}{\beta\Gamma(\alpha)} x^{\alpha-1} e^{-x+\beta x} x > 0$ [½]

the upper bound of this function is obtained at the same value for x as that of

$$\log\left(\frac{f(x)}{h(x)}\right) = \log\left(\frac{1}{\beta\Gamma(\alpha)}\right) + (\alpha-1)\log x - x + \beta x .$$ [1]

However the derivative of $(\alpha-1)\log x - x + \beta x$ is $\dfrac{\alpha-1}{x} - 1 + \beta$ [1]

which becomes zero at $x_0 = \dfrac{\alpha - 1}{1-\beta}$ [½]

and since the second derivative is $-\dfrac{\alpha-1}{x^2}$<0 implies that the maximum value of $f\,/\,h$ is attained for $x_0 = \dfrac{\alpha-1}{1-\beta}$. [½]

So $C = \dfrac{1}{\beta\Gamma(\alpha)} {x_0}^{\alpha-1} e^{-x+\beta x} x > 0$ [½]

(ii) The function $g(x) = \dfrac{f(x)}{C\,h(x)} = \dfrac{x^{\alpha-1}e^{(\beta-1)x}}{{x_0}^{\alpha-1}e^{(\beta-1)x_0}}$ [1]

So the rejection algorithm looks like this

1 – Simulate $U_1 \sim U(0,1)$ and set $Y = -\dfrac{1}{\beta}\log U_1$. [1½]

2 – Simulate $U_2 \sim U(0,1)$ if $U_2 < g(Y)$ set accept the value by setting $X = Y$ otherwise go back to stage 1. [1½]

(iii) The algorithm is most efficient choosing the value β making $\sup\left(\dfrac{f}{h}\right)$ the smallest (or minimizing C, or maximizing g(x)) [1]

$$\sup\left(\frac{f}{h}\right) = \frac{1}{\Gamma(\alpha)}\beta^{-1}x_0^{\alpha-1}e^{(\beta-1)x_0} = \frac{1}{\Gamma(\alpha)}\beta^{-1}\left(\frac{\alpha-1}{1-\beta}\right)^{\alpha-1} e^{1-\alpha} \quad [1]$$

Which is minimized if $\log\left(\beta^{-1}\left(\dfrac{\alpha-1}{1-\beta}\right)^{\alpha-1}\right)$ is minimized i.e. [1]

$$\left(-\log\beta + (1-\alpha)\log(1-\beta)\right)' = -\frac{1}{\beta} + \frac{\alpha-1}{1-\beta} = \frac{-1+\beta+\alpha\beta-\beta}{\beta(1-\beta)} = 0 \quad [2]$$

i.e. $\alpha\beta = 1$ so $\beta = \dfrac{1}{\alpha}$ [1]

[Total 14]

> Candidates mostly scored well in part (i), although those unfamiliar with the technique of using logs often struggled. Part (ii) was relatively well answered, although only the best candidates were able to score well in part (iii).

Q10 (i) $E(S_1) = 15*3500 = 52,500,\ E(S_2) = 25*3600 = 90,000,$ so $E(S) = 142,500$

[1]

$$\text{Var}(S_1) = 15*\left(\frac{1}{12}(4000-3000)^2 + 3500^2\right) = 185,000,000$$ [½]

$$\text{Var}(S_2) = 25*\left(2*3600^2\right) = 648,000,000$$ [½]

By independence

$$\text{Var}(S) = \text{Var}(S_1) + \text{Var}(S_2) = 833000000$$ [½]

And so $sd(S) = \sqrt{\text{Var}(S)} = 28,862$ [½]

(ii) $U(1) = U + c*1 - S(1) = U + 1.07E(S) - S$ [1]

$$P\{U(1) < 0\} = P\{S > U + 1.07E(S)\} = P\left(Z > \frac{U + 0.07E(S)}{sd(S)}\right)$$ [1½]

For this to be less than 0.015, need

$$\frac{U + 0.07E(S)}{sd(S)} > 2.1701, U_m > 2.1701*28,862 - 0.07*142,500 = 52,658$$ [1½]

(iii) Now

$$U'(1) = U_m + c_{net}*1 - S_I(1) = U_m + 1.07E(S) - 1.17(1-\alpha)E(S) - S_I =$$

$$U_m + (1.17\alpha - 0.1)E(S) - S_I$$ [1½]

Also $S_I \sim N\left(\alpha E(S), \alpha^2 \text{Var}(S)\right)$ [1]

So $P\left(U'(1)<0\right) = P\left\{S_I > U_m + \left(1.17\alpha - 0.1\right)E\left(S\right)\right\}$

$$= P\left\{Z > \frac{U_m + \left(0.17\alpha - 0.1\right)E\left(S\right)}{\alpha * sd\left(S\right)}\right\}$$ [1]

We need

$$\frac{U_m + \left(0.17\alpha - 0.1\right)E\left(S\right)}{\alpha * sd\left(S\right)} > 2.5758, \frac{52,655 + \left(0.17\alpha - 0.1\right)*142,500}{\alpha * 28,862} > 2.5758$$

[1]

So

$$2.5758 * 28,862 * \alpha - 142,500 * 0.17\alpha < 52,655 - 0.1 * 142,500$$

So

$\alpha < 76.6\%$ [1½]

(iv) Use a higher initial surplus, [1]
increase its premium loading, [1]
insure more type I claims relative to Type II claims since they have a smaller variance. [1]

[Total 16]

Candidates familiar with reinsurance theory were able to score highly here.

END OF EXAMINERS' REPORT