INSTITUTE AND FACULTY OF ACTUARIES

EXAMINERS' REPORT

September 2016

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
December 2016

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 diet was stronger than in the recent past. Well prepared candidates could score highly on both the mathematical and more descriptive elements of the questions.

2. Candidates' performance on the questions relating to time series and Bayes' theory was notably better, but many candidates were not able to recall the bookwork for Question 7.

3. The slightly higher pass mark reflects the fact that a few more marks than usual were available for relatively straightforward calculation work.

C. Pass Mark

The Pass Mark for this exam was 62.

Solutions

Q1 (i) Not sufficiently diversified / not independent / not large enough sample [1]
Not easily quantifiable [1]
Risk is potentially quite large (not necessarily a remote risk) [1]
Policyholders may not have financial interest in risk [1]
Moral hazard [1]
[Max 3]

(ii) Increase number of policies [1]
ensure definition of bad weather is extreme [1]
Diversify policies between different cities [1]
Diversify policies between different days [1]
Introduce policy excess [1]
[Max 2]

Most candidates were able to relate the theory of insurable risks to this question. A few candidates failed to sufficiently distinguish between their points, or did not consider the actual context presented.

Q2 (i) Let θ_i be the state of nature when the roll of the die = i.

Then the payoff matrix is:

	θ_1	θ_2	θ_3	θ_4
a_1	1.5	3	4.5	6
a_2	0.5	4	1.5	8
a_3	2	2	6	4

[1 mark for first correct row, ½ mark thereafter]

(ii) None of the decision functions is dominated. [1]

(iii) Since each number is equally likely, this is equivalent to summing up the payoffs for each decision function. [1]

This is 15, 14 and 14. [1]

So a_1 is the optimal decision under the Bayes criterion. [1]

This straightforward question was very well answered by most candidates, with many scoring full marks.

Q3 (i) Overall mean is $\bar{X} = \dfrac{46.8+30.2+74.5+60.7}{4} = 53.05$ [1]

$$E(s^2(\theta)) = \frac{1}{4}\sum_{i=1}^{4} S_i^2 = \frac{1227.4+1161.4+1340.3+1414.7}{4} = 1285.95 \quad [1]$$

$$\text{Var}\big(m(\theta)\big) = \frac{1}{3}\sum_{i=1}^{4}\left(\overline{X_i} - \bar{X}\right)^2 - \frac{1}{6}E\left(S^2(\theta)\right) \quad [1]$$

$$= \frac{(46.8-53.05)^2 + (30.2-53.05)^2 + (74.5-53.05)^2 + (60.7-53.05)^2}{3}$$

$$-\frac{1285.95}{6}$$

= 145.6 [1]

So the credibility factor is $Z = \dfrac{6}{6 + {}^{1285.95}\!/\!_{145.6}} = 0.4045$ [1]

And the credibility premia are:

(1) 0.4045 * 46.8+ 0.5955 * 53.05 = 50.5 [½]
(2) 0.4045* 30.2 + 0.5955 * 53.05 = 43.8 [½]
(3) 0.4045 * 74.5 + 0.5955 * 53.05 = 61.7 [½]
(4) 0.4045 * 60.7 + 0.5955 * 53.05 = 56.1 [½]

(ii) The variation within risks is much bigger relative to the variation between risks. This suggests that the variability is more explained by claim variability than in the underlying parameter, so we put more weight on the information provided by the data set as a whole, and less on the individual risks, resulting in a low credibility factor. [2]

[Total 9]

Many candidates scored full marks on part (i), but scored less well on part (ii). Stronger candidates were able to explain in words what was happening, beyond simply using mathematical formulae.

Q4 (i) If Standard type the algorithm is:

Sample U from $U(0,1)$, if $U <= 0.8$ remain Standard; if $U > 0.8$ the new state is Premium. [1]

If remain Standard repeat, otherwise if move to Premium algorithm: [½]
If type Premium the algorithm is:

Sample U from $U(0,1)$, if $U <= 0.5$ moves to Standard, if $0.5 < U$ the new state is Elite. [1]

If move to Standard move to the Standard algorithm, otherwise move to the Elite algorithm: [½]

If type Elite the algorithm is:

Sample U from $U(0,1)$, if $U <= 0.9$ moves to Premium type; if $0.9 < U$ stay as Elite type. [1]

If remain Elite type repeat, otherwise move to the Standard algorithm. [½]

Start in required state and move between algorithms as required until three years have been simulated. [½]

(ii) Start Standard, since $U = 0.89$ the new type is Premium (hence no claim in the previous year!). [1]

Now type Premium, since $U = 0.64$ the new type is Elite (no claim in the previous year). [1]

Now type Premium, since $U = 0.12 < 0.9$ then we have a claim and the new state is Premium. [1]

There is only one simulated claim. [1]
[Total 9]

The unfamiliar application of the inverse transform theory confused a few candidates, but most scored well on both parts. Only the strongest candidates demonstrated their algorithm simulated *three* years, rather than just one.

Q5 (i) (a) Each realisation of the variable is unaffected by previous outcomes and in turn does not affect future outcomes. [1]

The variables all come from the same distribution with the same parameters. [1]

(b) E.g. rolling a fair die, tossing a fair coin etc. [1]

(ii) Let $S = \sum_{i=1}^{n} X_i$, then

$$M_S(t) = \prod_{i=1}^{n} M_{X_i} = [M_X(t)]^n = \left[\left(1-\frac{t}{\lambda}\right)^{-1}\right]^n = \left[\left(1-\frac{t}{\lambda}\right)^{-n}\right]$$ [1]

By independence of claim amounts and uniqueness property of MGFs [½]

This is a gamma distribution with parameters n and λ. [½]

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

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

$$= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \lambda^\alpha \exp\{-(x+\beta)\lambda\}\,d\lambda$$ [1½]

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

$$= \beta^\alpha \setminus \Gamma(\alpha)\ \frac{\Gamma(\alpha+1)}{(x+\beta)^{\alpha+1}} = \frac{\alpha\beta^\alpha}{(x+\beta)^{\alpha+1}}, x > 0$$ [1½]

Since the final integral is the PDF of a Gamma distribution and so equals 1.

This is the PDF of a Pareto distribution with parameters α and β. [1]

[Total 10]

Part (i) was poorly answered, with many candidates simply repeating the words *independent* and *identical*. Part (ii) was well answered, although part (iii) was relatively poorly answered.

Q6 (i) For risk A with rate μ_1 the log-likelihood function is:

$$\log L_1 = \log \mu_1 \sum_{i=1}^{5} y_{1i} - 5\mu_1 - \sum_{i=1}^{5} \log y_{1i}!$$

$$= 12\log\mu_1 - 5\mu_1 - \sum_{i=1}^{5}\log y_{1i}!$$ [1½]

And therefore the mle for μ_1 is obtained for

$$\frac{\partial \log L}{\partial \mu_1} = \frac{12}{\mu_1} - 5 = 0$$ [1]

i.e. $\widehat{\mu_1} = 2.4$ [½]

Similarly we have that $\widehat{\mu_2} = \frac{22}{5} = 4.4$ and $\widehat{\mu_3} = \frac{28}{5} = 5.6$. [2]

(ii) Under the assumption that these risks share the same rate i.e. $\mu_1 = \mu_2 = \mu_3 = \mu$ then the mle estimate for this is simply $\hat{\mu} = \frac{62}{15}$. [½]

In order to compare these models we can use the scaled deviances to compare these models and using the chi-squared test.

The difference in the scaled deviance is chi-square with 3 − 1 = 2 degrees of freedom. [1]

$$2\left(\log L_1 + \log L_2 + \log L_3 - \log L\right)$$

$$= 2(12\log\widehat{\mu_1} - 5\widehat{\mu_1} + 22\log\widehat{\mu_2} - 5\widehat{\mu_2} + 28\log\widehat{\mu_3} - 5\widehat{\mu_3} - 62\log\hat{\mu} + 15\hat{\mu})$$ [1]

With the $\sum_{i=1}^{5}\log y_{1i}! + \sum_{i=1}^{5}\log y_{2i}! + \sum_{i=1}^{5}\log y_{3i}!$ cancelling out in the difference.

Hence

$$2\left(\log L_1 + \log L_2 + \log L_3 - \log L\right)$$

$$= 2\left(12\log 2.4 + 22\log 4.4 + 28\log 5.6 - 62\log\frac{62}{15} - \frac{5(12+22+28)}{5} + 15\frac{62}{15}\right)$$

$$= 2\left(12\log 2.4 + 22\log 4.4 + 28\log 5.6 - 62\log\frac{62}{15}\right) = 6.71034.$$ [1½]

This value is above 5.991 which is the critical value at the upper 5% level and therefore conclude that mean claim rates are different. [1]

[Total 10]

Part (i) was very well answered by most candidates. Fewer candidates scored full marks in part (ii), despite similar questions having been asked in several recent examinations.

Q7 (i) Let Z be the reinsurer claim distribution.

Then $g(z) = \dfrac{f(z+M)}{1-F(M)}$ where $f(x)$ and $F(x)$ refer to the underlying claim distribution [1]

$$f(z+M) = \frac{\alpha\lambda^{\alpha}}{(\lambda+z+M)^{\alpha+1}}; F(M) = 1 - \frac{\lambda^{\alpha}}{(\lambda+M)^{\alpha}} \quad [1½]$$

$$\text{So } g(z) = \frac{\alpha\lambda^{\alpha}}{(\lambda+z+M)^{\alpha+1}} \frac{(\lambda+M)^{\alpha}}{\lambda^{\alpha}} = \frac{\alpha(\lambda+M)^{\alpha}}{(\lambda+z+M)^{\alpha+1}} \quad [1]$$

This is in the form of a Pareto distribution with parameters α and $\lambda + M$. [½]

(ii) Let $Y = kX; P(Y < y) = P(kX < y) = \displaystyle\int_0^{\frac{y}{k}} \frac{\alpha\lambda^{\alpha}}{(\lambda+z)^{\alpha+1}} dz$ [1½]

$$\text{Let } x = kz; = \int_0^{y} \frac{\alpha\lambda^{\alpha}}{\left(\lambda+\frac{x}{k}\right)^{\alpha+1}} \frac{dx}{k} = \int_0^{y} \frac{\alpha\lambda^{\alpha}k^{\alpha}}{(k\lambda+x)^{\alpha+1}} dx. \quad [2]$$

Which is in the form of a Pareto distribution with parameters α and $k\lambda$. [½]

(iii) The reinsurer's distribution of claims is therefore Pareto with parameters α and $k\lambda + M$. [1]

(iv) The average claim retained by the insurer has increased by a factor less than k since the retention M is unchanged, so on average a greater proportion of claims get passed on to the reinsurer. [2]

For strong candidates familiar with the bookwork this question was very straightforward, but few candidates were able to score well here.

Q8 Adjusting for past inflation to 2015 prices gives

Underwriting Year	*Development Year* 0	1	2	3
2012	504 * 106/100 = 534.24	286 * 106/103 = 294.33	110 * 106/105 = 111.05	35
2013	621 * 106/103 = 639.09	302 * 106/105 = 304.88	120	
2014	685 * 106/105 = 691.52	340		
2015	801			

[2]

Cumulative figures (2015 prices)

Underwriting Year	*Development Year* 0	1	2	3
2012	534.24	828.57	939.62	974.62
2013	639.09	943.97	1,063.97	
2014	691.52	1,031.52		
2015	801			

[1]

Development factors:

Year 2 to 3: 974.62 / 939.62 = 1.037249 [1]
Year 1 to 2: (939.62 + 1063.97) / (828.57 + 943.97) = 1.130350 [1]
Year 0 to 1: (828.57 + 943.97 + 1031.52) / (534.24 + 639.09 + 691.52) = 1.503638 [1]

Projected cumulative figures (2015 prices)

Underwriting Year	*Development Year* 0	1	2	3
2012	534.24	828.57	939.62	974.62
2013	639.09	943.97	1,063.97	**1,103.60**
2014	691.52	1,031.52	**1,165.98**	**1,209.41**
2015	801	**1,204.41**	**1,361.40**	**1,412.12**

[2]

Projected incremental figures (2015 prices)

Underwriting Year	*0*	*1*	*2*	*3*
2012				
2013				39.63
2014			134.46	43.43
2015		403.41	156.99	50.72

[1]

Adjusting for future inflation

Underwriting Year	*0*	*1*	*2*	*3*
2012				
2013				39.63 * 105 / 106 = 39.26
2014			134.46 * 105 / 106 = 133.19	43.43 * 107 / 106 = 43.84
2015		403.41* 105 / 106 = 399.60	156.99 * 107 / 106 = 158.47	50.72 * 110 / 106 = 52.63

[2]

The estimated reserve is the sum of these: 827.0 [1]

[Total 12]

Most candidates scored very well on this straightforward chain ladder question.

Q9 (i) The first term in the equation has period 12 and so this removes the periodic effect. [1]

(ii) The characteristic polynomial will be $1-(\alpha+\beta)B+\alpha\beta B^2$ [1]

with roots $1/\alpha$ and $1/\beta$. [1]

Hence the stationarity holds for $|\alpha|<1$ and $|\beta|<1$. [1]

(iii) Y_t is an AR(2) where $a_1 = \alpha + \beta$ and $a_2 = -\alpha\beta$. Since from the Yule-walker equations for AR(2) we have

$\rho_1 = a_1 + a_2\rho_1$ [1]

and

$\rho_2 = a_1\rho_1 + a_2$ [1]

which imply that $a_1 = (1 - a_2)\rho_1 = 0$ since $\rho_1 = 0$. [1]

This implies that $\alpha + \beta = 0, \alpha = -\beta$ [1]

and the second equation $0.09 = \rho_2 = a_1\rho_1 + a_2 = a_2 = \alpha^2$ i.e. $\alpha = -\beta = \pm 0.3$. [1]

(iv) Since $Y_t = X_t - X_{t-12}$ we have that

$X_{T+1} = Y_{T+1} + X_{T-11}$ [½]

$X_{T+2} = Y_{T+2} + X_{T-10}$ [½]

With the forecasted values

$\hat{x}_{T+1} = \hat{y}_{T+1} + x_{T-11}$ [½]

and

$\hat{x}_{T+2} = \hat{y}_{T+2} + x_{T-10}$ [½]

where

$\hat{y}_{T+1} = 0 * y_T + 0.09\, y_{T-1} = 0.09(x_{T-1} - x_{T-13})$ [1]

And similarly

$\hat{y}_{T+2} = 0.09\ (x_T - x_{T-12})$ [1]

[Total 13]

The performance on this time series question was very good, although only the stronger candidates were able to score well on part (iv).

Q10 (i) The adjustment coefficient is the unique positive root of the equation

$$\lambda M_X(R) = \lambda + cR$$ [1]

(ii) $c = 125 * \left(\dfrac{20}{0.5}\right) * 1.15 = 5750$ and

$$M_X(R) = (1-2R)^{-20}$$ [1]

So R is the root of

$$f(R) = 125(1-2R)^{-20} - 125 - 5750R = 0$$ [1½]

When R is 0.006475 then $f(R) = -0.00282$

When R is 0.006485 then $f(R) = 0.005433$ [1½]

Since the function changes sign between 0.006475 and 0.006485 the unique positive root must lie between these values hence R is 0.00648 to 3 sf [1]

(iii) By Lundberg's inequality $\Psi(300) < \exp(-300 * .00648) = 0.143$ [2]

Total claims have a mean claim amount of 125 * 40 = 5000 [½]

And variance $125 * (80 + 40^2) = 210\,000$ [½]

So approximately

$$\Psi_1(300) = P\big(300 + 5750 - N(5000, 210000) < 0\big)$$

$$= P\left(N(0,1) > \frac{6050 - 5000}{\sqrt{210000}}\right)$$ [1]

$$= P\big(N(0,1) > 2.291\big) = 0.011$$ [1]

(iv) The probability would increase, since both the mean and variance of claim amounts are higher. [2]

(v) They could use a higher initial surplus or a higher premium loading. [2]
[Total 15]

This question was typically answered very well. Candidates who struggled with part (ii) should note the method used in the answer. Most candidates were able to give good explanations for parts (iv) and (v) and therefore scored well.

END OF EXAMINERS' REPORT