INSTITUTE AND FACULTY OF ACTUARIES

EXAMINERS' REPORT

September 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
December 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*

Students generally scored well on application type questions that were variants of or similar to past exam questions, and tended to underperform on more bookwork style questions or application questions that do not tend to appear frequently.

Unfortunately there was an ambiguity towards the end of Question 11 which meant many students struggled to answer the last part of Question 11, this has been reflected in the lower pass mark of 56.

C. Pass Mark

The Pass Mark for this exam was 56.

Solutions

Q1 $1-e^{-c1000^{\gamma}}=0.5$ [½]

$1-e^{-c5000^{\gamma}}=0.9$ [½]

$c1000^{\gamma}=-\ln 0.5$ [½]

$c5000^{\gamma}=-\ln 0.1$ [½]

$0.2^{\gamma}=\dfrac{\ln 0.5}{\ln 0.1}$ [½]

$\gamma=0.74594$ [½]

$c=\dfrac{-\ln 0.5}{1000^{0.74594}}=0.004009$ [1]

> *Most students scored very well on this straightforward question. Unfortunately there was a typographical error in the question which referred to 'method of moments' rather than 'method of percentiles'; although this did not affect many candidates, full credit was given to any reasonable attempt at applying the method of moments.*

Q2 (i) Pseudo random numbers can be reproduced [1]

Only single routine required, rather than lengthy table/hardware [1]

Difficult to generate very large set of truly random numbers [1]

[Max 2]

(ii) (a) $v_1=2*0.062-1=-.876, v_2=2*0.293-1=-.414,$

$s=v_1^2+v_2^2=0.938772$ [1]

$z_1=v_1*\sqrt{-2\ln(s)/s}=-.3214$

$z_2=v_2*\sqrt{-2\ln(s)/s}=-.1519$ [1]

(b) $(2*0.984-1)^2+(2*0.794-1)^2>1$, so no variate [½]

(c) $(2*0.008-1)^2+(2*0.961-1)^2>1$, so no variate [½]

Many candidates scored well here, although a number failed to make the adjustment to convert a U(0,1) into a U(-1,1).

Q3 A is false since there cannot be a claim until time 2 [2]

B is false since the insurance company could be ruined at time 3 if there is a claim, if *U* is sufficiently small. [2]

C is false since the insurance company cannot be ruined in year 4, since by that stage it will have sufficient premiums to cover any loss. [2]

D is true since if it is not ruined by time 4, the insurance company cannot be ruined. [2]

Stronger candidates were able to apply the information given in the question to ruin theory to score well, but a number of candidates failed to do so.

Q4 (i)

		Shivon						
		1	**2**	**3**	**4**	**5**	**6**	**7**
Craig	**1**	0	1	0	0	0	0	0
	2	1	0	1	0	0	0	0
	3	0	1	0	1	0	0	0
	4	0	0	1	0	1	0	0
	5	0	0	0	1	0	1	0
	6	0	0	0	0	1	0	1
	7	0	0	0	0	0	1	0

Where each element represents the payoff to Shivon. [2]

(ii) Clearly for Shivon, picking 1 and 7 are dominated, since she can do better by picking 3 or 5 respectively. So now A =

		Shivon				
		2	**3**	**4**	**5**	**6**
Craig	**1**	1	0	0	0	0
	2	0	1	0	0	0
	3	1	0	1	0	0
	4	0	1	0	1	0
	5	0	0	1	0	1
	6	0	0	0	1	0
	7	0	0	0	0	1

[1½]

But now for Craig, 1 dominates 3, 7 dominates 5, and 2 (or 6) dominate 4, so now A =

		Shivon				
		2	**3**	**4**	**5**	**6**
Craig	**1**	1	0	0	0	0
	2	0	1	0	0	0
	6	0	0	0	1	0
	7	0	0	0	0	1

[1½]

But now 4 is dominated for Shivon, so A =

		Shivon			
		2	**3**	**5**	**6**
Craig	**1**	1	0	0	0
	2	0	1	0	0
	6	0	0	1	0
	7	0	0	0	1

[1]

(iii) Clearly by symmetry Shivon should pick 2, 3, 5 or 6 a quarter of the time, and the value of the game to her is ¼. [2]

Candidates who were well prepared in decision theory were typically able to score well in all parts of this question.

Q5 (i) Cumulative amounts & claims

Accident Year	**0**	**1**	**2**
2014	5,419	6,327	6,566
2015	6,234	7,322	
2016	7,719		

[1]

Accident Year	**0**	**1**	**2**
2014	760	858	895
2015	819	912	
2016	881		

[1]

Average cost per claim (cumulative)

	Development Year		
Accident Year	**0**	**1**	**2**
2014	97.191% 7.130 263	100.515% 7.374 126	100% 7.336 313
2015	95.297% 7.611 722	100.515% 8.028509	**100%** **7.987 3**
2016	96.244% 8.761 635		**100%** **9.103 5**

[2]

Claim numbers (cumulative)

Accident Year	**0**	**1**	**2**
2014	84.916% 760	95.866% 858	100% 895
2015	86.090% 819	95.866% 912	**100%** **951.33**
2016	85.503% 881		**100%** **1,030.4**

[1½]

Total claims 6,566+7.98734*951.329 +9.103564*1030.373 = 23,545

[1]

Claims to date 21,607 so reserve €1,938k [½]

(ii) Claims fully developed after DY2. [½]
The proportions of claim numbers relating to each DY remain constant in different AYs. [1]
Cost of claims settled equals amount actually paid out. [½]
The average cost per claim figures relating to each DY remain constant in different AYs. [½]
Inflation has been allowed for. [½]
[Max 2]

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

Q6 (i) A fixed numbers of risks are assumed within the studied portfolio [½]

The number of risks does not change over the period of insurance cover [½]

These risks are independent [½]

Claim amounts from these risks are not (necessarily) identically distributed random variables [½]

(ii) The number of claims arising is either 0 or 1. [1]

(iii) e.g. Term assurance (yes), household contents insurance (no) [2]

(iv) In the individual risk model the number of risks is specified and fixed; and the number of claims from each risk is restricted. This is not the case for the collective risk model. [2]

Individual risks are independent in the individual risk model, whilst the individual claims amounts are independent under the collective risk model. [1]

Claims in collective risk model are identical whereas the claims in each policy in the individual risk model are not. [1]
[Max 3]

(v) When claim amounts are in fact identically distributed. [1]

Well prepared candidates were able to score highly on this bookwork question, but a number of students were unable to demonstrate sufficient knowledge of the individual risk model.

Q7 (i) The density of the Poisson distribution is then:

$$f(y) = e^{-\lambda}\frac{\lambda^y}{y!} = e^{\frac{y\log\lambda - \lambda}{1} - \log y!}$$ [1]

Which to the general form $g(y) = e^{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)}$ it corresponds to [1]

$$\theta = \log\lambda, b(\theta) = \lambda = e^{\theta}, a(\phi) = \phi = 1, c(y,\phi) = -\log y!$$

[½ per element]

(ii) From the theory $E(Y) = b'(\theta) = \left(e^{\theta}\right)' = \lambda$ and $var(Y) = b''(\theta) = \lambda$ [2]

(iii) A GLM consists of three components:

- a distribution for the data (Poisson, exponential, gamma, normal or binomial) [1]
- a linear predictor (a function of the covariates that is linear in the parameters) [1]

- a link function (that links the mean of the response variable to the linear predictor). [1]

Most candidates scored very well on this straightforward GLM question.

Q8 (i) Under excess of loss, the reinsurer covers claims above a certain retention amount M [½]

Under proportional reinsurance, the reinsurer covers a proportion of claims α [½]

(ii) $$\frac{1}{e^{-\lambda M}}\int_M^\infty x\lambda e^{-\lambda x}dx$$ [1]

Let $t = x - M$

Then want

$$\frac{1}{e^{-\lambda M}}\int_0^\infty (t+M)\lambda e^{-\lambda(t+M)}dt = \frac{e^{-\lambda M}}{e^{-\lambda M}}\left\{\int_0^\infty t\lambda e^{-\lambda t}dt + M\int_0^\infty \lambda e^{-\lambda t}dt\right\} = E(X)+M$$

[2]

(iii) Premium = $5\% * (85\% * 1000 + 15\% * 1000) * 1.15 = \$57.50$ [2]

(iv) (a) Expected reinsurance cost per policy for A is $35\% * 10\% * 1000 * 5\% = \$1.75$ [½]

For B only need to consider Type II

$E(X|X > 1500) = 2500$ from part (i) [½]

So expected cost per policy for B is $45\% * 15\% * e^{-1.5} * (2500 - 1500) * 5\% = \$0.75$ [½]

So B is better under the Bayes criterion [½]

(b) B is clearly better under minimax, since total claims are capped [1]

(v) Unlikely that the assumption of independence holds, since events that lead to travel cancellations are likely to affect more than one policy holder. [2]
The reinsurer is a malevolent opponent so minimax is appropriate. [2]

[Max 2]

The first part aside, this question proved to be the hardest on the paper, with only the strongest candidates able to score well, particularly on parts (iv) and (v).

Q9 (i) The cumulative distribution function for $h(x)$ is:

$$F_h(x) = \int_0^x 2(1-u)\,du = 2x - x^2$$ [1]

the equation $F_h(X) = U$ is then equivalent to $2X - X^2 = U$ [½]

so $(X-1)^2 = 1-U$, $X = 1 \pm (1-U)^{\frac{1}{2}}$ [1½]

so the required inversion is then obtained. Algorithm is therefore:

1. Sample u from $U(0,1)$ [½]
2. Return $X = 1 - \sqrt{(1-U)}$ [½]

Since $0 < X < 1$

(ii) In this case we need to find first:

$$C = \max \frac{f(x)}{h(x)} = \frac{6x(1-x)}{2(1-x)} = 3x \qquad \text{i.e. } C = 3$$ [1½]

The rejection algorithm is then:

Sample x from $h(x)$ as in (i) [1]

Sample u from $U(0,1)$ [½]

If $u < 1/3\, f(x)/h(x) = x$ then set $y = x$, otherwise go to step 1. [1]

(iii) If we are to use the standard uniform $U(0,1)$ as envelope, consider $C_1 = \max(f(x))$.

$C_1 = f(1/2) = 3/2$. (max at $6 - 12x = 0$) [1½]

Therefore, since the acceptance rate is $1/C_1 = 2/3$ which is greater than

$C = \frac{1}{3}\ h(x)$ is less efficient. [1½]

> *Most candidates were able to score well here, although a number of students failed to convert to a cumulative distribution in part (i). There was a markedly improved performance in answers to the question on efficiency, relative to previous diets.*

Q10 i) Since $E(Y_t) = \mu_Y$ is constant for each t, $E(X_t) = a + bt + \mu_Y$. [1]

Since this mean depends on t then the time series X_t is not stationary. [1]

(ii) $\Delta X_t = b + \Delta Y_t$ (and since Y_t is stationary ΔX_t is) [1]

(iii) And the covariance function is

$\text{Cov}(\Delta X_t, \Delta X_{t-s}) = \text{Cov}(Y_t - Y_{t-1}, Y_{t-s} - Y_{t-s-1})$ [1]

$= \text{Cov}(Y_t,Y_{t-s}) + \text{Cov}(Y_{t-1},Y_{t-1-s}) - \text{Cov}(Y_t,Y_{t-s-1}) - \text{Cov}(Y_{t-1},Y_{t-s})$ [1]

$= \gamma^Y(s) + \gamma^Y(s) - \gamma^Y(s+1) - \gamma^Y(s-1)$ [1]

$= 2\gamma^Y(s) - \gamma^Y(s+1) - \gamma^Y(s-1)$ [1]

Where $\gamma^Y(s)$ represents the autocovariance of Y at s.

(iv) $Y_t = \varepsilon_t + \beta\varepsilon_{t-1}$, where ε_t is white noise with variance σ^2 then
$\Delta X_t = b + \varepsilon_t + \beta\varepsilon_{t-1} - \varepsilon_{t-1} - \beta\varepsilon_{t-2}$ [1]
or
$\Delta X_t = b + (1 - L)(1 + \beta L)\,\varepsilon_t$. [1]
[Max 1]

(v) In particular $Y_t = \varepsilon_t + \beta\varepsilon_{t-1}$ the corresponding auto-covariance function is $\gamma^Y(0) = (1+\beta^2)\sigma^2$ and $\gamma^Y(1) = \beta\sigma^2$. [2]

So from (iii) $var(\Delta(X)) = 2\gamma^Y(0) - 2\gamma^Y(1) = 2(1+\beta^2)\sigma^2 - 2\beta\sigma^2 = (1+\beta^2)\sigma^2 + (1-\beta)^2\sigma^2 > (1+\beta^2)\sigma^2 = \gamma^Y(0)$ [2]

> *Most candidates were able to score well on parts (i) and (ii), and although only stronger candidates were able to score well on parts (iii) and (v), this question presented few problems for those well prepared on this topic.*

Q11 (i) A prior distribution is a conjugate prior if the resulting posterior distribution belongs to the same family as the prior distribution. [1]

(ii) Assume $f_{\text{prior}}(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}$

likelihood given by $L(p) = \prod_{i=1}^{n} p(1-p)^{x_i} = p^n (1-p)^{\sum x_i}$ [1]

posterior $\propto p^{n+\alpha-1}(1-p)^{\sum x_i+\beta-1}$ [1]

which is in the form of a beta distribution with parameters $n+\alpha$ and $\sum x_i + \beta$ [1]

(iii) $E\left(\frac{1-p}{p}\right) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 \left(\frac{1-p}{p}\right) p^{\alpha-1}(1-p)^{\beta-1}\,dp$ [½]

$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha-1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta)} \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha-1)\Gamma(\beta+1)} p^{\alpha-2}(1-p)^{\beta}\,dp$ [1½]

since the integrand is 1 (as it is a beta distribution), this is

$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha-1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta)} = \frac{\beta}{(\alpha-1)}$ [1]

(iv) From part (ii), know that the posterior distribution of p is beta with parameters $n+\alpha$ and $\sum x_i + \beta$

from part (iii), know the posterior mean is $\frac{\sum x_i + \beta}{(n+\alpha-1)}$ [½]

We can express this as $\frac{\sum x_i}{n} \frac{n}{n+a-1} + \frac{\beta}{(\alpha-1)} \frac{a-1}{n+a-1}$ [1]

Which is in the form of a credibility estimate with $Z = \frac{n}{n+a-1}$ [½]

(v) For Amit, $\beta_A = E\left(\frac{1-p}{p}\right) * (\alpha-1) = 10*4 = 40$ [½]

For Bonnie, $\beta_B = E\left(\frac{1-p}{p}\right) * (\alpha-1) = 5*4 = 20$ [½]

Want the difference in posterior mean to be less than 0.01, and the first term in the credibility estimate is the same for both since it is independent of β. [1]

So want $\left| \frac{\beta_A}{(\alpha-1)} \frac{a-1}{n+a-1} - \frac{\beta_B}{(\alpha-1)} \frac{a-1}{n+a-1} \right| < 0.5$ [1]

i.e. $\left| \frac{40}{n+4} - \frac{20}{n+4} \right| < 0.5$ [½]

i.e. $20 < 0.5(n+4)$ [½]

so $n > 36$ i.e. $n = 37$ [1]

Most candidates were able to score well on parts (i) and (ii), and well-prepared candidates also scored well on part (iii).
Unfortunately, there was an ambiguity in part (iv), which should have referred explicitly to the posterior distribution derived in part (ii). Full credit was given to any reasonable attempt whilst using the original distribution.
Part (v) was particularly difficult for those who had not derived the intended credibility estimate in part (iv), and only the very strongest candidates were able to score highly here. This has been reflected in the lower pass mark for this diet.

END OF EXAMINERS' REPORT