INSTITUTE AND FACULTY OF ACTUARIES

EXAMINERS' REPORT

September 2014 examinations

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 at the date the examination was set. Candidates should take into account the possibility that circumstances may have changed if using these reports for revision.

F Layton
Chairman of the Board of Examiners

November 2014

General comments on Subject CT6

The examiners for CT6 expect candidates to be familiar with basic statistical concepts from CT3 and so to be comfortable computing probabilities, means, variances etc. for the standard statistical distributions. Candidates are also expected to be familiar with Bayes' Theorem, common types of reinsurance, and risk models, and to be able to apply it to given situations. Many of the weaker candidates are not familiar with this material.

The examiners will accept valid approaches that are different from those shown in this report. In general, slightly different numerical answers can be obtained depending on the rounding of intermediate results, and these will still receive full credit. Numerically incorrect answers will usually still score some marks for method, providing candidates set their working out clearly.

Comments on the September 2014 paper

The examiners felt that this paper was generally better answered than recent papers. The quality of solutions was often good, with questions 3 and 7 providing the greatest challenge to most students.

There was a slight issue with Question 8 where the wording was not completely clear. All sensible attempts by candidates were given full credit. In addition the examiners reviewed scripts carefully to ensure that this issue would not have adversely affected candidates' final grades.

1 $E(X) = \sum_{i=1}^{240} p_i \times E(X_i)$

$$= \sum_{i=1}^{240} p_i \times \frac{100}{p_i}$$

$$= \sum_{i=1}^{240} 100$$

$$= 24000$$

Let Y_i denote the claim in the i^{th} policy. Then

$$Y_i = \begin{cases} 0 & \text{with probability } 1 - p_i \\ \text{Exp}\left(\dfrac{p_i}{100}\right) & \text{with probability } p_i \end{cases}$$

so $E(Y_i) = p_i \times \dfrac{100}{p_i} = 100$

and $E(Y_i^2) = p_i \times 2 \times \dfrac{100^2}{p_i^2} = \dfrac{20000}{p_i}$

so $\text{Var}(Y_i) = \dfrac{20000}{p_i} - 100^2$

$$= 10,000\left(\frac{2}{p_i} - 1\right)$$

$$= 10,000\left(\frac{2 - p_i}{p_i}\right)$$

and $\text{Var}(X) = \sum_{i=1}^{240} 10,000\left(\dfrac{2 - p_i}{p_i}\right)$

$$= 10,000 \sum_{i=1}^{240}\left(\frac{2 - p_i}{p_i}\right)$$

[6]

Full credit was also given to candidates who used standard individual risk model results. Many candidates scored well here although a disappointing number struggled to derive the variance.

2 (i) The three main components are:

- the distribution of the responsible variable
- a linear predictor of the covariates
- a link function between the response variable and the linear predictor

[3]

Other sensible points received full credit.

(ii) A saturated model has as many parameters as there are data points and is therefore a perfect fit to the data.

It is not useful from a predictive point of view which is why it is not used in practice.

It is, however, a useful benchmark against which to compare the fit of other models.

[3]
[Total 6]

This standard bookwork question was reasonably well answered.

3 (i) The four decision functions are:

d_1 – choose the gearbox regardless
d_2 – choose the gearbox if the car stops and the engine otherwise
d_3 – choose the engine if the car stops and the gearbox otherwise
d_4 – choose the engine regardless [2]

(ii) Let θ_1 = state of nature where gearbox is at fault
θ_2 = state of nature where engine is at fault

Let $R(d_i, \theta_j) = E(L(d_i, \theta_j))$

Then the expected loss matrix is:

	d_1	d_2	d_3	d_4
θ_1	0	3	2	5
θ_2	1	0.9	0.1	0

where

$R(d_2, \theta_1) = 0.6 \times 5 = 3$
$R(d_2, \theta_2) = 0.9 \times 1 = 0.9$
$R(d_3, \theta_1) = 0.4 \times 5 = 2$
$R(d_4, \theta_2) = 0.1 \times 1 = 0.1$

It is clear that d_2 is dominated by d_3. [3]

(iii) Under Bayes criteria, we need to minimise the expected loss.

Expected losses are $E(L(d_1)) = 0.p + 1.(1 - p) = 1 - p$
$E(L(d_3)) = 2p + 0.1(1 - p) = 1.9p + 0.1$
$E(L(d_4)) = 5p + 0.(1 - p) = 5p$

We need to choose p so that d_3 has the lowest expected loss, i.e.

$1.9p + 0.1 < 1 - p$ i.e. $2.9p < 0.9$ i.e. $p < 0.3103$

and

$1.9p + 0.1 < 5p$ i.e. $0.1 < 3.1p$ i.e. $p > 0.03226$

so we need $0.03226 < p < 0.3103$

$$\left[\frac{1}{31} < p < \frac{9}{29}\right].$$ [4]

[Total 9]

This Bayes' Criterion question was very disappointingly answered, with only the best candidates managing to calculate the correct answer to part (iii).

4 (i) Using the inversion method, set

$$u = F(x) = 1 - e^{-\frac{x}{10}}$$

i.e. $1 - u = e^{-\frac{x}{10}}$

i.e. $-\log(1 - u) = \frac{x}{10}$

i.e. $x = -10\log(1 - u)$

so the algorithm is:

Step 1 Generate a sample u from a $U(0,1)$ distribution.
Step 2 Set $x = -10\log(1 - u)$.

[3]

(ii) Again using the inversion method, set

$$u = F(x) = 1 - e^{-x^4}$$

i.e. $1 - u = e^{-x^4}$

i.e. $-x^4 = \log(1 - u)$

i.e. $x = [-\log(1 - u)]^{1/4}$

so the algorithm is

Step 1 Generate a sample u from a $U(0,1)$ distribution.
Step 2 Set $x = [-\log(1 - u)]^{1/4}$.

[3]

(iii) Our algorithm is as follows:

Step 1 Generate a sample u from a $U(0,1)$ distribution.

Step 2 If $0 \le u \le 0.4$ then total claim amount $X = 0$ and stop
Else continue to step 3.

Step 3 If $0.4 < u \le 0.8$ then simulate a claim from Exp(1/10) distribution using the algorithm in (i) and set X = this value and stop.

Else go to step 4.

Step 4 Simulate claims using the algorithms in (i) and (ii) and set X = total of the two simulated claims.

[4]
[Total 10]

This question was well answered.

5 (i) The cumulative cost of claims is given by:

Accident year	*Development year* *0*	*1*	*2*
2011	2,233	3,622	4222
2012	3,380	5,188	
2013	4,996		

Dividing by cumulative claim numbers:

Accident year	*Development year* 0	1	2
2011	15.950	17.842	18.848
2012	18.778	22.557	
2013	19.516		

using grossing up factors to estimate the ultimate average cost per claim for each accident year:

Accident year	*Development year* 0	1	2
2011	84.623% 15.950	94.663% 17.842	100% 18.848
2012	78.805% 18.778	94.663% 22.557	 23.828
2013	81.714% 19.516		 23.883

Taking the same approach for the claim numbers gives:

Accident year	*Development year* 0	1	2
2011	62.5% 140	90.625% 203	100% 224
2012	70.924% 180	90.625% 230	 253.8
2013	66.712% 256		 383.7

Total outstanding claims are therefore

$$253.8 \times 23.828 + 383.7 \times 23.883 - 5188 - 4996$$
$$= \underline{5028.2}$$ [7]

(ii) **Assumptions**

- The number of claims relating to each development year is a constant proportion of the total claim numbers from the relevant accident year.
- Claim amounts for each development year are a constant proportion of the total claim amount for the relevant accident year.
- Claims are fully run off after development year 2.

[3]
[Total 10]

Alternative valid points received full credit. This question was well answered, although many candidates scored poorly on part (ii).

6 $\bar{X}_1$ = (8130 + 9210 + 8870) / 3 = 8736.67

$\bar{X}_2$ = (7420 + 6980 + 8130) / 3 = 7510

$\bar{X}_3$ = (9070 + 8550 + 7730) / 3 = 8450

$\bar{X}$ = (8736.67 + 7510 + 8450) / 3 = 8232.22

$$s_1^2 = \sum_{j=1}^{3}(X_{ij}-\bar{X}_1)^2 = (8130-8736.67)^2+(9210-8736.67)^2 + (8870-8736.67)^2 = 609866.65$$

Similarly s_2^2 = 673,400

s_3^2 = 912,800

$$E(s^2(\theta)) = \frac{1}{3\times 2}(s_1^2+s_2^2+s_3^2) = 366{,}011.11$$

$$\text{Var}[m(\theta)] = \frac{1}{2}\left\{(8736.67-8232.22)^2+(7510-8232.22)^2 + (8450-8232.22)^2\right\} - \frac{1}{3}E(s^2(\theta))$$

= 411,749.83 – 366,011.11 / 3

= 289,746.13

so $Z = \dfrac{3}{3+\dfrac{366011.11}{289746.13}} = 0.7037$

so we have	Town 1	0.7037 × 8736.67	+ 0.2963 × 8232.22	= 8587.2
	Town 2	0.7037 × 7510	+ 0.2963 × 8232.22	= 7724.0
	Town 3	0.7037 × 8450	+ 0.2963 × 8232.22	= 8385.5

[10]

This question was well answered with many candidates scoring full marks.

7 (i) $\int_d^{L+d} xf(x)dx$

$$= \int_d^{L+d} \frac{\alpha\lambda^\alpha x}{(\lambda+x)^{\alpha+1}}\,dx$$

$$= \left[-\frac{\lambda^\alpha x}{(\lambda+x)^\alpha}\right]_d^{L+d} + \int_d^{L+d} \frac{\lambda^\alpha}{(\lambda+x)^\alpha}\,dx$$

$$= \lambda^\alpha\left[\frac{d}{(\lambda+d)^\alpha} - \frac{(L+d)}{(\lambda+L+d)^\alpha}\right] + \left[-\frac{\lambda^\alpha}{(\alpha-1)(\lambda+x)^{\alpha-1}}\right]_d^{L+d}$$

$$= \lambda^\alpha\left[\frac{d}{(\lambda+d)^\alpha} - \frac{(L+d)}{(\lambda+L+d)^\alpha} - \frac{1}{(\alpha-1)(\lambda+L+d)^{\alpha-1}} + \frac{1}{(\alpha-1)(\lambda+d)^{\alpha-1}}\right]$$

$$= \lambda^\alpha\left\{\frac{d(\alpha-1)+(\lambda+d)}{(\alpha-1)(\lambda+d)^\alpha} - \frac{(L+d)(\alpha-1)+(\lambda+L+d)}{(\alpha-1)(\lambda+L+d)^\alpha}\right\}$$

$$= \frac{\lambda^\alpha}{\alpha-1}\left\{\frac{d\alpha+\lambda}{(\lambda+d)^\alpha} - \frac{\alpha(L+d)+\lambda}{(\lambda+L+d)^\alpha}\right\}.$$

[5]

(ii) We first solve for the parameter values:

$$\frac{\lambda}{\alpha-1} = 16{,}000$$

$$\frac{\alpha\lambda^2}{(\alpha-1)^2(\alpha-2)} = 20{,}000^2$$

so $\frac{\alpha}{\alpha-2}\times\left(\frac{\lambda}{\alpha-1}\right)^2 = 20{,}000^2$

so $\frac{\alpha}{\alpha-2} = \frac{20{,}000^2}{16{,}000^2} = 1.5625$

so $\alpha = 1.5625(\alpha - 2)$

so $\alpha = 2 \times \frac{1.5625}{0.5625} = 5.555$

and $\lambda = 16{,}000 \times (\alpha - 1)$

$= 72{,}888.89$

Now denote by Z the amount paid by the reinsurer.

Then $P(Z > 0) = P(X > 40{,}000) = 1 - F(40{,}000)$

$$= \left(\frac{\lambda}{\lambda + 40{,}000}\right)^{\alpha}$$

$$= \left(\frac{72{,}888.89}{112{,}888.89}\right)^{5.5555}$$

$= 0.088004$

Now $E(Z) = \int_{40000}^{65000} (x - 40000) f(x)dx + \int_{65000}^{\infty} 25000 f(x)dx$

$$= \int_{40000}^{65000} xf(x)dx - 40000\int_{40000}^{65000} f(x)dx + 25000P(X > 65000)$$

$$\frac{72888.89^{5.5555}}{4.5555}\left\{\frac{40000 \times 5.5555 + 72888.89}{112888.89^{5.5555}} - \frac{65000 \times 5.5555 + 72888.89}{137888.89^{5.5555}}\right\} - 40000(F(65000) - F(40000))$$

$$+25000(1 - F(65000))$$

$$= 2941.71 - 40000\left(\left(\frac{72888.89}{112888.89}\right)^{5.5555} - \left(\frac{72888.89}{137888.89}\right)^{5.5555}\right)$$

$$+25000 \times \left(\frac{72888.89}{137888.89}\right)^{5.5555}$$

$= 2941.71 - 2361.67 + 724.10$

$= 1304.14$

and so $E[Z|Z > 0] = \dfrac{1304.14}{0.088004} = 14{,}819.10$

[8]

(iii) Pareto with parameters $\alpha = 5.5555$ and $\lambda = 72{,}888.89 \times 1.05$

$= \underline{76{,}533.33}$

[1]
[Total 14]

Along with question 3, candidates typically found this the hardest question on the paper to answer. Although many candidates were able to calculate the parameters in part (ii), only the better candidates were able to work through the integration and calculate the final result.

8 *Note: the question should have read "… premiums of c per claim per year", rather than "per policy". This would have meant the equation in (i) simplified to $1 + cR = M_x(R)$.*

(i) R is the solution to

$$\lambda + ncR = \lambda M_x(R), \text{ where } n \text{ is the number of policies}$$

Note: Full credit also given for $\lambda + cR = \lambda M_x(R)$ and $1 + cR = M_x(R)$

Note: The solution shown in part (ii) is based on the equation $1 + cR = M_x(R)$

[1]

(ii) $1 + cR = E(e^{XR})$

$$= E\left(1 + RX + \frac{R^2X^2}{2} + \ldots\right)$$

$$= 1 + RE(X) + \frac{R^2}{2}E(X^2) + \ldots$$

Now $E(X) = \mu$

and $E(X^2) = \text{Var}(X) + E(X)^2 = \sigma^2 + \mu^2$ so we have

$$1 + cR = 1 + R\mu + \frac{R^2}{2}(\mu^2 + \sigma^2) + \ldots$$

truncating at the term involving R^2 gives

$$\cancel{1} + c\hat{R} = \cancel{1} + \mu\hat{R} + \frac{\hat{R}^2}{2}(\mu^2 + \sigma^2)$$

i.e. $c = \mu + \frac{\hat{R}}{2}(\mu^2 + \sigma^2)$

$$\hat{R} = \frac{2(c-\mu)}{\mu^2 + \sigma^2}$$

[4]

Note: if candidates assumed that $\lambda + cR = \lambda M_x(R)$, the alternative correct solution receiving full credit is $\hat{R} = \frac{2(c-\lambda\mu)}{\lambda(\mu^2+\sigma^2)}$.

If candidates assumed that $\lambda + ncR = \lambda M_x(R)$, the alternative correct solution receiving full credit is $\hat{R} = \frac{2(nc-\lambda\mu)}{\lambda(\mu^2+\sigma^2)}$.

For part (iii), most candidates used the formula given in the question, although full credit was given to candidates who used the alternative formulae above and then correctly worked through the reinsurance outcomes, whether or not they left their answers in terms of λ and n, or set them to be some sensible value.

(iii) (A) We have $E(X) = 10 \times 0.3 + 20 \times 0.5 + 50 \times 0.15 + 100 \times 0.05$

$= 25.5$

and $E(X^2) = 10^2 \times 0.3 + 20^2 \times 0.5 + 50^2 \times 0.15 + 100^2 \times 0.05$

$= 1105$

Here $c = 25.5 \times 1.3 = 33.15$

and so $\hat{R} = \frac{2(33.15 - 25.5)}{1105} = 0.013846$

(B) We now have $\mu = 0.7 \times 25.5 = 17.85$

$\mu^2 + \sigma^2 = E((0.7X)^2) = 0.7^2 \times 1105 = 541.45$

and $c = 33.15 - 0.3 \times 25.5 \times 1.2$

$= 33.15 - 9.18$

$= 23.97$

and so $\hat{R} = \frac{2(23.97 - 17.85)}{541.45} = 0.02261$

(C) We now have $\mu = 10 \times 0.3 + 20 \times 0.5 + 50 \times 0.15 + 70 \times 0.05$

$= 24$

and $\mu^2 + \sigma^2 = 10^2 \times 0.3 + 20^2 \times 0.5 + 50^2 \times 0.15 + 70^2 \times 0.05$

$= 850$

the reinsurer charges premiums of $30 \times 0.05 \times 1.4 = 2.1$

so $c = 33.15 - 2.1 = 31.05$

and $\hat{R} = \dfrac{2(31.05 - 24)}{850} = 0.01659$

The higher the adjustment coefficient the lower the probability of ruin, so approach B gives the lowest probability of ruin. [10]

(iv) It is clear that B is better than A since the reinsurer's premium loading is lower than the insurer's. So we have a 30% reduction in claims but a lower than 30% reduction in premiums.

The excess of loss reinsurance in C does reduce risk relative to A but not as much as B. This will be a combination of the relatively high retention and the reinsurer's premium loading being higher than the insurer's. [2]
[Total 17]

Full credit was given for alternative comments reflecting the answers derived by candidates using the alternative formulae.

Despite the issue with the wording in the question, many candidates scored well on this question.

9 (i) The three main steps are:

- Model identification
- Parameter estimation
- Diagnostic checking

[3]

(ii) $\hat{\gamma}_0 = \dfrac{35.4}{200} = 0.177$

$\hat{\gamma}_1 = \dfrac{28.4}{200} = 0.142$

$\hat{\gamma}_2 = \dfrac{17.1}{200} = 0.0855$

[3]

(iii) $\hat{\phi}_1 = \hat{\rho}_1 = \dfrac{\hat{\gamma}_1}{\hat{\gamma}_0} = \dfrac{0.142}{0.177} = 0.8023$

$$\hat{\phi}_2 = \frac{\hat{\rho}_2 - \hat{\rho}_1^2}{1-\hat{\rho}_1^2} = \frac{\frac{0.0855}{0.177} - 0.8023^2}{1-0.8023^2} = -0.4506$$

[3]

(iv) Firstly $\hat{\mu} = \overline{x} = \dfrac{83.7}{200} = 0.4185.$

The Yule-Walker equations for this model give

$$\gamma_0 = a_1\gamma_1 + \sigma^2$$

$$\gamma_1 = a_1\gamma_0$$

so we have $\hat{a}_1 = \dfrac{\hat{\gamma}_1}{\hat{\gamma}_0} = \hat{\phi}_1 = 0.8023$

and $\hat{\sigma}^2 = \hat{\gamma}_0 - \hat{a}_1\hat{\gamma}_1 = 0.177 - 0.8023 \times 0.142 = 0.0631$ [5]

(v) The number of turning points T is approximately Normally distributed with

$$E(T) = \frac{2}{3}(N-2) = \frac{2}{3}\times 198 = 132$$

$$\text{Var}(T) = \frac{16N-29}{90} = \frac{16\times 200-29}{90} = 35.2333 = 5.936^2$$

so a 95% confidence interval for T is

$$[132 - 1.96 \times 5.936,\ 132 + 1.96 \times 5.936] = [120.4,\ 143.6]$$

We are testing

H_0 : observed $\hat{e}_t$ are from a white noise process
H_1 : observed $\hat{e}_t$ are <u>not</u> from a white noise process

Our observed value T = 110 does not lie within the 95% confidence interval. Therefore we have evidence to reject the H_0 and conclude that the observed $\hat{e}_t$ to not come from a white noise process.

A different model is required. [4]

[Total 18]

Full credit was given for considering p-values or significant values and also to candidates who applied a continuity correction.

Unusually for a time series question this was well answered by many candidates.

END OF EXAMNERS' REPORT