INSTITUTE AND FACULTY OF ACTUARIES

EXAMINERS' REPORT

April 2015 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

June 2015

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 April 2015 paper

The examiners felt that this paper was generally not as well answered as recent papers. A disappointing number of candidates were seemingly unfamiliar with core concepts such as Bayes' Theorem or EBCT Model 2. Well prepared candidates were typically able to score well, although some struggled with unfamiliar applications of theory.

1 (i) III will dominate II if $X \le -6$.

III will dominate IV if $X \le Y$.

I will dominate IV if $Y \ge -8$

I does not dominate II and vice versa (since I is better if Player B chooses 1 and II is better if Player B chooses 2).

III cannot dominate I (since I is better if Player B chooses 1).

IV cannot dominate any strategy since it gives the worst outcome if Player B chooses 3.

Similarly II cannot dominate any strategy since it gives the worst result if Player B chooses 1.

So there will exist dominated strategies if $X \le -6$ or $X \le Y$ or $Y \ge -8$.

(ii) A saddle point exists if an entry is both the largest in its column and the smallest in its row.

This can only occur in row 2 (the smallest values in rows 1 and 3 are not the largest in their columns).

X cannot give a saddle since this would require $X \le -8$ (to be the smallest in row 2) but then X would not be the largest in its column.

Equally Y cannot give a saddle point as this would require that $Y \le -8$ in which case Y would not be the largest in its column.

So there are no values of X and Y which give a saddle point.

This straightforward question was relatively poorly answered, with many candidates seemingly put off by the unfamiliar nature of the question.

2 The development factors are:

$$DF_{2,3} = \frac{2207}{2106} = 1.0480$$

$$DF_{1,2} = \frac{(2106+2985)}{(1969+2186)} = 1.2253$$

$$DF_{0,1} = \frac{(1969+2186+1924)}{(1509+1542+1734)} = 1.2704$$

The initial ultimate loss for 2014 is 0.935 × 4013 = 3752.16.

Total emerging liability = initial UL × $\left(1-\dfrac{1}{f}\right)$

$$= 3752.16 \times \left(1-\frac{1}{1.0480\times1.2253\times1.2704}\right)$$

$$= 1452.01$$

Total claims = 1452.01 + 1773

= 3225.01

This standard chain ladder question was very well answered by the majority of candidates.

3 (i) (a) To protect itself from the risk of large claims.

(b)

- Excess of loss reinsurance where the reinsurer pays any amount of a claim above the retention.
- Proportional reinsurance where the reinsurer pays a fixed proportion of any claim.

(ii) We must first find the parameters α and λ of the Pareto distribution.

$$\frac{\lambda}{\alpha-1} = 270 \text{ and } \frac{\alpha\lambda^2}{(\alpha-1)^2(\alpha-2)} = 340^2$$

$$\frac{\alpha}{\alpha-2}\times\frac{\lambda^2}{(\alpha-1)^2} = 340^2$$

so $\dfrac{\alpha}{\alpha-2} = \dfrac{340^2}{270^2} = 1.585733882$

so $\alpha = \dfrac{2\times1.585733882}{1.585733882-1}$

= 5.4145

and λ = 270 × 4.4145 = 1191.920375

We need to find M such that $P(X > M) = 0.05$

i.e. $$\left(\frac{\lambda}{\lambda+M}\right)^{\alpha} = 0.05$$

$$\frac{\lambda}{\lambda+M} = 0.05^{\frac{1}{\alpha}}$$

$$\lambda = 0.05^{\frac{1}{\alpha}}(\lambda+M)$$

$$M = \frac{\lambda\left(1-0.05^{\frac{1}{\alpha}}\right)}{0.05^{\frac{1}{\alpha}}}$$

$$= \frac{1191.920375\times\left(1-0.05^{\frac{1}{5.4145}}\right)}{0.05^{\frac{1}{5.4145}}}$$

$= \underline{\underline{880.8}}$

This question was the best answered on the paper with most candidates scoring well. Some candidates were unable to manipulate the Pareto distribution.

4 (i) X has an exponential distribution with parameter 1.

Let u be a sample from a $U(0,1)$ distribution.

Then using the inverse transform method we set

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

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

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

so the algorithm is

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

(ii) We first find $C = \max_{x>0} \frac{h(x)}{f(x)}$

$$= \max_{x>0} \frac{2xe^{-x^2}}{e^{-x}}$$

$$= \max_{x>0} 2xe^{x-x^2}$$

To find the maximum consider $g(x) = 2xe^{x-x^2}$

then $\log g(x) = \log 2 + \log x + x - x^2$

$$\frac{d \log g(x)}{dx} = \frac{1}{x} + 1 - 2x$$

setting this equal to zero we have

$$\frac{1}{x} + 1 - 2x = 0$$

i.e. $1 + x - 2x^2 = 0$

$2x^2 - x - 1 = 0$

$(2x + 1)(x - 1) = 0$

$x = 1$ or $x = -½$

but $x > 0$ so $C = 2 \times 1 \times e^{1-1} = 2$

so now set $w(x) = \frac{h(x)}{2f(x)} = xe^{x-x^2}$

and our algorithm is

Step 1 Generate u from $U(0,1)$ distribution.
Step 2 Generate $x = -\log(1 - u)$.
Step 3 Generate v from $U(0,1)$.
Step 4 If $v > w(x)$ return to step 1 else return x.

(iii) We accept a proportion $\frac{1}{c} = \frac{1}{2}$ so on average 2 simulations are needed to return 1 value, so in this case we need 20 simulations. Each simulation requires 2 pseudo-random numbers so on average we will need 40 pseudo-random numbers.

Most candidates were able to produce good quality answers to part (i), although the majority struggled with part (ii). Candidates with the confidence to apply the acceptance-rejection method scored well.

5 $\bar{P}_1 = 17 + 23 + 21 + 29 + 35 = 125$

$\bar{P}_2 = 42 + 51 + 60 + 55 + 37 = 245$

$\bar{P}_3 = 43 + 31 + 62 + 98 + 107 = 341$

$\bar{P} = 125 + 245 + 341 = 711$

$$\bar{X} = \frac{(850\times125+720\times245+900\times341)}{711} = 829.18$$

expected claims per policy for risk 1 next year = $\frac{25,200}{30} = 840$

so $840 = Z_1 \times 850 + (1 - Z_1) \times 829.18$

$$Z_1 = \frac{840-829.18}{850-829.18} = 0.519594$$

so $$\frac{125}{125+\frac{E(s^2(\theta))}{\text{Var}[m(\theta)]}} = 0.51969$$

so $$\frac{E(s^2(\theta))}{\text{Var}[m(\theta)]} = \frac{125-0.51969\times125}{0.51969} = 115.57217$$

so $$Z_2 = \frac{245}{245+115.528} = 0.6794756$$

$$Z_3 = \frac{341}{341+115.528} = 0.74686$$

So credibility premium per policy are

Type 2: 0.67956 × 720 + (1 – 0.67956) × 829.18 = 755.0
Type 3: 0.74694 × 900 + (1 – 0.74694) × 829.18 = 882.1

so overall expected claims

Type 2: 754.98 × 40 = 30,200
Type 3: 882.08 × 110 = 97,028

Candidates with good knowledge of EBCT Model 2 scored well here, however a disappointing number of candidates were apparently unfamiliar with this method.

6 (i) The likelihood for type 1 policies is given by

$$l = \frac{e^{-4\mu_1}\mu_1^{11}}{\prod_{j=1}^{4} y_{ij}!}$$

Taking logarithms gives

$$L = \log l = -4\mu_1 + 11\log\mu_1 - \sum_{j=1}^{4}\log y_{ij}!$$

so $\dfrac{\partial L}{\partial \mu_1} = -4 + \dfrac{11}{\mu}$

and setting $\dfrac{\partial L}{\partial \mu_1} = 0$ we have

$$-4 + \frac{11}{\hat{\mu}_1} = 0$$

i.e. $\hat{\mu}_1 = \dfrac{11}{4} = 2.75$

similarly $\hat{\mu}_2 = \dfrac{16}{4} = 4$

$\hat{\mu}_3 = \dfrac{20}{4} = 5$

(ii) Assuming $\mu = \mu_1 = \mu_2 = \mu_3$ we have $\hat{\mu} = \dfrac{11+16+20}{12} = 3.916667$

The difference in scaled deviance is given by

$$\Delta = 2(\log L_1 + \log L_2 + \log L_3 - \log L)$$

$$= 2(-4\hat{\mu}_1 + 11\log\hat{\mu}_1 - 4\hat{\mu}_2 + 16\log\hat{\mu}_2 - 4\hat{\mu}_3 + 20\log\hat{\mu}_3$$
$$+12\hat{\mu} - 47\log\hat{\mu})$$

[The logarithms of factorials cancel]

$= 2(–47 + 11 \log 2.75 + 16 \log 4 + 20 \log 5 + 47 – 47 \log 3.91667)$

$= 2.6615$

Under H_0: $\mu_1 = \mu_2 = \mu_3$ we have that Δ comes from a χ^2 distribution with $3 – 1 = 2$ degrees of freedom.

The upper 5% point of the χ^2_2 distribution is 5.991. The observed value is below this and so there is no evidence to suggest that the underlying parameters are different for each risk.

Most candidates were able to score well on part (i), although only the better prepared candidates were able to complete part (ii).

7 (i) Set $X_t = (1 - B^{12})(1 - B)\, Y_t$ where B is the background shift operator

i.e. $X_t = Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13}$

then we have $X_t = e_t + \beta_1 e_{t-1} + \beta_{12} e_{t-12} + \beta_1\beta_{12}e_{t-13}$

$= (1 + \beta_1 B)(1 + \beta_{12}B^{12})e_t$

which is a moving average process [of order 13].

(ii) This is called seasonal differencing because it compares the monthly change in Y_t with the corresponding monthly change at the same time last year.

(iii) We can see that

$$\gamma_0 = \text{Cov}(X_t, X_t) = (1+\beta_1^{\,2}+\beta_{12}^{\,2}+\beta_1^{\,2}\beta_{12}^{\,2})\sigma^2 = (1+\beta_1^{\,2})(1+\beta_{12}^{\,2})\sigma^2$$

$$\gamma_1 = \text{Cov}(X_t, X_{t-1}) = \text{Cov}(e_t + \beta_1 e_{t-1} + \beta_{12} e_{t-12} + \beta_1\beta_{12} e_{t-13};\; e_{t-1} + \beta_1 e_{t-2} + \beta_{12} e_{t-13} + \beta_1\beta_{12} e_{t-14})$$

$$= (\beta_1 + \beta_1\beta_{12}^{\,2})\sigma^2 = \beta_1(1+\beta_{12}^{\,2})\sigma^2$$

$$\gamma_{11} = \text{Cov}(X_t, X_{t-11}) = \text{Cov}(e_t + \beta_1 e_{t-1} + \beta_{12} e_{t-12} + \beta_1\beta_{12} e_{t-13};\; e_{t-11} + \beta_1 e_{t-12} + \beta_{12} e_{t-23} + \beta_1\beta_{12} e_{t-24})$$

$$= \beta_1\beta_{12}\,\sigma^2$$

$$\gamma_{12} = (\beta_{12} + \beta_1^{\,2}\beta_{12})\sigma^2 = \beta_{12}(1+\beta_1^{\,2})\sigma^2$$

$$\gamma_{13} = \beta_1\beta_{12}\,\sigma^2$$

and $\gamma_s = 0$ for all other values of s.

so $$\rho_1 = \frac{\beta_1(1+\beta_{12}^2)}{(1+\beta_1^2)(1+\beta_{12}^2)} = \frac{\beta_1}{1+\beta_1^2}$$

$$\rho_{11} = \rho_{13} = \frac{\beta_1\beta_{12}}{(1+\beta_1^2)(1+\beta_{12}^2)}$$

$$\rho_{12} = \frac{\beta_{12}(1+\beta_1^2)}{(1+\beta_1^2)(1+\beta_{12}^2)} = \frac{\beta_{12}}{1+\beta_{12}^2}$$

and $\rho_0 = 1$ and $\rho_s = 0$ for all other s.

Most candidates were able to identify this as a moving average process, however only the strongest candidates were able to work through the algebra to derive the auto-correlation function.

8 (i) First note that N has a type 2 negative binomial distribution with parameters $p = 0.8$ and $k = 1$. Hence

$$E(N) = \frac{0.2}{0.8} = 0.25$$

$$\text{Var}(N) = \frac{0.2}{0.8^2} = 0.3125$$

Let X denote the distribution of an individual claim. Then

$$E(X) = \frac{\lambda}{\alpha - 1} = \frac{1000}{4} = 250$$

$$\text{Var}(X) = 250^2 \times \frac{5}{3} = 104{,}166.666 = (322.75)^2$$

Now let S denote aggregate annual claims. Then

$$E(S) = E(N)E(X) = 0.25 \times 250 = 62.5$$

$$\text{Var}(S) = E(N)\,\text{Var}(X) + \text{Var}(N)\,E(X)^2$$

$$= 0.3125 \times 250^2 + 0.25 \times 104{,}166.666$$

$$= 45{,}572.92 = 213.478^2$$

(ii) (a) $P(S > 400) = P(N\,(62.5,\, 213.478^2) > 400)$

$$= P\left(N(0,1) > \frac{400 - 62.5}{213.478}\right)$$

$$= P(N(0,1) > 1.581)$$

$$= 1 - [0.94295 \times 0.9 + 0.1 \times 0.94408]$$

$$= \underline{0.0569}$$

(b) Let μ and σ be the parameters of the underlying Normal distribution. Then

$$e^{\mu + \frac{\sigma^2}{2}} = 62.5 \qquad \text{(A)}$$

$$e^{2\mu + \sigma^2}(e^{\sigma^2} - 1) = 213.478^2 \qquad \text{(B)}$$

$$(B) \div (A)^2 \Rightarrow e^{\sigma^2} - 1 = \frac{213.478^2}{62.5^2} = 11.66665$$

$$\sigma^2 = \log 12.66665 = 2.53897 = 1.5934^2$$

substituting into (A) $\quad \mu + \dfrac{2.53897}{2} = \log 62.5$

so $\mu = \log 62.5 - \dfrac{2.53897}{2} = 2.8657$

and so $P(S > 400) = P(N(2.8657, 1.5934^2) > \log 400)$

$$= P\left(N(0,1) > \frac{\log 400 - 2.8657}{1.5934}\right)$$

$$= P(N(0,1) > 1.9617)$$

$$= 1 - (0.17 \times 0.97558 + 0.83 \times 0.97500)$$

$$= 0.0249$$

(iii) The Pareto distribution is significantly skewed and the Normal approximation is not. The Normal approximation in (ii)(b) has variance 213.48^2 and mean 62.5, so negative values of S (which are impossible in reality) are less than 1 standard deviation from the mean.

The approximation in (ii)(b) will therefore be more reliable.

This question was well answered by the majority of candidates. Full credit was given to alternative correct answers in part (iii).

9 (i) Consider the loss function given by:

$$L(g(\underline{x})|p) = \begin{cases} 0 & p - \varepsilon \le g(\underline{x}) \le p + \varepsilon \\ 1 & \text{otherwise} \end{cases}$$

The expected loss is

$$E(L) = 1 - \int_{g-\varepsilon}^{g+\varepsilon} f(p \mid \underline{x})\,dp$$

$$= 1 - 2\varepsilon f(g|\underline{x})$$

To minimise this expression we need to maximise $f(g|\underline{x})$ which is done by choosing g at the mode of the posterior distribution.

(ii) The likelihood function is $L(p) \propto p^{13}(1-p)^{87}p^{20}(1-p)^{80+g}$

$$\propto p^{33}(1-p)^{167+g}$$

the prior likelihood of p is given by $f(p) \propto p(1-p)^7$

so $f(p|\underline{x}) \propto f(\underline{x}|p)\, f(p)$

$$\propto p^{33}(1-p)^{167+g}\, p(1-p)^7$$

$$\propto p^{34}(1-p)^{174+g}$$

so we have a Beta distribution with $\alpha = 35$ and $\beta = 175 + g$.

(iii) Under all-or-nothing loss we need to find the mode of

$$f(p|\underline{x}) \propto p^{34}(1-p)^{174+g}$$

$$\frac{\partial f}{\partial p} = 34p^{33}(1-p)^{174+g} - (174+g)\, p^{34}(1-p)^{173+g}$$

$$= p^{33}(1-p)^{173+g}[34(1-p) - (174+g)p]$$

And setting this equal to zero we have

$$34(1-\hat{p}) = (174+g)\hat{p}$$

$$34 = (208+g)\hat{p}$$

$$\hat{p} = \frac{34}{208+g}$$

Under quadratic loss the Bayes estimate is given by mean of the posterior distribution which is given by

$$\hat{p} = \frac{35}{35+175+g} = \frac{35}{210+g}$$

The two are equal if

$$\frac{34}{208+g} = \frac{35}{210+g}$$

i.e. $34(210 + g) = 35(208 + g)$

i.e. $7140 + 34g = 7280 + 35g$

i.e. $g = 7140 - 7280 = -140$

This would imply there were –40 policies in year 2 which is impossible.

Disappointingly few candidates knew the bookwork for part (i). Stronger candidates familiar with Bayes Theory were able to score very well here.

10 (i) A Poisson process is characterised by the probability of a single claim arising in a small time interval dt being λdt (with no probability of more than one claim).

For the reinsurer, the probability of a claim arising in a small time interval dt is given by

$$\lambda dt \times P(X_i > M)$$

$$= \lambda dt \times \int_M^\infty \frac{1}{\mu} e^{-\frac{1}{\mu}x} dx$$

$$= \lambda dt \left[-e^{-\frac{x}{\mu}} \right]_M^\infty$$

$$= \lambda dt \times e^{-\frac{M}{\mu}}$$

$$= \left(\lambda e^{-\frac{M}{\mu}} \right) dt$$

so we have a Poisson process with parameter $\lambda e^{-\frac{M}{\mu}}$.

(ii) $M_{X_i}(t) = E(e^{tX_i})$

$$= e^{0t} \times \left(1-e^{-\frac{M}{\mu}}\right) + \int_M^{\infty} e^{t(x-M)} \frac{1}{\mu} e^{-\frac{x}{\mu}} dx$$

$$= 1-e^{-\frac{M}{\mu}} + \int_M^{\infty} \frac{1}{\mu} e^{-Mt} e^{-x\left(\frac{1}{\mu}-t\right)} dx$$

$$= 1-e^{-\frac{M}{\mu}} + e^{-Mt} \left[-\frac{\frac{1}{\mu}}{\frac{1}{\mu}-t} e^{-x\left(\frac{1}{\mu}-t\right)} \right]_M^{\infty}$$

$$= 1-e^{-\frac{M}{\mu}} + e^{-Mt} \left\{ \frac{\frac{1}{\mu}}{\frac{1}{\mu}-t} \times e^{-M\left(\frac{1}{\mu}-t\right)} \right\}$$

$$= 1-e^{-\frac{M}{\mu}} + e^{-\frac{M}{\mu}} \times \frac{1}{1-t\mu}$$

$$= 1-e^{-\frac{M}{\mu}} \left(1 - \frac{1}{1-t\mu}\right)$$

$$= 1-e^{-\frac{M}{\mu}} \left(\frac{-t\mu}{1-t\mu}\right)$$

$$=1-e^{-\frac{M}{\mu}} \times \frac{\mu t}{1-\mu t}$$

(iii) (a) Now $M_{X_i}(t) = 1+e^{-\frac{M}{\mu}} \times \frac{\mu t}{1-\mu t}$

So $M'_{X_i}(t) = e^{-\frac{M}{\mu}} \left[\frac{\mu}{1-\mu t} + \frac{\mu t}{(1-\mu t)^2} \times \mu \right]$

and so $E(X_i) = M'_{X_i}(0) = \mu e^{-\frac{M}{\mu}}$

(b) The reinsurers annual rate of premium income is given by $\lambda(1 + \theta) E(X_i)$. So the adjustment coefficient satisfies

$$\lambda + \lambda(1 + \theta) E(X_i)R = \lambda M_{X_i}(R)$$

i.e. $1 + (1 + \theta) E(X_i)R = M_{X_i}(R)$

i.e. $1 + (1 + \theta)\, \mu e^{-\frac{M}{\mu}} R = 1 + e^{-\frac{M}{\mu}} \times \dfrac{\mu R}{1 - \mu R}$

i.e. $(1 + \theta)\, \mu R = \dfrac{\mu R}{1 - \mu R}$

i.e. $(1 + \theta)\, \mu R(1 - \mu R) = \mu R$

i.e. $(1 + \theta)(1 - \mu R) = 1$

i.e. $1 + \theta - \mu R - \theta(\mu R) = 1$

i.e. $\mu R(1 + \theta) = \theta$

i.e. $R = \dfrac{\theta}{(1 + \theta)\mu}$

(iv) R does not depend on the retention M.

This is a surprising result at first glance, but arises because of the memoryless feature of the exponential distribution

i.e. $X_i - M | X_i > M$ is exponential with parameter $\dfrac{1}{\mu}$

so the reinsurers claim process is just a slower version of the insurers.

Full credit was given for alternative solutions to part (i) and part (iv). This question was relatively poorly answered, although those candidates who were able to make a good attempt tended to score well.

END OF EXAMINERS' REPORT