INSTITUTE AND FACULTY OF ACTUARIES

EXAMINERS' REPORT

September 2012 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.

D C Bowie
Chairman of the Board of Examiners

December 2012

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, and 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 2012 paper

The examiners had felt that this paper contained a slightly greater proportion of more routine questions than the April 2012 paper and this was backed up by some good solutions to most of questions 1 to 8. However, many candidates struggled with the longer questions 9 and 10 and this was the main reason why the pass rate was lower than in April 2012. In particular, many candidates could not deal with impact of reinsurance in Q10 and the examiners expect well prepared candidates to understand this topic.

1 (i) The maximum losses are:

D1	17
D2	15
D3	16
D4	12

So the minimax decision is to choose D4.

(ii) The expected losses of the decisions are:

D1	8.3
D2	10.9
D3	11.5
D4	7.2

So the Bayes' decision is also D4.

This fairly standard question was well answered.

2 Firstly

$$E\left[X\right]=E\left[E\left[X\left|\alpha\right.\right]\right]=E\left[200+\alpha\right]=200+E\left[\alpha\right]=220$$

And secondly

$$Var\left(X\right)=Var\left[E\left[X\left|\alpha\right.\right]\right]+E[Var[X\left|\alpha\right.]]$$

Now

$$Var[E\left[X\left|\alpha\right.\right]]=Var[200+\alpha]=Var\left[\alpha\right]=4$$

And

$$E[Var\left[X\left|\alpha\right.\right]]=E[10+2\alpha]=10+2\times 20=50$$

Hence

$$Var\left[X\right]=4+50=54$$

Again, a fairly routine question that was well answered by most candidates.

3 (i) Polar algorithm:

(1) Generate independently U_1 and U_2 from $U(0,1)$

(2) Set $V_1 = 2U_1 - 1$, $V_2 = 2U_2 - 1$ and $S = V_1^2 + V_2^2$
(3) If $S > 1$ go to step 1

Otherwise set:

$$Z_1 = \sqrt{-\frac{2\ln S}{S}}V_1 \text{ and } Z_2 = \sqrt{-\frac{2\ln S}{S}}V_2$$

(ii) The acceptance probability is obtained from the condition $S < 1$. So the required probability is obtained as $P(V_1^2 + V_2^2 < 1)$ where V_i are independently drawn from $U(-1,1)$.

Simple geometrical arguments show that the required probability is equivalent to the event that a uniform draw from the points of the square defined by $V_1 \in [-1,1]$ and $V_2 \in [-1,1]$ falls within the circle with centre at the origin of coordinates $(0,0)$, and radius 1.

The probability of this event is equivalent to the ratios of the areas:

$$P = \frac{\pi 1^2}{2^2} = \frac{\pi}{4} = 0.7854\,.$$

Part (i) was mostly well answered, though some candidates lost marks as they did not specify how to transform U(0,1) random samples into U(–1,1) random samples. Very few candidates adopted the geometric approach in (ii).

4 (i) For the Pareto distribution with parameters α, λ as per the tables we have:

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

And

$$Var(X) = \frac{\alpha\lambda^2}{(\alpha-1)^2(\alpha-2)} = E(X)^2\frac{\alpha}{\alpha-2}$$

And so

$$E\left(X^2\right) = Var\left(X\right) + E(X)^2 = E\left(X\right)^2\left(\frac{\alpha}{\alpha-2}+1\right) = E\left(X\right)^2\left(\frac{2\alpha-2}{\alpha-2}\right)$$

The observed values we are trying to fit are

$$E\left(X\right) = 170$$
$$E\left(X^2\right) = 400^2 + 170^2 = 434.626^2$$

So we have

$$\frac{2\alpha-2}{\alpha-2} = \frac{E(X^2)}{E(X)^2} = \frac{434.626^2}{170^2} = 6.53633$$

And so

$$\alpha = \frac{2-2\times 6.53633}{(2-6.53633)} = 2.441$$

And finally $\lambda = 1.441 \times 170 = 244.95$

(ii) We must solve

$$0.5 = 1-\left(\frac{244.95}{244.95+x}\right)^{2.441}$$

Re-arranging and taking roots gives

$$0.5^{\frac{1}{2.441}} = 0.7527965 = \frac{244.95}{244.95+x}$$

And so

$$x = \frac{244.95 - 244.95\times 0.7527965}{0.7527965} = 80.44$$

So the median is significantly lower than the mean. This demonstrates how skew the Pareto distribution is.

Alternative correct (and in some cases quicker) solutions are possible and received full credit. This question was well answered with many candidates scoring full marks.

5 (i) From the definition

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

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

$$= \exp\left[\frac{y\theta - b(\theta)}{a(\varphi)} + c(y,\varphi) \right]$$

Which is the right form for a member of an exponential family where

$$\theta = \log\left(\frac{\mu}{1-\mu} \right)$$

$$\varphi = n$$

$$a(\varphi) = \frac{1}{\varphi}$$

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

$$c(y,\varphi) = \log\binom{\varphi}{\varphi y}$$

Hence the distribution does belong to an exponential family.

(ii) The three main components are:

- the <u>distribution of the response variable</u>
- a <u>linear predictor</u> of the covariates
- <u>link function</u> between the response variable and the linear predictor

(iii) In this case we have a binomial distribution and therefore the natural choice of link function is $g(\mu) = \log\left(\frac{\mu}{1-\mu} \right)$.

(iv) We could apply a <u>log</u> transform to the response and then apply a simple linear regression. Hence the link function is $\log(\mu)$.

This was well answered, though a number of candidates lost some marks through failing to carefully define all of the parameters involved in the characterisation as a member of the exponential family.

6 (i) The annual premium charged is $0.25 \times 150 \times 1.7 = 63.75$

(ii) Let X be an individual claim. Then

$$P(X < 200) = P\left(N\left(150, 30^2\right) < 200\right)$$

$$= P\left(N(0,1) < \frac{200-150}{30}\right)$$

$$= P(N(0,1) < 1.667)$$

$$= (0.95154 \times 0.3 + 0.7 \times 0.95254)$$

$$= 0.95224$$

(iii) We need to calculate:

$$p = \sum_{j=0}^{\infty} P(j \text{ claims}) \times P(\text{all claims below retention}) \text{ [1]}$$

$$= \sum_{j=0}^{\infty} e^{-0.25} \frac{(0.25)^j}{j!} \times (0.95224)^j$$

$$= e^{-0.25} \times \sum_{j=0}^{\infty} \frac{(0.25 \times 0.95224)^j}{j!}$$

$$= e^{-0.25} \times e^{0.25 \times 0.95224}$$

$$= 0.9881$$

(iv) We need to first calculate the mean claim amount paid by the reinsurer. This is given by

$$I = \int_{200}^{\infty} (x - 200) f(x) dx$$

Where $f(x)$ is the pdf of the Normal distribution with mean 150 and standard deviation 30.

Using the formula on p18 of the tables, we have:

$$I = \int_{200}^{\infty} xf(x)\,dx - 200P(X > 200)$$

$$= 150 \times \left[\Phi(\infty) - \Phi(1.667)\right] - 30 \times \left(\phi(\infty) - \phi(1.667)\right) - 200 \times (1 - 0.95224)$$

$$= 150(1 - 0.95224) - 30 \times (0 - 0.09942) - 200 \times 0.0.04776$$

$$= 0.5946$$

So the reinsurer charges $0.25 \times 0.5946 \times 2.2 = 0.32703$

(v) The direct insurers expected profit is given by:

$$63.75 - 0.32703 - 0.25 \times (150 - 0.5946) = 26.07$$

Comment: Answers were mixed here. Parts (i) and (ii) were generally well done. Only the best candidates completed part (iii) with most being unable to condition on the number of claims. On part (iv) most candidates wrote down the integral that needed to be evaluated, but only the better candidates were able to use the formula from the tables to evaluate it. A number of candidates struggled to compute the values of the probability density function of the Normal distribution.

7 (i) The aggregated claims are:

Underwriting year	*Development Year*			
	0	1	2	3
2008	85	127	157	164
2009	103	168	193	
2010	93	140		
2011	111			

Hence the development factors are given by:

$$DF_{0,1} = \frac{127 + 168 + 140}{85 + 103 + 93} = 1.548043$$

$$DF_{1,2} = \frac{157 + 193}{127 + 168} = 1.186441$$

$$DF_{2,3} = \frac{164}{157} = 1.044586$$

The completed triangle of cumulative claims is:

Underwriting year	*Development Year*			
	0	*1*	*2*	*3*
2008	85	127	157	164
2009	103	168	193	201.61
2010	93	140	166.10	173.51
2011	111	171.83	203.87	212.96

Dis-accumulating gives:

Underwriting year	*Development Year*			
	0	*1*	*2*	*3*
2008	85	42	30	7
2009	103	65	25	8.61
2010	93	47	26.10	7.41
2011	111	60.83	32.04	9.09

And so the outstanding claims are:

$$8.61+7.41+9.09+26.1+32.04+60.83 = 144.08$$

(ii) Applying the development factors to the claims in development year 0 gives:

Underwriting year	*Development Year*			
	0	*1*	*2*	*3*
2008	85	131.58	156.12	163.08
2009	103	159.45	189.18	
2010	93	143.97		
2011	111			

Dis-accumulating gives:

Underwriting year	*Development Year*			
	0	*1*	*2*	*3*
2008	85	46.58	24.53	6.96
2009	103	56.45	29.73	
2010	93	50.97		
2011	111			

And computing the difference between predicted and actual gives:

Underwriting year	*Development Year*			
	0	*1*	*2*	*3*
2008	0	–4.58	5.47	0.04
2009	0	8.55	–4.73	
2010	0	–3.97		
2011	0			

(iii) Overall the model seems a reasonable fit, though some of the individual differences are quite large in percentage terms – for example the difference of 5.47 is 18% of the observed value.

Part (i) was well answered, though a small number of candidates continue to throw away simple marks by not computing the single figure for outstanding claims. This was the first time in some years that the material in part (ii) has been tested, and a number of candidates performed the comparison on a cumulative basis rather than the incremental basis that the question asked for.

8 (i) Let N_i be the number of type 1 buildings covered in year *i*. Set $N = N_1 + \cdots + N_5 = 697$. Let the number of claims in year *i* be denoted by M_i and set $M = M_1 + \cdots + M_5$. Then under the conditions in the question $M \sim Poisson(N\lambda)$.

The likelihood is given by

$$L = Ce^{-697\lambda}\frac{(697\lambda)^m}{m!}$$

Where m=158 is the total number of claims over the 5 years. The log-likelihood is given by

$$l = \log L = D - 697\lambda + m\log 697\lambda$$

Differentiating gives

$$\frac{dl}{d\lambda} = -697 + \frac{m}{\lambda}$$

And setting this equal to zero we get

$$\hat{\lambda} = \frac{m}{697} = \frac{158}{697} = 0.226686$$

This is a maximum since $\dfrac{d^2 l}{d\lambda^2} = -\dfrac{m}{\lambda^2} < 0$.

(ii) We first need to calculate

$$P^* = \frac{1}{5\times 3 - 1}\sum_{i=1}^{3}\bar{P}_i\left(1 - \frac{\bar{P}_i}{\bar{P}}\right)$$

$$= \frac{1}{14}\left[697\left(1 - \frac{697}{1507}\right) + 295\left(1 - \frac{295}{1507}\right) + 515\left(1 - \frac{515}{1507}\right)\right]$$

$= 67.9207$

The estimators are given by

$$E\left(m\left(\theta\right)\right) = \bar{X} = 0.264101$$

$$E\left(s^2\left(\theta\right)\right) = \frac{1}{3}\sum_{i=1}^{3}\frac{1}{5-1}\sum_{j=1}^{5}P_{ij}\left(\frac{Y_{ij}}{P_{ij}} - \bar{X}_i\right)^2$$

$$= \frac{1}{12}\times\left(1.527016 + 0.96605 + 4.53253\right) = 0.585466$$

$$Var\left(m\left(\theta\right)\right) = \frac{1}{P^*}\left(\frac{1}{3\times 5-1}\sum_{i=1}^{3}\sum_{j=1}^{5}P_{ij}\left(\frac{Y_{ij}}{P_{ij}} - \bar{X}\right)^2 - E(s^2\left(\theta\right))\right)$$

$$= \frac{1}{67.9207}\times\left(\frac{1}{14}\left(2.502737 + 1.178133 + 6.775614\right) - 0.585466\right)$$

$$= 0.00237668$$

And the credibility factor for type 1 policies is given by

$$Z_1 = \frac{\bar{P}_1}{\bar{P}_1 + \frac{E(s^2\left(\theta\right))}{Var(m\left(\theta\right))}} = \frac{697}{697 + \frac{0.585466}{0.00237668}} = 0.73887$$

Number of claims per unit risk is then given by
$0.73887\times 0.226686 + \left(1 - 0.73887\right)\times 0.264101 = 0.2364571$

And so expected claims are $0.2364571\times 191 = 45.16$

(iii) The main differences are:

- The approach in (i) uses only the data from type 1 policies; the approach in (ii) uses a weighted average of the data from type 1 policies and the overall data.

- The approach in (i) makes a precise distributional assumption about claims (i.e. that they are Poisson distributed). This assumption is not used in approach (ii).

Part (i) was often not well answered, with many weaker candidates not reflecting the fact that the number of buildings covered impacts the parameter of the Poisson distribution for the *number of claims. Parts (ii) and (iii) were generally well answered, which was pleasing given that this was the first appearance of EBCT Model 2 since its return to the syllabus.*

9 (i) The order s will be 3 i.e. $Y_t = \nabla_3 X_t = X_t - X_{t-3}$

The characteristic polynomial will be $1-(\alpha+\beta)z+\alpha\beta z^2$ with roots $1/\alpha$ and $1/\beta$.

Hence the process is stationary for $|\alpha|<1$ and $|\beta|<1$.

(ii) The Yule-Walker equations for the differenced equations give:

$$\rho_1 - (\alpha+\beta) + \alpha\beta\rho_1 = 0$$
$$\rho_2 - (\alpha+\beta)\rho_1 + \alpha\beta = 0$$

Substituting the observed values of the auto-correlation gives:

$$0.2 - (\alpha+\beta) + 0.2\alpha\beta = 0$$
$$0.7 - 0.2(\alpha+\beta) + \alpha\beta = 0$$

Let $X = \alpha+\beta$ and let $Y = \alpha\beta$ then we have

$$0.2 - X + 0.2Y = 0$$
$$0.7 - 0.2X + Y = 0$$

The first equation gives $X = 0.2 + 0.2Y$ and substituting into the second gives:

$$0.7 - 0.04 - 0.04Y + Y = 0$$

So $0.96Y = -0.66$ and so Y=-0.6875 and X = 0.0625 [1]

This means that α and β are the roots of the quadratic equation

$$x^2 - 0.0625x - 0.6875 = 0$$

Which are

$$\frac{0.0625 \pm \sqrt{0.0625^2 + 4 \times 0.6875}}{2}$$

i.e. 0.860995 and −0.79849

(iii) Since $Y_t = X_t - X_{t-3}$ we have that

$$X_{101} = Y_{101} + X_{98}$$

and

$$X_{102} = Y_{102} + X_{99}$$

With the forecasted values

$$\hat{x}_{101} = \hat{y}_{101} + x_{98}$$

and

$$\hat{x}_{102} = \hat{y}_{102} + x_{99}$$

where

$$\hat{y}_{101} = 0.0625\,y_{100} + 0.6875\,y_{99} = 0.0625\,(x_{100} - x_{97}) + 0.6875(x_{99} - x_{96})$$

and

$$\hat{y}_{102} = 0.0625\,\hat{y}_{101} + 0.6875\,(x_{100} - x_{97})$$

Many candidates struggled with this question. In particular many failed to identify quickly that s=3 in part (i) leads to difficult algebra in part (ii). Those who did identify that s=3 were generally able to write down the Yule Walker equations and make some progress in part (ii) though only the better candidates were able to find the numerical values required.

10 (i) The insurer charges a premium of $\lambda \times (1 \times 0.7 + 8 \times 0.3) \times 1.6 = 4.96\lambda$

Where λ is the rate of the Poisson process. Expected claims outgo (net of reinsurance) is given by $\lambda \times (1 \times 0.7 + M \times 0.3) = \lambda(0.7 + 0.3M)$

The premiums charged by the reinsurer are

$$\lambda \times \left(0.3 \times (8 - M) \times 2.2\right) = 0.66\lambda(8 - M)$$

So the expected profit is positive if:

$$4.96\lambda - \lambda(0.7 + 0.3M) - 0.66\lambda(8 - M) > 0$$

i.e.

$$-1.02 + 0.36M > 0$$

i.e.

$$M > \frac{1.02}{0.36} = 2.833$$

(ii) The adjustment coefficient is equation is:

$$M_X(R) - 1 - cR = 0$$

Comment: *Or alternatively* $\lambda + c_{net}R = \lambda M_Y(R)$, *where* c_{net} *is the overall net premium.*

Where X is the distribution of net claim payments by the direct insurer. This gives:

$$0.7e^R + 0.3e^{MR} - 1 - \left(4.96 - 0.66\left(8 - M\right)\right)R = 0$$

(iii) With $M = 4$ this equation becomes:

$$f(R) := 0.7e^R + 0.3e^{4R} - 1 - 2.32R = 0$$

We shall find R by trial and error

$f(0.1) = -0.0108<0$
$f(0.2) = 0.058644209>0$
$f(0.15) = 0.01191961>0$
$f(0.125) = -0.002179<0$
$f(0.135\ ) = 0.002778077>0$

So the root lies between 0.125 and 0.135 and so R=0.13 (to 2 decimal places)

(iv) The premium charged by the reinsurer for the proportional reinsurance is

$$\lambda \times \alpha \times 2.2 \times \left(0.7 + 0.3 \times 8\right) = 6.82\alpha\lambda\ .$$

Equating the premiums for the two types of reinsurance we get

$$6.82\alpha\lambda = 0.66\lambda(8 - M)$$

i.e.

$$\alpha = \frac{0.66(8-M)}{6.82} = \frac{3(8-M)}{31}$$

(v) In this case $\alpha = 0.387096774$ and the premium charged by the reinsurer is 2.64λ .

The adjustment coefficient equation for the insurer is given by

$$g\left(R\right) := 0.7e^{0.612903226R} + 0.3e^{4.903225806R} - 1 - 2.32R = 0$$

Again by trial and error

$$g\left(0.125\right) = 0.019471559 > 0$$
$$g\left(0.135\right) = 0.028745229 > 0$$
$$g\left(0.001\right) = -0.00041625635 < 0$$

So the root lies between 0.001 and 0.125 and is therefore *less* than in the excess of loss case.

(vi) By Lundberg's inequality the adjustment coefficient is an inverse measure of risk – that is, the higher the coefficient the lower the probability of ruin. The excess of loss reinsurance is therefore more effective at reducing the probability of ruin than the proportional reinsurance.

Many candidates really struggled with this question, and in particular with the re-insurance arrangement and its impact on the claims paid and net premiums received by the insurer. A not insignificant number assumed that the insurer would reduce the premiums it charged the customer as a result of the reinsurance. Only the best candidates managed to accurately produce the equations satisfied by the adjustment coefficient and go on to find the numerical values.

END OF EXAMINERS' REPORT