EXAMINATION

20 April 2010 (am)

Subject CT6 — Statistical Methods
Core Technical

Time allowed: Three hours

INSTRUCTIONS TO THE CANDIDATE

1. *Enter all the candidate and examination details as requested on the front of your answer booklet.*

2. *You must not start writing your answers in the booklet until instructed to do so by the supervisor.*

3. *Mark allocations are shown in brackets.*

4. *Attempt all 11 questions, beginning your answer to each question on a separate sheet.*

5. *Candidates should show calculations where this is appropriate.*

Graph paper is NOT required for this paper.

AT THE END OF THE EXAMINATION

Hand in BOTH your answer booklet, with any additional sheets firmly attached, and this question paper.

In addition to this paper you should have available the 2002 edition of the Formulae and Tables and your own electronic calculator from the approved list.

1 A coin is biased so that the probability of throwing a head is an unknown constant p. It is known that p must be either 0.4 or 0.75. Prior beliefs about p are given by the distribution:

$$P(p = 0.4) = 0.6 \qquad P(p = 0.75) = 0.4$$

The coin is tossed 6 times and 4 heads are observed.

Find the posterior distribution of p. [5]

2 An insurance company is modelling claim numbers on its portfolio of motor insurance policies using a Poisson distribution, whose mean depends on the age and gender of the policyholder.

(i) Suggest a link function for fitting a generalised linear model for the mean of the Poisson distribution. [1]

(ii) Specify the corresponding linear predictor used for modelling the age and gender dependence as:

(a) age + gender
(b) age + gender + age × gender

[4]
[Total 5]

3 The following two models have been suggested for representing some quarterly data with underlying seasonality.

Model 1 $Y_t = \alpha Y_{t-4} + e_t$
Model 2 $Y_t = \beta e_{t-4} + e_t$

Where e_t is a white noise process in each case.

(i) Determine the auto-correlation function for each model. [4]

The observed quarterly data is used to calculate the sample auto-correlation.

(ii) State the features of the sample auto-correlation that would lead you to prefer Model 1. [1]
[Total 5]

4 The number of claims N on a portfolio of insurance policies follows a binomial distribution with parameters n and p. Individual claim amounts follow an exponential distribution with mean $1/\lambda$. The insurer has in place an individual excess of loss reinsurance arrangement with retention M.

(i) Derive an expression, involving M and λ, for the probability that an individual claim involves the reinsurer. [2]

Let I_i be an indicator variable taking the value 1 if the ith claim involves the reinsurer and 0 otherwise.

(ii) Evaluate the moment generating function $M_{I_i}(t)$. [1]

Let K be the number of claims involving the re-insurer so that $K = I_1 + \cdots + I_N$.

(iii) (a) Find the moment generating function of K.

(b) Deduce that K follows a binomial distribution with parameters that you should specify.

[4]
[Total 7]

5 An insurance company has issued life insurance policies to 1,000 individuals. Each life has a probability q of dying in the coming year. In a warm year, $q = 0.001$ and in a cold year $q = 0.005$. The probability of a warm year is 50% and the probability of a cold year is 50%. Let N be the aggregate number of claims across the portfolio in the coming year.

(i) Calculate the mean and variance of N. [4]

(ii) Calculate the alternative values for the mean and variance of N assuming that q is a constant 0.003. [2]

(iii) Comment on the results of (i) and (ii). [2]
[Total 8]

6 Observations $y_1, y_2, \ldots, y_n$ are made from a random walk process given by:

$$Y_0 = 0 \text{ and } Y_t = a + Y_{t-1} + e_t \text{ for } t > 0$$

where e_t is a white noise process with variance σ^2.

(i) Derive expressions for $E(Y_t)$ and $Var(Y_t)$ and explain why the process is not stationary. [3]

(ii) Show that $\gamma_{t,s} = Cov(Y_t, Y_{t-s})$ for $s < t$ is linear in s. [2]

(iii) Explain how you would use the observed data to estimate the parameters a and σ. [3]

(iv) Derive expressions for the one-step and two-step forecasts for Y_{n+1} and Y_{n+2}. [2]

[Total 10]

7 The truncated exponential distribution on the interval $(0, c)$ is defined by the probability density function:

$$f(x) = ae^{-\lambda x} \quad \text{for } 0 < x < c$$

where λ is a parameter and a is a constant.

(i) Derive an expression for a in terms of λ and c. [1]

(ii) Construct an algorithm for generating samples from this distribution using the inverse transform method. [3]

Suppose that $0 < \lambda < c$ and consider the truncated Normal distribution defined by the probability density function:

$$g(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}[\Phi(c) - 0.5]} \quad \text{for } 0 < x < c$$

where $\Phi(z)$ is the cumulative density function of the standard Normal distribution.

(iii) Extend the algorithm in (ii) to use samples generated from $f(x)$ to produce samples from $g(x)$ using the acceptance / rejection method. [6]

[Total 10]

8 The table below shows the incremental claims paid on a portfolio of insurance policies together with an extract from an index of prices. Claims are fully paid by the end of development year 3.

	Development Year					*Price index*
Accident Year	*0*	*1*	*2*	*3*	*Year*	*(mid year)*
2006	103	32	29	13	2006	100
2007	88	21	16		2007	104
2008	110	35			2008	109
2009	132				2009	111

Calculate the reserve for unpaid claims using the inflation-adjusted chain ladder approach, assuming that future claims inflation will be 3% p.a. [11]

9 On the tapas menu in a local Spanish restaurant customers can order a dish of 20 roasted chillies for £5. There is always a mixture of hot chillies and mild chillies on the dish and these cannot be distinguished except by taste. The restaurant produces two types of dish: one containing 4 hot and 16 mild chillies and one containing 8 hot and 12 mild chillies. When the dish is served, the waiter allows the customer to taste one chilli and then offers a 50% discount to customers who correctly guess whether the dish contains 4 hot chillies or 8 hot chillies.

A hungry actuary who regularly visits the restaurant is trying to work out the best strategy for guessing the number of hot chillies.

(i) List the four possible decision functions the actuary could use. [3]

(ii) Calculate the values of the risk function for the two different chilli dishes and each decision function. [6]

(iii) Determine the optimum strategy for the actuary using the Bayes criterion and work out the average price he will pay for a dish of chillies if the restaurant produces equal numbers of the two types of dish. [3]
[Total 12]

10 Claims on a portfolio of insurance policies arrive as a Poisson process with annual rate λ. Individual claims are for a fixed amount of 100 and the insurer uses a premium loading of 15%. The insurer is considering entering a proportional reinsurance agreement with a reinsurer who uses a premium loading of 20%. The insurer will retain a proportion α of each risk.

(i) Write down and simplify the equation defining the adjustment coefficient R for the insurer. [3]

(ii) By considering R as a function of α and differentiating show that

$$(120\alpha - 5)\frac{dR}{d\alpha} + 120R = \left(100R + 100\alpha\frac{dR}{d\alpha}\right)e^{100\alpha R}$$ [3]

(iii) Explain why setting $\frac{dR}{d\alpha} = 0$ and solving for α may give an optimal value for α. [3]

(iv) Use the method suggested in part (iii) to find an optimal choice for α. [4]

[Total 13]

11 An actuary has, for three years, recorded the volume of unsolicited advertising that he receives. He believes that the number of items that he receives follows a Poisson distribution with a mean which varies according to which quarter of the year it is. He has recorded Y_{ij} the number of items received in the ith quarter of the jth year (i = 1, 2, 3, 4 and j = 1, 2, 3). The actuary wishes to estimate the number of items that he will receive in the 1st quarter of year 4. He has recorded the following data:

	Y_{i1}	Y_{i2}	Y_{i3}	$\bar{Y}_i = \frac{1}{3}\sum_j Y_{ij}$	$\sum_j \left(Y_{ij} - \bar{Y}_i\right)^2$
$i = 1$	98	117	124	113	362
$i = 2$	82	102	95	93	206
$i = 3$	75	83	88	82	86
$i = 4$	132	152	148	144	224

(i) Estimate $Y_{1,4}$ the number of items that the actuary expects to receive in the first quarter of year 4 using the assumptions of EBCT model 1. [5]

The actuary believes that, in fact, the volume of items has been increasing at the rate of 10% per annum.

(ii) Suggest how the approach in (i) can be adjusted to produce a revised estimate taking this growth into account. [2]

(iii) Calculate the maximum likelihood estimate of $Y_{1,4}$ (based on the quarter 1 data already observed and the 10% p.a. increase described above). [5]

(iv) Compare the assumptions underlying the approach in (i) and (ii) with those underlying the approach in (iii). [2]

[Total 14]

END OF PAPER

EXAMINERS REPORT

April 2010 examinations

Subject CT6 — Statistical Methods
Core Technical

Introduction

The attached subject report has been written by the Principal Examiner with the aim of helping candidates. The questions and comments are based around Core Reading as the interpretation of the syllabus to which the examiners are working. They have however given credit for any alternative approach or interpretation which they consider to be reasonable.

R D Muckart
Chairman of the Board of Examiners

July 2010

Comments

These are given in italics at the end of each question.

1 $P(p=0.4|4H)=\dfrac{P(4H|p=0.4)P(p=0.4)}{P(4H)}$

But $P(4H)$ = $P(4H|p=0.4)P(p=0.4)+P(4H|p=0.75)P(p=0.75)$

$= \binom{6}{4}0.4^4 0.6^2\times 0.6+\binom{6}{4}0.75^4 0.25^2\times 0.4$

= 0.082944 + 0.11865

= 0.201596

So $P(p=0.4|4H)=\dfrac{0.082944}{0.201596}=0.411436$

So the posterior distribution of p is given by $P(p = 0.4) = 0.411436$ and $P(p = 0.75)$ $= 0.588564$

Comment: *This question was intended to be a straightforward application of Bayes' Theorem. However, the question was generally not well answered, with many candidates unable to find the posterior distribution in this slightly unfamiliar scenario.*

2 (i) The link function here is $g(\mu)=\log\mu$.

(ii) (a) The linear predictor is $\alpha_i+\beta x$ where the intercept α_i for $i=1,2$ depends on gender.

(b) The linear predictor is $\alpha_i+\beta_i x$ so that both parameters depend on gender.

Comment: *This straightforward question was generally well answered.*

3 (i) **Model 1**

In general we have $\gamma_k=\alpha\gamma_{k-4}$

Taking covariance with $Y_t, Y_{t-1}, Y_{t-2}, Y_{t-3}$ we get:

$\gamma_0=\alpha\gamma_4$

$\gamma_1=\alpha\gamma_3$

$\gamma_2=\alpha\gamma_2$

$\gamma_3=\alpha\gamma_1$

For $\alpha\neq 0$ these equations imply that $\rho_k=0$ unless k is divisible by 4.

So we have $\rho_{4k} = \alpha^k$ and all other autocorrelations are zero.

Model 2

Here we have $\gamma_k = 0$ unless k=4 and k=0. In these cases

$$\gamma_0 = cov(e_t + \beta e_{t-4}, e_t + \beta e_{t-4}) = (1+\beta^2)\sigma^2$$

$$\gamma_4 = cov(e_t + \beta e_{t-4}, e_{t-4} + \beta e_{t-8}) = \beta\sigma^2.$$

So $\rho_0 = 1$, $\rho_4 = \frac{\beta}{1+\beta^2}$ and all the other autocorrelations are zero.

(ii) Model 1 is preferred in situations where the sample auto-correlation is non-zero and decays exponentially.

Comment: *This question was reasonably well attempted, although a number of candidates dropped marks especially in calculating the ACF of Model 1.*

4 (i) Let X represent the distribution of individual claims. Let π denote the probability that an individual claim involves the reinsurer. Then

$$\pi = P(X > M) = \int_M^\infty \lambda e^{-\lambda x} dx$$

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

$$= e^{-\lambda M}$$

(ii) $M_{I_i}(t) = E(e^{tI_i}) = \pi e^t + 1 - \pi = e^{t-\lambda M} + 1 - e^{-\lambda M}$

(iii) Using the results for the moment generating function of a compound distribution, we have

$$M_K(t) = M_N(\log M_{I_i}(t))$$

$$= \left(pM_{I_i}(t) + 1 - p\right)^n$$

$$= \left(p(\pi e^t + 1 - \pi) + 1 - p\right)^n$$

$$= \left(p\pi e^t + p - p\pi + 1 - p\right)^n$$

$$= \left(p\pi e^t + 1 - p\pi\right)^n$$

$$= \left(pe^{t-\lambda w} + 1 - pe^{-\lambda w}\right)^n$$

Which is the MGF of a binomial distribution with parameters n and $p\pi$.

Hence, by the uniqueness of MGFs K has a binomial distribution with parameters n and $p\pi$.

Comment: *This question was well answered.*

5 (i) $E(N) = E\left[E(N|q)\right]$

$$= E\left[1000q\right]$$
$$= 1000 \times \left(0.5 \times 0.001 + 0.5 \times 0.005\right)$$
$$= 1000 \times 0.003 = 3$$

$$Var(N) = Var\left[E(N|q)\right] + E\left[Var\left(N|q\right)\right]$$

$$= Var(1000q) + E(1000q(1-q))$$

Now $E(q) = 0.003$ and $E(q^2) = 0.5 \times 0.001^2 + 0.5 \times 0.005^2 = 0.000013$

So $Var(q) = 0.000013 - 0.003^2 = 0.000004$

$$E(q(1-q)) = 0.5 \times 0.001 \times 0.999 + 0.5 \times 0.005 \times 0.995 = 0.002987$$

So $Var(N) = 1000^2 \times 0.000004 + 1000 \times 0.002987 = 6.987$

(ii) In this case $N \sim B(1000, 0.003)$ and so

$$E(N) = 1000 \times 0.003 = 3$$

and

$$Var(N) = 1000 \times 0.003 \times 0.997 = 2.991$$

(iii) The simplification in (ii) results in the same mean number of deaths, but a very significantly lower variance.

This is because in (i) there is a tendency for deaths to occur at the same time, or not at all (as a result of the weather) whereas in (ii) deaths are genuinely independent.

Comment: *Many good answers here although some candidates struggled to articulate why the variance in (ii) was lower.*

6 (i) We can deduce that $Y_t = at + \sum_{i=1}^{t} e_i$

and so $E(Y_t) = at$ and

$Var(Y_t) = t\sigma^2$.

Since these expressions depend on t the process is not stationary.

(ii) As $s < t$ we have

$$Cov(Y_t, Y_{t-s}) = Cov(at + \sum_{i=1}^{t} e_i, as + \sum_{j=1}^{t-s} e_j) = Var(\sum_{j=1}^{t-s} e_j) = (t-s)\sigma^2$$

Which is linear in s as required.

(iii) First note that the differenced series:

$$X_t = Y_t - Y_{t-1} = a + e_t$$

is essentially a white noise process. So estimates of a and σ^2 can be found by constructing the sample differences series $x_i = y_i - y_{i-1}$ for $i = 1, 2, \ldots, n$ and taking the mean and sample variance(or its square for estimating σ) respectively.

(iv) In this case $\hat{y}_n(1) = \hat{a} + y_n + 0 = \hat{a} + y_n$

And $\hat{y}_n(2) = \hat{a} + \hat{y}_n(1) + 0 = 2\hat{a} + y_n$

Comment: *This was a rather hard question with many candidates confirming the non-stationarity in (i) but not finding the general solution. In part (ii) a good number of answers failed to score full marks. Alternative answers to (i) and (ii) could have been obtained by increasing t iteratively and noticing a pattern developing.*

7 (i) $1 = \int_0^c ae^{-\lambda x}dx = \left[-\frac{a}{\lambda}e^{-\lambda x} \right]_0^c = \frac{a}{\lambda}\left(1-e^{-\lambda c}\right)$

So $a = \frac{\lambda}{1-e^{-\lambda c}}$

(ii) The distribution function is

$$F(x) = \int_0^x ae^{-\lambda y}dy = \left[-\frac{a}{\lambda}e^{-\lambda y} \right]_0^x$$

$$= \frac{a}{\lambda}\left(1-e^{-\lambda x}\right) = \frac{1-e^{-\lambda x}}{1-e^{-\lambda c}}$$

The required transformation is therefore given by:

$$u = F(x) \Leftrightarrow 1-e^{-\lambda x} = u(1-e^{-\lambda c}) \Leftrightarrow x = -\frac{\log\left(1-u(1-e^{-c\lambda})\right)}{\lambda}$$

So the algorithm is:

- Generate u from U(0, 1)
- Set $x = -\frac{\log\left(1-u(1-e^{-c\lambda})\right)}{\lambda}$

(iii) We need to find $M = \underset{0<x<c}{Max}\frac{g(x)}{f(x)} = \underset{0<x<c}{Max}\frac{(1-e^{-c\lambda})e^{-x^2/2+\lambda x}}{\sqrt{2\pi}[\Phi(c)-0.5]\lambda}$

Let $h(x) = e^{-x^2/2+\lambda x}$ then we can simply find the maximum of *h(x)* since it differs only by a constant.

Then $h'(x) = h(x)\times(-x+\lambda)$

So $h'(x) = 0$ when $x = \lambda$

$$h''(x) = h'(x)(\lambda - x) - h(x)$$

And so $h''(\lambda) = -h(\lambda) < 0$ so we do have a maximum.

Hence $M = \frac{(1-e^{-c\lambda})e^{\lambda^2/2}}{\sqrt{2\pi}[\Phi(c)-0.5]\lambda}$

And so $\dfrac{g(x)}{Mf(x)} = e^{-x^2/2+\lambda x-\lambda^2/2}$

So the algorithm is:

- Generate u from U(0,1)
- Set $x = -\dfrac{\log\left(1-u(1-e^{-c\lambda})\right)}{\lambda}$
- Generate v from U(0,1)
- If $v < e^{-x^2/2+\lambda x-\lambda^2/2}$ return x as the random sample, otherwise begin again.

Comment: *This was found particularly hard especially part (iii), where some harder calculations are involved. Many candidates just described the general theory of rejection algorithms without being able to apply it explicitly here.*

8 (i) Adjusting the incremental data for inflation to mid 2008 prices gives:

	Development Year			
Accident Year	*0*	*1*	*2*	*3*
2006	114.3	34.2	29.5	13
2007	93.9	21.4	16	
2008	112.0	35		
2009	132			

Cumulating gives

	Development Year			
Accident Year	*0*	*1*	*2*	*3*
2006	114.3	148.5	178.0	191.0
2007	93.9	115.3	131.3	
2008	112.0	147.0		
2009	132.0			

The development factors are:

Year 0 to year 1 $\dfrac{148.5+115.3+147.0}{114.3+93.9+112.0} = 1.2827$

Year 1 to year 2 $\dfrac{178.0+131.3}{148.5+115.3} = 1.1726$

Year 2 to year 3 $\frac{191.0}{178.0} = 1.0730$

The completed table at mid 2008 prices is:

		Development Year		
Accident Year	*0*	*1*	*2*	*3*
2006				
2007				140.9
2008			172.4	185.0
2009		169.3	198.5	213.0

Differencing gives:

		Development Year		
Accident Year	*0*	*1*	*2*	*3*
2006				
2007				9.6
2008			25.4	12.6
2009		37.3	29.2	14.5

And so the total reserve is

$$(37.3+25.4+9.6)\times 1.03+(29.2+12.6)\times 1.03^2+14.5\times 1.03^3=134.7$$

Comment: *This was a straightforward question with many candidates scoring full marks here.*

9 (i) The possibilities are (where H denotes trying a hot chilli and M denotes trying a mild chilli)

$d_1(H) = 4H$ and $d_1(M) = 4H$
$d_2(H) = 4H$ and $d_2(M) = 8H$
$d_3(H) = 8H$ and $d_3(M) = 4H$
$d_4(H) = 8H$ and $d_4(M) = 8H$

(ii) Under 4H we have $P(H) = 0.2$ and $P(M) = 0.8$
Under 8H we have $P(H) = 0.4$ and $P(M) = 0.6$

We can present the game so that the loss to the actuary is what he has to pay for the plate of chillis (i.e. the loss is either 2.5 or 5). Under this approach we have:

$$R(d_1, 4H) = P(H|4H) \times L(d_1(H), 4H) + P(M|4H) \times L(d_1(M), 4H) = 0.2 \times 2.5 + 0.8 \times 2.5 = 2.5$$
$$R(d_1, 8H) = P(H|8H) \times L(d_1(H), 8H) + P(M|8H) \times L(d_1(M), 8H) = 0.4 \times 5 + 0.6 \times 5 = 5$$
$$R(d_2, 4H) = P(H|4H) \times L(d_2(H), 4H) + P(M|4H) \times L(d_2(M), 4H) = 0.2 \times 2.5 + 0.8 \times 5 = 4.5$$
$$R(d_2, 8H) = P(H|8H) \times L(d_2(H), 8H) + P(M|8H) \times L(d_2(M), 8H) = 0.4 \times 5 + 0.6 \times 2.5 = 3.5$$
$$R(d_3, 4H) = P(H|4H) \times L(d_3(H), 4H) + P(M|4H) \times L(d_3(M), 4H) = 0.2 \times 5 + 0.8 \times 2.5 = 3$$
$$R(d_3, 8H) = P(H|8H) \times L(d_3(H), 8H) + P(M|8H) \times L(d_3(M), 8H) = 0.4 \times 2.5 + 0.6 \times 5 = 4$$
$$R(d_4, 4H) = P(H|4H) \times L(d_4(H), 4H) + P(M|4H) \times L(d_4(M), 4H) = 0.2 \times 5 + 0.8 \times 5 = 5$$
$$R(d_4, 8H) = P(H|8H) \times L(d_4(H), 8H) + P(M|8H) \times L(d_4(M), 8H) = 0.4 \times 2.5 + 0.6 \times 2.5 = 2.5$$

(iii) The payoff matrix for the player is:

	d_1	d_2	d_3	d_4
4H	2.5	4.5	3	5
8H	5.0	3.5	4	2.5
Expected Loss	3.75	4.0	3.5	3.75

So the Bayes criterion strategy is d_3.

Under this approach, the average price paid is £3.50.

Comment: *Alternative solutions are possible here. The question was not answered as well as those relating to the same material in previous years. In particular, many weaker candidates were unable to fully specify the possible decision functions and therefore made little headway with this question.*

10 (i) Let X denote the individual claim amounts net of re-insurance. Then $X = 100\alpha$ and $M_X(t) = e^{100\alpha t}$.

The insurer's annual net premium income is

$$100 \times \lambda \times 1.15 - 100 \times (1-\alpha) \times \lambda \times 1.2 = \lambda(120\alpha - 5)$$

So the adjustment coefficient R satisfies

$$\lambda + \lambda(120\alpha - 5)R = \lambda e^{100\alpha R}$$

That is $1 + (120\alpha - 5)R = e^{100\alpha R}$

(ii) Differentiating this equation with respect to α we get

$$120R + (120\alpha - 5)\frac{dR}{d\alpha} = \frac{d}{d\alpha}\left[e^{100\alpha R}\right]$$

and $\frac{d}{d\alpha}\left[e^{100\alpha R}\right] = e^{100\alpha R}\frac{d}{d\alpha}\left[100\alpha R\right]$

$$= e^{100\alpha R}\left(100R + 100\alpha\frac{dR}{d\alpha}\right)$$

So putting these together, we have:

$$(120\alpha - 5)\frac{dR}{d\alpha} + 120R = \left(100R + 100\alpha\frac{dR}{d\alpha}\right)e^{100\alpha R}$$

(iii) Firstly, by Lundberg's inequality the higher the value of R the lower the upper bound on the probability of ruin.

So we wish to choose α so that R is a maximum.

That is, we need $\frac{dR}{d\alpha} = 0$.

(iv) Putting $\frac{dR}{d\alpha} = 0$ in the equation in (ii) we get

$$120R = 100\,\mathrm{Re}^{100\alpha R}$$

i.e. $e^{100\alpha R} = 1.2$

i.e. $R = \frac{\log 1.2}{100\alpha}$

substituting into the equation in (i) gives

$$1 + (120\alpha - 5)\times\frac{\log 1.2}{100\alpha} = 1.2$$

i.e. $100\alpha + (120\alpha - 5)\log 1.2 = 120\alpha$

$$-20\alpha + 120\alpha\log 1.2 = 5\log 1.2$$

So $\alpha = \frac{5\log 1.2}{120\log 1.2 - 20} = 0.48526$

Comment: *This question on material new to the syllabus for 2010 was not answered well with many candidates struggling, especially in the last part.*

11 (i) The overall mean is given by $\bar{Y} = \dfrac{113+93+82+144}{4} = 108$

$$E\left(s^2(\theta)\right) = \frac{1}{4}\sum_{i=1}^{4}\left(\frac{1}{2}\sum_{j=1}^{3}(Y_{ij}-\bar{Y}_i)^2\right) = \frac{362+206+86+224}{8} = 109.75$$

$$Var(m(\theta)) = \frac{1}{3}\sum_{i=1}^{4}\left(\bar{Y}_i-\bar{Y}\right)^2 - \frac{1}{3}E(S^2(\theta))$$

$$= \frac{(113-108)^2+(93-108)^2+(82-108)^2+(144-108)^2}{3} - \frac{109.75}{3}$$

$$= 704.083$$

So the credibility factor is $Z = \dfrac{3}{3+\dfrac{109.75}{704.083}} = 0.950608$

And the estimate for next quarter is

$$0.950608\times 113 + (1-0.950608)\times 108 = 112.75$$

(ii) The average number of pieces of mail is assumed to be growing each year. We need to adjust the data to take account of this. Two approaches are:

- Convert the data into "Year 4" values by increasing by 10% p.a. and then applying the methodology above; OR

- Recognise the lower volume of data in earlier years, by applying a risk volume to each year and using EBCT model 2. If the risk volume for year 4 is 1, then the risk volume for year 3 is 1/1.1 and year 2 is 1/1.21 etc.

(iii) Let the mean number of items in quarter 1 of year 1 be given by λ. Then the likelihood is given by:

$$L \propto e^{-\lambda}\lambda^{Y_{11}}e^{-1.1\lambda}(1.1\lambda)^{Y_{12}}e^{-1.1^2\lambda}(1.1^2\lambda)^{Y_{13}}$$

And so the log likelihood is

$$l = \log L = C - \lambda(1+1.1+1.1^2) + (Y_{11}+Y_{12}+Y_{13})\log\lambda$$

Differentiating $\frac{\partial l}{\partial \lambda} = -(1+1.1+1.1^2) + \frac{Y_{11}+Y_{12}+Y_{13}}{\lambda}$

And setting this equal to zero gives:

$$\hat{\lambda} = \frac{Y_{11}+Y_{12}+Y_{13}}{1+1.1+1.1^2} = \frac{98+117+124}{3.31} = 102.417$$

So the estimate for Q1 in year 4 is $1.1^3 \times \hat{\lambda} = 1.331 \times 102.417 = 136.32$

(iv) The main difference is that the maximum likelihood estimate approach considers the data for Q1 in isolation, whereas the EBCT approach assumes that data from other quarters come from a related distribution and so can tell us something about Q1.

Specifically, the EBCT approach assumes that the mean volume of unsolicited mail for each quarter is itself a sample from a common distribution. Hence whilst each quarter has a different mean, they provide some information about the population from which the mean is drawn.

Comment: *The same comment as in the previous question is valid here. This question was not very well answered question but with various alternative answers in (ii) and (iv).*

END OF EXAMINERS' REPORT

EXAMINATION

30 September 2010 (am)

Subject CT6 — Statistical Methods
Core Technical

Time allowed: Three hours

INSTRUCTIONS TO THE CANDIDATE

1. *Enter all the candidate and examination details as requested on the front of your answer booklet.*

2. *You must not start writing your answers in the booklet until instructed to do so by the supervisor.*

3. *Mark allocations are shown in brackets.*

4. *Attempt all 11 questions, beginning your answer to each question on a separate sheet.*

5. *Candidates should show calculations where this is appropriate.*

Graph paper is NOT required for this paper.

AT THE END OF THE EXAMINATION

Hand in BOTH your answer booklet, with any additional sheets firmly attached, and this question paper.

In addition to this paper you should have available the 2002 edition of the Formulae and Tables and your own electronic calculator from the approved list.

1 An actuary is using a simulation technique to estimate the probability θ that a claim on an insurance policy exceeds a given amount. The actuary has carried out 50 simulations and has produced an estimate that $\hat{\theta} = 0.47$. The variance of the simulated values is 0.15.

Calculate the minimum number of simulations that the actuary will have to perform in order to estimate θ to within 0.01 with 95% confidence. [4]

2 Claims on a portfolio of insurance policies follow a compound Poisson process with annual claim rate λ. Individual claim amounts are independent and follow an exponential distribution with mean μ. Premiums are received continuously and are set using a premium loading of θ. The insurer's initial surplus is U.

Derive an expression for the adjustment coefficient, R, for this portfolio in terms of μ and θ. [4]

3 An underwriter has suggested that losses on a certain class of policies follow a Weibull distribution. She estimates that the 10^{th} percentile loss is 20 and the 90^{th} percentile loss is 95.

(i) Calculate the parameters of the Weibull distribution that fit these percentiles. [3]

(ii) Calculate the 99.5^{th} percentile loss. [2]

[Total 5]

4 An office worker receives a random number of e-mails each day. The number of emails per day follows a Poisson distribution with unknown mean μ. Prior beliefs about μ are specified by a gamma distribution with mean 50 and standard deviation 15. The worker receives a total of 630 e-mails over a period of ten days.

Calculate the Bayesian estimate of μ under all or nothing loss. [7]

5 The table below shows aggregate annual claim statistics for three risks over a period of seven years. Annual aggregate claims for risk i in year j are denoted by X_{ij}.

Risk, i	$\bar{X}_i = \frac{1}{7}\sum_{j=1}^{7} X_{ij}$	$S_i^2 = \frac{1}{6}\sum_{j=1}^{7}\left(X_{ij} - \bar{X}_i\right)^2$
i = 1	127.9	335.1
i = 2	88.9	65.1
i = 3	149.7	33.9

(i) Calculate the credibility premium of each risk under the assumptions of EBCT Model 1. [6]

(ii) Explain why the credibility factor is relatively high in this case. [2]

[Total 8]

6 The probability density function of a gamma distribution is given in the following parameterised form:

$$f(x) = \frac{\alpha^{\alpha}}{\mu^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x\alpha}{\mu}} \quad \text{for } x > 0.$$

(i) Express this density in the form of a member of the exponential family, specifying all the parameters. [6]

(ii) Hence show that the mean and variance of the distribution are given by μ and $\frac{\mu^2}{\alpha}$ respectively. [3]

[Total 9]

7 An insurance company has a portfolio of policies under which individual loss amounts follow an exponential distribution with mean $1/\lambda$. There is an individual excess of loss reinsurance arrangement in place with retention level 100. In one year, the insurer observes:

- 85 claims for amounts below 100 with mean claim amount 42; and
- 39 claims for amounts above the retention level.

(i) Calculate the maximum likelihood estimate of λ. [5]

(ii) Show that the estimate of λ produced by applying the method of moments to the distribution of amounts paid by the insurer is 0.011164. [5]

[Total 10]

8 Claims on a portfolio of insurance policies arrive as a Poisson process with rate λ. The claim sizes are independent identically distributed random variables $X_1, X_2, \ldots$with:

$$P(X_i = k) = p_k \text{ for } k = 1, 2, \ldots, M \text{ and } \sum_{k=1}^{M} p_k = 1.$$

The premium loading factor is θ.

(i) Show that the adjustment coefficient R satisfies:

$$\frac{1}{M}\log(1+\theta) < R < \frac{2\theta m_1}{m_2}$$

where $m_i = E(X_1^i)$ for $i = 1, 2$. [7]

[The inequality $e^{Rx} \leq \frac{x}{M}e^{RM} + 1 - \frac{x}{M}$ for $0 \leq x \leq M$ may be used without proof.]

(ii) (a) Determine upper and lower bounds for R if $\theta = 0.3$ and X_i is equally likely to be 2 or 3 (and cannot take any other values).

(b) Hence derive an upper bound on the probability of ruin when the initial surplus is U. [3]

[Total 10]

9 An actuarial student has been working on some claims projections but some of her workings have been lost. The cumulative claim amounts and projected ultimate claims are given by the following table:

Accident Year	*Development Year* 0	1	2	3	*Ultimate*
1	1001	1485	1762	W	X
2	1250	Y	1820		1862.3
3	1302	1805			2122.5
4	Z				2278.8

All claims are paid by the end of development year 3.

It is known that ultimate claims for accident years 2 and 3 have been estimated using the Basic Chain Ladder method.

(i) Calculate the values of W, X and Y. [5]

For accident year 4 the student has used the Bornhuetter-Ferguson method using an earned premium of 2,500 and an expected loss ratio of 90%.

(ii) Calculate the value of Z. [4]

(iii) Calculate the outstanding claims reserve for all accident years implied by the completed table. [1]

[Total 10]

10 An insurance company has a portfolio of 10,000 policies covering buildings against the risk of flood damage.

(i) State the conditions under which the annual number of claims on the portfolio can be modelled by a binomial distribution $B(n, p)$ with $n = 10,000$. [3]

These conditions are satisfied and $p = 0.03$. Individual claim amounts follow a normal distribution with mean 400 and standard deviation 50. The insurer wishes to take out proportional reinsurance with the retention α set such that the probability of aggregate payments on the portfolio after reinsurance exceeding 120,000 is 1%.

(ii) Calculate α assuming that aggregate annual claims can be approximated by a normal distribution. [7]

This reinsurance arrangement is set up with a reinsurer who uses a premium loading of 15%.

(iii) Calculate the annual premium charged by the reinsurer. [2]

As an alternative, the reinsurer has offered an individual excess of loss reinsurance arrangement with a retention of M for the same annual premium. The reinsurer uses the same 15% loading to calculate premiums for this arrangement.

(iv) Show that the retention M is approximately 358.50. [4]

[You may wish to use the following formula which is given on page 18 of the Tables:

If $f(x)$ is the PDF of the $N(\mu, \sigma^2)$ distribution then

$$\int_L^U xf(x)dx = \mu\left[\Phi(U') - \Phi(L')\right] - \sigma\left[\varphi(U') - \varphi(L')\right]$$

where $L' = \dfrac{L-\mu}{\sigma}$ and $U' = \dfrac{U-\mu}{\sigma}$.

Here $\Phi(z)$ is the cumulative density function of the $N(0, 1)$ distribution and

$\varphi(z) = \dfrac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}$.] [Total 16]

11 A time series model is specified by

$$Y_t = 2\alpha Y_{t-1} - \alpha^2 Y_{t-2} + e_t$$

where e_t is a white noise process with variance σ^2.

(i) Determine the values of α for which the process is stationary. [2]

(ii) Derive the auto-covariances γ_0 and γ_1 for this process and find a general recursive expression for γ_k for $k \geq 2$. [10]

(iii) Show that the auto-covariance function can be written in the form:

$$\gamma_k = A\alpha^k + kB\alpha^k$$

for some values of A, B which you should specify in terms of the constants α and σ^2. [5]

[Total 17]

END OF PAPER

INSTITUTE AND FACULTY OF ACTUARIES

EXAMINERS' REPORT

September 2010 examinations

Subject CT6 — Statistical Methods
Core Technical

Introduction

The attached subject report has been written by the Principal Examiner with the aim of helping candidates. The questions and comments are based around Core Reading as the interpretation of the syllabus to which the examiners are working. They have however given credit for any alternative approach or interpretation which they consider to be reasonable.

T J Birse
Chairman of the Board of Examiners

December 2010

1 We know that, approximately, $\theta - \hat{\theta} \approx N(0, \tau^2/n)$ where τ^2 can be approximated by 0.15.

Then $P\left[-1.96 \le \frac{\theta - \hat{\theta}}{\sqrt{0.15/n}} \le 1.96\right] = 0.95$

And we require $1.96 \times \sqrt{\frac{0.15}{n}} \le 0.01$

That is $n \ge \frac{1.96^2 \times 0.15}{0.01^2} = 5762.4$ i.e. n must be at least 5763

This question was generally poorly answered.

2 The adjustment coefficient satisfies the equation:

$$\lambda + \lambda\mu(1+\theta)R = \lambda M_X(R)$$

Where X is exponentially distributed with mean μ so that $M_X(t) = \frac{1/\mu}{1/\mu - t} = \frac{1}{1-\mu t}$

So we have $1 + \mu(1+\theta)R = \frac{1}{1-\mu R}$

and so $1 - \mu R + R\mu(1+\theta)(1-\mu R) = 1$

$$-\mu R + \mu R + \mu\theta R - \mu^2 R^2 (1+\theta) = 0$$

Dividing through by μR gives

$$\mu R(1+\theta) = \theta$$

So $R = \frac{\theta}{\mu(1+\theta)}$.

This question was well answered by most candidates.

3 (i) Let the parameters be c and γ as per the tables.

Then we have:

$$1-e^{-c\times 20^{\gamma}}=0.1 \text{ so } e^{-c\times 20^{\gamma}}=0.9 \text{ and so } c\times 20^{\gamma}=-\log 0.9 \text{ (A)}$$

And similarly $c\times 95^{\gamma}=-\log 0.1$ (B)

(A) divided by (B) gives $\left(\dfrac{20}{95}\right)^{\gamma}=\dfrac{\log 0.9}{\log 0.1}=0.0457575$

So $\gamma=\dfrac{\log 0.0457575}{\log\left(\dfrac{20}{95}\right)}=1.9795337$

And substituting into (A) we have $c=-\dfrac{\log 0.9}{20^{1.9795337}}=0.000280056$

(ii) The 99.5th percentile loss is given by

$$1-e^{-0.00280056x^{1.9795337}}=0.995$$

So that $-0.000280056x^{1.9795337}=\log 0.005$

$$\log x=\frac{\log\left(\dfrac{\log 0.005}{-0.000280056}\right)}{1.9795337}=4.97486366$$

So $x=e^{4.97486366}=144.73$

Most candidates scored well on this question.

4 Let the prior distribution of μ have a Gamma distribution with parameters α and λ as per the tables.

Then $\dfrac{\alpha}{\lambda}=50$ and $\dfrac{\alpha}{\lambda^2}=15^2$

Then dividing the first by the second $\lambda=\dfrac{50}{15^2}=0.22222$

And so $\alpha=50\times 0.22222=11.111111$

The posterior distribution of μ is then given by

$$f(\mu|x) \propto f(x|\mu)\, f(\mu)$$

$$\propto e^{-10\mu} \times \mu^{630} \times \mu^{10.11111} e^{-0.22222\mu}$$
$$\propto \mu^{640.11111} e^{-10.22222\mu}$$

Which is the pdf of a Gamma distribution with parameters $\alpha' = 641.11111$ and $\lambda' = 10.22222$

Now under all or nothing loss, the Bayesian estimate is given by the mode of the posterior distribution. So we must find the maximum of

$$f(x) = x^{640.11111} e^{-10.2222x} \text{ (we may ignore constants here)}$$

Differentiating:

$$f'(x) = e^{-10.22222x}\left(-10.2222x^{640.11111} + 640.1111x^{639.11111}\right)$$
$$= x^{639.1111} e^{-10.22222x}(-10.2222x + 640.11111)$$

And setting this equal to zero we get

$$x = \frac{640.111111}{10.22222} = 62.62$$

Alternatively, credit was given for differentiating the log of the posterior (which is simpler). This question was well answered by most candidates.

5 (i) The overall mean is given by $\bar{X} = \dfrac{127.9 + 88.9 + 149.7}{3} = 122.167$

$$E\left(s^2(\theta)\right) = \frac{1}{3}\sum_{i=1}^{3}\left(\frac{1}{6}\sum_{j=1}^{7}(X_{ij} - \bar{X}_i)^2\right) = \frac{335.1 + 65.1 + 33.9}{3} = 144.7$$

$$Var(m(\theta)) = \frac{1}{2}\sum_{i=1}^{3}\left(\bar{X}_i - \bar{X}\right)^2 - \frac{1}{7}E(S^2(\theta))$$
$$= \frac{(127.9 - 122.1)^2 + (88.9 - 122.1)^2 + (149.7 - 122.1)^2}{2} - \frac{144.7}{7}$$
$$= 928.14$$

So the credibility factor is $Z = \dfrac{7}{7 + {}^{144.7}\!/_{928.14}} = 0.978213$

And the credibility premia for the risks are:

For risk 1 : $0.978213 \times 127.9 + (1 - 0.978213) \times 122.167 = 127.8$
For risk 2 : $0.978213 \times 88.9 + (1 - 0.978213) \times 122.167 = 89.6$
For risk 3 : $0.978213 \times 149.7 + (1 - 0.978213) \times 122.167 = 149.1$

(ii) The data show that the variation within risks is relatively low (the S_i^2 are low, especially for the 2nd and 3rd risks) but there seems to be quite a high variation between the average claims on the risks.

With the S_i^2 being low, this variation cannot be explained just by variability in the claims, and must be due to variability in the underlying parameter.

This means that we can put relatively little weight on the information provided by the data set as a whole, and must put more on the data from the individual risks, leading to a relatively high credibility factor.

Most candidates scored well on part (i). Only the better candidates were able to give a clear explanation in part (ii).

6 (i) We must write *f(x)* in the form:

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

For some parameters θ, ϕ and functions *a*,*b* and *c*.

$$f(x) = \frac{\alpha^\alpha}{\mu^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x\alpha}{\mu}}$$

$$= \exp\left[\left(-\frac{x}{\mu} - \log\mu\right)\alpha + (\alpha-1)\log x + \alpha\log\alpha - \log\Gamma(\alpha)\right]$$

Which is of the required form with:

$$\theta = -\frac{1}{\mu}$$

$$\phi = \alpha$$

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

$$b(\theta) = -\log(-\theta) = \log \mu$$

$$c(x, \phi) = (\phi - 1)\log x + \phi \log \phi - \log \Gamma(\phi)$$

(ii) The mean and variance for members of the exponential family are given by $b'(\theta)$ and $a(\phi)b''(\theta)$.

In this case $b'(\theta) = -\dfrac{1}{\theta} = \mu$

$b''(\theta) = \theta^{-2} = \mu^2$ so the variance is μ^2 / α as required.

Generally well answered, though many candidates did not score full marks on part (i) because they failed to specify all the parameters involved.

7 (i) First note that the probability of a claim exceeding 100 is $e^{-100\lambda}$.

The likelihood function for the given data is:

$$L = C \times \lambda^{85} e^{-85\times 42\times\lambda} \times (e^{-100\times\lambda})^{39}$$

Where C is some constant. Taking logarithms gives

$$l = \log L = C' + 85\log\lambda - 85\times 42\times\lambda - 100\times 39\times\lambda$$

Differentiating with respect to λ gives

$$\frac{\partial l}{\partial \lambda} = \frac{85}{\lambda} - 85\times 42 - 100\times 39$$

Setting this expression equal to zero we get:

$$\hat{\lambda} = \frac{85}{85\times 42 + 100\times 39} = 0.011379$$

And this gives a maximum since $\dfrac{\partial^2 l}{\partial \lambda^2} = -\dfrac{85}{\lambda^2} < 0$

(ii) We must first calculate the mean amount paid by the insurer per claim. This is

$$\int_0^{100} x\lambda e^{-\lambda x}dx + 100P(X > 100) = \left[-xe^{-\lambda x}\right]_0^{100} + \int_0^{100} e^{-\lambda x}dx + 100e^{-100\lambda}$$

$$= -100e^{-100\lambda} + \left[-\frac{1}{\lambda}e^{-\lambda x}\right]_0^{100} + 100e^{-100\lambda}$$

$$= \frac{1}{\lambda}\left(1 - e^{-100\lambda}\right)$$

So we must show the given value of λ results in the actual average paid by the insurer. This is $\dfrac{85\times 42 + 39\times 100}{85+39} = 60.24$

Substituting for λ in the expression derived above, we get

$$\frac{1}{0.011164}\left(1 - e^{-100\times 0.0111654}\right) = \frac{0.6725435}{0.011164} = 60.24 \text{ as required.}$$

Stronger candidates scored well on this question, whereas the weaker candidates struggled with the calculations required in part (ii).

8 (i) The adjustment coefficient satisfies the equation

$$\lambda + \lambda(1+\theta)E(X_1)R = \lambda M_{X_1}(R)$$

That is $1 + (1+\theta)E(X_1)R = \sum_{j=1}^{M} e^{Rj} p_j$

Applying the inequality given in the question we have

$$1 + (1+\theta)E(X_1)R \le \sum_{j=1}^{M} p_j\left(\frac{j}{M}e^{RM} + 1 - \frac{j}{M}\right)$$

So $1 + (1+\theta)E(X_1)R \le \dfrac{e^{RM}}{M}\sum_{j=1}^{M} jp_j + 1 - \dfrac{1}{M}\sum_{j=1}^{M} jp_j = \dfrac{e^{RM}E(X_1)}{M} + 1 - \dfrac{E(X_1)}{M}$

So $(1+\theta)E(X_1)R \le \dfrac{E(X_1)}{M}(e^{RM} - 1)$

and so

$$(1+\theta)R \le \frac{1}{M}\left(1+RM+\frac{R^2M^2}{2!}+\frac{R^3M^3}{3!}+\cdots-1\right)=R\left(1+\frac{RM}{2!}+\frac{R^2M^2}{3!}+\cdots\right)$$

$$(1+\theta)R < R\left(1+RM+\frac{R^2M^2}{2!}+\cdots\right)=R\times e^{RM}$$

Taking logs, we have

$$\log(1+\theta) < RM$$

And so $R > \dfrac{\log(1+\theta)}{M}$ as required.

To get the other inequality, we go back to

$$1+(1+\theta)E(X_1)R=\sum_{j=1}^{M} e^{Rj}p_j$$

And so $1+(1+\theta)E(X_1)R > \displaystyle\sum_{j=1}^{M} p_j(1+Rj+\frac{R^2j^2}{2}) = 1+RE(X_1)+\frac{R^2}{2}E(X_i^2)$

So we have $(1+\theta)m_1R > m_1R+\dfrac{R^2m_2}{2}$

i.e. $\theta m_1 > \dfrac{Rm_2}{2}$

i.e. $R < \dfrac{2\theta m_1}{m_2}$ as required.

(ii) (a) In this case we have:

$$M=3$$

And $E(X_1)=2.5$ and $E(X_1^2)=(4+9)/2=6.5$

So the inequality in the question gives:

$$\frac{1}{3}\log 1.3 < R < \frac{2\times 0.3\times 2.5}{6.5}$$

That is $0.08745 < R < 0.23077$

(b) By Lundberg;s inequality $\psi(U) \le e^{-RU} \le e^{-0.08745U}$.

This question was not well answered, with relatively few candidates scoring more than 5 marks.

9 (i) The development ratio for development year 2 to development year 3 is given by 1862.3/1820 = 1.023242

Therefore W = 1762 × 1.023242 = 1803.0

Because there is no claims development beyond development year 3
X = 1803.0 also.

The development factor from development year 1 to ultimate is given by 2122.5/1805 = 1.1759003

So the ratio from development year 1 to development year 2 is given by 1.1759003/1.023242 = 1.149190785

But under the definition of the chain ladder approach, this is calculated as:

$$1.149190785 = \frac{1762+1820}{Y+1485} = \frac{3582}{Y+1485}$$

So $Y = \dfrac{3582}{1.149190785} - 1485 = 1632.0$

(ii) We require the development ratio from year 0 to year 1; this is given by:

$$\frac{1485+1632+1805}{1001+1250+1302} = \frac{4922}{3553} = 1.385308$$

The development factor to ultimate is therefore

$$1.385308 \times 1.149190785 \times 1.023242 = 1.628984285$$

And so $Z = 2278.8 - 2500 \times 0.9 \times \left(1 - \dfrac{1}{1.628984285}\right) = 1410.0$

(iii) The outstanding claims reserve is

$$1862.3 + 2122.5 + 2278.8 - 1820 - 1805 - 1410 = 1228.6$$

This slightly unusual question was nevertheless generally well answered, showing that candidates understood the principles underlying the calculations. Many candidates scored full marks here.

10 (i) We require:

- The risk of flood damage is a constant *p* for each building.
- There can only be one claim per policy per year.
- The risk of flood damage is independent from building to building.

(ii) Let the individual claim amounts net of re-insurance be *X*. Then

$$E(\alpha X) = \alpha E(X) = 400\alpha$$

And $Var(\alpha X) = \alpha^2 Var(X) = (50\alpha)^2$

So if *Y* represents the aggregate annual claims net of re-insurance, then we have:

$$E(Y) = 10,000 \times 0.03 \times 400\alpha = 120,000\alpha$$

and

$$Var(Y) = 10,000 \times 0.03 \times (50\alpha)^2 + 10,000 \times 0.03 \times 0.97 \times (400\alpha)^2 = 47,310,000\alpha^2$$
$$= (6,878.23\alpha)^2$$

We require α to be chosen so that

$$P(Y > 120,000) = 0.01$$

i.e. $P(N(0,1) > \dfrac{120,000 - 120,000\alpha}{6,878.23\alpha}) = 0.01$

$$\frac{120,000 - 120,000\alpha}{6,878.23\alpha} = 2.3263$$

i.e. $\alpha = \dfrac{120,000}{120,000 + 2.3263 \times 6,878.23} = 0.8823476; \alpha = 88.2\% \text{ } to \text{ } 3sf$

(iii) The mean claim amount for the re-insurer is $(1 - 0.882) \times 400 = 47.20$

The annual premiums for reinsurance are $10,000 \times 0.03 \times 47.20 \times 1.15 = 16,284$

(iv) We must show that using a retention of 358.50 to calculate the premium for the individual excess of loss arrangement gives the same result as the proportional reinsurance arrangement in part (ii).

We first calculate the mean claim amount paid by re-insurer. This is equal to

$$\int_{358.50}^{\infty} (x-358.50)f(x)dx$$

$$= 400\left[1-\Phi\left(\frac{358.50-400}{50}\right)\right] - 50\left[0-\varphi\left(\frac{358.50-400}{50}\right)\right] - 358.50\times\left(1-\Phi\left(\frac{358.50-400}{50}\right)\right)$$

This gives

$$400\left[1-\Phi(-0.83)\right] + 50\varphi(-0.83) - 358.50\times(1-\Phi(-0.83))$$

$$= 400\times 0.79673 + 50\times\frac{1}{\sqrt{2\pi}}e^{-\frac{0.83^2}{2}} - 358.50\times 0.79673$$

$$= 47.20$$

Then the aggregate premium charged will be $10,000\times 0.03\times 47.20\times 1.15$ = 16,284 which is the same as under the first arrangement as required.

Carrying forward more than 3 significant figures from the result in (ii) gives a slightly different value in (iii). To full accuracy, the solution in (iii) becomes 16,236 resulting in a minor discrepancy between the answers in (iii) and (iv). This appears not to have concerned candidates who were generally happy to observe that the results in (iii) and (iv) were approximately equal. The examiners gave credit for either approach.

This question was a good differentiator – the better prepared candidates were able to score well whilst weaker candidates struggled.

11 (i) Let B be the backward shift operator. Then the time series has the form:

$$(1-2\alpha B+\alpha^2B^2)Y_t = e_t$$

$$(1-\alpha B)^2 Y_t = e_t$$

And the roots of the characteristic equation will have modulus greater than 1 ands so the series will be stationary provided that $|\alpha| < 1$.

(ii) Firstly, note that $Cov(Y_t, e_t) = Cov(e_t, e_t) = \sigma^2$

So, taking the covariance of the defining equation with Y_t we get:

$$\gamma_0 = 2\alpha\gamma_1 - \alpha^2\gamma_2 + \sigma^2 \quad \text{(A)}$$

Taking the covariance with Y_{t-1} we get

$$\gamma_1 = 2\alpha\gamma_0 - \alpha^2\gamma_1$$

i.e. $(1+\alpha^2)\gamma_1 = 2\alpha\gamma_0$ (B)

Finally, taking the covariance with Y_{t-2} gives:

$$\gamma_2 = 2\alpha\gamma_1 - \alpha^2\gamma_0 \quad \text{(C)}$$

In general, for $k \geq 2$ we have $\gamma_k = 2\alpha\gamma_{k-1} - \alpha^2\gamma_{k-2}$

Substituting the expression for γ_2 in (C) into (A) gives:

$$\gamma_0 = 2\alpha\gamma_1 - \alpha^2(2\alpha\gamma_1 - \alpha^2\gamma_0) + \sigma^2$$

So that

$$(1-\alpha^4)\gamma_0 = 2\alpha(1-\alpha^2)\gamma_1 + \sigma^2$$

And now substituting the expression for γ_1 in (B) we get

$$(1-\alpha^4)\gamma_0 = 2\alpha(1-\alpha^2) \times \frac{2\alpha\gamma_0}{(1+\alpha^2)} + \sigma^2$$

$$\left(1-\alpha^4 - \frac{4\alpha^2(1-\alpha^2)}{1+\alpha^2}\right)\gamma_0 = \sigma^2$$

$$(1+\alpha^2-\alpha^4-\alpha^6-4\alpha^2+4\alpha^4)\gamma_0 = (1+\alpha^2)\sigma^2$$

So $\gamma_0 = \dfrac{(1+\alpha^2)}{(1-3\alpha^2+3\alpha^4-\alpha^6)}\sigma^2 = \dfrac{(1+\alpha^2)}{(1-\alpha^2)^3}\sigma^2$

And so $\gamma_1 = \dfrac{2\alpha\gamma_0}{1+\alpha^2} = \left(\dfrac{2\alpha(1+\alpha^2)}{(1-\alpha^2)^3(1+\alpha^2)}\right)\sigma^2 = \dfrac{2\alpha}{(1-\alpha^2)^3}\sigma^2$

And more generally $\gamma_k = 2\alpha\gamma_{k-1} - \alpha^2\gamma_{k-2}$ (D)

(iii) Suppose $\gamma_{k-1} = A\alpha^{k-1} + (k-1)B\alpha^{k-1}$ and $\gamma_{k-2} = A\alpha^{k-2} + (k-2)B\alpha^{k-2}$ and substitute into (D).

$$\gamma_k = 2\alpha A\alpha^{k-1} + 2\alpha(k-1)B\alpha^{k-1} - \alpha^2 A\alpha^{k-2} - (k-2)\alpha^2 B\alpha^{k-2}$$

$$= A(2\alpha^k - \alpha^k) + B(2\alpha^k(k-1) - (k-2)\alpha^k) = A\alpha^k + Bk\alpha^k$$

Which is of the correct form, so the general form of the expression holds.

Setting $k = 0$ we get $\gamma_0 = A$

So $A = \dfrac{(1+\alpha^2)}{(1-\alpha^2)^3}\sigma^2$

Setting $k = 1$ we get $\gamma_1 = (A+B)\alpha$

So $B = \dfrac{\gamma_1}{\alpha} - A = \left(\dfrac{2\alpha}{\alpha(1-\alpha^2)^3}\right)\sigma^2 - \dfrac{(1+\alpha^2)}{(1-\alpha^2)^3}\sigma^2 = \left(\dfrac{1-\alpha^2}{(1-\alpha^2)^3}\right)\sigma^2 = \dfrac{\sigma^2}{(1-\alpha^2)^2}$

[Alternatively, solve using the formula on page 4 of the Tables:

We have $g_k = 2\alpha g_{k-1} + \alpha^2 g_{k-2} = 0$

Using the Tables formula, the roots are $\lambda_1 = \lambda_2 = \alpha$ so we have a solution of the form $g_k = (A + Bk)\lambda^k = (A + Bk)\alpha^k$

Set $k = 0$ and $k = 1$ to get the same equations as before.]

Another good differentiator, with strong candidates scoring well, and weaker candidates struggling with parts (ii) and (iii) in particular.

END OF EXAMINERS' REPORT

INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

28 April 2011 (am)

Subject CT6 — Statistical Methods
Core Technical

Time allowed: Three hours

INSTRUCTIONS TO THE CANDIDATE

1. *Enter all the candidate and examination details as requested on the front of your answer booklet.*
2. *You must not start writing your answers in the booklet until instructed to do so by the supervisor.*
3. *Mark allocations are shown in brackets.*
4. *Attempt all 11 questions, beginning your answer to each question on a separate sheet.*
5. *Candidates should show calculations where this is appropriate.*

Graph paper is NOT required for this paper.

AT THE END OF THE EXAMINATION

Hand in BOTH your answer booklet, with any additional sheets firmly attached, and this question paper.

In addition to this paper you should have available the 2002 edition of the Formulae and Tables and your own electronic calculator from the approved list.

1 Give two examples of exercises where Monte-Carlo simulation should be performed using the same choice of random numbers, explaining your reasoning in each case. [4]

2 An insurance company has collected data for the number of claims arising from certain risks over the last 10 years. The number of claims in the jth year from the ith risk is denoted by X_{ij} for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, 10$. The distribution of X_{ij} for $j = 1\ 2, \ldots, 10$ depends on an unknown parameter θ_i and given θ_i the X_{ij} are independent identically distributed random variables.

(i) Give a brief interpretation of $E[s^2(\theta)]$ and $V[m(\theta))]$ under the assumptions of Empirical Bayes Credibility Theory Model 1. [2]

(ii) Explain how the value of the credibility factor Z depends on $E[s^2(\theta)]$ and $V[m(\theta)]$. [3]

[Total 5]

3 Let $y_1, \ldots, y_n$ be samples from a uniform distribution on the interval $[0, \theta]$ where $\theta > 0$ is an unknown constant. Prior beliefs about θ are given by a distribution with density

$$f(\theta) = \begin{cases} \alpha\beta^{\alpha}\theta^{-(1+\alpha)} & \theta > \beta \\ 0 & \text{otherwise} \end{cases}$$

where α and β are positive constants.

(i) Show that the posterior distribution of θ given y_1 is of the same form as the prior distribution, specifying the parameters involved. [4]

(ii) Write down the posterior distribution of θ given $y_1, \ldots, y_n$. [2]

[Total 6]

4 The annual number of claims on an insurance policy within a certain portfolio follows a Poisson distribution with mean μ. The parameter μ varies from policy to policy and can be considered as a random variable that follows an exponential distribution with mean $1/\lambda$.

Find the unconditional distribution of the annual number of claims on a randomly chosen policy from the portfolio. [6]

5 The number of claims under an insurance policy in a year is either 0 (with probability 40%) or 1 (with probability 20%) or 2 (with probability 40%). Individual claim amounts are equally likely to be 50 or 20. The insurance company calculates premiums using a premium loading of 50% and is considering operating one of the following arrangements:

(A) Making no changes.

(B) Introducing a policy excess of 10 (per claim) in return for a reduction of 5 in premiums.

(C) Effecting an individual excess of loss reinsurance arrangement with retention 30 for a premium of 10.

Construct a table of the insurance company's profits under all the possible outcomes for each of (A) (B) and (C) and hence determine the optimal arrangement using the Bayes criteria. [8]

6 The double exponential distribution with parameter $\lambda > 0$ has density given by

$$g(x) = \tfrac{1}{2}\lambda e^{-\lambda|x|} \quad x \in \mathbb{R}.$$

(i) Construct an algorithm for generating samples from this distribution. [3]

(ii) Construct an algorithm for producing samples from a $N(0,1)$ distribution using samples from the double exponential distribution and the acceptance-rejection method. [6]

[Total 9]

7 Consider the time series

$$Y_t = 0.7 + 0.4Y_{t-1} + 0.12Y_{t-2} + e_t$$

where e_t is a white noise process with variance σ^2.

(i) Identify the model as an ARIMA(p,d,q) process. [1]

(ii) Determine whether Y_t is a stationary process. [2]

(iii) Calculate $E(Y_t)$. [2]

(iv) Calculate the auto-correlations ρ_1, ρ_2, ρ_3 and ρ_4. [4]

[Total 9]

8 Suppose that Y is a random variable belonging to a special subset of the exponential family where the density function of Y has the form

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

For some constants θ and φ and functions b and c.

(i) Show that the moment generating function of Y is given by

$$M_Y(t)=\exp\left[\frac{b(\theta+t\varphi)-b(\theta)}{\varphi}\right] \qquad [3]$$

Hint: Note that the function $f(y, \theta + \varphi t, \varphi)$ is the density of another random variable of the same family and hence $\int_{-\infty}^{\infty} f(y,\theta+\varphi t,\varphi)dy=1$.

(ii) Show that $E(Y) = b'(\theta)$ and $\mathrm{Var}(Y) = \varphi b''(\theta)$ using the result in (i). [4]

(iii) Verify that the result in (i) holds if Y has a Poisson distribution. [4]
[Total 11]

9 Claims on a portfolio of insurance policies arise as a Poisson process with parameter λ. Individual claim amounts are taken from a distribution X and we define $m_i = E(X^i)$ for i = 1, 2, …. The insurance company calculates premiums using a premium loading of θ.

(i) Define the adjustment coefficient R. [1]

(ii) (a) Show that R can be approximated by $\dfrac{2\theta m_1}{m_2}$ by truncating the series expansion of $M_X(t)$.

(b) Show that there is another approximation to R which is a solution of the equation $m_3y^2 + 3m_2y - 6\theta m_1 = 0$. [6]

Now suppose that X has an exponential distribution with mean 10 and that θ = 0.3.

(iii) Calculate the approximations to R in (ii) and (iii) and compare them to the true value of R. [6]
[Total 13]

10 The number of claims on a portfolio of insurance policies has a Poisson distribution with mean 200. Individual claim amounts are exponentially distributed with mean 40. The insurance company calculates premiums using a premium loading of 40% and is considering entering into one of the following re-insurance arrangements:

(A) No reinsurance.

(B) Individual excess of loss insurance with retention 60 with a reinsurance company that calculates premiums using a premium loading of 55%.

(C) Proportional reinsurance with retention 75% with a reinsurance company that calculates premiums using a premium loading of 45%.

(i) Find the insurance company's expected profit under each arrangement. [6]

(ii) Find the probability that the insurer makes a profit of less than 2000 under each of the arrangements using a normal approximation. [8]

[Total 14]

11 The table below shows cumulative claims paid on a portfolio of insurance policies.

	Development Year			
Accident Year	*0*	*1*	*2*	*3*
2007	240	281.4	302	305
2008	260	320	322	
2009	270	312.9		
2010	276			

All claims are fully run off by the end of development year 3.

(i) Calculate the total reserve for outstanding claims using the basic chain ladder technique. [7]

An actuary is considering modelling the future claims assuming that individual development factors are lognormally distributed with the following parameters:

	Development Year		
Parameter	*0 to 1*	*1 to 2*	*2 to 3*
μ	0.171251	0.035850	0.008787
σ	0.032148	0.045606	0.046853

(ii) Show that under these assumptions the cumulative development factor to ultimate is also lognormally distributed. [3]

(iii) Calculate a 99% upper confidence limit for the outstanding claims relating to the 2010 accident year. [5]

[Total 15]

END OF PAPER

INSTITUTE AND FACULTY OF ACTUARIES

EXAMINERS' REPORT

April 2011 examinations

Subject CT6 — Statistical Methods
Core Technical

Introduction

The attached subject report has been written by the Principal Examiner with the aim of helping candidates. The questions and comments are based around Core Reading as the interpretation of the syllabus to which the examiners are working. They have however given credit for any alternative approach or interpretation which they consider to be reasonable.

T J Birse
Chairman of the Board of Examiners

July 2011

1 Example 1: Testing sensitivity to parameter variation – we want the results to change as a result of changes to the parameter not as a result of variations in the random numbers.

Example 2: Performance evaluation. When comparing two or more schemes which might be adopted we want differences in results to arise from differences between the schemes rather than as a result of variations in the random numbers.

Alternative Example: The same set of simulations could be used for the numerical evaluation of derivatives

$$\theta'(\alpha) = \frac{\theta(\alpha+\delta) - \theta(\alpha)}{\delta}$$

This question was generally answered well, although weaker candidates explained how to obtain random numbers or perform Monte-Carlo simulation instead of explaining why you would want to use the <u>*same*</u> *random numbers.*

2 (i) $E(s^2(\theta))$ represents the average variability of claim numbers from year to year for a single risk.

$V(m(\theta))$ represents the variability of the average claim numbers for different risks i.e. the variability of the means from risk to risk.

(ii) The credibility factor is given by

$$Z = \frac{n}{n + \dfrac{E(s^2(\theta))}{V(m(\theta))}}$$

We can see that it is the relative values of $E(s2(\theta))$ and $V(m(\theta))$ that matter. In particular, if $E(s2(\theta))$ is high relative to $V(m(\theta))$, this means that there is more variability from year to year than from risk to risk. More credibility can be placed on the data from other risks leading to a lower value of Z.

On the other hand, if $V(m(\theta))$ is relatively higher this means there is greater variation from risk to risk, so that we can place less reliance on the data as a whole leading to a higher value of Z.

This question was generally well answered. Weaker students were not able to give clear, concise descriptions of the quantities in part (i).

3 (i) First note that $f(y1|\theta) = \dfrac{1}{\theta}$ for $\theta > y1$

$$f(\theta|y1) \propto f(y1|\theta)\, f(\theta)$$

$$= \frac{1}{\theta}\,\alpha\beta\alpha\ \theta\text{–}(1+\alpha) \text{ for } \theta > \beta \text{ and } \theta > y1$$

$$\propto \theta\text{–}(2+\alpha) \text{ for } \theta > \max(\beta, y1)$$

$$\propto (\alpha + 1)\ \bar{\beta}^{\alpha+1}\theta\text{–}(1+1+\alpha) \text{ for } \theta > \bar{\beta} \text{ where } \bar{\beta} = \max(\beta, y1)$$

Which is of the same form with parameters $\alpha + 1$ and $\bar{\beta}$.

Alternatively, we can derive this formally as:

$$f(\theta|y_1) = \frac{f(y_1|\theta)f(\theta)}{\int_\theta f(y_1|\theta)f(\theta)d\theta}$$

This gives:

$$f(\theta|y_1) = \frac{\frac{1}{\theta}\times\alpha\beta^\alpha\theta^{-(1+\alpha)}}{\int_\beta^\infty \alpha\beta^\alpha\theta^{-\alpha}d\theta} = \frac{\alpha\beta^\alpha\theta^{-(2+\alpha)}}{\frac{\alpha}{\alpha+1}\beta^{-1}} = (\alpha+1)\beta^{\alpha+1}\theta^{-(2+\alpha)}$$

(ii) The posterior distribution has the same form with parameters $\alpha + n$ and $\max(\beta, y1, \ldots, yn)$.

This question received a wide range of quality of answers. Only the strongest candidates stated the correct range for the posterior in part (i). A number of candidates assumed a sample of size n *in part (i) and therefore failed to differentiate between parts (i) and (ii).*

4 Let the annual number of claims be denoted by N. Then

$$P(N = k) = \int_0^\infty P(N = k|\mu)\, f(\mu)\, d\mu$$

$$= \int_0^\infty e^{-\mu}\frac{\mu^k}{k!}\lambda e^{-\lambda\mu}d\mu$$

$$= \frac{\lambda}{k!}\int_0^\infty \mu^k e^{-(1+\lambda)\mu}d\mu$$

$$= \frac{\lambda}{k!}\times\frac{\Gamma(k+1)}{(1+\lambda)^{k+1}}\int_0^\infty\frac{(1+\lambda)^{k+1}}{\Gamma(k+1)}\mu^k e^{-(1+\lambda)\mu}d\mu$$

$$= \frac{\lambda}{(1+\lambda)^{k+1}}\times 1$$

Where the final integral is 1 since the integrand is the pdf of a Gamma distribution.

So

$$P(N=k) = \frac{\lambda}{(1+\lambda)^{k+1}} = \frac{\lambda}{1+\lambda} \times \frac{1}{(1+\lambda)^k}, \text{ for k = 0, 1, 2, …}$$

Which means that N has a geometric distribution with parameter $p = \frac{\lambda}{1+\lambda}$. This is equivalent to a Type II negative binomial with k=1

This was the worst answered question on this paper, with few candidates able to write down the first integral. Candidates who attempted the algebra often did not recognise the resultant distribution.

5 Premiums charged to the policyholder are

$1.5 \times (0.2 \times (0.5 \times 50 + 0.5 \times 20) + 0.4 \times (0.25 \times 100 + 0.5 \times 70 + 0.25 \times 40)) = 52.5$

The completed table is

		Claims					
Arrangement	*Net Premium*	*None*	*1 L*	*1 S*	*1S 1L*	*2 S*	*2 L*
A	52.5	0	50	20	70	40	100
B	47.5	0	40	10	50	20	80
C	42.5	0	30	20	50	40	60

So the completed table of profits for the insurer is:

	Profit						
	None	*1 L*	*1 S*	*1S 1L*	*2 S*	*2 L*	*E(P)*
Probability	0.4	0.1	0.1	0.2	0.1	0.1	
A	52.5	2.5	32.5	–17.5	12.5	–47.5	17.5
B	47.5	7.5	37.5	–2.5	27.5	–32.5	22.5
C	42.5	12.5	22.5	–7.5	2.5	–17.5	17.5

So the insurer should introduce the policy excess (arrangement **B**)

This question was answered well. Some candidates gave incorrect completed tables without showing any working, and the examiners were therefore unable to give partial credit in these cases.

6 (i) We first note that the distribution function for this double exponential is given by

$$G(x) = \begin{cases} 0.5e^{\lambda x} & x < 0 \\ 0.5(1 - e^{-\lambda x}) + 0.5 & x \geq 0 \end{cases}$$

And so the inverse function is given by

$$G^{-1}(u) = \begin{cases} \dfrac{\log 2u}{\lambda} & u < 0.5 \\ -\dfrac{\log(2(1-u))}{\lambda} & u \geq 0.5 \end{cases}$$

And so our algorithm is:

(A) Generate u from $U(0,1)$

(B) If $u < 0.5$ set $x = \dfrac{\log 2u}{\lambda}$ otherwise set $x = -\dfrac{\log(2(1-u))}{\lambda}$

(ii) We must first find $M = \operatorname{Sup} \dfrac{f(x)}{g(x)}$ where $f(x)$ is the pdf of the $N(0,1)$ distribution.

$$\operatorname{Sup} \frac{f(x)}{g(x)} = \operatorname{Sup} \frac{2}{\lambda\sqrt{2\pi}} e^{-\frac{x^2}{2} + \lambda|x|}$$

And using the symmetry around 0 we can concentrate on positive values of x

$$\operatorname{Sup} \frac{f(x)}{g(x)} = \operatorname{Sup} \frac{2}{\lambda\sqrt{2\pi}} e^{-\frac{x^2}{2} + \lambda x}$$

And the exponential expression is maximised when $-\dfrac{x^2}{2} + \lambda x$ is maximised.

Differentiating, this occurs when $-\lambda + x = 0$ i.e. $x = \lambda$

Hence $M = \operatorname{Sup} \dfrac{f(x)}{g(x)} = \dfrac{2}{\lambda\sqrt{2\pi}} e^{\frac{\lambda^2}{2}}$. So define

$$h(x) = \frac{f(x)}{Mg(x)} = e^{-\frac{x^2}{2} + \lambda|x| - \frac{\lambda^2}{2}}$$

So algorithm is as follows:

(A) Generate x as in part (i)
(B) Generate u from $U(0,1)$
(C) If $u \le h(x)$ then set $y = x$ otherwise return to (A)

This question was poorly answered. Only the strongest candidates treated the modulus in the pdf correctly. There was also difficulty deriving both parts of the inverse function for both the u ≥ 0.5 and the u < 0.5 case.

7 (i) The model is ARIMA(2,0,0) provided that the model is stationary.

(ii) The lag polynomial is $1 - 0.4L - 0.12L2 = (1 - 0.6L)(1 + 0.2L)$

Since the roots $\frac{1}{0.6}$ and $-\frac{1}{0.2}$ are both greater than one in absolute value the process is stationary.

(iii) Since the process is stationary we know that $E(Yt)$ is equal to some constant μ independent of t.

Taking expectations on both sides of the equation defining Yt gives

$E(Yt) = 0.7 + 0.4E(Yt-1) + 0.12E(Yt-2)$

$\mu = 0.7 + 0.4\mu + 0.12\mu$

$$\mu = \frac{0.7}{1-0.4-0.12} = 1.45833333$$

(iv) The auto-covariance function is not affected by the constant term of 0.7 in the equation, and this term can be ignored.

The Yule-Walker equations are

$\gamma0 = 0.4\gamma1 + 0.12\gamma2 + \sigma2$

$\gamma1 = 0.4\gamma0 + 0.12\gamma1$ (A)

$\gamma2 = 0.4\gamma1 + 0.12\gamma0$ (B)

$\gamma s = 0.4\gamma s-1 + 0.12\gamma s-2$ for $s > 2$

Dividing both sides of (A) by $\gamma0$ and noting that $\rho_s = \frac{\gamma_s}{\gamma_0}$ we have

$$\rho1 = 0.4 + 0.12\rho1 \text{ so that } \rho1 = \frac{0.4}{0.88} = 0.45454545$$

Substituting this result into (B) we have

$\rho 2 = 0.4 \times 0.45454545 + 0.12 = 0.30181818$

And using the final result we have

$\rho 3 = 0.4 \times 0.3018181818 + 0.12 \times 0.45454545 = 0.1752727$

and

$\rho 4 = 0.4 \times 0.1752727 + 0.12 \times 0.30181818 = 0.10632727$

Expressed as fractions:

$\rho_1 = \frac{5}{11} \quad \rho_2 = \frac{83}{275} \quad \rho_3 = \frac{241}{1375} \quad \rho_4 = \frac{731}{6875}$

This straightforward question was answered well.

8 (i) $MY(t) = \int e^{ty} f(y,\theta,\varphi)\, dy$

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

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

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

$$= \exp\left[\frac{b(\theta + t\varphi) - b(\theta)}{\varphi}\right] \times 1$$

using the hint to evaluate the second integral.

(ii) $$\frac{dM_Y(t)}{dt} = \frac{d}{dt}\left(\frac{b(\theta + t\varphi) - b(\theta)}{\varphi}\right) M_Y(t)$$

$$= \frac{\varphi b'(\theta + t\varphi)}{\varphi} M_Y(t)$$

$$= b'(\theta + t\varphi) M_Y(t)$$

And $E(Y) \;=\; M'_Y(0) \;=\; b'(\theta)M_Y(0) \;=\; b'(\theta)\times 1 = \; b'(\theta)$

$$\frac{d^2 M_Y(t)}{dt^2} \;=\; \varphi b''(\theta + t\varphi)M_Y(t) + b'(\theta + t\varphi)M'_Y(t)$$

So $E(Y2) \;=\; M''_Y(0) \;=\; \varphi b''(\theta)M_Y(0) + b'(\theta)M'_Y(0)$

$$= \; \varphi b''(\theta) + b'(\theta)^2$$

So $\text{Var}(Y) \;= E(Y2) - E(Y)2 = \; \varphi b''(\theta) + b'(\theta)^2 - b'(\theta)^2 \;=\; \varphi b''(\theta)$

Credit given for alternative approaches (e.g. CGF).

(iii) For the Poisson distribution with parameter μ we have

$$f(y, \theta, \varphi) = \frac{\mu^y e^{-\mu}}{y!} = \exp[y \log \mu - \mu - \log y!]$$

Which is of the form in the question with $\theta = \log \mu$, $\varphi = 1$ and $b(\theta) = e\theta$ and $c(y, \varphi) = -\log y!$

So the result from (i) gives

$$MY(t) \;=\; \exp\left[\frac{b(\theta + t\varphi) - b(\theta)}{\varphi}\right]$$

$$= \; \exp\left[\frac{e^{\log\mu + t} - e^{\log\mu}}{1}\right] = \exp[\mu(et - 1)]$$

which is indeed the MGF of the Poisson distribution as shown on p7 of the tables.

This question was generally well done, with many candidates who could not complete the derivation in part (i) nevertheless able to use the result to score well in parts (ii) and (iii). For part (iii) some candidates calculated the first two moments rather than showing that the MGF of the Poisson distribution has the form given in part (i).

9 (i) The adjustment coefficient is the unique positive solution to

$\lambda M_X(R) - \lambda - \lambda(1 + \theta) E(X) R = 0$

(ii) Cancelling the λ terms we have

(a) $M_X(R) = E(e^{RX}) = 1 + (1 + \theta) E(X)R$

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

And truncating the expression we get

$E(1 + RX + R^2X^2/2) = 1 + (1 + \theta) E(X)R$

i.e. $1 + Rm_1 + R^2m_2/2 = 1 + (1 + \theta) m_1R$

i.e. $R^2m_2 = 2\theta m_1R$

i.e. $R = \dfrac{2\theta m_1}{m_2}$

(b) Once more we have

$$E\left(1 + RX + \frac{R^2X^2}{2!} + \frac{R^3X^3}{3!} + \ldots.\right) = 1 + (1 + \theta) E(X)R$$

And truncating the expression we get

$$E\left(1 + RX + \frac{R^2X^2}{2} + \frac{R^3X^3}{6}\right) = 1 + (1 + \theta) E(X)R$$

i.e. $1 + Rm_1 + \dfrac{R^2m_2}{2} + \dfrac{R^3m_3}{6} = 1 + (1 + \theta)m_1R$

i.e. $3R^2m_2 + R^3m_3 = 6\theta m_1R$

i.e. $m_3R^2 + 3Rm_2 - 6\theta m_1 = 0$

As required

(iii) In this case $m_1 = 10$ and $m_2 = 200$ and $m_3 = 6000$

So the estimate from (ii) (a) is $R = \dfrac{2\theta m_1}{m_2} = \dfrac{2\times 0.3\times 10}{200} = \dfrac{6}{200} = 0.03$

The estimate from (ii) (b) is the solution to $6000R^2 + 600R - 18 = 0$

Which is given by

$$\frac{-b\pm\sqrt{b^2-4ac}}{2a} = \frac{-600+\sqrt{600^2+4\times 6000\times 18}}{12000} = 0.024161984$$

[The negative root of the equation is –0.12416]

The true value of R is given by the solution to

$M_X(R) = E(e^{RX}) = 1 + (1 + \theta)\, E(X)R$

That is $\dfrac{\mu}{\mu - R} = 1 + \dfrac{(1+\theta)R}{\mu}$ where $\mu = {}^1\!/_{10}$ is the parameter of the exponential distribution.

And so

$\mu^2 = \mu(\mu - R) + (1 + \theta)\, R(\mu - R)$

$\mu^2 = \mu^2 - \mu R + \mu R - R^2 + \mu\theta R - \theta R^2$

$0 = \mu\theta - R(1 + \theta)$

$R = \dfrac{\mu\theta}{1+\theta} = \dfrac{0.1\times 0.3}{1.3} = 0.0230769$

So the first estimate gives a greater error than the second (the error is 30% for the first approximation and only about 4.7% for the second). This is as we would expect since we took more terms before truncating.

Candidates generally answered part (i) and part (ii) well, although part (ii) b) caused problems and many candidates did not give sufficient detail. In (iii) many candidates produced an answer wrongly using a denominator of 100, or calculated the estimate but then did not explain the difference in estimates adequately.

10 (i) Denote the insurers profits by Z

Under A:

Premium income = 200 × 40 × 1.4 = 11200

Expected claims = 200 × 40 = 8000

So $E(Z)$ = 11200 − 8000 = 3200

Under B

We need first to calculate the expected loss for the insurer. Denote the insurer's loss by X. Then

$$E(X) = \int_0^{60} 0.025xe^{-0.025x}dx + 60 \times \int_{60}^{\infty} 0.025e^{-0.025x}dx$$

$$= [-xe^{-0.025x}]_0^{60} + \int_0^{60} e^{-0.025x}dx + 60 \times [-e^{-0.025x}]_{60}^{\infty}$$

$$= -60e^{-1.5} + [-40e^{-0.025x}]_0^{60} + 60e^{-1.5}$$

= 40 − 40e–1.5 = 31.07479359

So the expected loss for the re-insurer is 40 − 31.07479359 = 8.925206406

Premium income = 11200 − 200 × 1.55 × 8.925206406 = 8433.186014

Expected claims = 200 × 31.07479359 = 6214.958718

So $E(Z)$ = 8433.186014 − 6214.958718 = 2218.227

Under C

Premium income = 200 × 40 × 1.4 − 200 × 40 × 0.25 × 1.45 = 8300

Expected claims = 200 × 40 × 0.75 = 6000

So $E(Z)$ = 8300 − 6000 = 2300

(ii) We now need to find the variance of the total claim amount paid by the insurer. Denote this by Y. Then

<u>Under A</u>

Var(Y) = 200Var(X) + 200E(X)2

= 200 × 402 + 200 × 402 = 640,000 = 8002

So

$$\Pr(Z < 2000) = \Pr(Y > 9200) = \Pr\left(N(0,1) > \frac{9200-8000}{800}\right)$$

= Pr(N(0,1) > 1.5) = (1 – 0.93319) = 0.06681

<u>Under B</u>

We first need to find E(X2) as defined above.

$$E(X2) \;=\; \int_0^{60} 0.025x^2e^{-0.025x}dx + 60^2\int_{60}^{\infty}0.025e^{-0.025x}dx$$

$$= [-x^2e^{-0.025x}]_0^{60} + \int_0^{60} 2xe^{-0.025x}dx + 3600e^{-1.5}$$

$$= -3600e^{-1.5} + \frac{2}{0.025}\int_0^{60} 0.025xe^{-0.025x}dx + 3600e^{-1.5}$$

$$= \frac{2}{0.025}(E(X) - 60e^{-1.5}) \; = 1414.958718$$

And so

Var(X) = 1414.958718 – 31.07479359^2 = 449.3159219

And therefore

Var(Y) = 200Var(X) + 200E(X)2

= 200 × 449.3159219 + 200 × 31.074793592 = 282991.7438

Finally

Pr(Z < 2000) = Pr(Y > 6433.186014)

$$= \Pr\left(N(0,1) > \frac{6433.186014 - 6214.958718}{531.97}\right)$$

= Pr(N(0,1) > 0.41023) = 1 – 0.65918 = 0.34082

<u>Under C</u>

Var(Y) = 200Var(X) + 200E(X)2

Var(Y) = 200 × 0.752 × 402 + 200 × (0.75 × 30)2 = 360000 = 6002

So

Pr(Z < 2000) = Pr(Y > 6300) = $\Pr\left(N(0,1) > \dfrac{6300-6000}{600}\right)$

= Pr(N(0,1) > 0.5) = (1 – 0.69146) = 0.30854

This question received a wide range of quality of answers. Most candidates calculated A and C correctly, but many failed to produce a reasonable answer for B. Common errors for A and C included not re-calculating the variance and using the wrong claim amount when calculating the probability. Some candidates were unable to calculate the normal distribution probability correctly after deriving the correct claim and variance values. For arrangement B many struggled to evaluate the integral correctly.

11 (i) The development factors are:

r0,1 = $\dfrac{281.4+320+312.9}{240+260+270} = \dfrac{914.3}{770} = 1.187403$

r1,2 = $\dfrac{302+322}{281.4+320} = \dfrac{624}{601.4} = 1.037579$

r2,3 = $\dfrac{305}{302}$ = 1.009934

And the ultimate claims are:

For AY 2008: 322 × 1.009934 = 325.20

For AY 2009: 312.9 × 1.037579 × 1.009934 = 327.88

For AY 2010: 276 × 1.187403 × 1.037579 × 1.009934 = 343.42

So outstanding claims reserve is

325.20 + 327.88 + 343.42 – 322 – 312.9 – 276 = 85.60

(ii) Suppose that $R_{i,i+1} \sim \log N(\mu_i, \sigma_i^2)$ for $i = 0, 1, 2$.

Then $\log R_{i,i+1} \sim N(\mu_i, \sigma_i^2)$.

So (for example)

$\log R_{i,i+1} R_{i+1,i+2} = \log R_{i,i+1} + \log R_{i+1,i+2} \sim N(\mu_i + \mu_{i+1}, \sigma_i^2 + \sigma_{i+1}^2)$

Which means that the product $R_{i,i+1} R_{i+1,i+2}$ is also log-normally distributed. Since any product of log-normally distributed development factors is also log-normally distributed the development factors to ultimate must also be log-normally distributed.

(iii) Using the results from (ii) the development factor to ultimate for AY 2010 is log-normally distributed with parameters:

$\mu = 0.171251 + 0.035850 + 0.008787 = 0.215889$

$\sigma 2 = 0.0321482 + 0.0456062 + 0.0468532 = 0.0728602$

So an upper 99% confidence limit for the development factor to ultimate is given by $\exp(0.215889 + 0.07286 \times 2.3263) = 1.47018$

So an upper 99% confidence limit for total claims is
$1.47018 \times 276 = 405.77$

So an upper 99% confidence limit for outstanding claims is
$405.77 - 276 = 129.77$

Part (i) was answered very well by the majority of candidates. Parts (ii) and (iii) were answered poorly. Many candidates failed to produce sufficient detail in part (ii) for instance calculating the parameters of the distribution rather than explaining why it was a log-normal. For part (iii) few candidates were able to calculate the parameters of the distribution correctly. A further common mistake was to calculate a two-sided confidence interval.

END OF EXAMINERS' REPORT

INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

6 October 2011 (am)

Subject CT6 — Statistical Methods
Core Technical

Time allowed: Three hours

INSTRUCTIONS TO THE CANDIDATE

1. *Enter all the candidate and examination details as requested on the front of your answer booklet.*

2. *You must not start writing your answers in the booklet until instructed to do so by the supervisor.*

3. *Mark allocations are shown in brackets.*

4. *Attempt all 11 questions, beginning your answer to each question on a separate sheet.*

5. *Candidates should show calculations where this is appropriate.*

Graph paper is NOT required for this paper.

AT THE END OF THE EXAMINATION

Hand in BOTH your answer booklet, with any additional sheets firmly attached, and this question paper.

In addition to this paper you should have available the 2002 edition of the Formulae and Tables and your own electronic calculator from the approved list.

1 The loss function for a decision problem is given below.

	θ_1	θ_2	θ_3
D1	30	20	15
D2	20	25	10
D3	17	19	24
D4	35	20	17

(i) Explain which strategies, if any, are dominated. [3]

(ii) Find the minimax solution to this decision problem. [1]
[Total 4]

2 An accountant is using a psychic octopus to predict the outcome of tosses of a fair coin. He claims that the octopus has a probability $p > 0.5$ of successfully predicting the outcome of any given coin toss. His actuarial colleague is extremely sceptical and summarises his prior beliefs about p as follows: there is an 80% chance that $p = 0.5$ and a 20% chance that p is uniformly distributed on the interval [0.5,1]. The octopus successfully predicts the results of 7 out of 8 coin tosses.

Calculate the posterior probability that $p = 0.5$. [4]

3 Loss amounts under a class of insurance policies follow an exponential distribution with mean 100. The insurance company wishes to enter into an individual excess of loss reinsurance arrangement with retention level M set such that 8 out of 10 claims will not involve the reinsurer.

(i) Find the retention M. [2]

For a given claim, let X_I denote the amount paid by the insurer and X_R the amount paid by the reinsurer.

(ii) Calculate $E(X_I)$ and $E(X_R)$. [3]
[Total 5]

4 Claims on a portfolio of insurance policies follow a Poisson process with parameter λ. The insurance company calculates premiums using a premium loading of θ and has an initial surplus of U.

(i) Define the surplus process $U(t)$. [1]

(ii) Define the probabilities $\psi(U, t)$ and $\psi(U)$. [2]

(iii) Explain how $\psi(U, t)$ and $\psi(U)$ depend on λ. [2]
[Total 5]

5 An insurance company covers pedigree cats against the costs of medical treatment. The cost of claims from a policy in a year is assumed to have a normal distribution with mean μ (which varies from policy to policy) and known variance 25^2. It is assumed that $\mu = \alpha + \beta x$ where α and β are fixed constants and x is the age of the cat. You are given the following data for the pairs (y_i, x_i) for $i = 1, 2, \ldots, 50$ where y_i is the cost of claims last year for the ith policy and x_i is the age of the corresponding cat.

$$\sum_{i=1}^{50} x_i = 637 \quad \sum_{i=1}^{50} y_i = 5,492 \quad \sum_{i=1}^{50} y_i x_i = 74,532 \quad \sum_{i=1}^{50} x_i^2 = 8,312$$

Calculate the maximum likelihood estimates of α and β. [6]

6 Let X_1 and X_2 be two independent exponentially distributed random variables with parameters λ_1 and λ_2 respectively. The random variable X is related to X_1 and X_2 such that a single observation from X is chosen from X_1 with probability p and from X_2 with probability $1 - p$.

(i) Show that the density function of X is

$$pf_1(x) + (1 - p) f_2(x).$$

where $f_i(x)$ is the density function of X_i. [2]

(ii) Construct an algorithm for generating samples from X. [4]

(iii) Describe how the algorithm in (ii) could be generalised for k independent components $p_1 f_1(x) + \ldots + p_k f_k(x)$ where $p_1 + \ldots + p_k = 1$, each $p_i \geq 0$ and $f_i(x)$ is the density of an exponential distribution with parameter λ_i. [2]

[Total 8]

7 A portfolio of insurance policies contains two types of risk. Type I risks make up 80% of claims and give rise to loss amounts which follow a normal distribution with mean 100 and variance 400. Type II risks give rise to loss amounts which are normally distributed with mean 115 and variance 900.

(i) Calculate the mean and variance of the loss amount for a randomly chosen claim. [3]

(ii) Explain whether the loss amount for a randomly chosen claim follows a normal distribution. [2]

The insurance company has in place an excess of loss reinsurance arrangement with retention 130.

(iii) Calculate the probability that a randomly chosen claim from the portfolio results in a payment by the reinsurer. [3]

(iv) Calculate the proportion of claims involving the reinsurer that arise from Type II risks. [2]
[Total 10]

8 Consider the time series

$$Y_t = 0.1 + 0.4Y_{t-1} + 0.9e_{t-1} + e_t$$

where e_t is a white noise process with variance σ^2.

(i) Identify the model as an ARIMA(p,d,q) process. [1]

(ii) Determine whether Y_t is:

(a) a stationary process
(b) an invertible process

[2]

(iii) Calculate $E(Y_t)$ and find the auto-covariance function for Y_t. [6]

(iv) Determine the MA(∞) representation for Y_t. [4]
[Total 13]

9 Claim events on a portfolio of insurance policies follow a Poisson process with parameter λ. Individual claim amounts follow a distribution X with density

$$f(x) = 0.01^2 x e^{-0.01x} \qquad x > 0.$$

The insurance company calculates premiums using a premium loading of 45%.

(i) Derive the moment generating function $M_X(t)$. [3]

(ii) Determine the adjustment coefficient and hence derive an upper bound on the probability of ruin if the insurance company has initial surplus U. [5]

(iii) Find the surplus required to ensure the probability of ruin is less than 1% using the upper bound in (ii). [2]

Suppose instead that individual claims are for a fixed amount of 200.

(iv) Determine whether the adjustment coefficient is higher or lower than in (ii) and comment on your conclusion. [4]
[Total 14]

10 The table below shows cumulative claims paid on a portfolio of motor insurance policies.

	Development Year			
Accident Year	*0*	*1*	*2*	*3*
2007	120	134	146	148
2008	140	180	185	
2009	135	149		
2010	138			

All claims are fully run off by the end of development year 3.

(i) Calculate the total reserve for outstanding claims using the basic chain ladder technique. [7]

An actuarial student suggests an alternative approach to projecting the claims as follows:

- For each of development years 1 to 3 calculate the observed development factor separately for each accident year.
- Then project claims assuming the development factor for a given year is the maximum of the observed development factors for the relevant accident year.
- For example for the development factor from development year 1 to development year 2 we can observe actual factors for accident years 2007 and 2008. To project claims, we assume that the development factor for development year 1 to development year 2 is the maximum of the two observed factors.

(ii) Calculate the increase in the reserve for outstanding claims if claims are projected in this way. [5]

(iii) Discuss why the method in (ii) may not be appropriate. [2]
[Total 14]

11 Five years ago, an insurance company began to issue insurance policies covering medical expenses for dogs. The insurance company classifies dogs into three risk categories: large pedigree (category 1), small pedigree (category 2) and non-pedigree (category 3). The number of claims n_{ij} in the ith category in the jth year is assumed to have a Poisson distribution with unknown parameter θ_i. Data on the number of claims in each category over the last 5 years is set out as follows:

	Year						
	1	*2*	*3*	*4*	*5*	$\sum_{j=1}^{5} n_{ij}$	$\sum_{j=1}^{5} n_{ij}^2$
Category 1	30	43	49	58	60	240	12,144
Category 2	37	49	58	52	64	260	13,934
Category 3	26	31	18	37	32	144	4,354

Prior beliefs about θ_1 are given by a gamma distribution with mean 50 and variance 25.

(i) Find the Bayes estimate of θ_1 under quadratic loss. [5]

(ii) Calculate the expected claims for year 6 of each category under the assumptions of Empirical Bayes Credibility Theory Model 1 [6]

(iii) Explain the main differences between the approach in (i) and that in (ii). [3]

(iv) Explain why the assumption of a Poisson distribution with a constant parameter may not be appropriate and describe how each approach might be generalised. [3]

[Total 17]

END OF PAPER

INSTITUTE AND FACULTY OF ACTUARIES

EXAMINER'S REPORT

September 2011 examinations

Subject CT6 — Statistical Methods
Core Technical

Purpose of Examiners' Reports

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 who are 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. Although Examiners have access to the Core Reading, which is designed to interpret the syllabus, the Examiners are not required to examine the content of Core Reading. Notwithstanding that, the questions set, and the following comments, will generally be based on Core Reading.

For numerical questions the Examiners' preferred approach to the solution is reproduced in this report. Other valid approaches are always given appropriate credit; where there is a commonly used alternative approach, this is also noted in the report. For essay-style questions, and particularly the open-ended questions in the later subjects, this report contains all the points for which the Examiners awarded marks. This is much more than a model solution – it would be impossible to write down all the points in the report in the time allowed for the question.

T J Birse
Chairman of the Board of Examiners

December 2011

General comments on Subject CT6

The examiners for CT6 expect candidates to be familiar with basic statistical concepts from CT3 and so to be able to 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 2011 paper

The difficulty of this paper was in line with where the Examiner's seek to set the typical CT6 paper. Well prepared candidates were able to score well. Amongst the questions candidates struggled most with were Q's 2 and 6 reflecting a consistent theme across a number of sittings of candidates struggling with Bayes' Theorem and simulation techniques. The questions on time series and ruin theory were answered well, continuing a trend of better answers on these topics.

1 (i) We can see that D4 is dominated by D1.
D3 is not dominated since it gives the best results under θ_1.

D2 is not dominated since it gives the best results under $\theta_3\theta_3$..

D1 is not dominated by D2 since D1 is better under $\theta_2\theta_2\theta_3$.. Similarly D1 is not dominated by D3 since D1 is better under $\theta_3\theta_3\theta_3$..

(ii) The maximum losses are:

D1	30
D2	25
D3	24

So the minimax solution is D3.

This was a straightforward question and the majority of candidates scored well.

2 By Bayes theorem

$$\Pr(p = 0.5 \mid X = 7) = \frac{\Pr(X = 7 \mid p = 0.5) \times \Pr(p = 0.5)}{\Pr(X = 7 \mid p = 0.5) \times \Pr(p = 0.5) + \int_{0.5}^{1} f(x) \Pr(X = 7 \mid p = x) dx}$$

And $\int_{0.5}^{1} f(x) \Pr(X = 7 | p = x) dx = \int_{0.5}^{1} 0.4 \times 8 \times x^7 (1 - x) dx$

$$= 3.2 \left[\frac{x^8}{8} - \frac{x^9}{9} \right]_{0.5}^{1}$$

$$= 3.2 \left(\frac{1}{8} - \frac{1}{9} \right) - 3.2 \left(\frac{0.5^8}{8} - \frac{0.5^9}{9} \right)$$

$$\int_{0.5}^{1} f(x) \Pr(X = 7 \Big| p = x) dx = \int_{0.5}^{1} 0.4 \times 8 \times x^7 (1 - x) dx$$

= 0.043576389

And so

$$\Pr(p = 0.5 | X) = \frac{0.8 \times 8 \times 0.5^8}{0.8 \times 8 \times 0.5^8 + 0.043576389} = \frac{0.025}{0.025 + 0.043576389} = 0.364557$$

Many candidates struggled to apply Bayes' theorem, and many of those that did struggled with the mixed prior distribution. Candidates found this one of the harder questions on the paper.

3 (i) We must solve

$$\int_0^M 0.01 e^{-0.01x} dx = 0.8$$

$$[-e^{-0.01x}]_0^M = 0.8$$

$$1 - e^{-0.01M} = 0.8$$

$$M = \frac{\log 0.2}{-0.01} = 160.9437912$$

(ii) We have

$$E(X_I) = \int_0^M 0.01x e^{-0.01x} dx + MP(X > M)$$

$$= [-xe^{-0.01x}]_0^M + \int_0^M e^{-0.01x} dx + Me^{-0.01M}$$

$$= -Me^{-0.01M} + \left[\frac{-e^{-0.01x}}{0.01}\right]_0^M + Me^{-0.01M}$$

$$= -100e^{-0.01M} + 100$$

$$= -100e^{-1.6094} + 100 = 80$$

And hence $E(X_R) = E(X) - E(X_I) = 100 - 80 = 20$

This standard question was generally well answered. Alternatively, one could calculate $E(X_R)$ first and then apply $E(X_I) = E(X) - E(X_R)$.

4 (i) Let $S(t)$ denote the total claims up to time t and suppose individual claim amounts follow a distribution X.

Then $U(t) = U + \lambda t(1 + \theta) E(X) - S(t)$.

(ii) $\psi(U, t) = \Pr(U(s) < 0 \text{ for some } s \in [0, t])$

$\psi(U) = \Pr(U(t) < 0 \text{ for some } t > 0)$

(iii) The probability of ruin by time t will increase as λ increases. This is because claims and premiums arrive at a faster rate, so that if ruin occurs it will occur earlier, which leads to an increase in $\psi(U, t)$.

The probability of ultimate ruin does not depend on how quickly the claims arrive. We are not interested in the time when ruin occurs as we are looking over an infinite time horizon.

This is another standard theory question. Many candidates lost marks by not specifying the probabilities carefully enough in part (ii) – for example $\psi(U) = Pr(U(t) < 0)$ does not fully specify the probability since no information is given about t.

5 The likelihood is given by

$$l = C \times \prod_{i=1}^{50} e^{-\frac{1}{2}\left(\frac{y_i - \mu_i}{25}\right)^2}$$

So the log-likelihood is given by

$$L = \log l = D - \frac{1}{1250}\sum_{i=1}^{50}(y_i - \alpha - \beta x_i)^2$$

$$= D - \frac{1}{1250}\left(\sum_{i=1}^{50} y_i^2 - 2\alpha\sum_{i=1}^{50} y_i + 2\alpha\beta\sum_{i=1}^{50} x_i + 50\alpha^2 - 2\beta\sum_{i=1}^{50} x_i y_i + \beta^2\sum_{i=1}^{50} x_i^2\right)$$

We can ignore the factor of 1,250.

$$\frac{\partial L}{\partial \alpha} = 2\sum_{i=1}^{50} y_i - 2\beta\sum_{i=1}^{50} x_i - 100\alpha = 10,984 - 1,274\beta - 100\alpha$$

$$\frac{\partial L}{\partial \beta} = -2\alpha\sum_{i=1}^{50} x_i + 2\sum_{i=1}^{50} x_i y_i - 2\beta\sum_{i=1}^{50} x_i^2 = 149,064 - 1,274\alpha - 16,624\beta$$

Setting both partial derivatives to zero and solving:

$100\alpha + 1{,}274\beta = 10{,}984$ (AA)
$-1{,}274\alpha - 16{,}624\beta = -149{,}064$ (BB)

(AA) $\times$ 12.74 + (BB) gives $-393.24\beta = -9{,}127.84$ so that $\beta = 23.212$
And so $\alpha = 0.01(10{,}984 - 1274 \times 23.212) = -185.88$

This requires some calculations to produce the mle estimates and only the stronger candidates were able to carry the algebra through to the end. Alternatively, solutions for α and β could also be obtained using the least-squares linear regression expressions given in the tables. This approach gave full credit provided it was accompanied by an explanation of why it produces the same estimates.

6 (i) $F_X(x) = P(X \le x) = P\left(X = X_1 \cap X_1 \le x\right) + P\left(X = X_2 \cap X_2 \le x\right)$
$= pF_1(x) + (1-p)F_2(x)$

and so $f_X(x) = F_X'(x) = pF_1'(x) + (1-p)F_2'(x) = pf_1(x) + (1-p)f_2(x)$

(ii) We need to combine an algorithm for determining whether to sample from X_1 or X_2 with an algorithm for generating a sample from the appropriate exponential distribution.

If u is generated from a $U(0,1)$ distribution then $F_i^{-1}(u)$ is exponentially distributed with mean $1/\lambda_i$. But $F_i(x) = 1 - e^{-\lambda_i x}$ so that $F_i^{-1} = -\frac{\log(1-u)}{\lambda_i}$

So the algorithm is as follows:

(A) Generate u_1 and u_2 from $U(0,1)$

(B) If $u_1 < p$ then set $i = 1$ otherwise set $i = 2$.

(C) Set $x =$ $x = -\frac{\log 1 - u_2}{\lambda_i}$ $-\frac{\log(1-u_2)}{\lambda_i}$

(iii) The algorithm will be as follows:

(A) Generate u_1 and u$_2 u_2$ from $U(0,1)$

(B) Set $q_0 = 0, q_j = p_1 + ... + p_j$ for $j = 1, 2, \ldots, k$

(C) If $q_{j-1} \le u_1 < q_j$ then set $i = j$.

(D) Set $x = -\frac{\log(1-u_2)}{\lambda_i}$ $x = -\frac{\log 1 - u_2}{\lambda_i}$

A number of candidates struggled to generate the correct algorithm. Some attempted to use the inversion method in parts (ii) and (iii) but the method shown above is much easier.

7 (i) Let the loss amount be X. Then

$E(X) = 0.8 \times 100 + 0.2 \times 115 = 103$

$E(X^2) = 0.8 \times (100^2 + 400) + 0.2\,(115^2 + 900) = 11{,}145$

$\text{Var}(X) = E(X^2) - E(X)^2 = 11{,}145 - 103^2 = 536$

(ii) No, the loss distribution is not Normal. To see this, note that (for example) the pdf of the combined distribution will have local maxima at both 100 and 115. [Consider the case where the variances are very small to see this]

(iii) $\Pr(X > 130) = 0.8 \times \Pr(N(100, 20^2) > 130) + 0.2 \times \Pr(N(115,30^2) > 130)$

$= 0.8 \times \Pr\left(N(0,1) > \frac{130-100}{20}\right) + 0.2 \times \Pr\left(N(0,1) > \frac{130-115}{30}\right) \Pr(X > 130) = 0.8 \times \Pr(N(100, 20^2) > 130) + 0.2 \times \Pr(N(115, 30^2) > 130)$

$= 0.8 \times \Pr(N(0,1) > 1.5) + 0.2 \times \Pr(N(0,1) > 0.5)$
$= 0.8 \times (1 - 0.93319) + 0.2 \times (1 - 0.69146)$

$= 0.115156$

(iv) The relevant proportion is given by:

$$\frac{0.2 \times (1 - 0.69146)}{0.115156} = 53.6\%$$

Many weaker candidates struggled with this question, with a large number incorrectly asserting the loss distribution was Normal in part (ii).

8 (i) The model is ARIMA(1,0,1) if Y_t is stationary.

(ii) (a) The characteristic polynomial for the AR part is $A(z) = 1 - 0.4z$ the root of which has absolute value greater than 1 so the process is stationary.

(b) The characteristic polynomial for the MA part is $B(z) = 1 + 0.9z$ the root of which has absolute value greater than 1 so the process is invertible.

(iii) Since the process is stationary we know that $E(Y_t)$ is equal to some constant μ independent of t.

Taking expectations on both sides of the equation defining Y_t gives

$E(Y_t) = 0.1 + 0.4E(Y_{t-1})$

$\mu = 0.1 + 0.4\mu$

$\mu = \dfrac{0.1}{1 - 0.4} = 0.1666666$

Note that

$\mathrm{Cov}(Y_t, e_t) = \mathrm{Cov}(0.1 + 0.4Y_{t-1} + 0.9e_{t-1} + e_t, e_t)$

$= 0.4\mathrm{Cov}(Y_{t-1}, e_t) + 0.9\mathrm{Cov}(e_{t-1}, e_t) + \mathrm{Cov}(e_t, e_t) = 0 + 0 + \sigma^2 = \sigma^2$
Similarly

$\mathrm{Cov}(Y_t, e_{t-1}) = 0 + 0.4\mathrm{Cov}(Y_{t-1}, e_{t-1}) + 0.9\mathrm{Cov}(e_{t-1}, e_{t-1}) + \mathrm{Cov}(e_t, e_{t-1})$
$= 0.4\sigma^2 + 0.9\sigma^2 + 0 = 1.3\sigma^2$

So

$\gamma_0 = \text{Cov}(Y_t, Y_t) = \text{Cov}(Y_t, 0.1 + 0.4Y_{t-1} + 0.9e_{t-1} + e_t)$

$= 0.4\gamma_1 + 0.9 \times 1.3\sigma^2 + \sigma^2 = 0.4\gamma_1 + 2.17\sigma^2$ (A)

And

$\gamma_1 = \text{Cov}(Y_{t-1}, Y_t) = \text{Cov}(Y_{t-1}, 0.1 + 0.4Y_{t-1} + 0.9e_{t-1} + e_t)$

$= 0.4\gamma_0 + 0.9\sigma^2$ (B)

Substituting for γ_1 in (A) gives

$\gamma_0 = 0.4 \times 0.4\gamma_0 + 0.4 \times 0.9\sigma^2 + 2.17\sigma^2 = 0.16\gamma_0 + 2.53\sigma^2$

$$\gamma_0 = \frac{2.53}{0.84}\sigma^2 = 3.011905\sigma^2$$

Substituting into (B) gives

$\gamma_1 = 0.4 \times 3.011905\sigma^2 + 0.9\sigma^2 = 2.104762\sigma^2$

And in general

$\gamma_s = 0.4\gamma_{s-1}$ for $s \geq 2$

So $\gamma_s = 0.4^{s-1} \times 2.104762\sigma^2$.

(iv) We have $(1 - 0.4B)\, Y_t = 0.1 + 0.9e_{t-1} + e_t$

so $Y_t = (1 - 0.4B)^{-1}\,(0.1 + 0.9e_{t-1} + e_t)$

$$= \sum_{i=0}^{\infty} 0.4^i B^i (0.1 + 0.9e_{t-1} + e_t)$$

$$= \frac{0.1}{1-0.4} + 0.9\sum_{i=0}^{\infty} 0.4^i e_{t-i-1} + \sum_{i=0}^{\infty} 0.4^i e_{t-i}$$

$$= 0.16667 + e_t + 1.3\sum_{i=1}^{\infty} 0.4^{i-1} e_{t-i}$$

Overall, this time series question was reasonably well answered, consistent with the improvement in the standard of answers to this type of question in recent sittings. Weaker candidates could not generate the correct auto-covariance function here.

9 (i) $M_X(t) = E(e^{tX}) = \int_0^\infty e^{tx} f(x)dx$

$= \int_0^\infty 0.01^2 xe^{(t-0.01)x} dx$

$= \left[\frac{0.01^2 xe^{(t-0.01)x}}{t-0.01}\right]_0^\infty - \int_0^\infty \frac{0.01^2 e^{(t-0.01)x}}{t-0.01} dx$

$= 0 - 0 - \left[\frac{0.01^2 e^{(t-0.01)x}}{(t-0.01)^2}\right]_0^\infty$ provided that $t < 0.01$

$= \frac{0.01^2}{(t-0.01)^2}$ again provided that $t < 0.01$

(ii) The adjustment coefficient is the unique positive solution of

$M_X(R) = 1 + 1.45E(X)R$

But $E(X) = M'_X(0) = \frac{d}{dt}\left[\frac{0.01^2}{(t-0.01)^2}\right]\Bigg|_{t=0}$

$= \frac{-2 \times 0.01^2}{(t-0.01)^3}\Bigg|_{t=0} = \frac{-2}{-0.01} = 200$

So we need to solve $\frac{0.01^2}{(R-0.01)^2} = 1 + 290R$

i.e. $0.01^2 = (1 + 290R)(R - 0.01)^2 = (1 + 290R)(0.01^2 - 0.02R + R^2)$

i.e. $0.012 = 0.01^2 + 0.029R - 0.02R - 5.8R^2 + R^2 + 290R^3$

i.e. $290R^2 - 4.8R + 0.009 = 0$

$R = \frac{4.8 \pm \sqrt{4.8^2 - 4 \times 290 \times 0.009}}{2 \times 290}$

i.e. $R = 0.00215578$ or $R = 0.0143959$

So taking the smaller root we have $R = 0.00215578$ since that is less than 0.01

The upper bound for the probability of ruin is given by Lundberg's inequality as

$\psi(U) \le e^{-RU} = e^{-0.00215578U}$

(iii) We want $\psi(U) \le e^{-0.00215578U} \le 0.01$

i.e. $-0.00215578U \le \log 0.01$

i.e. $U \ge \dfrac{\log 0.01}{-0.00215578} = 2136.20$

(iv) This time the adjustment coefficient is the solution to:

$e^{200R} = 1 + 290\text{R}$

So the question is whether $y = e^{200R}$ crosses the line $y = 1 + 290R$ before or after $y = 0.01^2(0.01 - R)^{-2}$ crosses the same line
But when $R = 0.00215578$ we have
$e^{200R} = e^{200\times 0.00215578} = 1.539 < 1 + 290R = 1.625.$

So $y = e^{200R}$ has not yet crossed the given line, and the second scenario has a larger adjustment coefficient that the first.

This means the second risk has a lower probability of ruin, which is to be expected since although the mean claim amounts are the same in each scenario, the claim amounts in the first scenario are more variable suggesting a greater risk.

This was found one of the more challenging questions on the paper. In part (i), the final expression could be quoted from the tables but for full marks candidates had to show it from the definitions. Special care is needed here in calculations as decimal places of R can affect the final figures.

10 (i) The development factors are:

$r_{0,1} = \dfrac{134+180+149}{120+140+135} = \dfrac{463}{395} = 1.172151899$

$r_{1,2} = \dfrac{146+185}{134+180} = \dfrac{331}{314} = 1.054140127$

$r_{2,3} = \dfrac{148}{146} = 1.01369863$

The ultimate claims are therefore:

For AY2008: $185 \times 1.01369863 = 187.53$

For AY2009: 149 × 1.05414027 × 1.01369863 = 159.22
For AY2010: 138 × 1.172151899 ×1.054140127 × 1.01369863 = 172.85

So the outstanding claim reserve is

187.53 + 159.22 + 172.85 – 185 – 149 – 138 = 47.60

(ii) The individual development factors are as follows:

	Development Factor		
Accident Year	*0 to 1*	*1 to 2*	*2 to 3*
2007	1.1167	1.0896	1.0137
2008	1.2857	1.0278	
2009	1.1037		
Max	1.2857	1.0896	1.0137

The ultimate claims are therefore:

For AY2008: 185 × 1.0137 = 187.53
For AY2009: 149 × 1.0896 × 1.0137 = 164.57
For AY2010: 138 × 1.2857 × 1.0896 × 1.0137 = 195.97

So the outstanding claim reserve is

187.53 + 164.57 + 195.97 – 185 – 149 – 138 = 76.07

This is an increase of 28.47 which is 59.8% higher.

(iii) Selecting the maximum DF in each column increases the reserves by 60%. Better to take a weighted average of each column as per usual chain ladder approach, UNLESS we know something in particular why we should give full credence to the 1.286 factor (which is much larger than the other two factors in column 2/1) and the 1.09 factor (which is much larger than the 1.028 factor in column 3/2)

This question was well answered. Some candidates dropped marks in part (iii).

11 (i) We need to find the parameters of the Gamma distribution, say α and λ. Then

$$\frac{E(X)}{Var(X)} = \frac{\alpha/\lambda}{\alpha/\lambda^2} = \lambda = \frac{50}{25} = 2$$

And hence $\alpha = E(X) \times \lambda = 50 \times 2 = 100$

The posterior distribution is given by:

$f(\theta_1|x) \propto f(x|\theta_1)\, f(\theta_1)$

$$\propto \left(\prod_{j=1}^{5} e^{-\theta_1}\theta_1^{n_{1j}}\right) \times \theta_1^{\alpha-1} e^{-\lambda\theta_1}$$

$$\propto e^{-(\lambda+5)\theta_1}\theta_1^{\alpha+\sum_{j=1}^{5} n_{1j}-1}$$

Which is the pdf of a gamma distribution with parameters
$\alpha + \sum_{j=1}^{5} n_{1j} = 100 + 240 = 340$ and $\lambda + 5 = 7$.

Under quadratic loss the Bayes estimate is the mean of the posterior distribution. So we have an estimate of $\frac{340}{7} = 48.57$.

(ii) We have $\bar{n}_1 = \frac{240}{5} = 48$ and $\bar{n}_2 = \frac{260}{5} = 52$ and $\bar{n}_3 = \frac{144}{5} = 28.8$.

This gives $\bar{n} = \frac{48+52+28.8}{3} = 42.9333$

$$\sum_{j=1}^{5}(n_{1j}-\bar{n}_1)^2 = \sum_{j=1}^{5} n_{1j} - 2\sum_{j=1}^{5} n_{1j} \times \bar{n}_1 + 5 \times \bar{n}_1^2$$

$$= 12{,}144 - 2 \times 240 \times 48 + 5 \times 48^2 = 624$$

Similarly

$$\sum_{j=1}^{5}(n_{2j}-\bar{n}_2)^2 = 13{,}934 - 2 \times 260 \times 52 + 5 \times 52^2 = 414$$

$$\sum_{j=1}^{5}(n_{3j}-\bar{n}_3)^2 = 4{,}354 - 2 \times 144 \times 28.8 + 5 \times 28.8^2 = 206.8$$

So

$E(s^2(\theta)) = \frac{1}{3}\times\frac{1}{4}(624 + 414 + 206.8) = 103.733$

and

$$\text{Var}(m(\theta)) = \tfrac{1}{2}((48 - 42.9333)^2 + (52 - 42.9333)^2 + (28.8 - 42.9333)^2) - \tfrac{1}{5}\times 103.7333 = 133.06667$$

$$\text{So } Z = \frac{5}{5+\dfrac{103.733}{133.06667}} = 0.86512$$

So expected claims for next year are:

Cat 1 $\quad 0.13488 \times 42.9333 + 0.86512 \times 48 = 47.32$
Cat 2 $\quad 0.13488 \times 42.9333 + 0.86512 \times 52 = 50.78$
Cat 3 $\quad 0.13488 \times 42.9333 + 0.86512 \times 28.8 = 30.71$

This question contained a minor typographical error in the summary statistics. Based on the figures given in the question a direct calculation of $\sum_{j=1}^{5} n_{1j}^2$ *gives the correct figure 12,114 and not 12,144 which is given in the question. Candidates who used 12114 will have produced slightly different results as follows:*

$$\sum_{j=1}^{5}\left(n_{1j} - \overline{n_1}\right)^2 = 594$$

$$E(s^2(\theta)) = 101.2333$$

$$Var\left(S\left(\theta\right)\right) = 133.5666$$

$$Z = 0.86837$$

And the final three figures will change from 47.32, 50.78, 30.71 to 47.33, 50.81 and 30.66 respectively. Candidates producing these figures scored full marks.

(iii) The main differences are that:

- The approach under (i) makes use of prior information about the distribution of θ_1 whereas the approach in (ii) does not.

- The approach under (i) uses only the information from the first category to produce a posterior estimate, whereas the approach under (ii) assumes that information from the other categories can give some information about category 1.

- The approach under (i) makes precise distributional assumptions about the number of claims (i.e. that they are Poisson distributed) whereas the approach under (ii) makes no such assumptions.

(iv) The insurance policies were newly introduced 5 years ago, and it is therefore likely that the volume of policies written has increased (or at least not been constant) over time. The assumption that the number of claims has a Poisson distribution with a fixed mean is therefore unlikely to be accurate, as one would expect the mean number of claims to be proportional to the number of policies.

Let P_{ij} be the number of policies in force for risk i in year j. Then the models can be amended as follows:

The approach in (i) can be taken assuming that that the mean number of claims in the Poisson distribution is $P_{ij}\theta_i$.

The approach in (ii) can be generalised by using EBCT Model 2 which explicitly incorporates an adjustment for the volume of risk.

This long question was answered well generally. A bit of care was needed in the final two parts where only the better candidates were able to give a full discussion of the assumptions underlying the models and how the models could be amended.

END OF EXAMINERS' REPORT

INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

17 April 2012 (am)

Subject CT6 – Statistical Methods
Core Technical

Time allowed: Three hours

INSTRUCTIONS TO THE CANDIDATE

1. *Enter all the candidate and examination details as requested on the front of your answer booklet.*

2. *You must not start writing your answers in the booklet until instructed to do so by the supervisor.*

3. *Mark allocations are shown in brackets.*

4. *Attempt all 11 questions, beginning your answer to each question on a separate sheet.*

5. *Candidates should show calculations where this is appropriate.*

Graph paper is NOT required for this paper.

AT THE END OF THE EXAMINATION

Hand in BOTH your answer booklet, with any additional sheets firmly attached, and this question paper.

In addition to this paper you should have available the 2002 edition of the Formulae and Tables and your own electronic calculator from the approved list.

1 (i) Define what it means for a random variable to belong to an exponential family. [1]

(ii) Show that if a random variable has the exponential distribution it belongs to an exponential family. [3]
[Total 4]

2 A statistician is told that one of two dice has been chosen and rolled, and he is told the result of the roll. One dice is a conventional dice, but the other has three sides numbered 2 and three sides numbered 4. If the statistician correctly identifies the dice he wins a prize of 1.

(i) Determine the total number of decision functions available to the statistician. [2]

(ii) (a) Identify the most natural candidate for the decision function.

(b) Calculate the expected payoff for this function assuming that each of the two dice are equally likely to be chosen. [3]
[Total 5]

3 Claim amounts on a certain type of insurance policy follow a distribution with density

$$f(x) = 3cx^2 e^{-cx^3} \text{ for } x > 0$$

where c is an unknown positive constant. The insurer has in place individual excess of loss reinsurance with an excess of 50. The following ten payments are made by the insurer:

Losses below the retention: 23, 37, 41, 11, 19, 33
Losses above the retention: 50, 50, 50, 50

Calculate the maximum likelihood estimate of c. [6]

4 Claims on a particular type of insurance policy follow a compound Poisson process with annual claim rate per policy 0.2. Individual claim amounts are exponentially distributed with mean 100. In addition, for a given claim there is a probability of 30% that an extra claim handling expense of 30 is incurred (independently of the claim size). The insurer charges an annual premium of 35 per policy.

Use a normal approximation to estimate how many policies the insurer must sell so that the insurer has a 95% probability of making a profit on the portfolio in the year. [6]

5 The total claim amount per annum on a particular insurance policy follows a normal distribution with unknown mean θ and variance 200^2. Prior beliefs about θ are described by a normal distribution with mean 600 and variance 50^2. Claim amounts $x_1, x_2, \ldots, x_n$ are observed over n years.

(i) State the posterior distribution of θ. [2]

(ii) Show that the mean of the posterior distribution of θ can be written in the form of a credibility estimate. [3]

Now suppose that n=5 and that total claims over the five years were 3,400.

(iii) Calculate the posterior probability that θ is greater than 600. [2]

[Total 7]

6 A proportion p of packets of a rather dull breakfast cereal contain an exciting toy (independently from packet to packet). An actuary has been persuaded by his children to begin buying packets of this cereal. His prior beliefs about p before opening any packets are given by a uniform distribution on the interval [0,1]. It turns out the first toy is found in the $n_1 th$ packet of cereal.

(i) Specify the posterior distribution of p after the first toy is found. [3]

A further toy was found after opening another n_2 packets, another toy after opening another n_3 packets and so on until the fifth toy was found after opening a grand total of $n_1 + n_2 + n_3 + n_4 + n_5$ packets.

(ii) Specify the posterior distribution of p after the fifth toy is found. [2]

(iii) Show the Bayes' estimate of p under quadratic loss is not the same as the maximum likelihood estimate and comment on this result. [5]

[Total 10]

7 The numbers of claims on three different classes of insurance policies over the last four years are given in the table below.

	Year 1	*Year 2*	*Year 3*	*Year 4*	*Total*
Class 1	1	4	5	0	10
Class 2	1	6	4	6	17
Class 3	5	6	4	9	24

The number of claims in a given year from a particular class is assumed to follow a Poisson distribution.

(i) Determine the maximum likelihood estimate of the Poisson parameter for each class of policy based on the data above. [5]

(ii) Perform a test on the scaled deviance to check whether there is evidence that the classes of policy have different mean claim rates and state your conclusion. [5]

[Total 10]

8 The table below shows claims paid on a portfolio of general insurance policies. You may assume that claims are fully run off after three years.

Underwriting year	*Development Year*			
	0	*1*	*2*	*3*
2008	450	312	117	41
2009	503	389	162	
2010	611	438		
2011	555			

Past claims inflation has been 5% p.a. However, it is expected that future claims inflation will be 10% p.a.

Use the inflation adjusted chain ladder method to calculate the outstanding claims on the portfolio. [10]

9 Consider the time series model

$$(1-\alpha B)^3 X_t = e_t$$

where B is the backwards shift operator and e_t is a white noise process with variance σ^2.

(i) Determine for which values of α the process is stationary. [2]

Now assume that $\alpha = 0.4$.

(ii) (a) Write down the Yule-Walker equations.

(b) Calculate the first two values of the auto-correlation function ρ_1 and ρ_2. [7]

(iii) Describe the behaviour of ρ_k and the partial autocorrelation function ϕ_k as $k \to \infty$. [3]

[Total 12]

10 Let X_1 and X_2 be random variables with moment generating functions $M_{X_1}(t)$ and $M_{X_2}(t)$ respectively. A new random variable Y is formed by choosing a sample from X_1 with probability p or a sample from X_2 with probability $1-p$.

(i) Show that the moment generating function of Y is given by

$$M_Y(t) = pM_{X_1}(t) + (1-p)M_{X_2}(t)$$ [2]

A portfolio of insurance policies consists of two different types of policy. Claims on type 1 policies arrive according to a Poisson process with parameter λ_1 and claim amounts have a distribution X_1. Claims on type 2 policies arrive according to a Poisson process with parameter λ_2 and claim amounts have a distribution X_2.

(ii) Show that aggregate claims on the whole portfolio follow a compound Poisson distribution, specifying the claim rate and the claim size distribution. [6]

Now suppose that $\lambda_1 = 10$ and $\lambda_2 = 15$ and that the claim sizes are exponentially distributed with mean 50 for type 1 policies and mean 70 for type 2 policies.

(iii) Construct an algorithm for simulating total claims on the whole portfolio. [6]

[Total 14]

11 Claims on a portfolio of insurance policies arrive as a Poisson process with parameter 100. Individual claim amounts follow a normal distribution with mean 30 and variance 5^2. The insurer calculates premiums using a premium loading of 20% and has initial surplus of 100.

(i) Define carefully the ruin probabilities $\psi(100)$, $\psi(100,1)$ and $\psi_1(100,1)$. [3]

(ii) Define the adjustment coefficient *R*. [1]

(iii) Show that for this portfolio the value of *R* is 0.011 correct to 3 decimal places. [5]

(iv) Calculate an upper bound for $\psi\left(100\right)$ and an estimate of $\psi_1(100,1)$. [5]

(v) Comment on the results in (iv). [2]

[Total 16]

END OF PAPER

INSTITUTE AND FACULTY OF ACTUARIES

EXAMINERS' REPORT

April 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 who are 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. Although Examiners have access to the Core Reading, which is designed to interpret the syllabus, the Examiners are not required to examine the content of Core Reading. Notwithstanding that, the questions set, and the following comments, will generally be based on Core Reading.

For numerical questions the Examiners' preferred approach to the solution is reproduced in this report. Other valid approaches are always given appropriate credit; where there is a commonly used alternative approach, this is also noted in the report. For essay-style questions, and particularly the open-ended questions in the later subjects, this report contains all the points for which the Examiners awarded marks. This is much more than a model solution – it would be impossible to write down all the points in the report in the time allowed for the question.

T J Birse
Chairman of the Board of Examiners

July 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 April 2012 paper

Candidates found this paper to be slightly harder than the typical CT6 paper. Nevertheless, well prepared candidates were able to score well. Once again, the question on simulation techniques was poorly answered (or not attempted in many cases). Both this question and the decision theory question required no difficult mathematics, but did require a good understanding of the underlying ideas. Many candidates also struggled on questions using material from CT3.

The questions on ruin theory and time series were again well answered.

1 (i) A random variable Y belongs to an exponential family if the pdf of Y can be written in the form

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

Where a, b and c are functions.

(ii) Suppose that the parameter of the exponential distribution Y is λ. Then

$$f\left(y\right)=\lambda\exp\left(-\lambda y\right)$$

$$=\exp\left[\log\lambda-\lambda y\right]$$

$$=\exp\left[\frac{\lambda y-\log\lambda}{-1}\right]$$

Which is of the required form with

$$\theta=\lambda$$

$$a\left(\varphi\right)=-1$$

$$b(\theta)=\log\theta$$

$$c\left(y,\varphi\right)=0$$

Alternative solution: $\theta = -\lambda$; $a(\varphi) = 1$; $b(\theta) = -\log(-\theta)$; $c(y,\varphi) = 0$

This question was answered well.

2 (i) The decision function must nominate a choice of die for each potential outcome from the observation.

There are 6 possible outcomes from the die roll and hence 2 × 2 × 2 × 2 × 2 × 2 = 64 possible decision functions.

(ii) The most natural candidate is to nominate the conventional die on rolls of 1,3,5,6 and the special die on rolls of 2 or 4.

The expected payoff from this approach is:

$$0.5 \times \left(\frac{4}{6} \times 1 + \frac{2}{6} \times 0 \right) + 0.5 \times 1 = 0.83333$$

Candidates who understood what a decision function is scored well. However, many candidates struggled to make any headway with this question. For part (i) some candidates observed that if the dice roll is 1,3,5 or 6 it is obvious that you must chose the conventional dice. Therefore a choice is only needed on a roll of a 2 or a 4 giving a total of 2 × 2 = 4 functions. This was given full credit if carefully explained.

3 The likelihood function is given by:

$$L = D \times \prod_{i=1}^{6} 3cx_i^{\,2} e^{-cx_i^3} \times e^{-4\times 50^3 c}$$

where D is a constant.

Where the x_i are the claims below the retention.

$$l = \log L = \log D + \sum_{i=1}^{6} log 3cx_i^{\,2} - c\sum_{i=1}^{6} x_i^3 - 4\times 50^3 c$$

$$= \log D + 6\log 3 + 6\log c + \sum_{i=1}^{6} \log x_i^{\,2} - c\sum_{i=1}^{6} x_i^3 - 4\times 50^3 c$$

Differentiating we get

$$\frac{dl}{dc} = \frac{6}{c} - \sum_{i=1}^{6} x_i^3 - 500000$$

So our estimate is given by

$$\hat{c} = \frac{6}{\sum_{i=1}^{6} x_1^{\;3} + 500000} = \frac{6}{175868 + 500000} = 8.8775 \times 10^{-6}$$

This question was answered well.

4 Let the individual total claim costs be denoted by X. Then $X=Y+Z$ where Y is the cost of the claim and Z is the claim handling expense.

Then

$$E(X) = E(Y) + E(Z) = 100 + 0.3 \times 30 = 109$$

And

$$E\left(X^2\right) = E\left(Y^2 + 2YZ + Z^2\right) = E\left(Y^2\right) + 2E(Y)E(Z) + E(Z^2)$$

Using the independence of Y and Z. Now

$$E\left(Y^2\right) = 2E(Y)^2 = 2 \times 100^2 = 20000$$

and

$$E\left(Z^2\right) = 0.3 \times 30^2 = 270$$

So that

$$E\left(X^2\right) = 20000 + 2 \times 100 \times 9 + 270 = 22070 = 148.56^2$$

Now if there are n policies in the portfolio, total claim amounts S will have an approximately Normal distribution with mean $0.2 \times n \times 109 = 21.8n$ and variance $0.2 \times n \times 148.56^2$.

The premium income will be $35n$.

We need to solve for n in the following equation:

$$P\left(N\left(21.8n, 66.44^2 n\right) > 35n\right) < 0.05$$

i.e. $$P\left(N(0,1) > \frac{13.2n}{66.44\sqrt{n}}\right) < 0.05$$

So

$$0.198675496\sqrt{n} > 1.6449$$
$$n > 68.55$$

i.e. at least 69 policies must be sold.

Most candidates struggled with this question. Many did not calculate the variance correctly and a lot did not correctly use the number of policies, n, as a multiplier for the mean and variance of the claims. Others used n and the claim rate when calculating the additional claim handling expense.

5 (i) The posterior distribution of θ is Normal with variance given by

$$\sigma_*^2 = \frac{1}{\left(\frac{n}{200^2} + \frac{1}{50^2}\right)}$$

And mean given by

$$\mu_* = \sigma_*^2 \left(\frac{n\bar{x}}{200^2} + \frac{600}{50^2}\right)$$

(ii) Set

$$Z = \sigma_*^2 \frac{n}{200^2}$$

Then

$$Z = \frac{\frac{n}{200^2}}{\left(\frac{n}{200^2} + \frac{1}{50^2}\right)} = \frac{n}{(n+16)}$$

And

$$1 - Z = \frac{\frac{1}{50^2}}{\left(\frac{n}{200^2} + \frac{1}{50^2}\right)} = \sigma_*^2 \frac{1}{50^2}$$

And so

$$\mu_* = Z\bar{x} + (1 - Z)600$$

Which is in the form of a credibility estimate with 600 being the prior mean, $\overline{x}$ being the observed sample mean and Z being the credibility factor.

(iii) In this case we have

$$\sigma_*^2 = \frac{1}{\left(\frac{n}{200^2} + \frac{1}{50^2}\right)} = \frac{1}{\left(\frac{5}{200^2} + \frac{1}{50^2}\right)} = 43.64^2$$

and

$$\mu_* = \sigma_*^2 \left(\frac{n\overline{x}}{200^2} + \frac{600}{50^2}\right) = 43.64^2 \left(\frac{3400}{200^2} + \frac{600}{50^2}\right) = 619.0476$$

So

$$P(\theta > 600) = P\left(N\left(619.0476, 43.64^2\right) > 600\right)$$

$$= P\left(N(0,1) > \frac{600 - 619.0476}{43.64}\right) = P(N(0,1) > -0.436)$$

$$= 0.6 \times 0.67003 + 0.4 \times 0.66640$$

$$= 0.669$$

This question was well answered. Some candidates attempted to derive the answer to part (i) from first principles which was not required. Parts (ii) and (iii) were generally answered well.

6 (i) The posterior distribution has a likelihood given by

$$f(p|n_1) \propto f\left(n_1|p\right) f(p)$$

$$\propto (1-p)^{n_1-1}\, p \times 1$$

Which is the pdf of a Beta distribution with parameters α = 2 and $\beta = n_1$.

(ii) Now the posterior distribution has likelihood given by

$$f(p|n_1, n_2, \ldots, n_5) \propto f\left(n_1, n_2, \ldots, n_5|p\right) f(p)$$

$$\propto (1-p)^{n_1-1}\, p \times (1-p)^{n_2-1}\, p \times \cdots \times (1-p)^{n_5-1} \times p$$

$$\propto (1-p)^{n_1+n_2+\cdots+n_5-5} \times p^5$$

Which is the pdf of a Beta distribution with parameters α = 6 and $\beta = n_1 + n_2 + \cdots + n_5 - 4$.

(iii) Under squared error loss the Bayes estimate is given by the mean of the posterior distribution which in this case is

$$\hat{p} = \frac{\alpha}{\alpha+\beta} = \frac{6}{n_1 + n_2 + \cdots + n_5 + 2}$$

The maximum likelihood estimate is given by maximising the likelihood which is

$$L \propto (1-p)^{n_1+n_2+\cdots+n_5-5} \times p^5$$

The log-likelihood is given by

$$l = \log L = \log C + \left(n_1 + \cdots + n_5 - 5\right)\log(1-p) + 5\log p$$

And so $\dfrac{dl}{dp} = -\left(n_1 + \cdots + n_5 - 5\right) \times \dfrac{1}{1-p} + \dfrac{5}{p}$

And setting this expression to zero gives

$$\left(n_1 + \cdots + n_5 - 5\right)\hat{p} = 5(1-\hat{p})$$

And so $\left(n_1 + \cdots + n_5\right)\hat{p} = 5$

i.e. $\hat{p} = \dfrac{5}{n_1 + \cdots + n_5}$

So the two estimates are not the same. This is perhaps a little surprising given that we started with an uninformative prior, but arises because the estimates are calculated in two different ways – i.e. one maximises the likelihood and the other minimises the expected squared error. If we wanted the two to be the same we should use an "all-or-nothing" loss function.

A reasonably well answered question. Weaker candidates failed to identify the geometric distribution in part (i). Stronger candidates demonstrated a good understanding of loss functions in part (iii).

7 (i) Suppose that the Poisson rate for risk i is λ_i for =1,2,3.

For the first risk, the likelihood is given by:

$$L = e^{-4\lambda_1} \frac{(4\lambda_1)^{10}}{10!}$$

And so the log-likelihood is given by

$$l = \log L = -4\lambda_1 + 10\log 4\lambda_1 + \text{Constants}$$

Differentiating gives

$$\frac{dl}{d\lambda_1} = -4 + \frac{10}{\lambda_1}$$

And setting this equal to zero gives a maximum likelihood estimate of

$$\hat{\lambda}_1 = \frac{10}{4} = 2.5$$

Since $\dfrac{d^2 l}{d\lambda^2} = -\dfrac{10}{\lambda_i^{\,2}} < 0$ we do have a maximum.

In the same way $\hat{\lambda}_2 = \dfrac{17}{4} = 4.25$ and $\hat{\lambda}_3 = \dfrac{24}{4} = 6$.

(ii) Under the assumption that these risks share the same rate i.e. $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$ then the mle for this is simply

$$\hat{\lambda} = \frac{51}{12} = 4.25$$

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

The difference in the scaled deviance here should have a chi-square distribution with 3–1=2 degrees of freedom.

$$2(\log L_1 + \log L_2 + \log L_3 - \log L) = 10\log\hat{\lambda}_1 - 4\hat{\lambda}_1 + 17\log\hat{\lambda}_2 - 4\hat{\lambda}_2 + 24\log\hat{\lambda}_3 - 4\hat{\lambda}_3 - 51\log\hat{\lambda} + 12\hat{\lambda}$$

With the $\sum_{i=1}^{4}\log y_{1i}! + \sum_{i=1}^{4}\log y_{2i}! + \sum_{i=1}^{4}\log y_{3i}!$ cancelling out in the difference.

Hence

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

$$= 2\left(10\log 2.5 + 17\log 4.25 + 24\log 6 - 51\log\frac{51}{12} - \frac{4(10+17+24)}{4} + 12\frac{51}{12}\right)$$

$$= 2\left(10\log 2.5 + 17\log 4.25 + 24\log 6 - 51\log\frac{51}{12}\right) = 5.939778$$

This value is below 5.991 which is the critical value at the upper 5% level and therefore there is not a significant improvement by considering different rates for each risk.

Part (i) was answered very well. Most candidates struggled with part (ii).

8 The claims uplifted to 2011 prices are as follows:

Underwriting Year	*Development Year*			
	0	*1*	*2*	*3*
2008	520.93	343.98	122.85	41
2009	554.56	408.45	162	
2010	641.55	438		
2011	555			

Accumulating gives:

Underwriting Year	*Development Year*			
	0	*1*	*2*	*3*
2008	520.93	864.91	987.76	1028.76
2009	554.56	963.01	1125.01	
2010	641.55	1079.55		
2011	555			

Hence the development factors are given by:

$$DF_{0,1} = \frac{864.91 + 963.01 + 1079.55}{520.93 + 554.56 + 641.55} = 1.693304$$

$$DF_{1,2} = \frac{987.76 + 1125.01}{864.91 + 963.01} = 1.155833$$

$$DF_{2,3} = \frac{1028.76}{987.76} = 1.041508$$

The completed triangle of cumulative claims is:

Underwriting year	*Development Year*			
	0	*1*	*2*	*3*
2008	520.93	864.91	987.76	1028.76
2009	554.56	963.01	1125.01	1171.70
2010	641.55	1079.55	1247.78	1299.57
2011	555.00	939.78	1086.23	1131.32

Dis-accumlating gives (in 2011 prices):

Underwriting year	*Development Year*			
	0	*1*	*2*	*3*
2008				
2009				46.70
2010			168.23	51.79
2011		384.78	146.45	45.09

Inflating for future claims growth gives:

Underwriting year	*Development Year*			
	0	*1*	*2*	*3*
2008				
2009				51.37
2010			185.05	62.67
2011		423.26	177.20	60.01

And the outstanding claims are:

51.37+62.67+60.01+185.05+177.20+423.26 = 959.56

This question was tackled very well by most candidates

9 (i) The characteristic polynomial is $(1-\alpha Y)^3 = 0$.

This has a triple root at $1/\alpha$ and so the process is stationary when $\left|\frac{1}{\alpha}\right| > 1$ i.e. $|\alpha| < 1$.

(ii) Expanding the cubic equation and rearranging gives:

$$X_t - 3\alpha X_{t-1} + 3\alpha^2 X_{t-2} - \alpha^3 X_{t-3} = e_t$$

So the Yule-Walker equations give:

$$\rho_0 - 3\alpha\rho_1 + 3\alpha^2 \rho_2 - \alpha^3\rho_3 = \sigma^2$$
$$\rho_1 - 3\alpha + 3\alpha^2\rho_1 - \alpha^3\rho_2 = 0 \quad \text{(A)}$$
$$\rho_2 - 3\alpha\rho_1 + 3\alpha^2 - \alpha^3\rho_1 = 0 \quad \text{(B)}$$
$$\rho_3 - 3\alpha\rho_2 + 3\alpha^2 \rho_1 - \alpha^3\rho_0 = 0$$

So re-writing we have from (A) $\rho_1\left(1+3\alpha^2\right) - 3\alpha = \alpha^3\rho_2$

And substituting into (B) gives

$$\frac{\rho_1\left(1+3\alpha^2\right)-3\alpha}{\alpha^3} - 3\alpha\rho_1 + 3\alpha^2 - \alpha^3\rho_1 = 0$$

i.e. $$\frac{\rho_1(1+3\alpha^2-3\alpha^4-\alpha^6)}{\alpha^3} = \frac{3\alpha-3\alpha^5}{\alpha^3}$$

i.e. $\rho_1 = \dfrac{3\alpha(1-\alpha^4)}{(1+3\alpha^2-3\alpha^4-\alpha^6)} = 0.83573487$

And so

$$\rho_2 = \frac{3\alpha\left(1-\alpha^4\right)(1+3\alpha^2)}{(1+3\alpha^2-3\alpha^4-\alpha^6)\alpha^3} - \frac{3}{\alpha^2} = 0.576368876$$

Alternative solution:

Express the Yule-Walker equations in terms of the covariances:

$$X_t = 1.2X_{t-1} - 0.48X_{t-2} + 0.064X_{t-3} + e_t$$

$$\begin{aligned}
\gamma_0 &= 1.2\gamma_1 - 0.48\gamma_2 + 0.064\gamma_3 + \sigma^2 \\
\gamma_1 &= 1.2\gamma_0 - 0.48\gamma_1 + 0.064\gamma_2 \\
\gamma_2 &= 1.2\gamma_1 - 0.48\gamma_0 + 0.064\gamma_1 \\
\gamma_3 &= 1.2\gamma_2 - 0.48\gamma_1 + 0.064\gamma_0
\end{aligned}$$

Or in general:

$$\begin{aligned}
\gamma_0 &= 1.2\gamma_1 - 0.48\gamma_2 + 0.064\gamma_3 + \sigma^2 \\
\gamma_k &= 1.2\gamma_{k-1} - 0.48\gamma_{k-2} + 0.064\gamma_{k-3} \quad k \geq 1
\end{aligned}$$

Simplifying the second and third equations:

$$\begin{aligned}
148\gamma_1 &= 1.2\gamma_0 + 0.064\gamma_2 \Rightarrow \gamma_1 = \tfrac{30}{37}\gamma_0 + \tfrac{8}{185}\gamma_2 \\
\gamma_2 &= 1.264\gamma_1 + 0.064\gamma_1
\end{aligned}$$

To obtain:

$$\gamma_2 = \tfrac{200}{347}\gamma_0 \qquad \lambda_1 = \tfrac{290}{347}\gamma_0$$

Dividing both by γ_0 gives the same solutions as above.

(iii) The series is an AR(3) series. The asymptotic behaviour is therefore that ρ_k decays exponentially to zero
whilst ϕ_k is zero for k>3.

The latter parts of this question were not particularly well answered. Candidates generally showed an understanding of how to solve the problem, but made a number of arithmetic and algebraic slips.

10 (i) $M_Y(t) = E\left(e^{tY}\right) = pE\left(e^{tX_1}\right) + (1-p)E\left(e^{tX_2}\right)$

$= pM_{X_1}(t) + (1-p)M_{X_2}(t)$

(ii) Let S_1, S_2 denote aggregate claims on the type 1 and type 2 policies respectively, and let N_1, N_2 denote the number of claims from type 1 and type 2 policies respectively. Let $S = S_1 + S_2$ denote the aggregate claims on the combined portfolio. We know that S_1, S_2 follow compound Poisson processes and so

$$M_{S_i}(t) = M_{N_i}(\log M_{X_i}(t)) = \exp(\lambda_i (M_{X_i}(t) - 1))$$

Now

$$M_S(t) = M_{S_1+S_2}(t) = M_{S_1}(t) M_{S_2}(t)$$

$$= \exp\left(\lambda_1 \left(M_{X_1}(t) - 1\right)\right) \exp(\lambda_2 (M_{X_2}(t) - 1))$$

$$= \exp\left[(\lambda_1 + \lambda_2)\left(\frac{\lambda_1}{\lambda_1 + \lambda_2} M_{X_1(t)} + \frac{\lambda_2}{\lambda_1 + \lambda_2} M_{X_2(t)} - 1\right)\right]$$

$$= \exp((\lambda_1 + \lambda_2)(pM_{X_1}(t) + (1-p)M_{X_2}(t) - 1) \text{ where } p = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

$$= \exp((\lambda_1 + \lambda_2)(M_Y(t) - 1))$$

Where *Y* is defined as in part (i). This is of the form $M_N(logM_Y(t))$ where *N* is a Poisson distribution with parameter $\lambda_1 + \lambda_2$. Hence S has a compound Poisson distribution with rate $\lambda_1 + \lambda_2$ and where individual claim amounts are taken from distribution X_1 with probability $p = \dfrac{\lambda_1}{\lambda_1 + \lambda_2}$ and from distribution X_2 with probability $1 - p = \dfrac{\lambda_2}{\lambda_1 + \lambda_2}$.

(iii) **Step 1**

We first begin by generating a random sample from $N \sim P(25)$ as follows:

Let u be a random sample from a Uniform distribution on (0,1).

Find the positive integer i such that $P(N \le i-1) < u \le P(N \le i)$ (using the cumulative Poisson tables)

Then i is the simulated number of claims.

Step 2

Now we simulate the individual claim amount

Generate v a sample from a Uniform distribution on (0,1).

If $v \le \dfrac{10}{10+15} = \dfrac{10}{25} = 0.4$ then we have a type 1 claim otherwise we have a type 2 claim. Let the claim type be j.

Put $\mu_1 = 50$ and $\mu_2 = 70$. Generate w a sample from a uniform distribution on (0,1).

The simulated claim Z is given by setting

$$F_{X_j}(Z) = w$$

So $1 - \exp\left(\dfrac{1}{\mu_j} Z\right) = w$

So $Z = -\mu_j \ln(1-w)$

Step 3

Repeat Step 2 for a total of i samples and add the results.

Alternative algorithm: simulate the two results separately and add together at the end.

This question was not answered well. In particular, many candidates did not attempt part (iii). Of those that did, most had a good attempt at step 2, but very few got step 1 (to deduce the simulated number of claims).

11 (i) Let $S(t)$ denote cumulative claims to time *t*. Let the annual rate of premium income be *c* and let the insurer's initial surplus be *U*=100.

Then the surplus at time *t* is given by:

$$U(t) = U + ct - S(t)$$

And the relevant probabilities are defined by:

$$\psi(100) = P(U(t) < 0 \text{ for some } t > 0)$$
$$\psi(100,1) = P(U(t) < 0 \text{ for some } t \text{ with } 0 < t \le 1)$$
$$\psi_1(100,1) = P(U(1) < 0)$$

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

$$\lambda M_X(R) = \lambda + cR$$

Where λ is the rate of the Poisson process (i.e. 100) and *X* is the normal distribution with mean 30 and standard deviation 5.

(iii) In this case we have:

$$c = 100 \times 30 \times 1.2 = 3600$$

And

$$M_X(R) = \exp(30R + 12.5R^2)$$

So *R* is the root of

$$100\exp\left(30R + 12.5R^2\right) - 100 - 3600R = 0$$

Denote the left hand side of this equation by *f*(*R*).

When *R* = 0.0115 we have

$$f(0.0115) = 100\exp(0.346653125) - 100 - 41.4 = 0.032604592 > 0$$

And when *R* = 0.0105 we have

$$f(0.0105) = 100\exp(0.316378125) - 100 - 37.8 = -0.585099862 < 0$$

Since the function changes sign between 0.0105 and 0.0115 the unique positive root must lie between these values and hence the root is 0.011 correct to 3 decimal places.

(iv) By Lundberg's inequality $\psi(100) < \exp(-100 \times 0.011) = 0.33287$

Claims in the first year are approximately Normal, with mean $100 \times 30 = 3000$

And variance given by $100 \times \left(25 + 30^2\right) = 92500$

So approximately

$$\psi_1(100,1) = P(100 + 3600 - N(3000, 92500) < 0)$$

$$= P\left(N(3000, 92500) > 3700\right) = P\left(N(0,1) > \frac{3700 - 3000}{\sqrt{92500}}\right)$$

$$= P(N(0,1) > 2.302)$$

$$= 1 - (0.98928 \times 0.8 + 0.98956 \times 0.2)$$

$$= 0.0107.$$

(v) The probability of ruin is much smaller in the first year than the long-term bound provided by Lundberg's inequality. This suggests that either the bound in Lundberg's inequality may not be that tight or that there is significant probability of ruin at times greater than 1 year.

In part (i) many candidates lost straightforward marks by failing to give sufficiently precise definitions. In particular, many candidates gave solutions along the lines of P(U(t)<0, t>0). It isn't clear whether this refers to all positive values of t or just some positive value.

Most candidates got part (ii).

For part (iii), many candidates were able to show that when R=0.011 the two sides of the equation are approximately equal. Very few were able to give a precise demonstration that the root is at R=0.011 by considering where the curve cross the axis. Candidates for future exams should note this technique carefully.

For part (iv) most candidates got the upper bound for R.

Part (v) was well answered by stronger candidates.

END OF EXAMINERS' REPORT

INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

25 September 2012 (am)

Subject CT6 – Statistical Methods
Core Technical

Time allowed: Three hours

INSTRUCTIONS TO THE CANDIDATE

1. *Enter all the candidate and examination details as requested on the front of your answer booklet.*

2. *You must not start writing your answers in the booklet until instructed to do so by the supervisor.*

3. *Mark allocations are shown in brackets.*

4. *Attempt all 10 questions, beginning your answer to each question on a separate sheet.*

5. *Candidates should show calculations where this is appropriate.*

Graph paper is NOT required for this paper.

AT THE END OF THE EXAMINATION

Hand in BOTH your answer booklet, with any additional sheets firmly attached, and this question paper.

In addition to this paper you should have available the 2002 edition of the Formulae and Tables and your own electronic calculator from the approved list.

1 The potential losses from a decision problem are given in the table below

	D1	*D2*	*D3*	*D4*
θ_1	5	8	12	3
θ_2	10	15	7	8
θ_3	7	12	16	9
θ_4	17	4	10	12

(i) Find the optimal decision using the minimax criteria. [2]

Now suppose that $p(\theta_1) = p(\theta_2) = p(\theta_3) = 0.3$ and $p(\theta_4) = 0.1$.

(ii) Find the optimal decision using the Bayes criteria. [2]
[Total 4]

2 Claim amounts on a certain type of insurance policy depend on a parameter α which varies from policy to policy. The mean and variance of the claim amount X given α are specified by

$$E\left[X \middle| \alpha\right] = 200 + \alpha$$
$$V[X|\alpha] = 10 + 2\alpha$$

The parameter α follows a normal distribution with mean 20 and variance 4.

Find the unconditional mean and variance of X. [6]

3 An actuary needs to generate samples from the standard normal distribution for use in a simulation model he is constructing.

(i) Describe the polar algorithm for generating pairs of samples from the standard normal distribution given pairs of samples from a uniform distribution on [0,1]. [3]

(ii) Calculate the probability that a pair of samples from a uniform distribution on [0,1] results in an acceptable pair of samples from the standard normal distribution under the algorithm in (i). [3]
[Total 6]

4 Claims arising on a particular type of insurance policy are believed to follow a Pareto distribution. Data for the last several years shows the mean claim size is 170 and the standard deviation is 400.

(i) Fit a Pareto distribution to this data using the method of moments. [4]

(ii) Calculate the median claim using the fitted parameters and comment on the result. [3]

[Total 7]

5 A discrete probability distribution is defined by

$$f\left(y,\mu\right)=\binom{n}{ny}\mu^{ny}\left(1-\mu\right)^{n-ny} \qquad y=0,\frac{1}{n},\frac{2}{n},\ldots,1$$

where μ is a parameter between 0 and 1.

(i) Explain why this distribution belongs to an exponential family. [4]

(ii) State the three main components that need to be taken into account when constructing a generalised linear model. [3]

(iii) Suggest a natural choice of link function if the response variable followed the distribution defined above. [1]

(iv) Suggest a natural choice of link function if instead the response variable followed a lognormal distribution. [2]

[Total 10]

6 Individual claim amounts from a particular type of insurance policy follow a normal distribution with mean 150 and standard deviation 30. Claim numbers on an individual policy follow a Poisson distribution with parameter 0.25. The insurance company uses a premium loading of 70% to calculate premiums.

(i) Calculate the annual premium charged by the insurance company. [1]

The insurance company has an individual excess of loss reinsurance arrangement with a retention of 200 with a reinsurer who uses a premium loading of 120%.

(ii) Calculate the probability that an individual claim does not exceed the retention. [2]

(iii) Calculate the probability for a particular policy that in a given year there are no claims which exceed the retention. [2]

(iv) Calculate the premium charged by the reinsurer. [4]

(v) Calculate the insurance company's expected profit. [2]

[Total 11]

7 The table below shows claims paid on a portfolio of general insurance policies. Claims from this portfolio are fully run off after 3 years.

Underwriting year	*Development Year*			
	0	*1*	*2*	*3*
2008	85	42	30	7
2009	103	65	25	
2010	93	47		
2011	111			

(i) Estimate the outstanding claims using the basic chain ladder approach. [7]

You are asked to investigate the fit of the model by applying the development factors from part (i) to the claims paid in development year 0 and then comparing the fitted claim payments to the actual payments.

(ii) Construct a table showing the difference between the fitted payments and the actual payments in the table above. [3]

(iii) Comment on the results of the analysis in part (ii). [2]
[Total 12]

8 An insurer classifies the buildings it insures into one of three types. For Type 1 buildings, the number of claims per building per year follows a Poisson distribution with parameter λ. Data are available for the last five years as follows:

Year	*1*	*2*	*3*	*4*	*5*
Number of type 1 buildings covered	89	112	153	178	165
Number of claims	15	23	29	41	50

(i) Determine the maximum likelihood estimate of λ based on the data above. [5]

The insurer also has data for the other two types of building for all five years. Define

P_{ij} = the number of buildings insured in the jth year from type i and

Y_{ij} = the corresponding number of claims.

The five years of data can be summarised as follows:

Type(i)	$\bar{P}_i = \sum_{j=1}^{5} P_{ij}$	$\bar{X}_i = \sum_{j=1}^{5} \frac{Y_{ij}}{\bar{P}_i}$	$\sum_{j=1}^{5} P_{ij}\left(\frac{Y_{ij}}{P_{ij}} - \bar{X}_i\right)^2$	$\sum_{j=1}^{5} P_{ij}\left(\frac{Y_{ij}}{P_{ij}} - \bar{X}\right)^2$
Type 1	697	0.226686	1.527016	2.502737
Type 2	295	0.237288	0.96605	1.178133
Type 3	515	0.330097	4.53253	6.775614

$$\bar{X} = \sum_{i=1}^{3}\sum_{j=1}^{5} \frac{Y_{ij}}{\bar{P}} = 0.264101 \text{ where } \bar{P} = \sum_{i=1}^{3} \bar{P}_i$$

There are 191 buildings of Type 1 to be insured in year six.

(ii) Estimate the number of claims from Type 1 buildings in year six using Empirical Bayes Credibility Theory model 2. [6]

(iii) Explain the main differences between the approaches in parts (i) and (ii). [2]
[Total 13]

9 In order to model a particular seasonal data set an actuary is considering using a model of the form

$$(1-B^3)\left(1-(\alpha+\beta)B+\alpha\beta B^2\right)X_t = e_t$$

where B is the backward shift operator and e_t is a white noise process with variance σ^2.

(i) Show that for a suitable choice of s the seasonal difference series $Y_t = X_t - X_{t-s}$ is stationary for a range of values of α and β which you should specify. [3]

After appropriate seasonal differencing the following sample autocorrelation values for the series Y_t are observed: $\hat{\rho}_1 = 0.2$ and $\hat{\rho}_2 = 0.7$.

(ii) Estimate the parameters α and β based on this information. [7]

[HINT: let $X = \alpha + \beta$, $Y = \alpha\beta$ and find a quadratic equation with roots α and β.]

(iii) Forecast the next two observations $\hat{x}_{101}$ and $\hat{x}_{102}$ based on the parameters estimated in part (ii) and the observed values $x_1, x_2, \ldots, x_{100}$ of X_t. [4]
[Total 14]

10 Claims occur on a portfolio of insurance policies according to a Poisson process. Individual claim amounts are either 1 (with probability 0.7) or 8 (with probability 0.3). The insurance company uses a premium loading of 60% to calculate premiums and buys excess of loss reinsurance with a retention of M ($1<M<8$) from a reinsurer. The reinsurer uses a premium loading of 120%.

(i) Calculate the smallest value of M that the insurance company should consider if it wishes to expect to make a profit on this portfolio. [3]

(ii) Derive the adjustment coefficient equation for the insurance company. [2]

(iii) Calculate the adjustment coefficient (correct to 2 decimal places) if M=4. [4]

The same reinsurer also offers proportional reinsurance with the same premium loading such that the reinsurer pays a proportion α of each claim.

(iv) Show that the insurance company may either purchase excess of loss reinsurance with retention M or proportional reinsurance with $\alpha = \frac{3(8-M)}{31}$ for the same premium. [2]

(v) Determine whether the adjustment coefficient with proportional reinsurance is higher or lower than that with excess of loss reinsurance when M=4. [4]

(vi) Comment on the implications of part (v). [2]

[Total 17]

END OF PAPER

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

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

And secondly

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

Now

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

And

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

Hence

$$Var[X] = 4 + 50 = 54$$

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

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

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

$= 67.9207$

The estimators are given by

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

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

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

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

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

$$= 0.00237668$$

And the credibility factor for type 1 policies is given by

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

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

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

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:

$$M_X\left(R\right) - 1 - cR = 0$$

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:

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

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

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

Again by trial and error

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

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

INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

23 April 2013 (pm)

Subject CT6 – Statistical Methods
Core Technical

Time allowed: Three hours

INSTRUCTIONS TO THE CANDIDATE

1. *Enter all the candidate and examination details as requested on the front of your answer booklet.*

2. *You must not start writing your answers in the booklet until instructed to do so by the supervisor.*

3. *Mark allocations are shown in brackets.*

4. *Attempt all 11 questions, beginning your answer to each question on a separate sheet.*

5. *Candidates should show calculations where this is appropriate.*

Graph paper is NOT required for this paper.

AT THE END OF THE EXAMINATION

Hand in BOTH your answer booklet, with any additional sheets firmly attached, and this question paper.

In addition to this paper you should have available the 2002 edition of the Formulae and Tables and your own electronic calculator from the approved list.

1 Use the $U(0,1)$ random numbers 0.238 and 0.655 to generate two observations from the Weibull distribution with parameters $c = 0.002$ and $\gamma = 1.1$. [4]

2 Claims on a certain type of insurance policy are believed to follow an exponential distribution. The upper quartile claim size is 240.

Calculate the mean claim size. [4]

3 An actuary has a tendency to be late for work. If he gets up late then he arrives at work X minutes late where X is exponentially distributed with mean 15. If he gets up on time then he arrives at work Y minutes late where Y is uniformly distributed on [0,25]. The office manager believes that the actuary gets up late one third of the time.

Calculate the posterior probability that the actuary did in fact get up late given that he arrives more than 20 minutes late at work. [5]

4 (i) Explain what is meant by a two player zero-sum game. [2]

Sally and Fiona agree to play a game. The rules of the game are as follows:

- Each player chooses either the number 10 or the number 40.
- Neither player knows the other player's choice before selecting her number.
- If both players choose the same number, Fiona pays Sally the sum of the numbers.
- If the players choose differently, Sally pays Fiona the sum of the numbers.

Sally decides to adopt a randomised strategy where she chooses 10 with probability p and 40 with probability $1-p$.

(ii) (a) Determine the value of p for which Sally's expected payoff is the same regardless of what Fiona chooses.

(b) Explain why this strategy is optimal for Sally.

(c) Calculate Sally's expected payout each time the game is played, assuming that she follows this strategy.

[4]
[Total 6]

5 The following table shows incremental claims data from a portfolio of insurance policies for the accident years 2010, 2011 and 2012. Claims from this type of policy are fully run off after the end of development year two.

Incremental Claims		*Development year* 0	1	2
Accident year	2010	2,328	1,484	384
	2011	1,749	1,188	
	2012	2,117		

Estimate the total claims outstanding using the basic chain ladder technique. [7]

6 Claim numbers on a portfolio of insurance policies follow a Poisson process with parameter λ. Individual claim amounts X follow a distribution with moments $m_i = E(X^i)$ for $i = 1, 2, 3, \ldots$. Let S denote the aggregate claims for the portfolio. You may assume that the mean of S is λm_1 and the variance of S is λm_2.

(i) Derive the third central moment of S and show that the coefficient of skewness of S is $\dfrac{\lambda m_3}{(\lambda m_2)^{3/2}}$. [4]

(ii) Show that S is positively skewed regardless of the distribution of X. [2]

(iii) Show that the distribution of S tends to symmetry as $\lambda \to \infty$. [2]
[Total 8]

7 An insurance company believes that individual claim amounts from house insurance policies follow a gamma distribution with distribution function given by:

$$f(y) = \frac{\alpha^{\alpha}}{\mu^{\alpha}\Gamma(\alpha)} y^{\alpha-1} e^{-\frac{\alpha}{\mu}y} \text{ for } y > 0$$

where α and μ are positive parameters.

(i) Show that the gamma distribution can be written in exponential family form, giving the natural parameter and the canonical link function. [5]

The insurance company has data for claim amounts from previous claims. It believes that the claim amount is primarily influenced by two variables:

x_i the type of geographical area in which the house is situated. This can take one of 4 values.

y_i the category of the age of the house where the three categories are 0–29 years, 30-59 years and 60 years +.

It wishes to model claim amounts using this data and the generalised linear model from part (i) with canonical link function. The insurance company is investigating models which take into account these variables and has the following table of values:

Model	*Choice of predictor*	*Scaled Deviance*
A	1	900
B	Age	789
C	Age +location	544
D	Age * location	541

(ii) Explain, by analysing the scaled deviances, which model the insurance company should use. [6]
[Total 11]

8 An insurance company has a portfolio of 1,000 car insurance policies. Claims arise on individual policies according to a Poisson process with annual rate μ. The insurance company believes that μ follows a gamma distribution with parameters $\alpha = 2$ and $\lambda = 8$.

(i) (a) Show that the average annual number of claims per policy is 0.25.

(b) Show that the variance of the number of annual claims per policy is 0.28125.

[5]

Individual claim amounts follow a gamma distribution with density

$$f(x) = \frac{x}{1,000,000} e^{\frac{-x}{1000}} \quad \text{for } x > 0.$$

(ii) Calculate the mean and variance of the annual aggregate claims for the whole portfolio. [3]

The insurance company has agreed an aggregate excess of loss reinsurance contract with a retention of £0.55m (this means that the reinsurance company will pay the excess above £0.55m if the aggregate claims on the portfolio in a given year exceed £0.55m).

(iii) Calculate, using a Normal approximation, the probability of aggregate claims exceeding the retention in any year. [2]

For each of the last three years, the total claim amount has in fact exceeded the retention.

(iv) Comment on this outcome in light of the calculation in part (iii). [2]

[Total 12]

9 Claims on a portfolio of insurance policies arise as a Poisson process with rate λ. The mean claim amount is μ. The insurance company calculates premiums using a loading of θ and has an initial surplus of U.

(i) Explain how the parameters λ, μ, θ and U affect $\psi(U)$, the probability of ultimate ruin. [4]

Now suppose that $\lambda = 50$, $\mu = 200$ and $\theta = 30\%$. There are three models under consideration for the distribution of individual claim amounts:

A fixed claims of 200
B exponential with mean 200
C gamma with mean 200 and variance 800

Let the corresponding adjustment coefficients be R_A, R_B and R_C.

(ii) Find the numerical value of R_B and show that R_B is less than both R_A and R_C. [7]

(iii) Use the fact that $\left(1+\dfrac{x}{n}\right)^n \approx e^x$ for large n to show that R_A and R_C are approximately equal. [2]

[Total 13]

10 An insurance company has a portfolio of building insurance policies. The company classifies buildings into three types and believes that the number of claims on buildings of each type follows a Poisson distribution with parameters as shown:

Type	*Parameter*
1	λ
2	2λ
3	5λ

where λ is an unknown positive constant.

Actual claim numbers over the last five years have been as follows. Here X_{ij} represents the number of claims from the *i*th type in the *j*th year:

Number of claims X_{ij}

Year (j)

Type(i)	*5*	*4*	*3*	*2*	*1*	$\sum_{j=1}^{5}(X_{ij}-\bar{X}_i)^2$
1	23	17	9	21	12	139.2
2	56	39	44	29	35	417.2
3	87	115	141	92	84	2322.8

(i) Derive the maximum likelihood estimate of λ. [5]

(ii) Estimate the average number of claims per year for each type of building using EBCT Model 1. [7]

(iii) Comment on the results of parts (i) and (ii). [2]

(iv) Explain the main weakness of the model in part (ii). [1]

[Total 15]

11 An actuary is considering the time series model defined by

$$X_t = \alpha X_{t-1} + e_t$$

where e_t is a sequence of independent Normally distributed random variables with mean 0 variance σ^2. The series begins with the fixed value $X_0 = 0$.

(i) Show that the conditional distribution of X_t given X_{t-1} is Normal and hence show that the likelihood of making observations $x_1, x_2, \ldots, x_n$ from this model is:

$$L \propto \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \alpha x_{i-1})^2}{2\sigma^2}}.$$ [3]

(ii) Show that the maximum likelihood estimate of α can also be regarded as a least squares estimate. [2]

(iii) Find the maximum likelihood estimates of α and σ^2. [4]

(iv) Derive the Yule-Walker equations for the model and hence derive estimates of α and σ^2 based on observed values of the autocovariance function. [5]

(v) Comment on the difference between the estimates of α in parts (iii) and (iv). [1]

[Total 15]

END OF PAPER

INSTITUTE AND FACULTY OF ACTUARIES

EXAMINERS' REPORT

April 2013 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 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.

D C Bowie
Chairman of the Board of Examiners

July 2013

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 April 2013 paper

The examiners had felt that this paper contained a slightly greater proportion of more routine questions than previous papers and this was backed up by some good solutions to most of the questions. There was a marked improvement from previous sessions on topics such as Bayes' Theorem (Q3) and Ruin Theory (Q9).

1 Using the inverse transform method we need to set:

$$u = 1 - e^{-cx^{\gamma}}.$$

i.e.

$$-cx^{\gamma} = \log(1-u).$$

i.e.

$$x = \left(\frac{\log(1-u)}{-c}\right)^{\frac{1}{\gamma}}.$$

Using the equation above with the parameters $c = 0.002$ and $\gamma = 1.1$ we get:

u = 0.238 gives x = 86.96

u = 0.655 gives x = 300.73

This routine question was well answered, although a few candidates struggled with the algebra.

2 We need to solve:

$$1 - e^{-240\lambda} = 0.75$$

$$e^{-240\lambda} = 0.25$$

so

$$\lambda = \frac{\log(0.25)}{-240} = 0.005776$$

and so the mean is $\dfrac{1}{0.005776} = 173.12.$

This straightforward question was well answered.

3 Let L be the state getting up late and let M be the state of getting up on time.

Let Z be the number of minutes late.

According to Bayes' theorem:

$$P(L|Z>20)=\frac{P(Z>20|L)P(L)}{P(Z>20)}$$

but

$$P(Z>20|L)=e^{-\frac{20}{15}}=0.263597138$$

and

$$P(Z>20)=P(Z>20|L)P(L)+P(Z>20|M)P(M)$$

$$=0.263597138\times {}^{1}\!/_{3}+0.2\times {}^{2}\!/_{3}=0.221199046$$

and so

$$P(L|Z>20)=\frac{0.263597138\times {}^{1}\!/_{3}}{0.221199046}=0.3972.$$

This question was well answered by most candidates however weaker candidates were unable to apply Bayes' Theorem.

4 (i) A game with 2 players where whatever one player loses in the game the other player wins, and vice versa.

(ii) (a)

Value to Sally		**Sally**	
		10	**40**
Fiona	**10**	20	−50
	40	−50	80

Sally chooses 10 with probability p.

Then for the expected payoffs to be equal regardless of Fiona's choice we must have:

$$20p - 50(1-p) = -50p + 80(1-p)$$

so $200p = 130$

so $p = 0.65$

(b) This strategy is optimal for Sally because it produces the same expected payoff regardless of what Fiona does. Under any other randomized strategy Fiona can adopt a strategy that minimizes Sally's expected payoff.

(c) Value = 20 * 0.65 – 50 * 0.35 = –4.5.

This question was generally well answered, although a few candidates were thrown by a less familiar application of decision theory.

5 First accumulate claims:

Cumulative Claims		*Development year* 0	1	2
Accident year	*2010*	2,328	3,812	4,196
	2011	1,749	2,937	
	2012	2,117		

DY1 = (3,812 + 2,937) / (2,328 + 1,749) = 1.655 384
DY2 = 4,196 / 3,812 = 1.100 735

Now complete lower half of table:

Cumulative Claims		*Development year* 0	1	2
Accident year	*2009*	2,328	3,812	4,196
	2011	1,749	2,937	3,232.86
	2012	2,117	3,504.45	3,857.47

So estimated amount of outstanding claims is:

(3,232.86 – 2,937) + (3,857.47 – 2,117) = 2,036.3.

Most candidates scored full marks on this straightforward application of chain ladder theory.

6 (i) We have:

$$M_S(t) = M_N(\log M_X(t)) = e^{\lambda(M_X(t)-1)}$$

Let us work with the cumulant generating function:

$$C_S(t) = \log M_S(t) = \lambda M_X(t) - \lambda.$$

The third central moment is given by $C_S'''(0)$.

Now:

$$C_S'''(t) = \lambda M_X'''(t)$$

and so

$$C_S'''(0) = \lambda M_X'''(0) = \lambda m_3.$$

Hence the coefficient of skewness is given by:

$$\frac{E((S-E(S))^3}{(Var(S))^{3/2}} = \frac{\lambda m_3}{(\lambda m_2)^{\frac{3}{2}}}.$$

(ii) Since X takes only positive values we have $m_3 = E\left(X^3\right) > 0.$

Both λ and $m_2 = E(X^2)$ are also always positive.

This means the coefficient of skewness is always positive.

(iii) Re-writing the equation for the coefficient of skewness we have:

$$\frac{\lambda m_3}{(\lambda m_2)^{\frac{3}{2}}} = \frac{m_3}{\lambda^{0.5} m_2^{1.5}} \to 0 \text{ as } \lambda \to \infty.$$

Hence the distribution of S tends to symmetry as $\lambda \to \infty$.

Well prepared candidates who knew their bookwork were able to answer this question well, however weaker candidates struggled with part (i) and gave unconvincing answers to part (ii) & (iii).

7 (i) From the definition of the gamma density given in the question

$$f(y) = \frac{\alpha^{\alpha}}{\mu^{\alpha}\Gamma(\alpha)} y^{\alpha-1} e^{-\frac{\alpha}{\mu}y}$$

$$= \exp\left[\left(-\frac{y}{\mu} - \log\mu\right)\alpha + (\alpha-1)\log y + \alpha\log\alpha - \log\Gamma(\alpha)\,)\right]$$

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

where:

$$\theta = -\frac{1}{\mu}$$

$$\varphi = \alpha$$

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

$$b(\theta) = -\log(-\theta)$$

$$c(y,\varphi) = (\varphi-1)\log y + \varphi\log\varphi - \log\Gamma(\varphi).$$

Hence the distribution has the right form for a member of an exponential family.

The natural parameter is $-{}^{1}\!/_{\mu}$. The canonical link function is $\frac{1}{\mu}$.

(ii) Using the information given, we can calculate the deviance differences and compare that with the differences of the degrees of freedom for each of the nested models. If the decrease in the deviances is greater than twice the difference in degrees of freedom this suggests an improvement.

Model	*Scaled Deviance*	*Degrees of freedom*	*Difference in scaled deviance*
1	900	12	
Age	789	10	111
Age +location	544	7	245
Age * location	541	1	3

From the table we can see that the interaction model does not indicate any improvement hence the recommended model would be **Age +location**.

Again well prepared candidates were able to score highly on this question, however weaker candidates dropped marks as a result of not specifying a full parameterisation in part (i). In part (ii) full credit was given to candidates who used the chi-squared test rather than the approximation set out above..

8 (i) (a) $E(N) = E[E(N|\mu)]$
$= E[\mu] = 2/8 = 0.25$

(b) $\text{var}(N) = E[\text{var}(N|\mu)] + \text{var}[E(N|\mu)]$
$= E[\mu] + \text{var}[\mu]$
$= 2/8 + 2/8^2 = 0.28125$

(ii) Let Y be aggregate claims from one policy.

Individual claim is gamma with $\alpha = 2$ and $\lambda = 0.001$.

$E(Y) = E(X)E(N) = 2000 \times 0.25 = 500.$

$Var(Y) = E(N)Var(X) + Var(N)E(X)^2$

$= 0.25 \times 2000000 + 9/32 \times 2000^2 = 1,625,000.$

So the mean and variance of total claims are 500,000 and 1,625,000,000 respectively.

(iii) Our approximate distribution for S is $S \sim N(500{,}000\ ,\ 1625000000)$.

$$P(S > 550000) = P\left(Z > \frac{550000 - 500000}{\sqrt{1625000000}}\right) = P(Z > 1.24035) = 0.1074\,.$$

(iv) The prob three years in a row is $0.1074^3 = 0.00124$.

The probability of this happening is very low. It is more likely that the insurance company's belief about the distribution of claims amounts is incorrect.

The normal approximation tails off quickly and so underestimates the probability of extreme events

Part (i) was straightforward, however some candidates failed to show sufficient working to gain full marks. A surprising number of candidates were unfamiliar with the standard bookwork underlying part (ii). Credit was given for any sensible comments in part (iv).

9 (i) $\psi(U)$ does not depend on λ. This parameter affects the speed with which the process runs, but does not affect the ultimate probability of ruin.

$\psi(U)$ is higher for higher values of μ since the significance of the starting capital falls as μ rises, providing proportionately less of a buffer.
$\psi(U)$ is lower for higher values of θ since the higher θ is the higher the premiums with no change to claim amounts, so that there is a larger buffer against ruin.

$\psi(U)$ is lower for higher values of U since the higher U is the higher the larger the buffer against ruin given by the initial capital.

(ii) The adjustment coefficients are the solutions to:

$$M_X(R) = 1 + 200 \times 1.3 \times R = 1 + 260R$$

for the various choices of the moment generating function.

Our first task is to find the parameters in the gamma distribution in C.

Denoting these by α and β we have:

$$\frac{\alpha}{\beta} = 200 \text{ and } \frac{\alpha}{\beta^2} = 800$$

Dividing the second by the first we get $1/\beta = 4$ so $\beta = 0.25$ and $\alpha = 50$.

Solving for R_B we have:

$$\frac{0.005}{0.005 - R_B} = 1 + 260R_B.$$

$$1 = (1 + 260R_B)(1 - 200R_B).$$

$$1 = 1 + 60R_B - 52,000R_B^2$$

$$R_B = \frac{60}{52,000} = 0.001153846.$$

Consider the three functions:

$$A = e^{200R} - 1 - 260R$$

$$B = \frac{0.005}{0.005 - R} - 1 - 260R$$

$$C = \left(\frac{0.25}{0.25 - R}\right)^{50} - 1 - 260R.$$

We can tabulate the values of these functions as follows:

R	**A**	**B**	**C**
0.0001	–0.00579	–0.00559	–0.00579
0.0012	–0.04075	0.003789	–0.0400

So the second function has changed sign, but the first and third have not which gives the required result.

(iii) We know that R_C satisfies:

$$\left(\frac{0.25}{0.25 - R_C}\right)^{50} = 1 + 260R_C.$$

We can re-write this as:

$$\left(\frac{0.25 - R_C}{0.25}\right)^{-50} = 1 + 260R_C.$$

$$\frac{1}{(1 - \frac{R_C}{0.25})^{50}} = 1 + 260R_C.$$

$$\frac{1}{\left(1 - \frac{200R_C}{50}\right)^{50}} = 1 + 260R_C.$$

But due to the approximation given in the question, the denominator of the left hand side is approximately e^{-200R_C}.

So we have, approximately:

$$\frac{1}{e^{-200R_C}} = 1 + 260R_C.$$

i.e. $e^{200R_C} = 1 + 260R_C.$

Which is the equation satisfied by R_A. Hence R_C and R_A are approximately equal.

Most candidates scored well in part (i), although many simply stated how the probability of ruin changes without explaining why. Well prepared candidates scored well on part (ii), noting the method in previous examinations for finding the root of the equation; however very few candidates scored well on part (iii) which was stretching.

10 (i) The likelihood function is given by:

$$L \propto \prod_{i=1}^{5} e^{-\lambda}\lambda^{x_{1,i}} \prod_{i=1}^{5} e^{-2\lambda}(2\lambda)^{x_{2,i}} \prod_{i=1}^{5} e^{-5\lambda}(5\lambda)^{x_{3,i}}.$$

Where $x_{j,i}$ is the number of claims on the *j*th type in the *i*th year.

The log likelihood is given by:

$$l = \log L = C - 5\lambda - 10\lambda - 25\lambda + (\log\lambda)\sum_{i,j} x_{j,i}.$$

$$= C - 40\lambda + 804(\log\lambda).$$

Differentiating gives:

$$\frac{dl}{d\lambda} = -40 + \frac{804}{\lambda}.$$

and setting this equal to zero gives:

$$\hat{\lambda} = \frac{804}{40} = 20.1.$$

This is a maximum since:

$$\frac{d^2 l}{d\lambda^2} = -\frac{804}{\lambda^2} < 0.$$

(ii) The mean number of claims for the various types are:

$$\bar{X}_1 = 16.4 \text{ and } \bar{X}_2 = 40.6 \text{ and } \bar{X}_3 = 103.8.$$

With overall mean $\bar{X} = 53.6$.

So we have parameter estimates:

$$E\left(m(\theta)\right) = \bar{X} = 53.6.$$

$$E\left(s^2(\theta)\right) = \frac{1}{3}\sum_{i=1}^{3}\left[\frac{1}{4}\sum_{j=1}^{5}\left(X_{ij} - \bar{X}_i\right)^2\right]$$

$$= \frac{1}{12}\left(139.2 + 417.2 + 2322.8\right) = 239.9333333.$$

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

$$= 0.5\left[(16.4-53.6)^2 + (40.6-53.6)^2 + (103.8-53.6)^2\right] - 0.2\times 239.93333333$$

$$=1988.4533333.$$

And so:

$$Z = \frac{5}{5 + \dfrac{E(s^2(\theta))}{Var(m(\theta))}} = \frac{5}{5 + \dfrac{239.9333333}{1988.4533333}} = 0.976436003\,.$$

and the expected claims from the three types are:

Type	*Credibility Premium*
1	0.976436002 × 16.4 + 0.023563998 × 53.6 = 17.3
2	0.976436002 × 40.6 + 0.023563998 × 53.6 = 40.9
3	0.976436002 × 103.8 + 0.023563998 × 53.6 = 102.6

(iii) The corresponding estimates based on our computed $\hat{\lambda}$ are 20.1, 40.2 and 100.5.

The estimates are remarkably similar. The biggest difference is for type 1 buildings, where the maximum likelihood estimate gives a lower weight to the data from that risk, but the credibility estimate gives greater weight.

(iv) The main limitation is that the model in (ii) does not take account of the volume of buildings covered, which will probably vary from year to year.

Again well prepared candidates found this question relatively straightforward. Weaker candidates were unable to construct the likelihood function in part (i). A disappointing number of candidates were unable to accurately render the standard formulae in part (ii).

11 (i) $X_t - \alpha X_{t-1} = e_t \sim N(0, \sigma^2)$.

So $X_t \,|\, X_{t-1} \sim N(\alpha X_{t-1}, \sigma^2)$

and so the likelihood is given by:

$$L \propto \prod_{i=1}^{n} P(X_i = x_i \,|\, x_{i-1}) \times P(x_0)$$

$$L \propto \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \alpha x_{i-1})^2}{2\sigma^2}} \times 1$$

(ii) We can see that maximising the likelihood with respect to α is the same as minimising the expression:

$$L \propto \sigma^{-n} e^{-\frac{\sum_{i=1}^{n}(x_i - \alpha x_{i-1})^2}{2\sigma^2}}$$

$$\sum_{i=1}^{n}(x_i - \alpha x_{i-1})^2 .$$

(iii) The log-likelihood is given by:

$$l = logL = -n \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n}(x_i - \alpha x_{i-1})^2 + \text{Constant}$$

Differentiating with respect to α gives:

$$\frac{\partial l}{\partial \alpha} = \frac{1}{2\sigma^2} \sum_{i=1}^{n} 2x_{i-1}(x_i - \alpha x_{i-1}) \quad = \frac{1}{\sigma^2} \sum_{i=1}^{n} x_i x_{i-1} - \frac{\alpha}{\sigma^2} \sum_{i=1}^{n} x_{i-1}^2$$

and setting $\frac{\partial l}{\partial \alpha} = 0$ we have:

$$\hat{\alpha} = \frac{\sum_{i=1}^{n} x_i x_{i-1}}{\sum_{i=1}^{n} x_{i-1}^2} .$$

Differentiating with respect to σ we have:

$$\frac{\partial l}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i - \alpha x_{i-1})^2.$$

Setting this expression equal to zero we have:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\alpha} x_{i-1})^2.$$

(iv) The Yule Walker equations are:

$$\gamma_0 = \text{cov}\,(\alpha X_{t-1} + e_t, X_t) = \alpha\,\text{cov}(X_{t-1}, X_t) + \text{cov}\,(e_t\,, X_t) = \alpha\gamma_1 + \sigma^2$$
$$\gamma_1 = \text{cov}\,(\alpha X_{t-1} + e_t, X_{t-1}) = \alpha\,\text{cov}(X_{t-1}, X_{t-1}) + \text{cov}\,(e_t\,, X_{t-1}) = \alpha\gamma_0$$

Using these to estimate the parameters we get:

$$\hat{\sigma}^2 = \hat{\gamma}_0 - \hat{\alpha}\hat{\gamma}_1.$$

(v) $$\hat{\alpha} = \frac{\hat{\gamma}_1}{\hat{\gamma}_0} = \frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)(x_{i-1} - \overline{x})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}.$$

The difference between them is that in the second approach we need to centralise the data around the mean $\overline{x}$.

This question was relatively well answered for a time series question. It was clear that some candidates had learnt the bookwork, but struggled with this more unfamiliar application of time series. In particular only the best candidates accurately completed the differentiation needed in part (iii).

END OF EXAMINERS' REPORT

INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

1 October 2013 (pm)

Subject CT6 – Statistical Methods
Core Technical

Time allowed: Three hours

INSTRUCTIONS TO THE CANDIDATE

1. *Enter all the candidate and examination details as requested on the front of your answer booklet.*

2. *You must not start writing your answers in the booklet until instructed to do so by the supervisor.*

3. *Mark allocations are shown in brackets.*

4. *Attempt all 10 questions, beginning your answer to each question on a separate sheet.*

5. *Candidates should show calculations where this is appropriate.*

Graph paper is NOT required for this paper.

AT THE END OF THE EXAMINATION

Hand in BOTH your answer booklet, with any additional sheets firmly attached, and this question paper.

In addition to this paper you should have available the 2002 edition of the Formulae and Tables and your own electronic calculator from the approved list.

1 An insurance company has a portfolio of n policies. The probability of a claim in a given year on each policy is p independently from policy to policy, and the possibility of more than one claim can be ignored. Prior beliefs about p are specified by a Beta distribution with parameters α and β. In one year the insurance company has a total of k claims on the portfolio.

Calculate the posterior estimate of p under all or nothing loss and show that it can be written in the form of a credibility estimate. [5]

[You may use without proof the fact that the mode of a Beta distribution with parameters α and β is $\dfrac{\alpha-1}{\alpha+\beta-2}$.]

2 Claim amounts on a certain type of insurance policy follow an exponential distribution with mean 100. The insurance company purchases a special type of reinsurance policy so that for a given claim X the reinsurance company pays

0	if $0 < X < 80$;
$0.5X - 40$	if $80 \le X < 160$;
$X - 120$	if $X \ge 160$.

Calculate the expected amount paid by the reinsurance company on a randomly chosen claim. [6]

3 Andy is a famous weight lifter who will be competing at the Olympic Games. He has taken out special insurance which pays out if he is injured. If the injury is so serious that his career is ended the policy pays \$1m and is terminated. If he is injured but recovers the insurance payment is \$0.1m and the policy continues.

The insurance company's underwriters believe that the probability of an injury in any year is 0.2, and that the probability of more than one injury in a year can be ignored. If Andy is injured, there is a 75% chance that he will recover.

Annual premiums are paid in advance, and the insurance company pays claims at the end of the year. Assume that this is the only policy that the insurance company writes, and that it has an initial surplus of \$0.1m.

(i) Define what is meant by $\psi(\$0.1\text{m},1)$ and $\psi(\$0.1\text{m})$. [2]

(ii) Calculate the annual premium charged assuming the insurance company uses a premium loading of 30%. [2]

(iii) Determine $\psi(\$0.1\text{m},2)$. [4]

[Total 8]

4 The table below shows the probability distribution of a discrete random variable X.

Value	1	2	3
Probability	0.3	0.3	0.4

(i) Construct an algorithm to generate random samples from X. [2]

The random variable Z takes values from X with probability 0.2 and values from an exponential distribution Y with probability 0.8. The upper quartile point of the distribution of Y is 2.5.

(ii) Calculate the expected value of Y. [3]

(iii) Extend the algorithm in part (i) to generate random samples from Z. [4]
[Total 9]

5 An insurance company has a portfolio of life insurance policies for 2,000 workers at a factory. The policies pay out £5,000 if a worker dies in an industrial accident and £2,000 if a worker dies for any other reason. For each worker, the probability of death in any year is 0.02 and 25% of deaths are the result of industrial accidents. The insurance company charges an annual premium of £74.25 per worker.

(i) Calculate the premium loading used by the insurance company. [2]

The insurance company is considering adopting one of the following three approaches to reinsurance:

A None.

B 30% proportional reinsurance at a cost of £27 per worker.

C Individual excess of loss reinsurance with retention £3,000 and a premium of £15 per worker.

(ii) Find the optimal decision under the Bayes criterion. [4]

(iii) Find the optimal decision under the minimax criterion. [2]

(iv) Comment on your answer to part (iii). [2]
[Total 10]

6 The tables below show cumulative data for the number of claims and the total claim amounts arising from a portfolio of insurance policies.

Claim Numbers

	Development Year		
	0	*1*	*2*
2010	87	132	151
2011	117	156	
2012	99		

Total Claim Amounts

	Development Year		
	0	*1*	*2*
2010	43,290	87,430	126,310
2011	68,900	125,290	
2012	74,250		

Claims are fully run off after two development years.

Estimate the outstanding claims using the average cost per claim method with grossing up factors. [10]

7 An insurance company offers dental insurance to the employees of a small firm. The annual number of claims follows a Poisson process with rate 20. Individual loss amounts follow an exponential distribution with mean 100. In order to increase the take-up rate, the insurance company has guaranteed to pay a minimum amount of £50 per qualifying claim. Let S be the total claim amount on the portfolio for a given year.

(i) Show that the mean and variance of S are 2,213.06 and 413,918.40 respectively. [7]

[You may use without proof the result that if $I_n = \int_M^\infty y^n \lambda e^{-\lambda y} dy$

then $I_n = M^n e^{-\lambda M} + \frac{n}{\lambda} I_{n-1}$]

(ii) (a) Fit a log-normal distribution for S using the method of moments.

(b) Estimate the probability that S is greater than 4,000.

[3]

Sarah, the insurance company's actuary, has instead approximated S by a Normal distribution.

(iii) Explain, without performing any further calculations, whether the probability that she calculates that S exceeds 4,000 will be greater or smaller than the calculation in part (ii). [2]

[Total 12]

8 The number of claims per month Y arising on a certain portfolio of insurance policies is to be modelled using a modified geometric distribution with probability density given by

$$p(y|\alpha) = \frac{\alpha^{y-1}}{(1+\alpha)^y} \quad y = 1, 2, 3, \ldots$$

where α is an unknown positive parameter. The most recent four months have resulted in claim numbers of 8, 6, 10 and 9.

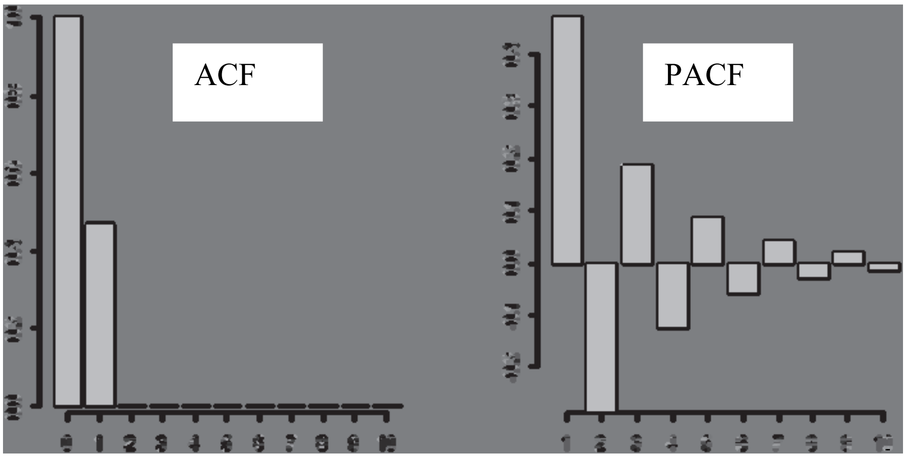
(i) Derive the maximum likelihood estimate of α. [5]

(ii) Show that Y belongs to an exponential family of distributions and suggest its natural parameter. [5]

[Total 10]

9 (i) State the three main stages in the Box-Jenkins approach to fitting an ARIMA time series model. [3]

(ii) Explain, with reasons, which ARIMA time series would fit the observed data in the charts below. [2]



Now consider the time series model given by

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \beta_1 e_{t-1} + e_t$$

where e_t is a white noise process with variance σ^2.

(iii) Derive the Yule-Walker equations for this model. [6]

(iv) Explain whether the partial auto-correlation function for this model can ever give a zero value. [2]

[Total 13]

10 The number of service requests received by an IT engineer on any given day follows a Poisson distribution with mean μ. Prior beliefs about μ follow a gamma distribution with parameters α and λ. Over a period of n days the actual numbers of service requests received are $x_1, x_2, \ldots, x_n$.

(i) Derive the posterior distribution of μ. [3]

(ii) Show that the Bayes estimate of μ under quadratic loss can be written as a credibility estimate and state the credibility factor. [2]

Now suppose that $\alpha = 10$, $\lambda = 2$ and that the IT worker receives 42 requests in 6 days.

(iii) Calculate the Bayes estimate of μ under quadratic loss. [1]

Three quarters of requests can be resolved by telling users to restart their machine, and the time taken to do so follows a Pareto distribution with density

$$f(x) = \frac{3 \times 20^3}{(20+x)^4} \quad \text{for } x>0.$$

One quarter of requests are much harder to resolve, and the time taken to resolve these follows a Weibull distribution with density

$$f(x) = 0.4 \times 0.5 x^{-0.5} e^{-0.4 x^{0.5}} \quad \text{for } x>0.$$

(iv) (a) Calculate the probability that a randomly chosen request takes more than 30 minutes to resolve.

(b) Calculate the average time spent on each request.

(c) Calculate the expected total amount of time the IT worker spends dealing with service requests each day, using the estimate of μ from part (iii).

[5]

The IT worker's line manager is carefully considering his staffing requirements. He decides to model the time taken on each request approximately using an exponential distribution.

(v) (a) Fit an exponential distribution to the time taken per request using the method of moments.

(b) Calculate the probability that a randomly chosen request takes more than 30 minutes to resolve using this approximation.

(c) Comment briefly on your answer to part (v)(b). [2]

The IT engineer needs to devote more of his time to a separate project, so his firm have hired an assistant to help him. The assistant is just as fast at dealing with the straightforward requests, and the time taken to resolve these still follows the Pareto distribution given above. He is significantly slower at dealing with the difficult requests, and the time taken to resolve these now follows a Weibull distribution with density:

$$f(x) = c \times 0.5x^{-0.5}e^{-cx^{0.5}} \quad \text{for } x>0$$

where c is a positive parameter. The line manager again fits an exponential distribution as an approximation to the time taken to service each request using the method of moments. His approximation results in an estimate that the probability that a random service request takes longer than 30 minutes to resolve is 10%.

(vi) Determine the value of c. [4]

[Total 17]

END OF PAPER

INSTITUTE AND FACULTY OF ACTUARIES

EXAMINERS' REPORT

September 2013 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 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.

D C Bowie
Chairman of the Board of Examiners

December 2013

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 2013 paper

The examiners felt that this paper was slightly less routine than the April paper, but broadly in line with other recent papers. The quality of solutions was often good, with questions 7 and 9 providing the greatest challenge to most students.

1 The posterior distribution of p is given by

$$f(p \mid k \text{ claims}) \propto f(k \text{ claims} \mid p) \times f(p)$$

$$\propto p^k (1-p)^{n-k} p^{\alpha-1} (1-p)^{\beta-1}$$

$$\propto p^{\alpha+k-1} (1-p)^{\beta+n-k-1}$$

which is the pdf of a Beta distribution with parameters $\alpha + k$ and $\beta + n - k$.

Using the fact given in the questions, the mode of the posterior distribution (which is the estimate of p under all or nothing loss is given by:

$$\hat{p} = \frac{\alpha + k - 1}{\alpha + k + \beta + n - k - 2} = \frac{\alpha + k - 1}{\alpha + \beta + n - 2}$$

$$= \frac{\alpha - 1}{\alpha + \beta - 2} \times \frac{\alpha + \beta - 2}{\alpha + \beta + n - 2} + \frac{k}{n} \times \frac{n}{\alpha + \beta + n - 2}$$

$$= (1 - Z) \times \frac{\alpha - 1}{\alpha + \beta - 2} + Z \times \frac{k}{n}$$

where $Z = \dfrac{n}{\alpha + \beta + n - 2}$.

This is in the form of a credibility estimate since $\dfrac{\alpha - 1}{\alpha + \beta - 2}$ is the prior estimate of p under all or nothing loss and k/n is the estimate of p derived from the data.

The first part of this question was answered well. Most candidates didn't recognise the need to base the prior estimate on the mode of the prior distribution and therefore didn't manage to express the posterior estimate as a credibility estimate.

2 The mean amount paid by the reinsurance company is given by:

$$\int_{80}^{160}\left(\frac{1}{2}x-40\right)\times 0.01e^{-0.01x}dx+\int_{160}^{\infty}(x-120)\times 0.01e^{-0.01x}dx\,.$$

The first integral is (using integration by parts):

$$\left[-\left(\frac{1}{2}x-40\right)e^{-0.01x}\right]_{80}^{160}+\int_{80}^{160}\frac{1}{2}e^{-0.01x}dx$$

$$=-40e^{-1.6}-\left[50e^{-0.01x}\right]_{80}^{160}$$

$$=50e^{-0.8}-90e^{-1.6}=4.2957.$$

The second integral is:

$$\left[-(x-120)e^{-0.01x}\right]_{160}^{\infty}+\int_{160}^{\infty}e^{-0.01x}dx$$

$$=40e^{-1.6}+100e^{-1.6}$$

$$=28.26551.$$

So total mean claim = 4.29576 + 28.26551 = 32.56.

This question was generally answered well. Weaker candidates could not integrate by parts accurately.

3 (i) If $U(t)=U+ct-S(t)$ where $U=U(0)=\$0.1$ m then

$$\Psi(\$0.1m,1)=\Pr\Big(U(t)\le 0 \text{ for some } t\in(0,1] \text{ given } U(0)=\$0.1m\Big)$$

and

$$\Psi(\$0.1m)=\Pr(U(t)\le 0 \text{ for some } t>0 \text{ given } U(0)=\$0.1m)$$

(ii) The premium charged will be:

$$1.3\times 0.2\times(0.25\times\$1\text{m}+0.75\times\$0.1\text{m})=\$0.0845\text{m}.$$

(iii) The possibilities are tabulated below, where N means not injured, R means injured but recovered and X means injured but career ending:

Year 1	**Year 2**	**Probability**	**Ruin?**
N	N	$0.8 \times 0.8 = 0.64$	No
N	R	$0.8 \times 0.15 = 0.12$	No
N	X	$0.8 \times 0.05 = 0.04$	Yes
R	N	$0.15 \times 0.8 = 0.12$	No
R	R	$0.15 \times 0.15 = 0.0225$	No
R	X	$0.15 \times 0.05 = 0.0075$	Yes
X	N/A	0.05	Yes

Summing the cases where ruin occurs we have:

$$\psi(\$0.1\text{m}, 2) = 0.04 + 0.0075 + 0.05 = 0.0975$$

Many candidates lost marks in part (i) by not giving a sufficiently precise definition to score full marks. For part (iii) candidates who worked through the possibilities methodically generally scored well. A number of candidates unnecessarily used approximate methods in part (iii).

4 (i) The algorithm is as follows:

Step 1 Generate u from the uniform distribution on [0,1].

Step 2 If $0 < u < 0.3$ set $X = 1$.

If $0.3 <= u < 0.6$ set $X = 2$.

Otherwise set $X = 3$.

(ii) We need to solve $P(Y < 2.5) = 0.75$

but $P(Y < 2.5) = 1 - e^{-2.5\lambda}$

so $1 - e^{-2.5\lambda} = 0.75$

so $e^{-2.5\lambda} = 0.25$

so $\lambda = \dfrac{\log(0.25)}{-2.5} = 0.554517744$ and the mean of Y is 1.803368801.

(iii) The extended algorithm is:

Step 1 Generate v from the uniform distribution on [0,1].

Step 2 If $v<0.2$ then generate a sample from X as in (i) and finish, otherwise go to step 3.

Step 3 Generate u from the uniform distribution on [0,1].

Step 4 Set $1-e^{-\lambda x}=u$

i.e. $x=\dfrac{\log(1-u)}{-0.55451744}$

This question was answered well.

5 (i) The premium loading θ is given by:

$$74.25=(1+\theta)\times(0.75\times 2000+0.25\times 5000)\times 0.02=55(1+\theta)$$

and so

$$\theta=\frac{74.25}{55}-1=35\%.$$

(ii) Under A expected profit is:

$$\begin{aligned}&2000\times 74.25-2000\times 0.02\times(0.75\times 2000+0.25\times 5000)\\&=38,500.\end{aligned}$$

Under B expected profit is:

$$\begin{aligned}&2,000\times 74.25-2,000\times 27-0.7\times 2,000\times 0.02\times(0.75\times 2,000+0.25\times 5,000)\\&=17,500.\end{aligned}$$

Under C expected profit is:

$$\begin{aligned}&2000\times 74.25-2000\times 15-2000\times 0.02\times(0.75\times 2000+0.25\times 3000)\\&=28,500\end{aligned}$$

so the optimal course under the Bayes criterion is no reinsurance.

(iii) Under the minimax we need to consider the worst case scenario – which is that all 2,000 workers die in industrial accidents.

Under this outcome, the losses are:

Under A: $2000\times 74.25 - 2000\times 5000 = -9,851,500$
Under B: $2000\times(74.25-27) - 2000\times 5000\times 0.7 = -6,905,500$
Under C: $2000\times(74.25-15) - 2000\times 3000 = -5,881,500$

so the optimal decision under the minimax criterion is C.

(iv) The approach in (iii) puts all the weight on what is at first seems a pretty unlikely scenario – so that our decision making is driven by something fairly remote.

That said, the workers are all in the same factory, so it is not inconceivable that a single catastrophe could result in a large number of claims all at the same time – i.e. the lives are not independent.

This question was well answered. There are a number of alternative approaches available (for example working on a per policy basis) which all give the same results, and all of which were given full credit. Candidates made a range of comments in part (iv) and all sensible answers were given credit.

6 The average cost per claim is given in the table:

	0	*1*	*2*
2010	497.59	662.35	836.49
2011	588.89	803.14	
2012	750.00		

The grossing up factors for average costs are given in the table below (the underlined figures are the simple averages):

	0	*1*	*2*	*Ult*
2010	497.6	662.3	836.5	836.5
	59.49%	79.18%	<u>100.00%</u>	
2011	588.9	803.1		1014.3
	58.06%	<u>79.18%</u>		
2012	750.0			1276.1
	<u>58.77%</u>			

The grossing up factors for claim numbers are as follows:

	0	*1*	*2*	*Ult*
2010	87.0	132.0	151.0	151.0
	57.62%	87.42%	<u>100.00%</u>	
2011	117.0	156.0		178.5
	65.56%	<u>87.42%</u>		
2012	99.0			160.7
	<u>61.59%</u>			

So the total claims are:

	Average amount	*Number*	*Total*
2010	836.5	151.0	126310
2011	1014.3	178.5	181006
2012	1276.1	160.7	205126
			512442

So the outstanding claims are:

512,442 – 126,310 – 125,290 – 74,250 = 186,592.

This question was well answered.

7 (i) Let X_i be the amount paid on the ith claim:

Then

$$E(X_i) = 50\int_0^{50} f(y)dy + \int_{50}^{\infty} yf(y)dy$$

$$= 50\int_0^{50} 0.01e^{-0.01y}dy + I_1$$

Using the notation given in the question.

Now

$$I_1 = 50e^{-0.5} + 100\int_{50}^{\infty} 0.01e^{-0.01y}\,dy$$

$$= 50e^{-0.5} + 100\left[-e^{-0.01y}\right]_{50}^{\infty}$$

$$= 50e^{-0.5} + 100e^{-0.5} = 150e^{-0.5}$$

So

$$E(X_i) = 50\left[-e^{-0.01y}\right]_0^{50} + 150e^{-0.5}$$

$$= -50e^{-0.5} + 50 + 150e^{-0.5} = 50 + 100e^{-0.5} = 110.653066$$

And

$$E(X_i^2) = 50^2\int_0^{50} f(y)\,dy + \int_{50}^{\infty} y^2 f(y)\,dy$$

$$= 2,500\int_0^{50} 0.01e^{-0.01y}\,dy + I_2$$

$$= 2,500\left[-e^{-0.01y}\right]_0^{50} + 50^2 e^{-0.5} + 200I_1$$

$$= -2,500e^{-0.5} + 2,500 + 2,500e^{-0.5} + 200\times 150e^{-0.5}$$

$$= 2,500 + 200\times 150e^{-0.5} = 20,695.91979$$

So finally we have:

$$E(S) = 20\times 110.653066 = 2213.06$$

$$\text{Var}(S) = 20\times 20695.91979 = 413918.40.$$

(ii) We need to solve:

$$e^{\mu+\sigma^2/2} = 2213.06 \quad (1)$$

and

$$e^{2\mu+\sigma^2}\left(e^{\sigma^2}-1\right) = 413918.40. \quad (2)$$

Dividing (2) by the square of (1) we have:

$$e^{\sigma^2}-1 = \frac{413918.40}{2213.06^2} = 0.084514$$

$$\sigma^2 = \log(1.084514) = 0.081132$$

and substituting into (1) we have:

$$\mu = \log(2213.06) - \frac{0.081132}{2} = 7.6615655.$$

Finally:

$$P(S > 4000) = P(N(7.6615655, 0.081132) > \log(4000))$$

$$= P\left(N(0,1) > \frac{8.29404964 - 7.6615655}{\sqrt{0.081132}}\right) = P(N(0,1) > 2.2205)$$

$$= 0.95 \times (1 - 0.98679) + 0.05 \times (1 - 0.98713)$$

$$= 0.01319$$

(iii) The probability will be lower.

This is because the log normal distribution has a "fat tail" and hence gives more weight to extreme outcomes.

Only the best candidates were able to derive the value of the variance in part (i) despite the formula for integration by parts being given in the question paper. The remaining parts were well answered.

8 (i) We have 4 years of observations such that $y_1 + y_2 + y_3 + y_4 = 33$. The likelihood function is then:

$$L = \prod_{i=1}^{4} \frac{\alpha^{y_i - 1}}{(1+\alpha)^{y_i}} = \frac{\alpha^{33-4}}{(1+\alpha)^{33}} = \frac{\alpha^{29}}{(1+\alpha)^{33}}$$

The log-likelihood is then:

$$l = 29\log\alpha - 33\log(1+\alpha)$$

Taking its derivative w.r.t. α and equation it to zero we have:

$$\frac{29}{\alpha} - \frac{33}{1+\alpha} = 0$$

$$29(1+\alpha) = 33\alpha$$

which implies that $29 = 4\alpha$

therefore $\hat{\alpha} = \dfrac{29}{4} = 7.25.$

Differentiating the log likelihood again gives $-\dfrac{29}{\alpha^2} + \dfrac{33}{(1+\alpha)^2}$ which is negative at $\hat{\alpha} = 7.25$.

(ii) We have:

$$p(y) = \frac{\alpha^{y-1}}{(1+\alpha)^{y}} = \exp[y\log\alpha - y\log(1+\alpha) - \log\alpha]$$

$$= \exp\left[y\log\left(\frac{\alpha}{1+\alpha}\right) - \log\alpha\right]$$

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

where

$$\theta = \log\left(\frac{\alpha}{1+\alpha}\right), \text{ the natural parameter}$$

$$\varphi = 1$$

$$a(\varphi) = \varphi$$

$$b(\theta) = \log \alpha = \theta - \log(1 - e^{\theta})$$

$$c(y, \varphi) = 0$$

This question was mostly well answered. Only the best candidates showed that the estimate was a maximum by evaluating the second derivative of the log-likelihood at the value of the estimate. In part (ii) some candidates failed to score full marks as a result of not specifying all the parameters.

9 (i) The three main stages are:

(a) tentative model identification
(b) model fitting
(c) diagnostics

(ii) Since the auto-correlation is non-zero for the first lag only and the partial auto-correlation function decays exponentially it is likely that the observed data comes from an MA(1) (or equivalently a ARMA(0,1) or ARIMA(0,0,1) model).

(iii) First note that for this model:

$$\mathrm{Cov}(X_t, e_t) = \sigma^2$$

and

$$\mathrm{Cov}(X_t, e_{t-1}) = \alpha_1 \mathrm{Cov}(X_{t-1}, e_{t-1}) + \beta_1 \sigma^2 = (\alpha_1 + \beta_1)\sigma^2.$$

Taking the covariance of the defining equation with X_t we get:

$$\gamma_0 = \alpha_1 \gamma_1 + \alpha_2 \gamma_2 + \beta_1(\alpha_1 + \beta_1)\sigma^2 + \sigma^2.$$

Taking the covariance with X_{t-1} we get:

$$\gamma_1 = \alpha_1 \gamma_0 + \alpha_2 \gamma_1 + \beta_1 \sigma^2.$$

Taking the covariance with X_{t-2} we get:

$$\gamma_2 = \alpha_1 \gamma_1 + \alpha_2 \gamma_0$$

and in general

$$\gamma_n = \alpha_1\gamma_{n-1} + \alpha_2\gamma_{n-2} \text{ for } n > 2.$$

(iv) The presence of the term $\beta_1 e_{t-1}$ means that the PACF will decay exponentially to zero, but it will never get there, so that the PACF will always be non-zero.

Many candidates struggled with this question, with only the best accurately calculating the covariance of X_t with e_{t-1}. The chart on the printed examination paper was not clear, and the examiners took a generous approach to marking part (ii) where candidates had struggled interpreting the chart.

10 (i) Firstly:

$$f(\mu | x_1, x_2, \ldots, x_n) \propto f(x_1, x_2, \ldots, x_n) f(\mu)$$

$$\propto \prod_{i=1}^{n} e^{-\mu} \frac{\mu^{x_i}}{x_i!} \times \frac{\lambda^{\alpha}}{\Gamma(\alpha)} \mu^{\alpha-1} e^{-\lambda\mu}$$

$$\propto \mu^{\sum_{i=1}^{n} x_i + \alpha - 1} e^{-(\lambda+n)\mu}$$

which is the pdf of another gamma distribution. So the posterior distribution is gamma with parameters $\alpha + \sum_{i=1}^{n} x_i$ and $\lambda + n$.

(ii) Under quadratic loss the Bayes estimate is the mean of the posterior distribution, so:

$$\hat{\mu} = \frac{\alpha + \sum_{i=1}^{n} x_i}{\lambda + n}$$

which can be written as

$$\hat{\mu} = \frac{\alpha}{\lambda} \times \frac{\lambda}{\lambda + n} + \frac{\sum_{i=1}^{n} x_i}{n} \times \frac{n}{\lambda + n}$$

$$= (1 - Z)\frac{\alpha}{\lambda} + Z\overline{x}$$

where $Z = \dfrac{n}{\lambda + n}$. This is in the form of a credibility estimate since the mean of the prior distribution is α/λ and we have written the posterior mean as a weighted average of the prior mean and the mean of the observed data.

(iii) In this case we have:

$$\hat{\mu} = \frac{10+42}{2+6} = 6.5.$$

(iv) (a) Let *S* be the time taken to resolve a single query. Then for a simple query:

$$P(S > 30 | \text{simple}) = \left(\frac{20}{20+30}\right)^3 = 0.4^3 = 0.064.$$

For a complicated query we have

$$P(S > 30 | \text{complicated}) = e^{-0.4 \times 30^{0.5}} = 0.111817.$$

And finally

$$P(S > 30) = 0.75 \times 0.064 + 0.25 \times 0.111817 = 0.07595.$$

(b) Mean time for simple calls is $\dfrac{20}{3-1} = 10$.

Mean time for complicated calls is

$$\Gamma\left(1 + \frac{1}{0.5}\right) \times 0.4^{-1/0.5} = \Gamma(3) \times 0.4^{-2} = 2 \times 0.4^{-2} = 12.5.$$

Overall mean is $0.75 \times 10 + 0.25 \times 12.5 = 10.625$.

(c) Overall total time is $6.5 \times 10.625 = 69.0625$.

(v) (a) The parameter of the exponential distribution is $\dfrac{1}{10.625} = 0.094117647$.

(b) The probability of taking more than 30 minutes using this approximation is $P(S > 30) = e^{-\frac{30}{10.625}} = 0.059396$.

(c) This compares to the true value of 0.07595. The exponential distribution underestimates this tail probability since it has less fat tails than the Pareto and Weibull distributions.

(vi) We first find the parameter β of the exponential distribution being used. This is given by:

$$P(S > 30) = e^{-30\beta} = 0.1$$

so

$$\beta = \frac{\log 0.1}{-30} = 0.076752836$$

and the mean of the exponential distribution is 13.0288.

The mean of the given Weibull distribution is $\Gamma\left(1+\frac{1}{0.5}\right)\times c^{-\frac{1}{0.5}} = 2c^{-2}$.

The overall mean is then given by $0.75\times 10 + 0.25\times 2c^{-2} = 7.5 + 0.5c^{-2}$.

Equating this to 13.0288 gives:

$$7.5 + 0.5c^{-2} = 13.0288$$

$$c^{-2} = 11.0576$$

$$c = 0.300725.$$

Parts (i) to (iv) were well answered. Parts (v) and (vi) were attempted by only the more able candidates.

END OF EXAMINERS' REPORT

INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

28 April 2014 (pm)

Subject CT6 – Statistical Methods
Core Technical

Time allowed: Three hours

INSTRUCTIONS TO THE CANDIDATE

1. *Enter all the candidate and examination details as requested on the front of your answer booklet.*

2. *You must not start writing your answers in the booklet until instructed to do so by the supervisor.*

3. *Mark allocations are shown in brackets.*

4. *Attempt all 12 questions, beginning your answer to each question on a new page.*

5. *Candidates should show calculations where this is appropriate.*

Graph paper is NOT required for this paper.

AT THE END OF THE EXAMINATION

Hand in BOTH your answer booklet, with any additional sheets firmly attached, and this question paper.

In addition to this paper you should have available the 2002 edition of the Formulae and Tables and your own electronic calculator from the approved list.

1 (i) List six of the characteristics that insurable risks usually have. [3]

(ii) List two key characteristics of a short term insurance contract. [1]
[Total 4]

2 Ruth takes the bus to school every morning. The bus company's ticket machine is unreliable and the amount Ruth is charged every morning can be regarded as a random variable with mean 2 and non-zero standard deviation. The bus company does offer a "value ticket" which gives a 50% discount in return for a weekly payment of 5 in advance. There are 5 days in a week and Ruth walks home each day.

Ruth's mother is worried about Ruth not having enough money to pay for her ticket and is considering three approaches to paying for bus fares:

A Give Ruth 10 at the start of each week.

B Give Ruth 2 at the start of each day.

C Buy the 50% discount card at the start of the week and then give Ruth 1 at the start of each day.

Determine the approach that will give the lowest probability of Ruth running out of money. [4]

3 The table below shows the payoff to a player from a decision problem with three uncertain states of nature θ_1, θ_2 and θ_3 and four possible decisions D_1, D_2, D_3 and D_4.

	D_1	D_2	D_3	D_4
θ_1	10	3	−7	9
θ_2	−5	12	6	−7
θ_3	−8	−3	13	−10

(i) Determine whether any of the decisions are dominated. [2]

(ii) Determine the optimal decision using the minimax criteria. [2]

Now suppose $P(\theta_1) = 0.5$ and $P(\theta_2) = 0.3$ and $P(\theta_3) = 0.2$.

(iii) Determine the optimal decision under the Bayes criterion. [2]
[Total 6]

4 Individual claim amounts on a portfolio of motor insurance policies follow a Gamma distribution with parameters α and λ. It is known that $\lambda = 3$ for all drivers, but the parameter α varies across the population. 70% of drivers have $\alpha = 300$ and the remaining 30% have $\alpha = 600$.

Claims on the portfolio follow a Poisson process with annual rate 500 and the likelihood of a claim arising is independent of the parameter α.

Calculate the mean and variance of aggregate annual claims on the portfolio. [6]

5 A particular portfolio of insurance policies gives rise to aggregate claims which follow a Poisson process with parameter $\lambda = 25$. The distribution of individual claim amounts is as follows:

Claim	50	100	200
Probability	30%	50%	20%

The insurer initially has a surplus of 240. Premiums are paid annually in advance.

Calculate approximately the smallest premium loading such that the probability of ruin in the first year is less than 10%. [7]

6 Claim amounts arising from a certain type of insurance policy are believed to follow a Lognormal distribution. One thousand claims are observed and the following summary statistics are prepared:

mean claim amount	230
standard deviation	110
lower quartile	80
upper quartile	510

(i) Fit a Lognormal distribution to these claims using:

(a) the method of moments.
(b) the method of percentiles.

[6]

(ii) Compare the fitted distributions from part (i). [2]
[Total 8]

7 The heights of adult males in a certain population are Normally distributed with unknown mean μ and standard deviation $\sigma = 15$.

Prior beliefs about μ are described by a Normal distribution with mean 187 and standard deviation 10.

(i) Calculate the prior probability that μ is greater than 180. [2]

A sample of 80 men is taken and the mean height is found to be 182.

(ii) Calculate the posterior probability that μ is greater than 180. [4]

(iii) Comment on your results from parts (i) and (ii). [2]
[Total 8]

8 (i) (a) Write down the Box-Muller algorithm for generating samples from a standard Normal distribution.

(b) Give an advantage and a disadvantage of the Box-Muller algorithm relative to the Polar method.
[3]

(ii) Extend the algorithm in part (i) to generate samples from a Lognormal distribution with parameters μ and σ^2. [1]

A portfolio of insurance policies contains n independent policies. The probability of a claim on a policy in a given year is p and the probability of more than one claim is zero. Claim amounts follow a Lognormal distribution with parameters μ and σ^2. The insurance company is interested in estimating the probability θ that aggregate claims exceed a certain fixed level M.

(iii) Construct an algorithm to simulate aggregate annual claims from this portfolio. [2]

The insurance company estimates that θ is around 10%.

(iv) Calculate the smallest number of simulations the insurance company should undertake to be able to estimate θ to within 1% with 95% confidence. [2]

The insurance company is considering the impact on θ of entering into a reinsurance arrangement.

(v) Explain whether the insurance company should use the same pseudo random numbers when simulating the impact of reinsurance. [1]
[Total 9]

9 The table below sets out incremental claims data for a portfolio of insurance policies.

Accident year	*Development year* 0	1	2
2011	1,403	535	142
2012	1,718	811	
2013	1,912		

Past and projected future inflation is given by the following index (measured to the mid point of the relevant year).

Year	*Index*
2011	100
2012	107
2013	110
2014	113
2015	117

Estimate the outstanding claims using the inflation adjusted chain ladder technique. [9]

10 For a certain portfolio of insurance policies the number of claims on the i^{th} policy in the j^{th} year of cover is denoted by Y_{ij}. The distribution of Y_{ij} is given by

$$P(Y_{ij} = y) = \theta_{ij}\,(1 - \theta_{ij})^y \qquad y = 0, 1, 2, \ldots$$

where $0 \leq \theta_{ij} \leq 1$ are unknown parameters with $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, l$.

(i) Derive the maximum likelihood estimate of θ_{ij} given the single observed data point y_{ij}. [4]

(ii) Write $P(Y_{ij} = y)$ in exponential family form and specify the parameters. [4]

(iii) Describe the different characteristics of Pearson and deviance residuals. [2]
[Total 10]

11 Let θ denote the proportion of insurance policies in a certain portfolio on which a claim is made. Prior beliefs about θ are described by a Beta distribution with parameters α and β.

Underwriters are able to estimate the mean μ and variance σ^2 of θ.

(i) Express α and β in terms of μ and σ. [3]

A random sample of n policies is taken and it is observed that claims had arisen on d of them.

(ii) (a) Determine the posterior distribution of θ.

(b) Show that the mean of the posterior distribution can be written in the form of a credibility estimate. [5]

(iii) Show that the credibility factor increases as σ increases. [3]

(iv) Comment on the result in part (iii). [1]
[Total 12]

12 A sequence of 100 observations was made from a time series and the following values of the sample auto-covariance function (SACF) were observed:

Lag	*SACF*
1	0.68
2	0.55
3	0.30
4	0.06

The sample mean and variance of the same observations are 1.35 and 0.9 respectively.

(i) Calculate the first two values of the partial correlation function $\hat{\phi}_1$ and $\hat{\phi}_2$. [1]

(ii) Estimate the parameters (including σ^2) of the following models which are to be fitted to the observed data and can be assumed to be stationary.

(a) $Y_t = a_0 + a_1 Y_{t-1} + e_t$

(b) $Y_t = a_0 + a_1 Y_{t-1} + a_2 Y_{t-2} + e_t$

In each case e_t is a white noise process with variance σ^2. [12]

(iii) Explain whether the assumption of stationarity is necessary for the estimation for each of the models in part (ii). [2]

(iv) Explain whether each of the models in part (ii) satisfies the Markov property. [2]
[Total 17]

END OF PAPER

INSTITUTE AND FACULTY OF ACTUARIES

EXAMINERS' REPORT

April 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 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.

D C Bowie
Chairman of the Board of Examiners

June 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, 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 April 2014 paper

The examiners felt that this paper was broadly in line with other recent papers. The quality of solutions was often good, with questions 2 and 8 providing the greatest challenge to most students.

1 (i)

- policyholder has an interest in the risk
- risk is of a financial nature and reasonably qualifiable
- independence of risks
- probability of event is relatively small
- pool large numbers of potentially similar risks
- ultimate limit on liability of insurer
- moral hazards eliminated as far as possible
- claim amount must bear some relationship to financial loss
- sufficient data to reasonably estimate extent of risk / likelihood of occurence

(ii)

- policy lasts for a fixed term
- policy lasts for a relatively short period of time
- policyholder pays a premium
- insurer pays claims that arise during the policy term
- option (but no obligation) to renew policy
- claim does not bring policy to an end

Other sensible points received full credit. This question was generally well answered.

2 A is better than B since Ruth has a capital buffer at the start of the week which can offset later journeys, whereas under B a high fare on Monday causes Ruth to run out of funds.

B and C are the same – the net funds available under C are always exactly ½ of those available under B.

So overall A gives the lowest probability of running out of cash.

Many candidates did not attempt this question which required a qualitative analysis of the situation set out. Those candidates who had a good understanding of the basic principles underlying the material on ruin theory were able to score well.

3 (i) D_1 dominates D_4 since D_1 gives a higher outcome for every state of nature.

D_1 has the best result under θ_1 and so is not dominated.

D_2 has the best result under θ_2 and so is not dominated.

Similarly D_3 has the best result under θ_3 and so is not dominated.

So only D_4 is dominated (by D_1).

(ii) The maximum losses are

D_1	–8
D_2	–3
D_3	–7
D_4	–10

The highest of these is –3 under D_2. Therefore D_2 is the optimal strategy under the minimax criterion

(iii) $E(D_1) = 10 \times 0.5 - 5 \times 0.3 - 8 \times 0.2 = 1.9$
$E(D_2) = 3 \times 0.5 + 12 \times 0.3 - 3 \times 0.2 = 4.5$
$E(D_3) = -7 \times 0.5 + 6 \times 0.3 + 13 \times 0.2 = 0.9$

So D_2 is the optimal decision under the Bayes criterion.

This question was well answered.

4 $E(X_i) = E(E(X_i|\alpha))$

$$= E\left(\frac{\alpha}{\lambda}\right) = \frac{1}{3}E(\alpha)$$

$$= \frac{1}{3}(0.7 \times 300 + 0.3 \times 600) = \frac{1}{3} \times 390$$

$$= 130$$

$\mathrm{Var}(X_i) = \mathrm{Var}(E(X_i|\alpha)) + E(\mathrm{Var}(X_i|\alpha))$

$$= \mathrm{Var}\left(\frac{\alpha}{\lambda}\right) + E\left(\frac{\alpha}{\lambda^2}\right)$$

$$= \frac{1}{\lambda^2}\mathrm{Var}(\alpha) + \frac{1}{\lambda^2}E(\alpha)$$

$$= \frac{1}{9}(0.7 \times 300^2 + 0.3 \times 600^2 - 390^2) + \frac{1}{9} \times 390$$

$$= 2100 + \frac{390}{9} = 2143.33$$

so overall

$$E(S) = \lambda E(X) = 500 \times 130 = 65{,}000$$
$$\text{Var}(S) = \lambda E(X^2) = 500 \times (2143.33 + 130^2)$$
$$= 9{,}521{,}665$$

There are other approaches which can be taken to calculating the variance, all of which were given full credit. Whilst most candidates were able to calculate the mean only the better candidates were able to accurately calculate the variance.

5 Mean claim is $50 \times 0.3 + 100 \times 0.5 + 200 \times 0.2$

$$= 15 + 50 + 40$$

$$= 105$$

Also $E(X^2) = 50^2 \times 0.3 + 100^2 \times 0.5 + 200^2 \times 0.2$

$$= 13{,}750$$

so over 1 year the mean aggregate claim amount is

$$25 \times 105 = 2625$$

and the variance of aggregate claims is

$$25 \times 13{,}750 = 586.30^2$$

Using a Normal approximation we need to find θ such that

$$P(N(2625, 586.3^2) > 240 + 25 \times 105 \times (1 + \theta)) = 0.1$$

i.e. $P(N(2625, 586.3^2) > 240 + 2625(1 + \theta)) = 0.1$

i.e. $$P\left(N(0,1) > \frac{240 + 2625\theta}{586.3}\right) = 0.1$$

so $$\frac{240 + 2625\theta}{586.3} = 1.2816$$

i.e. $$\theta = \frac{1.2816 \times 586.3 - 240}{2625}$$

$$= 0.1948$$

This question was well answered with many candidates scoring well.

6 (i) (a) Let the parameters of the Lognormal distribution be μ and σ.

Then we must solve

$$e^{\mu+\frac{\sigma^2}{2}} = 230 \qquad \text{(A)}$$

$$e^{2\mu+\sigma^2}(e^{\sigma^2}-1) = 110^2 \qquad \text{(B)}$$

$$\text{(B)} \div \text{(A)}^2 \Rightarrow e^{\sigma^2}-1 = \frac{110^2}{230^2}$$

so $$e^{\sigma^2} = 1+\frac{110^2}{230^2} = 1.22873$$

so $\sigma^2 = \log 1.22873 = 0.205984$

so $\sigma = 0.45385$

Substituting into (A) gives

$$e^{\mu+\frac{0.205984}{2}} = 230$$

$$\mu = \log(230) - \frac{0.205984}{2}$$

$$= 5.3351$$

(b) This time we have

$$e^{\mu+0.6745\sigma} = 510 \qquad \text{(A)}$$

$$e^{\mu-0.6745\sigma} = 80 \qquad \text{(B)}$$

$\log A + \log B \Rightarrow 2\mu = \log 510 + \log 80$

so $\mu = 5.30822$

and substituting into (A)

$$5.30822 + 0.6745\sigma = \log 510$$

$$\sigma = \frac{\log 510 - 5.30822}{0.6745} = 1.37315$$

(ii) Calculating the upper and lower quartiles from the parameter in (i)(a) gives

$$UQ = e^{5.3351+0.6745\times0.45385} = 282 \qquad \text{cf } 510$$

$$LQ = e^{5.3351-0.6745\times0.45385} = 153 \qquad \text{cf } 80$$

This is not a good fit, suggesting the underlying claims have greater weight in the tails than a Lognormal distribution.

Most candidates were able to apply the method of moments in part (i) but many struggled to apply the method of percentiles. In particular, it was clear that many candidates could not relate the lognormal distribution back to the underlying normal distribution (this was also a common issue in Q8). Alternative comments on the data were given credit in part (ii).

7 (i) $P(\mu > 180) = P(N(187, 10^2) > 180)$

$$= P\left(N(0,1) > \frac{180-187}{10}\right)$$

$$= P(N(0,1) > -0.7)$$

$$= 0.75804$$

(ii) We know that $\mu|x \sim N(\mu_*, \sigma_*^2)$

$$\text{Where } \mu_* = \left(\frac{80\times182}{15^2}+\frac{187}{10^2}\right)\Big/\left(\frac{80}{15^2}+\frac{1}{10^2}\right) = 182.14$$

$$\text{And } \sigma_*^2 = \frac{1}{\frac{80}{15^2}+\frac{1}{10^2}} = 2.73556 = 1.6540^2$$

so $P(\mu > 180) = P(N(182.14, 1.654^2) > 180)$

$$= P\left(N(0,1) > \frac{180-182.14}{1.654}\right)$$

$$= P(N(0,1) > -1.29192)$$

$$= 0.38 \times 0.9032 + 0.62 \times 0.90147$$

$$= 0.90180$$

(iii) The probability has risen, reflecting our much greater certainty over the value of μ as a result of taking a large sample.

This is despite the fact that our mean belief about μ has <u>fallen</u>, which a priori might make a lower value of μ more likely.

The posterior distribution has thinner tails / lower volatility, since we have increased credibility around the mean

This question was mostly well answered. A small number of candidates were not aware that the formulae for the Normal / Normal model are given in the tables, and therefore struggled with the algebra required to derive the posterior distribution.

8 (i) (a) Let u_1 and u_2 be independent samples from a $U(0,1)$ distribution.

Then

$$Z_1 = \sqrt{-2\log u_1}\cos(2\pi u_2)$$
$$Z_2 = \sqrt{-2\log u_1}\sin(2\pi u_2)$$

are independent standard normal variables.

(b) Advantage – generates a sample of every pair of u_1 and u_2 – no possibility of rejection.

Disadvantage – requires calculation of sin and cos functions which is more computationally intensive.

(ii) Generate Z as in (i). Then

$$Y = \exp(\mu + \sigma Z)$$

is a sample from the required Lognormal distribution.

(iii) Set $X = 0$, $k = 0$

Step 1 generate a sample u from $U(0,1)$, set $k = k + 1$

Step 2 If $u \le p$ then go to step 3 else go to step 4

Step 3 Generate a sample Y from the Lognormal distribution in (ii) and set $X = X + Y$

Step 4 If $k = n$ finish else go to step 1

X represents aggregate claims on the portfolio.

(iv) The standard error will be approximately $\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} = \sqrt{\frac{0.09}{n}}$.

We want $\sqrt{\frac{0.09}{n}} \times 1.96 < 0.01$

i.e. $\sqrt{n} > \frac{\sqrt{0.09} \times 1.96}{0.01} = 58.8$

i.e. $n > 3457.44$

so 3,458 simulations are needed.

(v) The insurer should use the same pseudo-random numbers so that any variation in simulation results is due to the impact of the reinsurance and not just due to random variation in the simulation process.

Parts (i) and (v) were well answered. The remaining parts were found by many candidates to be the hardest questions on the paper. In part (ii) many candidates could not relate the Lognormal distribution to the Normal distribution from which samples had been generated in (i). Only the best candidates attempted parts (iii) and (iv).

9 Incremental claims in mid 2013 prices are given by:

Accident year	*Development year* 0	1	2
2011	1543.3	550	142
2012	1766.17	811	
2013	1912		

Cumulative claims in mid 2013 prices:

Accident year	*Development year* 0	1	2
2011	1543.3	2093.3	2235.3
2012	1766.17	2577.17	
2013	1912		

$DF_{0,1}$ = (2093.3 + 2577.17) / (1543.3 + 1766.17) = 1.4112441

$DF_{1,2}$ = 2235.3 / 2093.3 = 1.067835

Completed cumulative claims

Accident year	*Development year* 0	1	2
2011			
2012			2751.99
2013		2698.30	2881.34

Incremental claims (mid 2013 prices)

Accident year	*Development year* 0	1	2
2011			
2012			174.82
2013		786.30	183.04

And projecting for inflation, outstanding claims = $(174.82 + 786.30) \times \frac{113}{110}$

$$+ 183.04 \times \frac{117}{110} = 1182.02$$

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

10 (i) The likelihood is given by

$$L = \theta_{ij}(1-\theta_{ij})^{y_{ij}}$$

Taking logs gives

$$l = \log L = \log\theta_{ij} + y_{ij}\log(1-\theta_{ij})$$

Differentiating with respect to θ_{ij} gives

$$\frac{\partial l}{\partial\theta_{ij}} = \frac{1}{\theta_{ij}} - \frac{y_{ij}}{(1-\theta_{ij})}$$

and setting $\frac{\partial l}{\partial\theta_{ij}} = 0$ we have

$$\frac{1}{\hat{\theta}_{ij}} = \frac{y_{ij}}{1-\hat{\theta}_{ij}}$$

so $\quad 1-\hat{\theta}_{ij} = y_{ij}\hat{\theta}_{ij}$

so $\quad 1 = (1+y_{ij})\hat{\theta}_{ij}$

i.e. $\quad \hat{\theta}_{ij} = \dfrac{1}{1+y_{ij}}$

and since $\dfrac{\partial^2 l}{\partial\theta_{ij}^2} = -\dfrac{1}{\theta_{ij}^2} - \dfrac{y_{ij}}{(1-\theta_{ij})^2} < 0$ (since $y_{ij} > 0$)

we do have a maximum.

(ii) $P(Y_{ij} = y) = \theta_{ij}(1-\theta_{ij})^y$

$= \exp[\log\theta_{ij} + y\log(1-\theta_{ij})]$

$= \exp[y\log(1-\theta_{ij})] + \log\theta_{ij}]$

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

where $\theta = \log(1-\theta_{ij})$ is the natural parameter

$b(\theta) = -\log\theta_{ij} = -\log[1-e^{\theta}]$

$\varphi = 1 \qquad a(\varphi) = 1$

$c(y, \varphi) = 0$

(iii) The Pearson residuals are often skewed for non normal data which makes the interpretation of residual plots difficult.

Deviance residuals are usually more likely to be symmetrically distributed and are preferred for actuarial applications.

This question was, for the most part, answered well. A common mistake in part (i) was to try to sum across either years or policies when the question specifically referred to a single data point.

11 (i) We have $\mu = \dfrac{\alpha}{\alpha+\beta}$ so $\beta = \dfrac{\alpha(1-\mu)}{\mu}$

and $\sigma^2 = \dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \mu \times \dfrac{\beta}{(\alpha+\beta)(\alpha+\beta+1)} = \dfrac{\mu(1-\mu)}{\alpha+\beta+1}$

substituting for $\alpha + \beta = \frac{\alpha}{\mu}$ gives

$$\sigma^2 = \frac{\mu(1-\mu)}{\frac{\alpha}{\mu}+1}$$

so $\sigma^2 = \dfrac{\mu^2(1-\mu)}{\alpha+\mu}$

so $\alpha = \dfrac{\mu^2(1-\mu)-\mu\sigma^2}{\sigma^2}$

and $\beta = \dfrac{(\mu^2(1-\mu)-\mu\sigma^2)(1-\mu)}{\mu\sigma^2} = \dfrac{(\mu(1-\mu)-\sigma^2)(1-\mu)}{\sigma^2}$

(ii) (a) $f(\theta|\underline{x}) \;\alpha\; f(\underline{x}|\theta)\, f(\theta)$

$\alpha\; \theta^d(1-\theta)^{n-d}\,\theta^{\alpha-1}(1-\theta)^{\beta-1}$

$\alpha\; \theta^{\alpha+d-1}(1-\theta)^{n-d+\beta-1}$

which is the pdf of a Beta distribution with parameters $\alpha + d$ and $\beta + n - d$.

(b) The posterior mean is given by

$$\frac{\alpha+d}{\alpha+d+\beta+n-d} = \frac{\alpha+d}{\alpha+\beta+n}$$

$$= \frac{\alpha}{\alpha+\beta}\times\frac{\alpha+\beta}{\alpha+\beta+n}+\frac{d}{n}\times\frac{n}{\alpha+\beta+n}$$

$$= \frac{\alpha}{\alpha+\beta}\times(1-Z)+\frac{d}{n}Z$$

where $Z = \dfrac{n}{\alpha+\beta+n}$.

This is in the form of a credibility estimate since $\frac{\alpha}{\alpha+\beta}$ is the mean of the prior distribution and $\frac{d}{n}$ is the MLE.

(iii) $Z = \frac{n}{\alpha+\beta+n}$ decreases as $\alpha + \beta$ increases and increases as $\alpha + \beta$ decreases.

From (i) $\alpha + \beta = \frac{\mu^2(1-\mu)-\mu\sigma^2+\mu(1-\mu)^2-(1-\mu)\sigma^2}{\sigma^2}$

$$= \frac{\mu(1-\mu)(\mu+1-\mu)-\mu\sigma^2-\sigma^2+\mu\sigma^2}{\sigma^2}$$

$$= \frac{\mu(1-\mu)}{\sigma^2}-1$$

so increasing σ^2 decreases $\alpha + \beta$ and increases Z.

(iv) Higher σ^2 implies less certainty in the prior estimate / prior is less reliable and so should lead to more weight on the observed data – which it does via a higher Z.

Only the best candidates were able to complete the algebra in part (i). Many candidates nevertheless scored well on parts (ii) and (iv).

12 (i) $\hat{\phi}_1 = \hat{\rho}_1 = \frac{0.68}{0.9} = 0.755556$

$$\hat{\phi}_2 = \frac{\hat{\rho}_2 - \hat{\rho}_1^2}{1-\hat{\rho}_1^2} = \frac{\frac{0.55}{0.9} - 0.755556^2}{1-0.755556^2} = 0.093786$$

(ii) (a) 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 $\hat{a}_1 = \frac{\hat{\gamma}_1}{\hat{\gamma}_0} = \hat{\rho}_1 = 0.755556$

and $\hat{\sigma}^2 = \hat{\gamma}_0 - \hat{a}_1\hat{\gamma}_1 = \hat{\gamma}_0(1-\hat{a}_1\hat{\rho}_1)$

$= 0.9 - 0.755556 \times 0.68 = 0.38622$

Finally we let $\mu = E(Y_t)$ and observe that

$\mu = a_0 + a_1\mu$

so $\mu = \dfrac{a_0}{1-a_1}$

so $\hat{a}_0 = \hat{\mu}(1-\hat{a}_1) = 1.35(1-0.755556) = 0.33000$

(b) For this model the Yule-Walker equations are

$\gamma_0 = a_1\gamma_1 + a_2\gamma_2 + \sigma^2$ (1)
$\gamma_1 = a_1\gamma_0 + a_2\gamma_1$ (2)
$\gamma_2 = a_1\gamma_1 + a_2\gamma_0$ (3)

substituting the observed values in (2) and (3) gives

$0.68 = 0.9\hat{a}_1 + 0.68\hat{a}_2$ (4)
$0.55 = 0.68\hat{a}_1 + 0.9\hat{a}_2$ (5)

(4) $\times 0.68$ – (5) $\times 0.9 \Rightarrow 0.68^2 - 0.55\times 0.9 = \hat{a}_2\left(0.68^2 - 0.9^2\right)$

So $\hat{a}_2 = \dfrac{-0.0326}{-0.3476} = 0.09379$

And $\hat{a}_1 = \dfrac{0.68(1-0.09378596)}{0.9} = 0.68470$

substituting into (1)

$\hat{\sigma}^2 = 0.9 - 0.68470 \times 0.68 - 0.09379 \times 0.55 = 0.38283$

and finally setting $\mu = E(Y_t)$ we have $\mu = a_0 + a_1\mu + a_2\mu$

so $\mu = \dfrac{a_0}{1-a_1-a_2}$

so $\hat{a}_0 = \hat{\mu}\times(1-\hat{a}_1-\hat{a}_2) = 1.35(1-0.68470-0.09379) = 0.29905$

(iii) Stationarity is necessary for both models since the Yule-Walker equations do not hold without the existence of the auto-covariance function.

(iv) Model (a) does satisfy the Markov property since the current value depends only on the previous value.

This does not hold for Model (b).

Most candidates were able to derive the Yule-Walker equations and therefore scored marks on this question. Only the best candidates were able to use these equations to derive numerical values of the parameters. Part (iv) was generally well answered.

Although the question stated that the given values were for the auto-covariance function, many candidates calculated as if the given values came from the auto-correlation function. The Examiners noted that the core reading does use the abbreviation ACF for the auto-correlation function, and therefore gave full credit to candidates who interpreted the question in this way. The numerical values of the estimated parameters taking this approach are as follows:

(i) $\hat{\phi}_1 = 0.68$

$\hat{\phi}_2 = 0.1629$

(ii) (a) $\hat{a}_0 = 0.432$

$\hat{a}_1 = 0.68$

$\hat{\sigma}^2 = 0.48384$

(b) $\hat{a}_0 = 0.3617$

$\hat{a}_1 = 0.5692$

$\hat{a}_2 = 0.1629$

$\hat{\sigma}^2 = 0.4710$

END OF EXAMINERS' REPORT

INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

25 September 2014 (am)

Subject CT6 – Statistical Methods
Core Technical

Time allowed: Three hours

INSTRUCTIONS TO THE CANDIDATE

1. *Enter all the candidate and examination details as requested on the front of your answer booklet.*

2. *You must not start writing your answers in the booklet until instructed to do so by the supervisor.*

3. *Mark allocations are shown in brackets.*

4. *Attempt all nine questions, beginning your answer to each question on a new page.*

5. *Candidates should show calculations where this is appropriate.*

Graph paper is NOT required for this paper.

AT THE END OF THE EXAMINATION

Hand in BOTH your answer booklet, with any additional sheets firmly attached, and this question paper.

In addition to this paper you should have available the 2002 edition of the Formulae and Tables and your own electronic calculator from the approved list.

1 An insurance company has a portfolio of 240 insurance policies. The probability of a claim on the i^{th} policy in a year is p_i independently from policy to policy and there is no possibility of more than one claim. Claim amounts on the i^{th} policy follow an exponential distribution with mean $100/p_i$.

Let X denote the aggregate annual claims on the portfolio.

Determine the mean and variance of X. [6]

2 (i) List the three main components of a generalised linear model. [3]

(ii) Explain what is meant by a saturated model and discuss whether such a model is useful in practice. [3]

[Total 6]

3 Sara is a car mechanic for a racing team. She knows that there is a problem with the car, but is unsure whether the fault is with the gearbox or the engine. Sara is able to observe one practice race.

If the underlying problem is with the gearbox there is a 40% chance the car will not complete the practice race. If the underlying problem is with the engine there is a 90% chance the car will not complete the practice race.

At the end of the practice race Sara must decide, on the basis of whether the car completes the practice race, whether the fault lies with the gearbox or the engine.

(i) Write down the four decision functions Sara could adopt. [2]

If Sara correctly identifies the fault there is no cost. The cost of incorrectly deciding the fault is with the gearbox is £1m. The cost of incorrectly deciding the fault is with the engine is £5m.

(ii) Show that one of the decision functions is dominated. [3]

The probability that the fault lies with the gearbox is p.

(iii) Determine the range of values of p for which Sara will, under the Bayes criterion, choose a decision function whose outcome is affected by whether or not the car completes the practice race. [4]

[Total 9]

4 As part of a simulation study an actuary is asked to design an algorithm for simulating claims from a particular type of insurance policy. The probability distribution of the annual number of claims on a policy is given by:

	No claims	*One claim*	*Two claims*
Probability	0.4	0.4	0.2

The claim size distributions of the first and second claims are different. The size of the first claim follows an exponential distribution with mean 10. The size of the second claim follows a Weibull distribution with parameters $c = 1$ and $\gamma = 4$.

(i) Construct an algorithm to simulate the first claim on a given policy. [3]

(ii) Construct an algorithm to simulate the second claim on a given policy. [3]

(iii) Construct an algorithm to simulate the total annual claims on a given policy. [4]

[Total 10]

5 The table below shows the incremental claims incurred for a certain portfolio of insurance policies.

Accident year	*Development year* 0	1	2
2011	2,233	1,389	600
2012	3,380	1,808	
2013	4,996		

Cumulative numbers of claims are shown in the following table:

Accident year	*Development year* 0	1	2
2011	140	203	224
2012	180	230	
2013	256		

(i) Calculate the outstanding claim reserve for this portfolio using the average cost per claim method with grossing up factors. [7]

(ii) State the assumptions underlying the calculations in part (i). [3]

[Total 10]

6 For three years an insurance company has insured buildings in three different towns against the risk of fire damage. Aggregate claims in the j^{th} year from the i^{th} town are denoted by X_{ij} for i = 1, 2, 3 and j = 1, 2, 3. The data is given in the table below.

Town i	*Year j*		
	1	*2*	*3*
1	8,130	9,210	8,870
2	7,420	6,980	8,130
3	9,070	8,550	7,730

Calculate the expected claims from each town for the next year using the assumptions of Empirical Bayes Credibility Theory model 1. [10]

7 The random variable X follows a Pareto distribution with parameters α and λ.

(i) Show that for L, $d > 0$

$$\int_d^{L+d} xf(x)dx = \frac{\lambda^\alpha}{\alpha-1}\left[\frac{d\alpha+\lambda}{(\lambda+d)^\alpha} - \frac{\alpha(L+d)+\lambda}{(\lambda+L+d)^\alpha}\right]$$ [5]

Claims on a certain type of motor insurance policy follow a Pareto distribution with mean 16,000 and standard deviation 20,000. The insurance company has an excess of loss reinsurance policy with a retention level of 40,000 and a maximum amount paid by the reinsurer of 25,000.

(ii) Determine the mean claim amount paid by the reinsurer on claims that involve the reinsurer. [8]

Claim amounts increase by 5%.

(iii) State the new distribution of claim amounts. [1]
[Total 14]

8 Claims on a portfolio of insurance policies follow a Poisson process with rate λ. Individual claim amounts follow a distribution X with mean μ and variance σ^2. The insurance company charges premiums of c per policy per year.

(i) Write down the equation satisfied by the adjustment coefficient R. [1]

(ii) Show that R can be approximated by

$$\hat{R} = \frac{2(c-\mu)}{\sigma^2+\mu^2}$$ [4]

Now suppose that individual claims follow a distribution given by

Value	10	20	50	100
Probability	0.3	0.5	0.15	0.05

The insurance company uses a premium loading of 30%. It is considering the following reinsurance arrangements:

A No reinsurance.

B Proportional reinsurance where the insurer retains 70% of all claims with a reinsurer using a 20% premium loading.

C Excess of loss reinsurance with retention 70 with a reinsurer using a 40% premium loading.

(iii) Determine which arrangement gives the insurance company the lowest probability of ultimate ruin, using the approximation in part (ii) [10]

(iv) Comment on your result in part (iii). [2]
[Total 17]

 PLEASE TURN OVER

9 (i) List the main steps in the Box-Jenkins approach to fitting an ARIMA time series to observed data. [3]

Observations $x_1, x_2, \ldots, x_{200}$ are made from a stationary time series and the following summary statistics are calculated:

$$\sum_{i=1}^{200} x_i = 83.7 \qquad \sum_{i=1}^{200} (x_i - \bar{x})^2 = 35.4 \qquad \sum_{i=2}^{200} (x_i - \bar{x})(x_{i-1} - \bar{x}) = 28.4$$

$$\sum_{i=3}^{200} (x_i - \bar{x})(x_{i-2} - \bar{x}) = 17.1$$

(ii) Calculate the values of the sample auto-covariances $\hat{\gamma}_0, \hat{\gamma}_1$ and $\hat{\gamma}_2$. [3]

(iii) Calculate the first two values of the partial correlation function $\hat{\phi}_1$ and $\hat{\phi}_2$. [3]

The following model is proposed to fit the observed data:

$$X_t - \mu = a_1 (X_{t-1} - \mu) + e_t$$

where e_t is a white noise process with variance σ^2.

(iv) Estimate the parameters μ, a_1 and σ^2 in the proposed model. [5]

After fitting the model in part (iv) the 200 observed residual values $\hat{e}_t$ were calculated. The number of turning points in the residual series was 110.

(v) Carry out a statistical test at the 95% significance level to test the hypothesis that $\hat{e}_t$ is generated from a white noise process. [4]

[Total 18]

END OF PAPER

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

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

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

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

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

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

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

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

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

[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 $s_2^2 = 673{,}400$

$s_3^2 = 912{,}800$

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

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

$= 411{,}749.83 – 366{,}011.11 / 3$

$= 289{,}746.13$

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

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

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

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

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

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

$$\frac{72888.89^{5.5555}}{4.5555}\left\{\frac{40000\times 5.5555 + 72888.89}{112888.89^{5.5555}}\right.$$

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

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

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

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

$= 2941.71 - 2361.67 + 724.10$

$= 1304.14$

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* $\hat{R} = \frac{2(c-\lambda\mu)}{\lambda\left(\mu^2+\sigma^2\right)}$.

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

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 $\hat{\mu} = \bar{x} = \dfrac{83.7}{200} = 0.4185$.

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