EXAMINATION

14 April 2005 (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.

1 List the three main perils typically covered by employer's liability insurance. [3]

2 An insurer wishes to estimate the expected number of claims, λ, on a particular type of policy. Prior beliefs about λ are represented by a Gamma distribution with density function

$$f(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \quad (\lambda > 0).$$

For an estimate, d, of λ the loss function is defined as

$$L(\lambda, d) = (\lambda - d)^2 + d^2.$$

Show that the expected loss is given by

$$E(L(\lambda, d)) = \frac{\alpha(\alpha+1)}{\beta^2} - \frac{2d\alpha}{\beta} + 2d^2$$

and hence determine the optimal estimate for λ under the Bayes rule. [5]

3 (i) Explain the disadvantages of using truly random, as opposed to pseudo-random, numbers. [3]

(ii) List four methods for the generation of random variates. [2]
[Total 5]

4 Y_t, $t = 1, 2, 3, \ldots$ is a time series defined by

$$Y_t - 0.8Y_{t-1} = Z_t + 0.2Z_{t-1}$$

where Z_t, $t = 0, 1, \ldots$ is a sequence of independent zero-mean variables with common variance σ^2.

Derive the autocorrelation ρ_k, $k = 0, 1, 2, \ldots$ [6]

5 An insurer believes that claim amounts, X, on its portfolio of pet insurance policies follow an exponential distribution with mean £200.

A reinsurance policy is arranged such that the reinsurer pays X_R, where

$$X_R = \begin{cases} 0 & \text{if } X \le £50 \\ X - 50 & \text{if } £50 < X \le M \\ M - 50 & \text{if } X > M \end{cases}$$

calculate M such that $E[X_R]$ = £100. [8]

6 On 1 January 2001 an insurer in a far off land sells 100 policies, each with a five year term, to householders wishing to insure against damage caused by fireworks. The insurer charges annual premiums of £600 payable continuously over the life of the policy.

The insurer knows that the only likely date a claim will be made is on the day of St Ignitius' feast on 1 August each year, when it is traditional to have an enormous fireworks display. The annual probability of a claim on each policy is 40%. Claim amounts follow a Pareto distribution with parameters $\alpha = 10$ and $\lambda = 9{,}000$.

(i) Calculate the mean and standard deviation of the annual aggregate claims. [4]

(ii) Denote by $\psi(U, t)$ the probability of ruin before time t given initial surplus U.

(a) Explain why for this portfolio $\psi(U, t_1) = \psi(U, t_2)$ if $7/12 < t_1, t_2 < 19/12$. [1]

(b) Estimate $\psi(15{,}000, 1)$ assuming annual claims are approximately Normally distributed. [4]

[Total 9]

7 The no claims discount (NCD) system operated by an insurance company has three levels of discount: 0%, 25% and 50%.

If a policyholder makes a claim they remain at or move down to the 0% discount level for two years. Otherwise they move up a discount level in the following year or remain at the maximum 50% level.

The probability of an accident depends on the discount level:

Discount Level	*Probability of accident*
0%	0.25
25%	0.2
50%	0.1

The full premium payable at the 0% discount level is 750.

Losses are assumed to follow a lognormal distribution with mean 1,451 and standard deviation 604.4.

Policyholders will only claim if the loss is greater than the total additional premiums that would have to be paid over the next three years, assuming that no further accidents occur.

(i) Calculate the smallest loss for which a claim will be made for each of the four states in the NCD system. [2]

(ii) Determine the transition matrix for this NCD system. [6]

(iii) Calculate the proportion of policyholders at each discount level when the system reaches a stable state. [3]

(iv) Determine the average premium paid once the system reaches a stable state. [1]

(v) Describe the limitations of simple NCD systems such as this one. [2]

[Total 14]

8 (i) Write down the general form of a statistical model for a claims run-off triangle, defining all terms used. [5]

(ii) The table below shows the cumulative incurred claims on a portfolio of insurance policies.

Accident Year	*Delay Year*		
2000	2,748	3,819	3,991
2001	2,581	4,014	
2002	3,217		

The company decides to apply the Bornhuetter-Ferguson method to calculate the reserves, with the assumption that the Ultimate Loss Ratio is 85%.

Calculate the reserve for 2002, if the earned premium is 5,012 and the paid claims are 1,472. [9]

[Total 14]

9 $Y_1, Y_2, \ldots, Y_n$ are independent claims, which are assumed to be exponentially distributed, with

$$E[Y_i] = \mu_i \,.$$

(i) Show that the canonical link function is the inverse link function. [3]

(ii) It is decided that the canonical link function should not be used, but that the mean claim sizes should be modelled as follows:

$$\log\mu_i = \begin{cases} \alpha & i = 1, 2, \ldots, m \\ \beta & i = m+1, m+2, \ldots, n \end{cases}$$

(a) Show that the log-likelihood can be written as

$$-\left[m\alpha + (n-m)\beta + e^{-\alpha}\sum_{i=1}^{m} y_i + e^{-\beta}\sum_{i=m+1}^{n} y_i \right]$$

(b) Derive the maximum likelihood estimators of α and β.

(c) Show that the scaled deviance for this model is

$$2\left(\sum_{i=1}^{m} \log\left(\frac{\frac{1}{m}\sum_{j=1}^{m} y_j}{y_i} \right) + \sum_{i=m+1}^{n} \log\left(\frac{\frac{1}{n-m}\sum_{j=m+1}^{n} y_j}{y_i} \right) \right)$$ [12]

(iii) For a particular data set, $m = 20$, $n = 44$,

$$\frac{1}{20}\sum_{i=1}^{20} y_i = 14.2, \frac{1}{24}\sum_{i=21}^{44} y_i = 18.7.$$

Calculate the deviance residual for $y_1 = 7$. [3]

[Total 18]

10 (i) Explain what a conjugate prior distribution is. [2]

(ii) The random variables $X_1, X_2, \ldots, X_n$ are independent and have density function

$$f(x) = \lambda e^{-\lambda x} \quad (x > 0).$$

Show that the conjugate prior distribution for λ is a Gamma distribution. [3]

(iii) (a) The density function of λ is

$$f(\lambda) = \frac{s^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-s\lambda} \ (\lambda > 0).$$

Show that $E(1/\lambda) = s/(\alpha - 1)$.

(b) Hence if $X_1, X_2, \ldots, X_n$ is an independent random sample from an exponential distribution with parameter λ, show that the posterior mean of $1/\lambda$ can be expressed as a weighted average of the prior mean of $1/\lambda$ and the sample average.

[5]

(iv) An insurer is considering introducing a new policy to provide insurance against the failure of toasters within the first five years of purchase. Alan and Beatrice are underwriters working for the insurer. Based on his experience of similar products, Alan believes that toasters last three years on average. Beatrice believes that six years is the average lifetime. Both are adamant and are prepared to express their uncertainties about the average lifetime in terms of standard deviations of six months and one year respectively. They decide to resolve their differences by testing a sample of toasters large enough to ensure the difference in their posterior expectations for the average lifetime will be less than one year.

Calculate how many toasters they should test, assuming the exponential distribution is a good model for toaster lifetimes.

You may use the fact that if $\lambda \sim \Gamma(\alpha, s)$ then

$$\text{Var}(1/\lambda) = [E(1/\lambda)]^2 \times \frac{1}{\alpha - 2}.$$ [8]

[Total 18]

END OF PAPER

EXAMINATION

April 2005

Subject CT6 — Statistical Methods
Core Technical

EXAMINERS' REPORT

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.

M Flaherty
Chairman of the Board of Examiners

15 June 2005

1 The three main perils are:

- accidents caused by the negligence of the employer or other employees
- exposure to harmful substances
- exposure to harmful working conditions

2 The expected loss is given by

$$E[L(\lambda, d)] = E[(\lambda - d)^2 + d^2]$$

$$= E[\lambda^2 - 2\lambda d + d^2 + d^2]$$

$$= E(\lambda^2) - 2dE(\lambda) + 2d^2.$$

Now we know that $\lambda \sim \Gamma(\alpha, \beta)$. From the tables, we know that $E(\lambda) = \alpha/\beta$ and $\text{Var}(\lambda) = \alpha/\beta^2$. Now

$$E(\lambda^2) = \text{Var}(\lambda) + E(\lambda)^2$$

$$= \frac{\alpha}{\beta^2} + \frac{\alpha^2}{\beta^2}$$

$$= \frac{\alpha(\alpha+1)}{\beta^2}.$$

So $$E[L(\lambda, d)] = \frac{\alpha(\alpha+1)}{\beta^2} - \frac{2d\alpha}{\beta} + 2d^2$$

as required.

First set $f(d) = E[(L(\lambda, d)].$

Then $$f'(d) = -\frac{2\alpha}{\beta} + 4d$$

and to minimise the expected loss, we must find the value of d^* of d for which $f'(d^*) = 0$. This occurs when

$$4d^* = \frac{2\alpha}{\beta}$$

so that $$d^* = \frac{\alpha}{2\beta}.$$

We can confirm this is a minimum since $f''(d) = 4 > 0$.

3 (i) The stored table of random numbers generated by a physical process may be too short — a combination of linear congruential generators (LCG) can produce a sequence which is infinite for practical purposes.

It might not be possible to reproduce exactly the same series of random numbers again with a truly random number generator unless these are stored. A LCG will generate the same sequence of numbers with the same seed.

Truly random numbers would require either a lengthy table or hardware enhancement compared with a single routine for pseudo random numbers.

(ii) Inverse Transform method.
Acceptance-Rejection Method
Box-Muller algorithm (from the standard normal distribution)
Polar algorithm (from the standard normal distribution)

4 Let

$$\gamma_k = \text{Cov}(Y_t, Y_{t-k})$$

$$Y_t = 0.8Y_{t-1} + Z_t + 0.2Z_{t-1}$$

$$\text{Cov}(Y_t, Z_t) = \sigma^2$$

$$\text{Cov}(Y_t, Z_{t-1}) = 0.8\text{Cov}(Y_{t-1}, Z_{t-1}) + 0.2\sigma^2$$

$$= 0.8\sigma^2 + 0.2\sigma^2 = \sigma^2$$

$$\gamma_0 = \text{Cov}(Y_t, Y_t) = \text{Cov}(Y_t, 0.8Y_{t-1} + Z_t + 0.2Z_{t-1})$$

$$= 0.8\gamma_1 + \sigma^2 + 0.2\sigma^2$$

$$= 0.8\gamma_1 + 1.2\sigma^2$$

$$\gamma_1 = \text{Cov}(Y_t, Y_{t-1}) = \text{Cov}(0.8Y_{t-1} + Z_t + 0.2Z_{t-1}, Y_{t-1})$$

$$= 0.8\gamma_0 + 0.2\sigma^2$$

$$\therefore \gamma_0 = 0.64\gamma_0 + 0.16\sigma^2 + 1.2\sigma^2$$

$$\therefore 0.36\gamma_0 = 1.36\sigma^2$$

$$\therefore\ \gamma_0 = \frac{1.36}{0.36}\sigma^2 = 3.78\sigma^2$$

$$\therefore \gamma_1 = 3.22\sigma^2$$

For $k \ge 2$, $\gamma_k = \text{Cov}(Y_t, Y_{t-k}) = 0.8\gamma_{k-1}$

$\therefore$ The autocorrelation function is

$$\rho_0 = 1,\ \rho_1 = \frac{3.22}{3.78} = 0.85294,$$

$$\rho_k = 0.8^{k-1}\rho_1\ (k \ge 2)$$

5 We know that M is to be chosen so that

$$\int_{50}^{M}(x-50)\lambda e^{-\lambda x}dx + \int_{M}^{\infty}(M-50)\lambda e^{-\lambda x}dx = 100$$

where $\lambda = \dfrac{1}{200} = 0.005$.

The LHS of the expression above can be written as

$$\int_{50}^{M} x\lambda e^{-\lambda x}dx - 50\int_{50}^{M}\lambda e^{-\lambda x}dx + \int_{M}^{\infty}(M-50)\lambda e^{-\lambda x}dx$$

$$= \left[-xe^{-\lambda x}\right]_{50}^{M} + \int_{50}^{M} e^{-\lambda x}dx - 50\left[-e^{-\lambda x}\right]_{50}^{M} + \left[-(M-50)e^{-\lambda x}\right]_{M}^{\infty}$$

$$= -Me^{-\lambda M} + 50e^{-50\lambda} + \left[-\frac{e^{-\lambda x}}{\lambda}\right]_{50}^{M} + 50e^{-\lambda M} - 50e^{-50\lambda} + (M-50)e^{-\lambda M}$$

$$= -Me^{-\lambda M} - 200e^{-\lambda M} + 200e^{-50\lambda} + 50e^{-\lambda M} + Me^{-\lambda M} - 50e^{-\lambda M}$$

$$= 200e^{-50\lambda} - 200e^{-\lambda M}.$$

So the equation for M becomes

$$100 = 200e^{-0.25} - 200e^{-0.005M}$$

so that

$$e^{-0.005M} = \frac{200e^{-0.25} - 100}{200} = 0.2788$$

and hence

$$M = \frac{-\log 0.2788}{0.005} = 255.45$$

6 (i) The number of annual claims N follows a binomial distribution:
$N \sim B(100, 0.4)$ then

$$E(N) = 100 \times 0.4 = 40$$

and

$$\text{Var}(N) = 100 \times 0.4 \times 0.6 = 24.$$

Let X denote the distribution of the individual claim amounts, so that $X \sim$ Pareto(10, 9,000). Then

$$E(X) = \frac{9,000}{10-1} = 1,000$$

and

$$\text{Var}(X) = \frac{9,000^2 \times 10}{9^2 \times 8} = 1,250,000.$$

The annual aggregate claim amount S has

$$E(S) = E(N)E(X) = 40 \times 1,000 = 40,000$$

and

$$\begin{aligned}\text{Var}(S) &= E(N)\text{Var}(X) + \text{Var}(N)E(X)^2 \\ &= 40 \times 1,250,000 + 24 \times 1,000^2 \\ &= 74,000,000 \\ &= (8,602.33)^2\end{aligned}$$

(ii) (a) Since claims can only fall on one day of the year, there is effectively only one day of the year on which ruin can occur, namely 1 August (or strictly shortly thereafter). For a year after 1 August, the insurer will be receiving premiums but paying no claims, and hence solvency will be improving. Hence

$$\psi(U, t_1) = \psi(U, t_2) \text{ if } 7/12 < t_1, t_2 < 19/12.$$

(b) We must find $\psi(15{,}000, 1)$. But ruin will have occurred before time 1 only if it occurs at $t = 7/12$. Just before the claims occur, the insurers assets will be $7/12 \times 100 \times 600 + 15{,}000 = 50{,}000$ and ruin will occur if the aggregate claims in the first year exceed this level. Assuming that S is approximately normally distributed, we have

$$P(\text{Ruin}) = P(N(40{,}000, (8{,}602.33)^2) > 50{,}000)$$

$$= P\left(N(0,1) > \frac{50{,}000 - 40{,}000}{8{,}602.33}\right)$$

$$= 1 - \Phi(1.162)$$

$$= 0.123.$$

7 (i) Denote:

0	just had a claim
0*	1 claim free year after accident or new customer
1	25%
2	50%

	Premiums if no claim	*Premiums if claim*	*Difference*
0	750, 562.50, 375	750, 750, 562.50	375
0*	562.50, 375, 375	750, 750, 562.50	750
1	375, 375, 375	750, 750, 562.50	937.50
2	375, 375, 375	750, 750, 562.50	937.50

So minimum claim in state 0 is 375, in state 0* is 750 and in states 1 and 2 is 937.50.

(ii) $P(\text{Claim}) = P(\text{Claim} \mid \text{Accident}) \,.\, P(\text{Accident})$
$= P(X > x) \times P(\text{Accident})$

Where X is the loss and x is the minimum loss for which a claim will be made.

$E(x) = \exp(\mu + \tfrac{1}{2}\sigma^2) = 1{,}451$

$\text{Var}(x) = \exp(2(\mu + \tfrac{1}{2}\sigma^2)) \exp((\sigma^2) - 1) = 604.4^2$

Therefore, $\exp(\sigma^2) - 1 = 604.4^2 / 1{,}451^2$

$\exp(\sigma^2) = 1.1735$

$\sigma^2 = 0.16$

$\sigma = 0.4$

$\mu = 7.2$

$$P(X > 375) = 1 - \Phi\left(\frac{\ln 375 - 7.2}{0.4}\right) = 0.99927$$

$$P(X > 750) = 1 - \Phi\left(\frac{\ln 750 - 7.2}{0.4}\right) = 0.9264$$

$$(X > 937.50) = 1 - \Phi\left(\frac{\ln 937.50 - 7.2}{0.4}\right) = 0.8138$$

So the transition matrix is

$$\begin{bmatrix} 0.2498 & 0.7502 & 0 & 0 \\ 0.2316 & 0 & 0.7684 & 0 \\ 0.1628 & 0 & 0 & 0.8372 \\ 0.0814 & 0 & 0 & 0.9186 \end{bmatrix}$$

(iii)

$$
\begin{aligned}
0.2498\pi_0 + 0.2316\pi_0^* + 0.1628\pi_1 + 0.0814\pi_2 &= \pi_0 \\
0.7502\pi_0 &= \pi_0^* \\
0.7684\pi_0^* &= \pi_1 \\
0.8372\pi_1 + 0.9186\pi_2 &= \pi_2 \\
\pi_0 + \pi_0^* + \pi_1 + \pi_2 &= 1 \\
\pi_1 = 0.7502 \times 0.7684 \times \pi_0 &= 0.5766\pi_0 \\
0.8372\pi_1 &= \pi_2(1 - 0.9186) \\
\pi_2 &= 10.2850\pi_1 \\
\pi_0 + 0.7502\pi_0 + 0.5766\pi_0 + 10.2850 \times (0.5766\pi_0) &= 1 \\
8.2556\pi_0 &= 1 \\
\pi_0 &= 0.1211 \\
\pi_0^* &= 0.0909 \\
\pi_1 = 0.7684 \times 0.0909 &= 0.0698 \\
\pi_2 = 1 - \pi_0 - \pi_0^* - \pi_1 &= 0.7182
\end{aligned}
$$

(iv) Average premium across portfolio

$$750 \times (0.1211 + 0.0909 + 0.0698 \times 0.75 + 0.7182 \times 0.5) = £467.59$$

(v) Intention is to automatically premium rate with NCD system. Small number of categories and the relatively low discount result in high proportion of policyholders in maximum discount category. Many more categories and higher discount levels would be required to correctly rate such a heterogeneous population.

8 (i) The general form can be written as

$$C_{ij} = r_j s_i x_{i+j} + e_{ij}$$

where C_{ij} is incremental claims

r_j is the development factor for year j, independent of origin year i, representing proportion of claims paid by development year j

s_i is a parameter varying by origin year, representing the exposure

x_{i+j} is a parameter varying by calendar year, representing inflation

e_{ij} is an error term

(ii) Development factors are

$$\frac{3{,}991}{3{,}819} = 1.04504$$

and $$\frac{7{,}833}{5{,}329} = 1.46988$$

$$1-\frac{1}{f} = 1 - \frac{1}{1.04504\times 1.46988}$$

$$= 0.3490$$

2002: Emerging liability

$$= 5{,}012 \times 0.85 \times 0.3490$$

$$= 1{,}487$$

Reported liability 3,217

$\therefore$ Ultimate liability is 4,704

$\therefore$ Reserve $= 4{,}704 - 1{,}472$

$= 3{,}232$

9 (i) $$f(y) = \frac{1}{\mu}e^{-y/\mu}$$

$$= \exp\left[-\frac{y}{\mu} - \log\mu\right]$$

which is in the form of an exponential family of distributions, with $\theta = -\frac{1}{\mu}$.

Hence the canonical link function is $\frac{1}{\mu}$.

(ii) (a) The likelihood is $\prod_{i=1}^{n} f(y_i) = \prod_{i=1}^{n} \frac{1}{\mu_i} e^{-y_i/\mu_i}$

The log-likelihood is $\ell = \sum_{i=1}^{n} \left(-\log \mu_i - \frac{y_i}{\mu_i} \right)$

Hence $\ell_c = -\sum_{i=1}^{m} (\alpha + e^{-\alpha} y_i) - \sum_{i=m+1}^{n} (\beta + e^{-\beta} y_i)$

$$= -\left[m\alpha + (n-m)\beta + e^{-\alpha} \sum_{i=1}^{m} y_i + e^{-\beta} \sum_{i=m+1}^{n} y_i \right]$$

(b) $\frac{\partial \ell_c}{\partial \alpha} = -m + e^{-\alpha} \sum_{i=1}^{m} y_i$

$$\frac{\partial \ell_c}{\partial \alpha} = 0 \Rightarrow -m + e^{-\hat{\alpha}} \sum_{i=1}^{m} y_i = 0$$

$$\therefore \ \hat{\alpha} = \log \left(\frac{\sum_{i=1}^{m} y_i}{m} \right)$$

$$\frac{\partial \ell_c}{\partial \beta} = -(n-m) + e^{-\beta} \sum_{i=m+1}^{n} y_i$$

$$\frac{\partial \ell_c}{\partial \beta} = 0 \Rightarrow -(n-m) + e^{-\beta} \sum_{i=m+1}^{n} y_i = 0$$

$$\therefore \ \hat{\beta} = \log \left(\frac{\sum_{i=m+1}^{n} y_i}{n-m} \right)$$

(c) The deviance is $2(\ell_f - \ell_c)$

$$\ell_f = \sum_{i=1}^{n} \left(-\log y_i - \frac{y_i}{y_i} \right) = -\sum_{i=1}^{n} \log y_i - n$$

Hence the deviance is

$$2\left[-\sum_{i=1}^{n}\log y_i - n + \sum_{j=1}^{m}\left(\log\left(\frac{\sum_{j=1}^{m} y_i}{m}\right) + \frac{y_i}{\left(\frac{1}{m}\sum_{j=1}^{m} y_j\right)}\right) + \sum_{i=m+1}^{n}\left(\log\left(\frac{\sum_{j=m+1}^{n} y_j}{n-m}\right) + \frac{y_i}{\left(\frac{1}{n-m}\sum_{j=m+1}^{n} y_j\right)}\right)\right]$$

$$= 2\left[-\sum_{i=1}^{n}\log y_i - n + \sum_{i=1}^{m}\log\left(\frac{1}{m}\sum_{j=1}^{m} y_j\right) + m + \sum_{i=m+1}^{n}\log\left(\frac{1}{n-m}\sum_{j=m+1}^{n} y_j\right) + n - m\right]$$

$$= 2\left(\sum_{i=1}^{m}\log\left(\frac{\frac{1}{m}\sum_{j=1}^{m} y_j}{y_i}\right) + \sum_{i=m+1}^{n}\log\left(\frac{\frac{1}{n-m}\sum_{j=m+1}^{n} y_j}{y_i}\right)\right)$$

(iii) The deviance residual is $\operatorname{sign}(y_i - \hat{y}_i)\sqrt{D_i}$ where the deviance is $\sum_{i=1}^{n} D_i$.

$\hat{y}_i = 14.2$

$$D_1 = 2\left[-\log y_1 - 1 + \log\left(\frac{1}{m}\sum_{i=1}^{m} y_i\right) + \frac{y_1}{\frac{1}{m}\sum_{i=1}^{m} y_i}\right]$$

$$= 2\left[-\log 7 - 1 + \log 14.2 + \frac{7}{14.2}\right]$$

$= 0.2$

Hence the deviance residual is $-\sqrt{0.2} = -0.4472$.

10 (i) If, having take a sample from the distribution parameterised by λ, the posterior distribution of λ belongs to the same family as the prior distribution then the prior is called a conjugate prior.

(ii) We know that the prior distribution of λ is $\Gamma(\alpha, s)$. If $\underline{X}$ is the sample taken from the exponential distribution, then the posterior density satisfies:

$$f(\lambda|\underline{X}) \propto f(\underline{X}|\lambda)f(\lambda)$$

$$= \left[\prod_{i=1}^{n} \lambda e^{\lambda x_i}\right] \times \frac{s^{\alpha}\lambda^{\alpha-1}e^{-s\lambda}}{\Gamma(\alpha)}$$

$$\propto \lambda^{\alpha+n-1} e^{-\lambda(s+\sum_{i=1}^{n} x_i)}$$

$$\propto \text{pdf of } \Gamma\left(\alpha+n,\, s+\sum_{i=1}^{n} x_i\right)$$

This means that the posterior distribution of λ also follows a Gamma distribution and therefore the Gamma distribution satisfies the definition of a conjugate prior.

(iii) (a) We know that $\lambda \sim \Gamma(\alpha, s)$. So

$$E(1/\lambda) = \int_0^{\infty} \frac{f(\lambda)}{\lambda} d\lambda$$

$$= \int_0^{\infty} \frac{s^{\alpha}\lambda^{\alpha-1}e^{-s\lambda}}{\lambda\Gamma(\alpha)} d\lambda$$

$$= \int_0^{\infty} \frac{s^{\alpha}\lambda^{\alpha-2}e^{-s\lambda}}{\Gamma(\alpha)} d\lambda$$

$$= \frac{s}{\alpha-1} \int_0^{\infty} \frac{s^{\alpha-1}\lambda^{\alpha-2}e^{-s\lambda}}{\Gamma(\alpha-1)} d\lambda$$

$$= \frac{s}{\alpha-1} \times 1$$

$$= \frac{s}{\alpha-1}$$

since the final integral is of the pdf of a $\Gamma(\alpha - 1, s)$ distribution.

(b) Posterior mean is $E(1/\lambda)$ where $\lambda \sim \Gamma\left(\alpha + n, s + \sum_{i=1}^{n} x_i\right)$. The prior mean is $\frac{s}{\alpha - 1}$. The previous result implies that the posterior mean is given by

$$\frac{s + \sum_{i=1}^{n} x_1}{\alpha + n - 1} = \frac{s}{\alpha + n - 1} + \frac{\sum_{i=1}^{n} x_1}{\alpha + n - 1}$$

$$= \frac{\alpha - 1}{\alpha + n - 1} \times \frac{s}{\alpha - 1} + \frac{n}{\alpha + n - 1} \times \frac{\sum_{i=1}^{n} x_1}{n}$$

and $$\frac{\alpha - 1}{\alpha + n - 1} + \frac{n}{\alpha + n - 1} = 1$$

(iv) First consider Alan's beliefs. We know from the formula given in the question that

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

Hence for Alan we have

$$0.5^2 = 3^2 \times \frac{1}{\alpha - 2}$$

which means that

$$\alpha - 2 = \frac{9}{0.25} = 36$$

and hence $\alpha = 38$. Using the result for the posterior mean, we have $\frac{s}{\alpha - 1} = 3$ and hence $s = 3 \times 37 = 111$. So Alan's prior distribution for λ is $\Gamma(38, 111)$.

Similarly for Beatrice, we have

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

Hence

$$1^2 = 6^2 \times \frac{1}{\alpha - 2}$$

which means that

$$\alpha - 2 = \frac{36}{1} = 36$$

and hence $\alpha = 38$ again. Using the results for the posterior mean, we have $\frac{s}{\alpha - 1} = 6$ and hence $s = 6 \times 37 = 222$. So Beatrice's prior distribution for λ is $\Gamma(38, 222)$.

We will use the weighted average formula above to calculate the difference in the posterior means. First note that since both Alan and Beatrice have the same α we have

$$Z_A = Z_B = \frac{\alpha - 1}{\alpha + n - 1} = \frac{37}{37 + n}.$$

So the difference in posterior means is given by

$$Z_A \times 3 + (1 - Z_A) \times \bar{x} - Z_B \times 6 - (1 - Z_B) \times \bar{x} = -3 \times Z.$$

So we need to ensure that n is large enough that

$$3Z < 1$$

$$Z < 1/3$$

$$\frac{37}{37 + n} < 1/3$$

$$37 < \frac{37 + m}{3}$$

$$3 \times 37 - 37 < n$$

$$n > 74.$$

END OF EXAMINERS' REPORT

EXAMINATION

15 September 2005 (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.

1 Claims occur on a portfolio of insurance policies according to a Poisson process at a rate λ. All claims are for a fixed amount d, and premiums are received continuously. The insurer's initial surplus is U ($< d$) and the annual premium income is $1.2\lambda d$. Show that the probability that ruin occurs at the first claim is:

$$1 - e^{-\frac{1}{1.2}\left(1-\frac{U}{d}\right)}.$$ [4]

2 $Y_1, Y_2, \ldots, Y_n$ are independent random variables, and

Y_i ~ Poisson, mean μ_i.

The fitted values for a particular model are denoted by $\hat{\mu}_i$. Derive the form of the scaled deviance. [5]

3 A manufacturer of specialist products for the retail market must decide which product to make in the coming year. There are three possible choices — basic, deluxe or supreme — each with different tooling up costs. The manufacturer has fixed overheads of £1,300,000.

The revenue and tooling up costs for each product are as follows:

	Tooling up costs £	*Revenue per item sold* £
Basic	100,000	1.00
Deluxe	400,000	1.20
Supreme	1,000,000	1.50

Last year the manufacturer sold 2,100,000 items and is preparing forecasts of profitability for the coming year based on three scenarios: Low sales (70% of last year's level); Medium sales (same as last year) and High sales (15% higher than last year).

(i) Determine the annual profits in £ under each possible combination. [3]

(ii) Determine the minimax solution to this problem. [2]

(iii) Determine the Bayes criterion solution based on the annual profit given the probability distribution: P(Low) = 0.25; P(Medium) = 0.6 and P(High) = 0.15. [2]

[Total 7]

4 XYZ bank are about to offer a new mortgage product to consumers with a poor creditrating. They currently offer a similar product to customers with normal credit ratings. The normal product charges all customers a Standard Variable Rate (SVR) of 6% which moves up and down in line with short term interest rates. In addition there is a maximum "loan to value" of 90% — in other words the customer cannot borrow more than 90% of the value of the property. For loans above this level an additional Mortgage Indemnity Guarantee insurance premium must be paid — this protects the bank in the event that the borrower defaults and the value of the property has fallen. There are no other charges on the normal product.

The bank intends to use its experience from the normal business as a basis for setting terms on the new product.

(i) List the factors the bank should take into account when setting terms on the new product compared with the "normal" business. [3]

(ii) Suggest ways in which the bank may mitigate the additional risks associated with this product. [4]

[Total 7]

5 An insurer believes that claims from a particular type of policy follow a Pareto distribution with parameters $\alpha = 2.5$ and $\lambda = 300$. The insurer wishes to introduce a deductible such that 25% of losses result in no claim on the insurer.

(i) Calculate the size of the deductible. [4]

(ii) Calculate the average claim amount net of the deductible. [6]

[Total 10]

6 The number of claims on a portfolio of washing machine insurance policies follows a Poisson distribution with parameter 50. Individual claim amounts for repairs are a random variable $100X$ where X has a distribution with probability density function

$$f(x) = \begin{cases} \frac{3}{32}(6x - x^2 - 5) & 1 < x < 5 \\ 0 & \text{otherwise} \end{cases}$$

In addition, for each claim (independently of the cost of the repair) there is a 30% chance that an additional fixed amount of £200 will be payable in respect of water damage.

(i) Calculate the mean and variance of the total individual claim amounts. [7]

(ii) Calculate the mean and variance of the aggregate claims on the portfolio. [3]

[Total 10]

7 An insurer operates a simple no claims discount system with 5 levels: 0%, 20%, 40%, 50% and 60%.

The rules for moving between levels are:

- An introductory discount of 20% is available to new customers.
- If no claims are made during a year the policyholder moves up to the next discount level or remains at the maximum level.
- If one or more claims are made during the year, a policyholder at the 50% or 60% discount level moves to the 20% level and a policyholder at 0%, 20% or 40% moves to or remains at the 0% level.

The full annual premium is £600.

When an accident occurs, the distribution of loss is exponential with mean £1,750. A policyholder will only claim if the loss is greater than the extra premiums over the next four years, assuming no further accidents occur.

(i) For each discount level, calculate the smallest cost for which a policyholder will make a claim. [3]

(ii) For each discount level, calculate the probability of a claim being made in the event of an accident occurring. [3]

(iii) Comment on the results of (ii). [2]

(iv) Currently, equal numbers of customers are in each discount level and the probability of a policyholder not having an accident each year is 0.9. Calculate the expected proportions in each discount level next year. [4]
[Total 12]

8 The following time series model is used for the monthly inflation rate (Y_t) in a particular country:

$$Y_t = 0.4Y_{t-1} + 0.2Y_{t-2} + Z_t + 0.025$$

where $\{Z_t\}$ is a sequence of uncorrelated identically distributed random variables whose distributions are normal with mean zero.

(i) Derive the values of p, d and q, when this model is considered as an ARIMA(p, d, q) model. [3]

(ii) Determine whether $\{Y_t\}$ is a stationary process. [2]

(iii) Assuming an infinite history, calculate the expected value of the rate of inflation over this. [1]

(iv) Calculate the autocorrelation function of $\{Y_t\}$. [5]

(v) Explain how the equivalent infinite-order moving average representation of $\{Y_t\}$ may be derived. [2]

[Total 13]

9 The claims paid to date on a motor insurance policy are as follows (Figures in £000s):

	Development year			
Policy Year	*0*	*1*	*2*	*3*
2001	1,256	945	631	378
2002	1,439	1,072	723	
2003	1,543	1,133		
2004	1,480			

Inflation for the 12 months to the middle of each year was as follows:

2002	2.10%
2003	1.20%
2004	–0.80%

Annual premiums written in 2004 were £5,250,000.

Future inflation from mid-2004 is estimated to be 2.5% per annum.

The ultimate loss ratio (based on mid-2004 prices) has been estimated at 75%.

Claims are assumed to be fully run-off at the end of development year 3.

Estimate the outstanding claims arising from policies written in 2004 only (taking explicit account of the inflation statistics in **both** cases), using:

(i) The chain ladder method. [9]

(ii) The Bornhuetter-Ferguson method. [7]

[Total 16]

 PLEASE TURN OVER

10 The total amounts claimed each year from a portfolio of insurance policies over n years were $x_1, x_2, \ldots, x_n$. The insurer believes that annual claims have a normal distribution with mean θ and variance σ_1^2, where θ is unknown. The prior distribution of θ is assumed to be normal with mean μ and variance σ_2^2.

(i) Derive the posterior distribution of θ. [4]

(ii) Using the answer in (a), write down the Bayesian point estimate of θ under quadratic loss. [2]

(iii) Show that the answer in (b) can be expressed in the form of a credibility estimate and derive the credibility factor. [2]

The claims experience over five years for two companies was as follows:

	Year	*1*	*2*	*3*	*4*	*5*
Company A	Amount	421	417	438	456	463
Company B	Amount	343	335	356	366	380

(iv) Determine the Bayes credibility estimate of the premiums the insurer should charge for each company based on the modelling assumptions of part (i), a profit loading of 25% and the following parameters:

	Company A	*Company B*
μ	400	300
σ_1^2	500	350
σ_2^2	800	600

[6]

(v) Comment on the effect on the result of increasing σ_1^2 and σ_2^2. [2]

[Total 16]

END OF PAPER

EXAMINATION

September 2005

Subject CT6 — Statistical Methods
Core Technical

EXAMINERS' REPORT

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.

M Flaherty
Chairman of the Board of Examiners

15 November 2005

EXAMINERS' COMMENTS

Comments on solutions presented to individual questions for this September 2005 paper are given below.

Q1 The general standard of answers was poor. Very few candidates correctly derived the condition for ruin to occur in terms of time until the first claim.

Q2 Reasonably well answered.

Q3 Well answered by the majority of candidates.

Q4 Very few candidates generated enough different points to score high marks on this question.

Q5 Most candidates were able to answer part (i). Answers to part (ii) were more varied, with many candidates struggling with the necessary integration, or evaluating the wrong integral.

Q6 Reasonably well answered, but a large number of candidates failed to cope with the possible additional fixed claim amounts. A common error was to treat the separate elements as entirely independent – which is incorrect since the fixed claim can only occur when a variable claim has already been made.

Q7 Answers were generally of a high standard. However, a surprisingly large number of candidates ignored the instruction in part (iv) of the question and instead derived the proportions in each state in a stationary population.

Q8 In general this was reasonably well answered, although many candidates failed to give any justification for their classification in part (i). In part (iv) a number of candidates claimed that the autocorrelation function was zero for k greater than 2.

Q9 The inflation adjustment needed to past claims data produced a number of common alternative approaches – credit was given for these providing they were sensible, consistent and clearly explained. Most candidates were able to derive the necessary development factors. However, a large number of candidates failed to adjust the outstanding claims for future inflation (ie the answer was given in 2004 prices) and so failed to obtain full marks. Many candidates did not show any working for the calculations in part (ii) making it difficult to give any credit for partially correct solutions.

Q10 Many candidates simply wrote down the result in part (i) (which is given in the tables) rather than deriving it as instructed. The rest of the question was very well answered in general.

1 Ruin will occur if the time of the first claim t is such that

$$U + 1.2\lambda dt < d$$

i.e. if

$$t < \frac{d-U}{1.2\lambda d}.$$

The time until the first claim follows an exponential distribution with parameter λ. So the probability of ruin is given by

$$\int_0^{\frac{d-U}{1.2\lambda d}} \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x}\right]_0^{\frac{d-U}{1.2\lambda d}}$$

$$= 1 - e^{-\lambda\left(\frac{d-U}{1.2\lambda d}\right)}$$

$$= 1 - e^{-\frac{1}{1.2}\left(1-\frac{U}{d}\right)}.$$

2 The deviance is $2(l_f - l_c)$, where l_f and l_c are the log-likelihoods of the full and current model, respectively.

$$f(y) = \frac{\mu^y e^{-\mu}}{y!}$$

$$l = \sum_{i=1}^{n}\left[y_i \log \mu_i - \mu_i - \log y_i!\right]$$

$$l_f = \sum_{i=1}^{n}\left[y_i \log y_i - y_i - \log y_i!\right]$$

$$l_c = \sum_{i=1}^{n}\left[y_i \log \hat{\mu}_i - \hat{\mu}_i - \log y_i!\right]$$

Hence the deviance is

$$2\left[\sum_{i=1}^{n} y_i \log \frac{y_i}{\hat{\mu}_i} - \sum_{i=1}^{n}(y_i - \hat{\mu}_i)\right]$$

3 (i) **Revenue (£)**

	Low	*Medium*	*High*
Basic	1,470,000	2,100,000	2,415,000
Deluxe	1,764,000	2,520,000	2,898,000
Supreme	2,205,000	3,150,000	3,622,500

Costs (£)

	Low	*Medium*	*High*
Basic	1,400,000	1,400,000	1,400,000
Deluxe	1,700,000	1,700,000	1,700,000
Supreme	2,300,000	2,300,000	2,300,000

Profit (£)

	Low	*Medium*	*High*	*Min*	*Max*	*Expected Profit*
Basic	70,000	700,000	1,015,000	70,000	1,015,000	589,750
Deluxe	64,000	820,000	1,198,000	64,000	1,198,000	687,700
Supreme	–95,000	850,000	1,322,500	–95,000	1,322,500	684,625

(ii) Minimax: Decision is "Basic".

(iii) Highest expected profit. Decision is "Deluxe".

4 (i) Higher risk of default.
Cost of MIG insurance.
Other insurers.
Cost of underwriting.
Profitability of normal product.
Adjust for underlying economic conditions.
Exceptional defaults.

(ii) Higher SVR.
Lower LTV.
Higher MIG.
Penalty Payments.
Increase charges.
Compulsory insurance.
Maximum loan amount.
Charges/SVR variable according to wish.

Other sensible suggestions were given credit.

5 (i) We must find D such that

$$\int_0^D f(x)dx = 0.25$$

this means that

$$0.25 = \int_0^D \frac{\alpha\lambda^\alpha}{(\lambda+x)^{\alpha+1}}dx$$

$$= \left[-\frac{\lambda^\alpha}{(\lambda+x)^\alpha}\right]_0^D$$

$$= 1-\frac{\lambda^\alpha}{(\lambda+D)^\alpha}$$

$$= 1-\left(\frac{300}{300+D}\right)^{2.5}$$

So

$$\frac{300}{300+D} = (1-0.25)^{\frac{1}{2.5}} = 0.8913$$

and hence

$$300 + D = \frac{300}{0.8913}$$

so that

$$D = \frac{300}{0.8913}-300 = 36.59.$$

(ii) The average net claim is given by $E[X - D \mid X > D]$

$$\int_D^\infty (x-D)f(x)dx = \int_D^\infty \frac{x\alpha\lambda^\alpha}{(\lambda+x)^{\alpha+1}}dx - D\int_D^\infty \frac{\alpha\lambda^\alpha}{(\lambda+x)^{\alpha+1}}dx$$

$$= \left[-\frac{x\lambda^\alpha}{(\lambda+x)^\alpha}\right]_D^\infty + \int_D^\infty \frac{\lambda^\alpha}{(\lambda+x)^\alpha}dx - D\left[-\frac{\lambda^\alpha}{(\lambda+x)^\alpha}\right]_D^\infty$$

$$= 0 + \frac{D\lambda^{\alpha}}{(\lambda+D)^{\alpha}} + \left[-\frac{\lambda^{\alpha}}{(\alpha-1)(\lambda+x)^{\alpha-1}} \right]_D^{\infty} + 0 - \frac{D\lambda^{\alpha}}{(\lambda+D)^{\alpha}}$$

$$= \frac{\lambda^{\alpha}}{(\alpha-1)(\lambda+D)^{\alpha-1}}$$

$$= \frac{300^{2.5}}{1.5\times(300+36.59)^{1.5}}$$

$$= 168.29.$$

Hence $E[X-D \mid X>D] = \dfrac{168.29}{0.75} = 224.39$

6 We first calculate $E(100X)$ and $\mathrm{Var}(100X)$. Taking the expectation first, we have

$$\begin{aligned} E(100X) &= 100E(X) \\ &= 100\times\frac{3}{32}\times\int_1^5 6x^2 - x^3 - 5x\,dx \\ &= \frac{300}{32}\left[2x^3 - \frac{x^4}{4} - \frac{5x^2}{2}\right]_1^5 \\ &= \frac{300}{32}\times(31.25+0.75) \\ &= 300. \end{aligned}$$

Now for the variance

$$\begin{aligned} E((100X)^2) &= 100^2\times E(X^2) \\ &= 100^2\times\frac{3}{32}\times\int_1^5 6x^3 - x^4 - 5x^2\,dx \\ &= \frac{30{,}000}{32}\left[\frac{6x^4}{4} - \frac{x^5}{5} - \frac{5x^3}{3}\right]_1^5 \end{aligned}$$

$$= \frac{30,000}{32} \times (104.166 + 0.366)$$

$$= 98,000$$

so that

$$\begin{aligned} \mathrm{Var}X &= E(X^2) - E(x)^2 \\ &= 98,000 - 300^2 \\ &= 8000 \\ &= (89.44)^2. \end{aligned}$$

Now let Y denote the cost (if any) of the associated repair. Then Y is independent of X and

$$E(Y) = 0.3 \times 200 = 60$$

and

$$\mathrm{Var}(Y) = 0.3 \times 200^2 - 60^2 = 8,400 = (91.65)^2.$$

So the mean individual claim amount Z is

$$E(X + Y) = E(X) + E(Y) = 300 + 60 = 360$$

and the variance of an individual claim is

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) = 7,998.75 + 8,400.00 = 16,398.75$$

The mean and variance of the aggregate claims S are given by the formulae

$$E(S) = E(N)E(Z)$$

and

$$\mathrm{Var}(S) = E(N)\,\mathrm{Var}(Z) + \mathrm{Var}(N)E(Z)^2$$

where N is the total number of claims. In our case

$$E(S) = 50 \times 360 = 18,000$$

and

$$\mathrm{Var}(S) = 50 \times 16,398.75 + 50 \times 360^2 = 7,299,937.5 = (2,701.84)^2$$

7 (i)

Discount level	*Premiums if no claim*	*Premiums if claim*	*Difference*
0%	480, 360, 300, 240	600, 480, 360, 300	360
20%	360, 300, 240, 240	600, 480, 360, 300	600
40%	300, 240, 240, 240	600, 480, 360, 300	720
50%/60%	240, 240, 240, 240	480, 360, 300, 240	420

(ii)

0% $P(\text{Cost} > 360) = \exp(-360/1{,}750) = 0.814$

20% $P(\text{Cost} > 600) = \exp(-600/1{,}750) = 0.710$

40% $P(\text{Cost} > 720) = \exp(-720/1{,}750) = 0.663$

50%/60% $P(\text{Cost} > 420) = \exp(-420/1{,}750) = 0.787$

(iii) The amount for which the 50%/60% discount drivers will claim is much lower than the 20% and 40% categories. It seems illogical for this to be the case as this leads to a higher probability of a claim in the event of an accident. Suggest that structure is altered to make claims for these categories less likely.

(iv) The transition matrix is

$$\begin{bmatrix} 0.0814 & 0.9186 & 0 & 0 & 0 \\ 0.0710 & 0 & 0.9290 & 0 & 0 \\ 0.0663 & 0 & 0 & 0.9337 & 0 \\ 0 & 0.0787 & 0 & 0 & 0.9213 \\ 0 & 0.0787 & 0 & 0 & 0.9213 \end{bmatrix}$$

This year the proportions at each level are

(0.2, 0.2, 0.2, 0.2, 0.2)

Next year, the expected proportions are

$0.2 \times (0.0814 + 0.0710 + 0.0663) = 0.04374$

$0.2 \times (0.9186 + 0.0787 + 0.0787) = 0.2152$

$0.2 \times 0.9290 = 0.1858$

$0.2 \times 0.9337 = 0.18674$

$0.2 \times (0.9213 + 0.9213) = 0.36852$

8 (i) $(1 - 0.4B - 0.2B^2)\, Y_t = Z_t + 0.025$

Characteristic equation

$$1 - 0.4z - 0.2z^2 = 0$$

has no root at $z = 1$, so $d = 0$.

No functional dependence on Z_{t-1}, Z_{t-2}, etc., so $q = 0$.

Hence this is an ARIMA(2, 0, 0).

(ii) Roots of characteristic equation are $-1 \pm \sqrt{6}$, which are outside (–1, +1), so $\{Y_t\}$ is stationary.

(iii) Mean is stationary over time

$$(1 - 0.4 - 0.2)E[Y_t] = 0.025$$

$$\therefore\ E[Y_t] = \frac{0.025}{0.4} = 0.0625.$$

(iv) $Y_t - 0.0625 = 0.4(Y_{t-1} - 0.0625) + 0.2(Y_{t-2} - 0.0625) + Z_t$

$$\rho_k = E[(Y_t - 0.0625)(Y_{t-k} - 0.0625)] = 0.4\rho_{k-1} + 0.2\rho_{k-2}$$

Put $k = 1$, and note that $\rho_0 = 1$ and $\rho_{-1} = \rho_1$

$$\therefore\ \rho_1 = 0.4 + 0.2\rho_1 \quad \therefore\ \rho_1 = \frac{0.4}{0.8} = 0.5$$

$$\rho_2 = 0.4\rho_1 + 0.2 = 0.4$$

$$\rho_3 = 0.4\rho_2 + 0.2\rho_1 = 0.26$$

and so on.

(v) $(1 - 0.4B - 0.2B^2)(Y_t - 0.0625) = z_t$

$$Y_t - 0.0625 = (1 - 0.4B - 0.2B^2)^{-1} Z_t$$

So we need to invert $(1 - 0.4B - 0.2B^2)$
and multiply by Z_t to obtain the equivalent moving average process.

9 (Figures in £000s)

Inflation factors for each development year

AY	*0*	*1*	*2*	*3*
2001	1.02499	1.00390	0.99200	1.00000
2002	1.00390	0.99200	1.00000	
2003	0.99200	1.00000		
2004	1.00000			

Other inflation adjustments were given credit providing they were sensible, consistent and some explanation was given.

Inflation adjusted claim payments in mid 2004 prices

AY	*0*	*1*	*2*	*3*
2001	1,287.39	948.69	625.95	378
2002	1,444.61	1,063.42	723	
2003	1,530.66	1,133		
2004	1,480			

Inflation adjusted cumulative claim payments in mid 2004 prices

AY	*0*	*1*	*2*	*3*
2001	1,287.39	2,236.08	2,862.03	3,240.03
2002	1,444.61	2,508.03	3,231.03	
2003	1,530.66	2,663.66		
2004	1,480			

Column sum		7,407.77	6,093.06	3,240.03
Column sum minus last entry	4,262.66	4,744.11	2,862.03	
Development factor	1.73783	1.28434	1.13207	

(i)

Outstanding amounts arising from 2004 policies

Accumulated	1,480.0	2,572.0	3,303.3	3,739.6
Disaccumulated		1,092.0	731.3	436.3
Inflation		1.025	1.050625	1.076891
Inflation adj by year		1,119.3	768.3	469.8
Total		2,357.4		

For example, 2,572.0 = 1,480 × 1.7378

(ii) Ultimate amount for 2004 policies 5,250 × 0.75 = 3,937.50

Outstanding amounts for 2004 policies

	1,149.8	770.0	459.4
Infl adj by year	1.025	1.050625	1.076781
Inflation adj by year	1,178.5	809.0	494.7
Total	2,482.2		

For example,

1,149.8 = 3,937.5 × (1 / (1.28434 * 1.13207) – 1 / (1.73783 * 1.28434 * 1.13207))

10 (i) $x_1, \ldots, x_n$ are the observed claims:

$$f(\theta) \propto e^{-\frac{(\theta-\mu)^2}{2\sigma_2^2}} \propto e^{-\frac{1}{2\sigma_2^2}(\theta^2-2\theta\mu)}$$

$$p(\underline{x}|\theta) \propto \prod_{i=1}^{x} e^{-\frac{(x_1-\theta)^2}{2\sigma_1^2}}$$

$$= e^{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n}(x_i-\theta)^2}$$

$$\propto e^{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n}(\theta^2-2\theta x_i)}$$

$$= e^{-\frac{1}{2\sigma_1^2}\left(n\theta^2-2\theta\sum_{i=1}^{n}x_i\right)}$$

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

$$\propto e^{-\frac{1}{2\sigma_2^2}(\theta^2-2\theta\mu)-\frac{1}{2\sigma_1^2}\left(n\theta^2-2\theta\sum_{i=1}^{n}x_i\right)}$$

$$= e^{-\left\{\left(\frac{1}{2\sigma_2^2}+\frac{n}{2\sigma_1^2}\right)\theta^2-\left(\frac{\mu}{2\sigma_2^2}+\frac{\sum_{i=1}^{n}x_i}{2\sigma_1^2}\right)2\theta\right\}}$$

$$= e^{-\left(\frac{\sigma_1^2+n\sigma_2^2}{2\sigma_1^2\sigma_2^2}\right)\theta^2-2\theta\left(\frac{\mu\sigma_1^2+\left(\sum_{i=1}^{n}x_i\right)\sigma_2^2}{2\sigma_1^2\sigma_2^2}\right)}$$

$$\propto e^{-\frac{\sigma_1^2+n\sigma_2^2}{2\sigma_1^2\sigma_2^2}\left(\theta-\left(\frac{\mu\sigma_1^2}{\sigma_1^2+n\sigma_2^2}+\frac{\sigma_2^2\sum_{i=1}^{n}x_i}{\sigma_1^2+n\sigma_2^2}\right)\right)^2}$$

$$\Rightarrow \theta|\underline{x} \sim N\left(\frac{\mu\sigma_1^2+\left(\sum_{i=1}^{n}x_i\right)\sigma_2^2}{\sigma_1^2+n\sigma_2^2}, \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+n\sigma_2^2}\right)$$

(ii) Point estimator under quadratic loss is the posterior mean:

$$E(\theta|\underline{x}) = \frac{\sigma_1^2}{\sigma_1^2+n\sigma_2^2}\mu+\frac{\sigma_2^2}{\sigma_1^2+n\sigma_2^2}\left(\sum_{i=1}^{n}x_i\right)$$

(iii) $E(\theta|\underline{x}) = (1-Z)\,\mu + Z\left(\frac{\sum_{i=1}^{n}x_i}{n}\right)$

which is in the form of a credibility estimate, and

$Z = \dfrac{n\sigma_2^2}{\sigma_1^2+n\sigma_2^2}$

is the credibility factor.

(iv)

	Company A	Company B
n	5	5
σ_1^2	500	350
σ_2^2	800	600
$Z = \dfrac{\sigma_2^2}{\sigma_2^2+\frac{\sigma_1^2}{n}}$	0.8889	0.8955
$\overline{x}$	439	356
μ	400	300

Credibility Premium $= Z\bar{x} + (1 - Z)\mu$	434.7	350.1
Premium (CP + 25%)	543.3	437.7

(v) σ_1^2 increases — Reduces credibility factor and hence credibility premium moves towards prior mean.

σ_2^2 increases — Increases credibility factor and hence credibility premium moves towards sample mean.

END OF EXAMINERS' REPORT

EXAMINATION

28 March 2006 (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.

1 List the main perils typically insured against under a household buildings policy. [3]

2 A No Claims Discount (NCD) system has 3 levels of discount

Level 0	no discount
Level 1	discount = p
Level 2	discount = $2p$

where $0 < p < 0.5$.

The probability of a policyholder NOT making a claim each year is 0.9.

In the event of a claim, the policyholder moves to, or remains at level 0. Otherwise, the policyholder moves to the next higher level (or remains at level 2).

The premium paid in level 0 is £1,000.

Derive an expression in terms of p for the average premium paid by a policyholder once the steady state has been reached. [6]

3 Based on the proposal form, an applicant for life insurance is classified as a standard life (1), an impaired life (2) or uninsurable (3). The proposal form is not a perfect classifier and may place the applicant into the wrong category.

The decision to place the applicant in state i is denoted by d_i, and the correct state for the applicant is θ_i.

The loss function for this decision is shown below:

	d_1	d_2	d_3
θ_1	0	5	8
θ_2	12	0	3
θ_3	20	15	0

(i) Determine the minimax solution when assigning an applicant to a category. [1]

(ii) Based on the application form, the correct category for a new applicant appears to be as an impaired life. However, of applicants which appear to be impaired lives, 15% are in fact standard lives and 25% are uninsurable. Determine the Bayes solution for this applicant. [4]
[Total 5]

4 (i) Derive the autocovariance and autocorrelation functions of the AR(1) process

$$X_t = \alpha X_{t-1} + e_t$$

where $|\alpha| < 1$ and the e_t form a white noise process. [4]

(ii) The time series Z_t is believed to follow a ARIMA(1, d, 0) process for some value of d. The time series $Z_t^{(k)}$ is obtained by differencing k times and the sample autocorrelations, $\{r_i : i = 1, \ldots, 10\}$, are shown in the table below for various values of k.

	$k = 0$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
r_1	100%	100%	83%	–3%	–45%	–64%
r_2	100%	100%	66%	–12%	–5%	13%
r_3	100%	100%	54%	–11%	–4%	–3%
r_4	100%	99%	45%	–1%	6%	4%
r_5	100%	99%	37%	–3%	4%	5%
r_6	100%	99%	30%	–12%	–12%	–12%
r_7	99%	98%	27%	3%	7%	9%
r_8	99%	98%	24%	3%	0%	–4%
r_9	99%	97%	19%	3%	5%	6%
r_{10}	99%	97%	13%	–7%	–5%	–4%

Suggest, with reasons, appropriate values for d and the parameter α in the underlying AR(1) process. [4]

[Total 8]

5 (i) Let n be an integer and suppose that $X_1, X_2, \ldots, X_n$ are independent random variables each having an exponential distribution with parameter λ. Show that $Z = X_1 + \ldots + X_n$ has a Gamma distribution with parameters n and λ. [2]

(ii) By using this result, generate a random sample from a Gamma distribution with mean 30 and variance 300 using the 5 digit pseudo-random numbers.

63293 43937 08513 [5]

[Total 7]

6 An insurance company has a set of n risks (i = 1, 2, …, n) for which it has recorded the number of claims per month, Y_{ij}, for m months (j = 1, 2, …, m).

It is assumed that the number of claims for each risk, for each month, are independent Poisson random variables with

$$E[Y_{ij}] = \mu_{ij} .$$

These random variables are modelled using a generalised linear model, with

$$\log\mu_{ij} = \beta_i \quad (i = 1, 2, \ldots, n)$$

(i) Derive the maximum likelihood estimator of β_i . [4]

(ii) Show that the deviance for this model is

$$2\sum_{i=1}^{n}\sum_{j=1}^{m}\left(y_{ij}\log\frac{y_{ij}}{\overline{y}_i} - (y_{ij} - \overline{y}_i)\right)$$

where $\overline{y}_i = \dfrac{1}{m}\sum_{j=1}^{m} y_{ij}$. [3]

(iii) A company has data for each month over a 2 year period. For one risk, the average number of claims per month was 17.45. In the most recent month for this risk, there were 9 claims. Calculate the contribution that this observation makes to the deviance. [3]

[Total 10]

7 (i) Let N be a random variable representing the number of claims arising from a portfolio of insurance policies. Let X_i denote the size of the ith claim and suppose that X_1, X_2, … are independent identically distributed random variables, all having the same distribution as X. The claim sizes are independent of the number of claims. Let $S = X_1 + X_2 + \ldots + X_N$ denote the total claim size. Show that

$$M_S(t) = M_N(\log M_X(t)).$$ [3]

(ii) Suppose that N has a Type 2 negative binomial distribution with parameters $k > 0$ and $0 < p < 1$. That is

$$P(N = x) = \frac{\Gamma(k+x)}{\Gamma(x+1)\Gamma(k)} p^k q^x \qquad x = 0, 1, 2, \ldots$$

Suppose that X has an exponential distribution with mean $1/\lambda$. Derive an expression for $M_s(t)$. [2]

(iii) Now suppose that the number of claims on another portfolio is R with the size of the ith claim given by Y_i. Let $T = Y_1 + \ldots + Y_R$. Suppose that R has a binomial distribution, with parameters k and $1 - p$, and that Y_i has an exponential distribution with mean $1/\theta$. Show that if θ is chosen appropriately then S and T have the same distribution. [6]

You may use any standard formulae for moment generating functions of specific distributions shown in the Formulae and Tables.

[Total 11]

8 An insurer has for 2 years insured a number of domestic animals against veterinary costs. In year 1 there were n_1 policies and in year 2 there were n_2 policies. The number of claims per policy per year follows a Poisson distribution with unknown parameter θ.

Individual claim amounts were a constant c in year 1 and a constant $c(1 + r)$ in year 2. The average total claim amount per policy was y_1 in year 1 and y_2 in year 2. Prior beliefs about θ follow a gamma distribution with mean α/λ and variance α/λ^2. In year 3 there are n_3 policies, and individual claim amounts are $c(1 + r)^2$. Let Y_3 be the random variable denoting average total claim amounts per policy in year 3.

(i) State the distribution of the number of claims on the whole portfolio over the 2 year period. [1]

(ii) Derive the posterior distribution of θ given y_1 and y_2. [5]

(iii) Show that the posterior expectation of Y_3 given y_1, y_2 can be written in the form of a credibility estimate

$$Z \times k + (1 - Z) \times \frac{\alpha}{\lambda} \times c(1+r)^2$$

specifying expressions for k and Z. [5]

(iv) Describe k in words and comment on the impact the values of n_1, n_2 have on Z. [3]

[Total 14]

9 (i) The general form of a run-off triangle can be expressed as:

Accident Year	*Development Year* 0	1	2	3	4	5
0	$C_{0,0}$	$C_{0,1}$	$C_{0,2}$	$C_{0,3}$	$C_{0,4}$	$C_{0,5}$
1	$C_{1,0}$	$C_{1,1}$	$C_{1,2}$	$C_{1,3}$	$C_{1,4}$	
2	$C_{2,0}$	$C_{2,1}$	$C_{2,2}$	$C_{2,3}$		
3	$C_{3,0}$	$C_{3,1}$	$C_{3,2}$			
4	$C_{4,0}$	$C_{4,1}$				
5	$C_{5,0}$					

Define a model for each entry, C_{ij}, in general terms and explain each element of the formula. [3]

(ii) The run-off triangles given below relate to a portfolio of motorcycle insurance policies.

The cost of claims paid during each year is given in the table below:

(Figures in £000s)

Accident Year	*Development Year* 0	1	2	3
2002	2,905	535	199	56
2003	3,315	578	159	
2004	3,814	693		
2005	4,723			

The corresponding number of settled claims is as follows:

Accident Year	*Development Year* 0	1	2	3
2002	430	51	24	7
2003	465	58	24	
2004	501	59		
2005	539			

Calculate the outstanding claims reserve for this portfolio using the average cost per claim method with grossing-up factors, and state the assumptions underlying your result. [9]

(iii) Compare the results from the analysis in (ii) with those obtained from the basic chain ladder method. [5]

[Total 17]

10 An insurance company has two portfolios of independent policies, on each of which claims occur according to a Poisson process. For the first portfolio, all claims are for a fixed amount of £5,000 and 10 claims are expected per annum. For the second portfolio, claim amounts are exponentially distributed with mean £4,000 and 30 claims are expected per annum.

Let S denote aggregate annual claims from the two portfolios together.

A check is made for ruin only at the end of the year.

The insurer includes a loading of 10% in the premiums, for all policies.

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

(ii) Use a normal approximation to the distribution of S to calculate the initial capital, u, required in order that the probability of ruin at the end of the first year is 0.01. [3]

The insurer is considering purchasing proportional reinsurance from a reinsurer that includes a loading of ξ in its premiums. The proportion of each claim to be retained by the direct insurer is $\alpha (0 \le \alpha \le 1)$.

Let S_I denote the aggregate annual claims paid by the direct insurer on the two portfolios together, net of reinsurance.

(iii) Use a normal approximation to the distribution of S_I to show that the initial capital, u', required in order that the probability of ruin at the end of the first year is 0.01 can be written as

$$u' = \alpha u + (1 - \alpha)(\xi - 0.1) E[S].$$ [6]

(iv) Show that $u > u'$, as long as $\xi < 0.476$. [3]

(v) Show that $u - u'$ decreases as ξ increases, and discuss the practical implications of this result. [3]

[Total 19]

END OF PAPER

EXAMINATION

April 2006

Subject CT6 — Statistical Methods
Core Technical

EXAMINERS' REPORT

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.

M Flaherty
Chairman of the Board of Examiners

June 2006

Comments

Individual comments are shown after each question.

1 Fire
Flood
Storm
Theft
Explosions
Lightning
Damage caused by measures taken to put out a fire.

Comments on question 1: *This straightforward bookwork question was poorly done with relatively few candidates scoring full marks. Credit was given for any other reasonable suggestion not included on the list above.*

2 The transition matrix is

$$\begin{bmatrix} 0.1 & 0.9 & 0 \\ 0.1 & 0 & 0.9 \\ 0.1 & 0 & 0.9 \end{bmatrix}$$

$$0.1\,(\pi_0 + \pi_1 + \pi_2) = \pi_0$$

$$\therefore \pi_0 = 0.1$$

$$0.9\pi_0 = \pi_1$$

$$\therefore \pi_1 = 0.09$$

$$\therefore \pi_2 = 0.81$$

Average premium paid is

$$[0.1 + 0.09(1 - p) + 0.81(1 - 2p)] \times 1{,}000$$

$$= [1 - 0.09p - 1.62p] \times 1{,}000 = [1 - 1.71p] \times 1{,}000$$

Comments on question 2: *Most candidates obtained full marks. A few incorrectly identified the transition matrix or failed to solve the simultaneous equations.*

3 (i)

	0	5	8	←
	12	0	3	
	20	15	0	
Maximum loss:	20	15	8	

Minimax is d_3.

(ii) $P(\theta_1) = 0.15$
$P(\theta_2) = 0.6$
$P(\theta_3) = 0.25$

$d_1 = 0.15 \times 0 + 0.6 \times 12 + 0.25 \times 20 = 12.2$
$d_2 = 0.15 \times 5 + 0.6 \times 0 + 0.25 \times 15 = 4.5$
$d_3 = 0.15 \times 8 + 0.6 \times 3 + 0.25 \times 0 = 3$

Hence the Bayes decision is d_3.

Comments on question 3: *No comments given.*

4 (i)

$$\begin{aligned}\gamma_k &= \mathrm{Cov}(X_t, X_{t-k}) \\ &= \mathrm{Cov}(\alpha X_{t-1} + e_t, X_{t-k}) \\ &= \alpha\gamma_{k-1}\end{aligned}$$

and

$$\begin{aligned}\gamma_0 &= \mathrm{Cov}(X_t, X_t) \\ &= \mathrm{Cov}(\alpha X_{t-1} + e_t, \alpha X_{t-1} + e_t) \\ &= \alpha^2\mathrm{Cov}(X_{t-1}, X_{t-1}) + \mathrm{Cov}(e_t, e_t) \\ &= \alpha^2\gamma_0 + \sigma^2\end{aligned}$$

and hence

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

i.e. $$\gamma_0 = \frac{\sigma^2}{1-\alpha^2}$$

So our solution is

$$\gamma_k = \frac{\alpha^k\sigma^2}{1-\alpha^2}$$

The autocorrelation function is given by

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \alpha^k$$

(ii) The autocorrelation should decay exponentially as i increases. Looking at the table this behaviour occurs after differencing 2 times, suggesting the value of $d = 2$.

We know that the ratio of successive r's should be α. We can form these ratios as follows:

r_2/r_1	80%
r_3/r_2	82%
r_4/r_3	83%
r_5/r_4	82%
r_6/r_5	81%
r_7/r_6	90%
r_8/r_7	89%
r_9/r_8	79%
r_{10}/r_9	68%
Average	81.6%

Alternatively we can take the ith root of the ith autocorrelation:

i	ith root
1	83%
2	81%
3	81%
4	82%
5	82%
6	82%
7	83%
8	84%
9	83%
10	82%
Average	82%

Both approaches suggest the value of alpha is around 82%.

Full credit should be given to any reasonable approach.

Comments on question 4: *This question was generally done poorly. Although most candidates made a reasonable attempt at (i), very few correctly identified appropriate values or sensible reasons in (ii).*

5 (i) Each X_i has moment generating function $M_{X_i}(t) = \dfrac{\lambda}{\lambda - t}$. Hence

$$M_Z(t) = M_{X_1+\ldots X_n}(t) = M_{X_i}(t)^n = \left(\frac{\lambda}{\lambda - t}\right)^n$$

which is the moment generating function of a gamma distribution with parameters n and λ and hence Z has this distribution.

(ii) $\dfrac{\alpha}{\beta} = 30$, $\dfrac{\alpha}{\beta^2} = 300$

$\therefore\ \alpha = 3$ and $\beta = 0.1$

The random sample can be generated by producing three independent samples from an Exponential distribution with parameter 0.1 and adding them together. To do this, we need to solve

$$F_X(x) = 1 - e^{-0.1x} = u$$

where u is a pseudo-random number from a $U(0, 1)$ distribution.

Solving, we have $x = \dfrac{-\log(1-u)}{0.1}$

So using our pseudo-random numbers to give the exponential samples we have:

$u = 0.63292$ $\quad x = 10.022$
$u = 0.43937$ $\quad x = 5.787$
$u = 0.08513$ $\quad x = 0.890$

and the sample from the gamma distribution is

$$10.022 + 5.787 + 0.890 = 16.699.$$

Comments on question 5: *Part (i) was well answered but most candidates failed to generate the required random sample in part (ii).*

6 (i) The likelihood is $\prod_{i,j} \frac{\mu_{ij}^{y_{ij}} e^{-\mu_{ij}}}{y_{ij}!}$

and the loglikelihood is

$$l = \sum_{i=1}^{n} \sum_{j=1}^{m} (y_{ij} \log \mu_{ij} - \mu_{ij} - \log y_{ij}!)$$

Hence

$$l = \sum_{i=1}^{n} \sum_{j=1}^{m} (y_{ij}\beta_i - e^{\beta_i} - \log y_{ij}!)$$

$$\frac{\partial l}{\partial \beta_i} = \sum_{j=1}^{m} y_{ij} - me^{\beta_i}$$

$$\frac{\partial l}{\partial \beta_i} = 0 \quad \Rightarrow \quad e^{\hat{\beta}_i} = \bar{y}_i, \text{ where } \bar{y}_i = \frac{1}{m}\sum_{j=1}^{m} y_{ij}$$

and $\hat{\beta}_i = \log \bar{y}_i$

(ii) The deviance is

$$2(l_f - l_c) = 2\left(\sum_{i=1}^{n} \sum_{j=1}^{m} (y_{ij} \log y_{ij} - y_{ij}) - \sum_{i=1}^{n} \sum_{j=1}^{m} (y_{ij} \log \bar{y}_i - \bar{y}_i) \right)$$

$$= 2\sum_{i=1}^{n} \sum_{j=1}^{m} \left(y_{ij} \log \frac{y_{ij}}{\bar{y}_i} - (y_{ij} - \bar{y}_i) \right)$$

(iii) The deviance is

$$D_{ij} = y_{ij} \log \frac{y_{ij}}{\bar{y}_i} - (y_{ij} - \bar{y}_i) = 9 \log \frac{9}{17.45} - (9 - 17.45)$$

$$= 2.491$$

Full credit should also be given to 2 x 2.491 = 4.98

Comments on question 6: *Most candidates scored well on (i). Only more able candidates scored well on parts (ii) and (iii). There were some relatively easy marks available in (iii) for applying data to the formula given in (ii).*

7 (i)

$$M_s(t) = E(e^{St})$$

$$= E(E(e^{(X_1+...+X_N)t}|N))$$

$$= E(E(e^{X_1t}e^{X_2t}...e^{X_Nt}|N))$$

$$= E(M_X(t)^N)$$

$$= E(e^{N\log M_X(t)})$$

$$= M_N(\log M_X(t))$$

(ii) From the tables

$$M_N(t) = \left(\frac{p}{1-qe^t}\right)^k$$

$$M_X(t) = \frac{\lambda}{\lambda - t}$$

So

$$M_s(t) = \left(\frac{p}{1-qM_X(t)}\right)^k$$

$$= \left(\frac{p}{1-q\left(\frac{\lambda}{\lambda-t}\right)}\right)^k$$

$$= \left(\frac{p(\lambda-t)}{\lambda-t-q\lambda}\right)^k$$

$$= \left(\frac{p(\lambda-t)}{p\lambda-t}\right)^k$$

(iii) We now have

$$M_Y(t) = \frac{\theta}{\theta - t}$$

$$M_R(t) = (p + qe^t)^k$$

and so

$$M_T(t) = \left(p + q\frac{\theta}{\theta - t}\right)^k$$

$$= \left(\frac{p\theta - pt + q\theta}{\theta - t}\right)^k$$

$$= \left(\frac{\theta - pt}{\theta - t}\right)^k$$

Thus if we choose $\theta = p\lambda$ then $M_T(t) = M_S(t)$ and by the uniqueness of Moment Generating Functions, S and T have the same distribution.

Comments on question 7: *This question was generally well answered although relatively few managed the final step of demonstrating that S and T have the same distribution.*

8 (i) The total number of claims has a Poisson distribution with parameter $(n_1 + n_2)\theta$.

(ii) Let Y_i denote the average total claim amount per policy in year i and let X_i denote the total number of claims in year i. Then X_i has a Poisson distribution with parameter $n_i\theta$ and

$$X_1 = \frac{n_1 Y_1}{c} \text{ and } X_2 = \frac{Y_2 n_2}{c(1+r)}.$$

$$f(\theta | y_1, y_2) \propto f(y_1, y_2 | \theta)\, f(\theta)$$

$$\propto e^{-n_1\theta}(n_1\theta)^{y_1 n_1/c} e^{-n_2\theta}(n_2\theta)^{y_2 n_2/c(1+r)} e^{-\lambda\theta}\theta^{\alpha-1}$$

$$\propto e^{-(\lambda+n_1+n_2)\theta}\, \theta^{\left(\alpha + \frac{n_1 y_1}{c} + \frac{n_2 y_2}{c(1+r)}\right)-1}$$

So the posterior distribution of θ is gamma with parameters $\alpha + \dfrac{n_1 y_1}{c} + \dfrac{n_2 y_2}{c(1+r)}$ and $\lambda + n_1 + n_2$.

(iii) $E(Y_3|y_1, y_2) = \dfrac{c(1+r)^2}{n_3} \times E(X_3|y_1, y_2)$

$$= \frac{c(1+r)^2}{n_3} \times n_3 \times \frac{\alpha + \dfrac{n_1 y_1}{c} + \dfrac{n_2 y_2}{c(1+r)}}{\lambda + n_1 + n_2}$$

$$= \frac{c\alpha(1+r)^2 + n_1 y_1 (1+r)^2 + n_2 y_2 (1+r)}{\lambda + n_1 + n_2}$$

$$= \left[c(1+r)^2 \times \frac{\alpha}{\lambda} \times \left(\frac{\lambda}{\lambda + n_1 + n_2} \right) + \left(\frac{n_1 y_1 (1+r)^2 + n_2 y_2 (1+r)}{n_1 + n_2} \right) \times \frac{n_1 + n_2}{\lambda + n_1 + n_2} \right]$$

$$k = \left(\frac{n_1 y_1 (1+r)^2 + n_2 y_2 (1+r)}{n_1 + n_2} \right) \text{ and}$$

$$Z = \frac{n_1 + n_2}{\lambda + n_1 + n_2}$$

(iv) k is effectively a weighted average of the inflation adjusted average claim amounts for the previous 2 years, weighted by the number of policies in force. As the number of policies in force increases, Z becomes closer to 1, and so more weight is placed on the actual experience, and less on the prior expectations.

Comments on question 8: *Candidates found this the most difficult question in the paper. Only those candidates with a methodical approach and an excellent grasp of the relevant bookwork managed to progress to the later parts of the question.*

9 (i) Each entry can be expressed as:

$$C_{ij} = r_j \,.\, s_i \,.\, x_{i+j} + e_{ij}$$

where:

r_j is the development factor for year j, representing the proportion of claim payments by year j. Each r_j is independent of the accident year i

s_i is a parameter varying by origin year, i, representing the exposure, for example the number of claims incurred in the accident year i

x_{i+j} is a parameter varying by calendar year, for example representing inflation

e_{ij} is an error term

(ii) The cumulative cost of claims paid is:

Accident Year	*Development Year* 0	1	2	3
2002	2,905	3,440	3,639	3,695
2003	3,315	3,893	4,052	
2004	3,814	4,507		
2005	4,723			

The number of accumulated settled claims is as follows:

(Figures in £000s)

Accident Year	*Development Year* 0	1	2	3	*Ult*
2002	430 (84.0%)	481 (93.9%)	505 (98.6%)	512 (100%)	512
2003	465 (83.8%)	523 (94.3%)	547 (98.6%)		554.8
2004	501 (84.2%)	560 (94.1%)			595.1
2005	539 (84.0%)				641.7

Average cost per settled claim:

Accident Year	*Development Year* 0	1	2	3	*Ult*
2002	6.756 (93.6%)	7.152 (99.1%)	7.206 (99.8%)	7.217 (100%)	7.217
2003	7.129 (96.0%)	7.444 (100.3%)	7.408 (99.8%)		7.419
2004	7.613 (94.3%)	8.048 (99.7%)			8.072
2005	8.763 (94.7%)				9.256

The total ultimate loss is therefore:

Accident Year	*ACPC*	*Claim Numbers*	*Projected Loss*
1	7.217	512	3,695
2	7.419	554.8	4,114
3	8.071	595.1	4,802
4	9.256	641.7	5,938
			18,550

Claims paid to date	16,977
Outstanding claims	1,573

Assumptions:

Claims fully run-off by end of development year 3.

Projections based on simple average of grossing up factors.

Number of claims relating to each development year are a constant proportion of total claim numbers from the origin year.

Similarly for average claim amounts, i.e. same proportion of total average claim amount for origin year.

(iii) Development table:

AF	*0*	*1*	*2*	*3*	
2002	2,905	3,440	3,639	3,695	3,695
2003	3,315	3,893	4,052		4,114
2004	3,814	4,507			4,800
2005	4,723				5,935

DF				
	11,840	7,691	3,695	Column sum
10,034	7,333	3,639		Column sum minus last entry
	1.17999	1.04882	1.01539	

Ultimate loss	18,544
Claims paid to date	16,977
Outstanding claims	1,567

Comments on question 9: It was encouraging to see so many candidates achieve full marks for the bookwork in (i). Parts (ii) and (iii) were generally well done. A small number of candidates obtained very different answers for (ii) and (iii) but failed to appreciate that this was due to an arithmetical error rather than the method used.

10 (i) $E(S) = E[S_1] + E[S_2]$

$= 10 \times 5{,}000 + 30 \times 4{,}000 = 170{,}000$

$\text{Var}[S] = \text{Var}[S_1] + \text{Var}[S_2]$

$= 10 \times 5{,}000^2 + 30 \times (4{,}000^2 + 4{,}000^2)$

$= 1.21 \times 10^9$

(ii) We require u such that

$$P(u + c < S) = 0.01$$

i.e. $$P\left(\frac{S - E(S)}{\sqrt{\text{Var}(S)}} > \frac{u + c - E(S)}{\sqrt{\text{Var}(S)}}\right) = 0.01$$

so $$\frac{u + c - E(S)}{\sqrt{\text{Var}(S)}} = 2.326$$

$$\begin{aligned}\therefore u &= 2.326\sqrt{\text{Var}(S)} + E(S) - 1.1E(S)\\ &= 2.326\sqrt{\text{Var}(S)} - 0.1E(S)\\ &= 63{,}922\end{aligned}$$

(iii) We require u' such that

$$P(u' + c - c_R < S_I) = 0.01$$

where c_R = reinsurance premium

$$\frac{u' + c - c_R - E[S_I]}{\sqrt{\text{Var}[S_I]}} = 2.326 = \frac{u + c - E[S]}{\sqrt{\text{Var}(S)}}$$

$E[S_I] = \alpha E[S]$ and $\text{Var}[S_I] = \alpha^2 \text{Var}[S]$.

$c_R = (1 + \xi)(1 - \alpha)\, E[S]$

Hence $$\frac{u' + 1.1E[S] - (1+\xi)(1-\alpha)E[S] - \alpha E[S]}{\alpha\sqrt{\text{Var}[S]}} = \frac{u + 1.1E[S] - E[S]}{\sqrt{\text{Var}(S)}}$$

$$\begin{aligned}\therefore u' &= \alpha(u + 0.1E[S]) - 1.1E[S] + (1 + \xi)(1 - \alpha)\, E[S] + \alpha E[S]\\ &= \alpha u + (0.1\alpha - 1.1 + 1 - \alpha + \xi(1 - \alpha) + \alpha)\, E(S)\\ &= \alpha u + (1 - \alpha)(\xi - 0.1)\, E(S)\end{aligned}$$

(iv) $u - u' = (1 - \alpha)[u - (\xi - 0.1)\, E(S)]$

$u - u' > 0 \Rightarrow u - (\xi - 0.1)\, E(S) > 0$

i.e. $\xi < \dfrac{u}{E(S)} + 0.1$

i.e. $\xi < \dfrac{63{,}922}{170{,}000} + 0.1 = 0.476$

(v) Since $(1 - \alpha)\, E[S] > 0$, $u - u'$ decreases as ξ increases.

The greater the premium loading required by the reinsurer, the smaller the reduction in capital required by the insurer, i.e. the less effective the reinsurance is in reducing P(ruin) and hence replacing the capital.

Comments on question 10: *After Q8, candidates found this the most difficult question on the paper. Parts (i) and (ii) were well answered but the inclusion of premium loadings confused most candidates and consequently very few scored well on (iii), (iv) and (v).*

END OF MARKING SCHEDULE

EXAMINATION

5 September 2006 (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.

1 The loss function under a decision problem is given by:

	θ_1	θ_2	θ_3
D1	11	9	19
D2	10	13	17
D3	7	13	10
D4	16	5	13

(i) State which decision can be discounted immediately and why. [2]

(ii) Explain what is meant by the minimax criterion and determine the minimax solution in this case. [2]

[Total 4]

2 A sequence of pseudo-random numbers from a uniform distribution over the interval [0, 1] has been generated by a computer.

(i) Explain the advantage of using pseudo-random numbers rather than generating a new set of random numbers each time. [1]

(ii) Use examples to explain how a sequence of pseudo-random numbers can be used to simulate observations from:

(a) a continuous distribution
(b) a discrete distribution [4]

[Total 5]

3 State the Markov property and explain briefly whether the following processes are Markov:

AR(4);
ARMA (1, 1).

[5]

4 An insurer insures a single building. The probability of a claim on a given day is p independently from day to day. Premiums of 1 unit are payable on a daily basis at the start of each day. The claim size is independent of the time of the claim and follows an exponential distribution with mean $1/\lambda$. The insurer has a surplus of U at time zero.

(i) Derive an expression for the probability that the first claim results in the ruin of the insurer. [6]

(ii) If $p = 0.01$ and $\lambda = 0.0125$ find how large U must be so that the probability that the first claim causes ruin is less than 1%. [2]

[Total 8]

5 (i) Let p be an unknown parameter, and let $f(p|\underline{x})$ denote the probability density of the posterior distribution of p given information $\underline{x}$. Show that under all-or-nothing loss the Bayes estimate of p is the mode of $f(p|\underline{x})$. [2]

(ii) Now suppose p is the proportion of the population carrying a particular genetic condition. Prior beliefs about p have a $U(0, 1)$ distribution. A sample of size N is taken from the population revealing that m individuals have the genetic condition.

(a) Suggest why the $U(0, 1)$ distribution has been chosen as the prior, and derive the posterior distribution of p.

(b) Calculate the Bayes estimate of p under all-or-nothing loss.

[6]
[Total 8]

6 The table below shows cumulative paid claims and premium income on a portfolio of general insurance policies.

Underwriting Year	*Development Year* 0	1	2	*Premium Income*
2002	38,419	77,112	91,013	120,417
2003	31,490	78,504		117,101
2004	43,947			135,490

(i) Assuming an ultimate loss ratio of 93% for underwriting years 2003 and 2004, calculate the Bornhuetter-Ferguson estimate of outstanding claims for this triangle. [8]

(ii) State the assumptions underlying this estimate. [2]
[Total 10]

7 The random variable W has a binomial distribution such that

$$P(W = w) = \binom{n}{w} \mu^w (1 - \mu)^{n-w} \quad (w = 0, 1, 2, \ldots, n)$$

Let $Y = \dfrac{W}{n}$.

(i) Write down an expression for $P(Y = y)$, for $y = 0, \frac{1}{n}, \frac{2}{n}, \ldots, 1$. [1]

(ii) Express the distribution of Y as an exponential family and identify the natural parameter and the dispersion parameter. [3]

(iii) Derive an expression for the variance function. [3]

(iv) For a set of n independent observations of Y, derive an expression of the scaled deviance. [3]
[Total 10]

8 (i) Let X denote the claim amount under an insurance policy, and suppose that X has a probability density $f_X(x)$ for $x > 0$. The insurer has an individual excess of loss reinsurance arrangement with a retention of £M. Let Y be the amount paid by the insurer net of reinsurance. Express Y in terms of X and hence derive an expression for the probability density function of Y in terms of $f_X(x)$. [3]

For a particular class of policy X is believed to follow a Weibull distribution with probability density function

$$f_X(x) = 0.75cx^{-0.25} e^{-cx^{0.75}} \quad (x > 0)$$

where c is an unknown constant. The insurer has an individual excess of loss reinsurance arrangement with retention £500. The following claims data are observed:

Claims below retention:	78, 104, 116, 135, 189, 243, 270, 350, 411, 491
Claims above retention:	3 in total
Total number of claims:	13

(ii) Estimate c using maximum likelihood estimation. [7]

(iii) Apply the method of percentiles using the median claim to estimate c. [4]
[Total 14]

9 An insurer operates a No Claims Discount system with three levels of discount:

	Discount
Level 0	0%
Level 1	20%
Level 2	50%

The annual premium in level 0 is £650.

If a policyholder makes no claims in a policy year, they move to the next high discount level (or remain at level 2). In all other cases they move to (or remain at) discount level 0.

For a policyholder who has not yet had an accident in a policy year, the probability of an accident occurring is 0.1. The time at which an accident occurs in the policy year is denoted by T, where

$0 \le T \le 1$;
$T = 0$ means that the accident occurs at the start of the policy year;
$T = 1$ means that the accident occurs at the end of the policy year.

It is assumed that T has a uniform distribution.

Given that a policyholder has had their first accident, the probability of them having a second accident in the same policy year is $0.4(1 - T)$. It is assumed that a policyholder will not have more than two accidents in a policy year.

The cost of each accident has an exponential distribution with mean £1,000.

After each accident, the policyholder decides whether or not to make a claim by comparing the increase in the premium they would have to pay in the next policy year with the claim size. In doing this, they assume that they will have no further accidents.

(i) Show that the distribution of the number of accidents, K, that a policyholder has in a year is:

$P(K = 0) = 0.9$
$P(K = 1) = 0.08$
$P(K = 2) = 0.02$

[4]

(ii) For each level of discount, calculate the probability that a policyholder makes n claims in a policy year, where $n = 0, 1, 2$. [8]

(iii) Write down the transition matrix. [2]

(iv) Derive the steady state distribution. [3]

[Total 17]

10 (i) Let $I_k = \int_m^\infty x^k e^{-\beta x} dx$

where k is a non-negative integer.

Show that $I_0 = \dfrac{1}{\beta} e^{-\beta m}$

and $I_k = \dfrac{m^k}{\beta} e^{-\beta m} + \dfrac{k}{\beta} I_{k-1} \quad (k = 1, 2, 3, \ldots)$ [3]

For a certain portfolio of insurance policies the number of claims annually has a Poisson distribution with mean 25. Claim sizes have a gamma distribution with mean 100 and variance 5,000 and the insurer includes a loading of 10% in its premium.

The insurer is considering purchasing individual excess of loss reinsurance with retention m from a reinsurer that includes a loading of 15% in its premium.

Let X_I and X_R denote the amounts paid by the direct insurer and the reinsurer, respectively, on an individual claim.

(ii) Calculate the premium, c, charged by the direct insurer for this portfolio. [1]

(iii) Show that $E[X_R] = \dfrac{1}{50^2} (I_2 - mI_1)$ and hence that

$$E[X_R] = (m + 100)\, e^{-m/50}.$$ [7]

(iv) Use the result in (iii) to derive an expression for $E[X_I]$. [1]

(v) Derive an expression for the direct insurer's expected annual profit. [3]

(vi) The table below shows the direct insurer's expected annual profit (Profit) and probability of ruin (P(ruin)), for various values of the retention level, m:

m	*Profit*	*P(ruin)*
36	1.8	0.002
50	*	0.01
100	148.5	0.05

Calculate the missing value in the table and discuss the issues facing the direct insurer when deciding on the retention level to use. [4]

[Total 19]

END OF PAPER

EXAMINATION

September 2006

Subject CT6 — Statistical Methods
Core Technical

EXAMINERS' REPORT

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.

M A Stocker
Chairman of the Board of Examiners

November 2006

Comments

Comments on solutions presented to individual questions for this September 2006 paper are given below.

Question 1

Consistently well answered.

Question 2

Generally well answered, though a number of candidates did not produce an example for the discrete case.

Question 3

Most candidates dealt with AR(4) well. However, most thought that ARMA(1,1) was Markov. Many of those who knew it is not Markov did not explain why it is not.

Question 4

Candidates found this the hardest question on the paper, with the vast majority struggling to score many marks. Most candidates did not make the first step of conditioning on the time to the first claim, and therefore made little if any progress with the question.

Question 5

Many candidates did not recall the bookwork in part (i). The standard of answers to part (ii) was nevertheless good.

Question 6

Generally very well answered, although a number of candidates failed to explicitly state the outstanding claims and therefore did not score full marks.

Question 7

Parts (i) and (ii) were consistently well answered. Stronger candidates also scored well on parts (iii) and (iv).

Question 8

This question was a good differentiator — whilst weaker candidates struggled, better candidates were able to score well, the main difficulties being specifying the distribution of Y in part (i) and dealing with the claims above the retention in part (ii).

Question 9

Only a few candidates were sufficiently methodical in their approach to part (ii) and therefore gained full marks on this question. Nevertheless, many candidates scored well on the question overall, in particular picking up the follow-on marks available in parts (iii) and (iv). A number of candidates found the link between losses and claims confusing and therefore interpreted the question in a way that that made it more complicated than it actually was.

Question 10

This was well answered overall, with many candidates scoring well, especially on parts (i) to (iv). Only the best candidates scored highly on part (vi).

1 (i) D2 can be eliminated since it is dominated by D3; that is under all circumstances the loss from D2 is greater than or equal to that from D3.

(ii) The minimax criterion is to choose D so that the loss, maximised with respect to theta, is a minimum. The relevant maximum losses are

D1	19
D3	13
D4	16

So we should chose D3.

2 (i) Using pseudo-random numbers removes the variability of using different sets of random numbers, which is helpful for comparing different models.

Only a single routine is required for generation of pseudo-random numbers whereas in the case of truly random numbers we need either a lengthy table or a hardware enhancement to a computer.

If we wish to use the same sequence of random numbers in 2 models we need only store the seed for the pseudo-random random numbers as opposed to a record of potentially millions of truly random numbers.

(ii) u is a random number from $U(0, 1)$

(a) Find x from $u = F(x)$

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

e.g. exponential $u = e^{-\lambda x}$

$$x = \frac{-\log u}{\lambda}$$

(b) Working with integers, find x such that $P(X \le x - 1) < u \le P(X \le x)$

e.g. toss a coin, X = number of heads

$X = 0$ if $u \le 0.5$, $X = 1$ otherwise

3 The Markov property for a process $\{Y_t\}$ states that the conditional distribution of $Y_t | Y_{t-1}$ is the same as the conditional distribution of

$$Y_t | Y_{t-1}, Y_{t-2}, \ldots$$

Development can be predicted from present state without any reference to past history.

AR(4)

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3} + \beta_4 Y_{t-4} + e_t$$

This is not Markov since the distribution of $Y_t | Y_{t-1}$ changes when Y_{t-2}, Y_{t-3}, Y_{t-4} are also given.

ARMA(1,1)

$$Y_t = \alpha + \beta Y_{t-1} + e_t - \theta e_{t-1}$$

$$Y_{t-1} = \alpha + \beta Y_{t-2} + e_{t-1} - \theta e_{t-2}$$

Hence $e_{t-1} = Y_{t-1} - \alpha - \beta Y_{t-2} + \theta e_{t-2}$, and substituting into the expression for Y_t, it can be seen that knowledge of Y_{t-2} changes the distribution of $Y_t | Y_{t-1}$. So this is not Markov.

4 (i)

$$P(\text{ruin on } 1^{\text{st}} \text{ claim}) = \sum_{t>0} P\left(\text{First claim at } t\right) \times P(\text{this claim causes ruin})$$

$$= \sum_{t>0} q^{t-1} \times p \times P(\text{this claim} > U + t)$$

$$= \sum_{t>0} pq^{t-1} \times \int_{U+t}^{\infty} \lambda e^{-\lambda x} dx$$

$$= \sum_{t>0} pq^{t-1} \times \left[-e^{-\lambda x}\right]_{U+t}^{\infty}$$

$$= \sum_{t>0} pq^{t-1} e^{-\lambda(U+t)}$$

$$= \sum_{t>0} pq^{t-1} e^{-\lambda U} e^{-\lambda t}$$

$$= pe^{-\lambda U} e^{-\lambda} \sum_{t>0} (qe^{-\lambda})^{t-1}$$

$$= \frac{pe^{-\lambda(U+1)}}{1 - qe^{-\lambda}}$$

(ii) We want

$$\frac{0.01 \times e^{-0.0125(U+1)}}{1 - 0.99 \times e^{-0.0125}} < 0.01$$

hence

$$\frac{0.009875778 \times e^{-0.0125U}}{0.022297977} < 0.01$$

i.e.

$$e^{-0.0125U} < 0.02257845$$

taking logarithms

$$-0.0125U < -3.79076$$

so that we require

$$U > 303.26$$

5 (i) Consider the loss function

$$L(g(x), p) = \begin{cases} 0 & \text{if } g - \varepsilon < p < g + \varepsilon \\ 1 & \text{otherwise} \end{cases}$$

Then the expected posterior loss is given by

$$1 - \int_{g-\varepsilon}^{g+\varepsilon} f(p|\underline{x}) dp$$

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

for small values of ε. This is minimised by setting g to be the maximum (i.e. the mode) of $f(p|\underline{x})$.

(ii) (a) Using $U(0, 1)$ as the prior for p suggests that no prior information or beliefs about p have been formed — it is equally likely to lie anywhere in the range [0, 1].

$f(p|m) \propto f(m|p)\, f(p)$

$$\propto p^m(1-p)^{N-m} \times 1$$

So posterior beliefs about p have a Beta distribution with parameters $m + 1$ and $N - m + 1$.

(b) We must find the mode of $(f(p|m)$.

Maximising this is the same as maximising

$$g(p) = \log f(p|m) = m \log p + (N-m) \log (1-p) + \text{constant}$$

$$g'(p) = \frac{m}{p} - \frac{N-m}{1-p}$$

and $g'(p) = 0$ when

$$\frac{m}{p} - \frac{N-m}{1-p} = 0 \text{ i.e.}$$

$$m(1-p) = (N-m)\,p$$

$$Np = m$$

$$p = m / N$$

6 (i)

Column totals:		155,616	91,013
Column totals excluding last entry:	69,909	77,112	
Development factors:		2.2260	1.1803
f:	2.6273	1.1803	

Underwriting Year	*Premium*	*Initial UL*
2003	117,101	108,904
2004	135,490	126,006

	Initial UL	*f*	*1 − 1/f*	*Emerging Liability*
2003	108,904	1.1803	0.1527	16,634
2004	126,006	2.6273	0.6194	78,045

So the estimate of outstanding claims is 94,678.

(ii) Assumptions:

Data already adjusted for inflation or past pattern of inflation will be repeated in future.
Payment pattern same for each underwriting year.
Estimated Loss Ratio is appropriate.
Claims from underwriting year 2002 are fully run off.

7 (i) $P(Y=y) = \binom{n}{ny} \mu^{ny} (1-\mu)^{n-ny}$

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

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

which is in the form of an exponential family.

The natural parameter is $\log\frac{\mu}{1-\mu}$.

The dispersion parameter is

either $\varphi = n$ and $a(\varphi) = \dfrac{1}{\varphi}$

or $\varphi = \dfrac{1}{n}$ and $a(\varphi) = \varphi$

(iii) $V(\mu) = b''(\theta)$

$$b(\theta) = -\log(1-\mu) = \log\frac{1}{1-\mu} = \log(1+e^{\theta})$$

$$b'(\theta) = \frac{e^{\theta}}{1+e^{\theta}}$$

$$b''(\theta) = \frac{(1+e^{\theta})e^{\theta} - e^{\theta}e^{\theta}}{(1+e^{\theta})^2} = \frac{e^{\theta}}{(1+e^{\theta})^2}$$

$$= \mu(1-\mu)$$

(iv) Scaled deviance is $-2(l_c - l_f)$

$$l_c = \sum_i \left[n\left(y_i \log\frac{\mu_i}{1-\mu_i} - \log\frac{1}{1-\mu_i} \right) + \log\binom{n}{ny_i} \right]$$

$$l_f = \sum_i \left[n\left(y_i \log\frac{y_i}{1-y_i} - \log\frac{1}{1-y_i} \right) + \log\binom{n}{ny_i} \right]$$

Hence the scaled deviance is

$$-2(l_c - l_f) = -2\sum_i n\left(y_i \log\left(\frac{\mu_i}{y_i}\frac{1-y_i}{1-\mu_i} \right) - \log\left(\frac{1-y_i}{1-\mu_i} \right) \right)$$

8 (i) $Y = \begin{cases} X & \text{if } X < M \\ M & \text{if } X \geq M \end{cases}$

Y has a mixed distribution given by

$f_Y(x) = f_X(x)$ for $x < M$ and

$P(Y = M) = 1 - F_X(M)$

where $F_X(x) = \int_0^x f_X(u)du$.

(ii) The probability of an individual claim being above the retention is given by

$$1 - F(500) = e^{-c500^{0.75}} = e^{-105.74 \times c}$$

The likelihood of the observed data is then (denoting by $x_1, \ldots, x_{10}$ the ten claims below the retention)

$$L = k \times \prod c x_i^{-0.25} e^{-c x_i^{0.75}} \times (e^{-105.74c})^3$$

and the log-likelihood is given by

$$l = \log L = \text{const} + 10 \log c - 0.25 \sum \log x_i - c \sum x_i^{0.75} - 3 \times 105.74c$$

Differentiating gives

$$l' = 10 / c - \sum x_i^{0.75} - 317.22$$

Equating this to zero gives

$$10 / \hat{c} = \sum x_i^{0.75} + 317.22$$

$$\hat{c} = \frac{10}{\sum x_i^{0.75} + 317.22} = \frac{10}{589.40 + 317.22} = 0.011$$

(iii) The median claim is for £270. We solve

$$F(270) = 0.5$$

$$1 - e^{-c270^{0.75}} = 0.5$$

$$e^{-66.61c} = 0.5$$

$$c = \frac{\log 0.5}{-66.61} = 0.0104$$

9 (i) $P(K = 0) = 0.9$

$P(K = 1) = 0.1 \times P(\text{no } 2^{\text{nd}} \text{ accident})$

$$P(2^{\text{nd}} \text{ accident}) = \int_0^1 0.4(1-t) f(t)dt$$

$$= 0.4\int_0^1 (1-t)dt = 0.4[t - ½t^2]_0^1$$

$$= 0.4 \times ½ = 0.2$$

$\therefore P(K = 1) = 0.1 \times (1 - 0.2) = 0.08$

$P(K = 2) = 1 - 0.9 - 0.08 = 0.02$

(ii) Let N = number of claims a policyholder makes.

Then $P(N = n) = \sum_{k=0}^{2} P(N = n | K = k)P(K = k)$

Level 0: Change in premium when first claim made = $650 - 0.8 \times 650 = 130$

$P(X > 130) = e^{-130/1,000} = 0.8781$

Levels 1, 2: Change in premium when first claim made = $650 - 0.5 \times 650 = 325$

$P(X > 325) = e^{-325/1,000} = 0.7225$

$$P(N = 0) = P(K = 0) + P(N = 0|K=1)\,P(K = 1) + P(N = 0|K = 2)\,P(K = 2)$$

Level 0: $P(N = 0) = 0.9 + 0.1219 \times 0.08 + 0.1219^2 \times 0.02$
$= 0.9100$

Levels 1, 2: $P(N = 0) = 0.9 + 0.2775 \times 0.08 + 0.2775^2 \times 0.02$
$= 0.9237$

Note that if one claim has already been made then the NCD has already been lost, and it is therefore certain that a second claim will be made, regardless of the size of the loss. Therefore, for two accidents to result in only one claim it must be that the first accident resulted in no claim, and the second resulted in a claim.

$$P(N = 1) = P(N = 1|K = 1)\, P(K = 1) + P(N = 1|K = 2)\, P(K = 2)$$

Level 0: $P(N = 1) = 0.8781 \times 0.08 + 0.1219\times 0.8781 \times 0.02$
$= 0.0724$

Levels 1, 2: $P(N = 1) = 0.7225 \times 0.08 + 0.2775 \times 0.7255 \times 0.02$
$= 0.0618$

Two accidents will result in two claims whenever the first accident results in a claim (since in this case the second accident will certainly result in a claim).

$P(N = 2) = P(N = 2|K = 2)\, P(K = 2)$

Level 0: $0.8781 \times 0.02 = 0.0176$

Levels 1, 2: $P(N = 2) = 0.7225 \times 0.02 = 0.0145$

(iii) The transition matrix is

$$\begin{pmatrix} 0.0900 & 0.9100 & 0 \\ 0.0762 & 0 & 0.9238 \\ 0.0762 & 0 & 0.9238 \end{pmatrix}$$

(iv) $\underline{\pi} = P\underline{\pi}$

$0.9100\pi_0 = \pi_1$

$0.9238(\pi_1 + \pi_2) = \pi_2$

$\therefore\ \pi_2 = 12.123\pi_1 = 11.032\pi_0$

Since $\pi_0 + \pi_1 + \pi_2 = 1$

$\pi_0 + 0.9100\pi_0 + 11.032\pi_0 = 1$

$\therefore\ \pi_0 = 0.0773,\ \pi_1 = 0.0703,\ \pi_2 = 0.8524$

10 (i) $$I_0 = \int_m^\infty e^{-\beta x}dx \quad = \left[-\frac{1}{\beta}e^{-\beta x}\right]_m^\infty = \frac{1}{\beta}e^{-\beta m}$$

$$I_k = \int_m^\infty x^k e^{-\beta x}dx = \left[-\frac{1}{\beta}x^k e^{-\beta x}\right]_m^\infty + \int_m^\infty \frac{kx^{k-1}}{\beta}e^{-\beta x}dx$$

$$= \frac{m^k}{\beta}e^{-\beta m} + \frac{k}{\beta}\int_m^\infty x^{k-1}e^{-\beta x}dx$$

$$= \frac{m^k}{\beta}e^{-\beta m} + \frac{k}{\beta}I_{k-1}$$

(ii) $c = 1.1 \times 25 \times 100 = 2{,}750$

(iii) $$E[X_R] = \int_m^\infty (x-m)\,f(x)dx$$

$f(x)$ is gamma, and $\dfrac{\alpha}{\beta} = 100$, $\dfrac{\alpha}{\beta^2} = 5{,}000$

$\therefore \beta = \dfrac{1}{50}$ and $\alpha = 2$

$$\therefore\ f(x) = \left(\frac{1}{50}\right)^2 xe^{-x/50} \qquad (x > 0)$$

$$E[X_R] = \frac{1}{50^2}\left[\int_m^\infty x^2 e^{-x/50}dx - m\int_m^\infty xe^{-x/50}dx\right]$$

$$= \frac{1}{50^2}[I_2 - mI_1]$$

$I_0 \quad = 50e^{-m/50}$

$I_1 \quad = 50me^{-m/50} + 50^2e^{-m/50}$

$\quad = 50(m + 50)\, e^{-m/50}$

$I_2 \quad = 50m^2e^{-m/50} + \dfrac{2}{\beta} I_1$

$\quad = 50m^2e^{-m/50} + 5{,}000(m + 50)\, e^{-m/50}$

$\quad = 50(m^2 + 100(m + 50))\, e^{-m/50}$

$$\therefore E[X_R] = \frac{1}{50^2}\left[50(m^2 + 100(m+50))e^{-m/50} - 50m(m+50)e^{-m/50}\right]$$

$$= \frac{1}{50}\left[m^2 + 100(m+50) - m(m+50)\right]e^{-m/50}$$

$$= \frac{1}{50}\left[m^2 + 100m + 5{,}000 - m^2 - 50m]e^{-m/50}\right]$$

$$= \frac{1}{50}(50m + 5{,}000)e^{-m/50}$$

$$= (m + 100)\, e^{-m/50}$$

(iv) $\quad E[X_I] \quad = 100 - E[X_R]$

$\quad = 100 - (m + 100)\, e^{-m/50}$

(v) $\quad$ Insurer's expected profit is $c - c_R - 25E[X_I]$

i.e. $2{,}750 - 1.15 \times 25 \times (m + 100)\, e^{-m/50}$

$\quad - 25(100 - (m + 100)\, e^{-m/50})$

$\quad = 250 - 0.15 \times 25(m + 100)\, e^{-m/50}$

(vi) $\quad$ The completed table is

m	*Profit*	*P(Ruin)*
36	1.8	0.002
50	43.1	0.01
100	148.5	0.05

As m increases (less reinsurance)

Profit increases
P(Ruin) increases

There is a level beyond which it is not sensible to go (when Profit becomes negative).

It is a trade-off between profit and security.

Other sensible points were given credit.

END OF EXAMINERS' REPORT

EXAMINATION

19 April 2007 (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.

1 (i) State two conditions for a risk to be insurable. [2]

(ii) Describe briefly three distinct examples of financial loss insurance policies. [3]
[Total 5]

2 (i) Explain the concept of cointegrated time series. [3]

(ii) Give two examples of circumstances when it is reasonable to expect that two processes may be cointegrated. [2]
[Total 5]

3 The random variable X has an exponential distribution with mean 1000. Individual claim amounts on a certain type of insurance policy, Y, are such that

$$Y = X \qquad \text{for } 0 < X < 2000$$

and $P(Y = 2000) = P(X \geq 2000)$.

The insurer applies a deductible of 100 on claims from this type of insurance.

Calculate the mean of the distribution of individual claim amounts paid by the insurer. [5]

4 A casino operator moving into a country for the first time must apply to the casino regulator for a licence. There are three types of licence to choose from — slots, dice and cards — each with different running costs. The casino operator has to pay a fixed amount annually (£1,300,000) to the regulator, plus a variable annual licence cost.

The variable licence cost and expected revenue per customer for each type of game are as follows:

	Variable Licence cost £	*Expected Revenue per customer* £
Slots	250,000	60
Dice	550,000	120
Cards	1,150,000	160

The casino operator is uncertain about the number of customers and decides to prepare a profit forecast based on cautious, best estimate and optimistic numbers of customers. The figures are 14,000; 20,000 and 23,000 respectively.

(i) Determine the annual profits under each possible combination. [2]

(ii) Determine the minimax solution for optimising the profits. [2]

(iii) Determine the Bayes criterion solution based on the annual profit given the probability distribution P(cautious) = 0.2, P(best estimate) = 0.7 and P(optimistic) = 0.1. [2]

[Total 6]

5 The delay triangles given below relate to a portfolio of motor insurance policies.

The cost of claims settled during each year is given in the table below:

(Figures in £000s)

Accident Year	*Development year* 0	1	2
2004	4,144	694	183
2005	4,767	832	
2006	5,903		

The corresponding number of settled claims is as follows:

Accident Year	*Development year* 0	1	2
2004	581	75	28
2005	626	71	
2006	674		

Calculate the outstanding claims reserve for this portfolio using the average cost per claim method with grossing-up factors, and state the assumptions underlying your result.

[7]

6 (i) Explain the main advantage of the Polar method compared with the Box-Muller method for generating pairs of uncorrelated pseudo-random values from a standard normal distribution. [2]

(ii) Pseudo-random numbers are generated using the Box-Muller method in order to simulate values of $Y = \sqrt{X}$, where X has a lognormal distribution with parameters $\mu = 5$ and $\sigma = 2$. The quantity of interest is $\theta = E[Y]$.

(a) Calculate the value of Y when the number generated by the Box-Muller method is 0.9095.

(b) The variance of Y has been estimated as 26.3. Calculate how many simulations should be performed in order to ensure that the discrepancy between Y and $\hat{\theta}$, measured by the absolute error, is less than 1 with probability at least 0.9. [5]

[Total 7]

7 The total claims arising from a certain portfolio of insurance policies over a given month is represented by

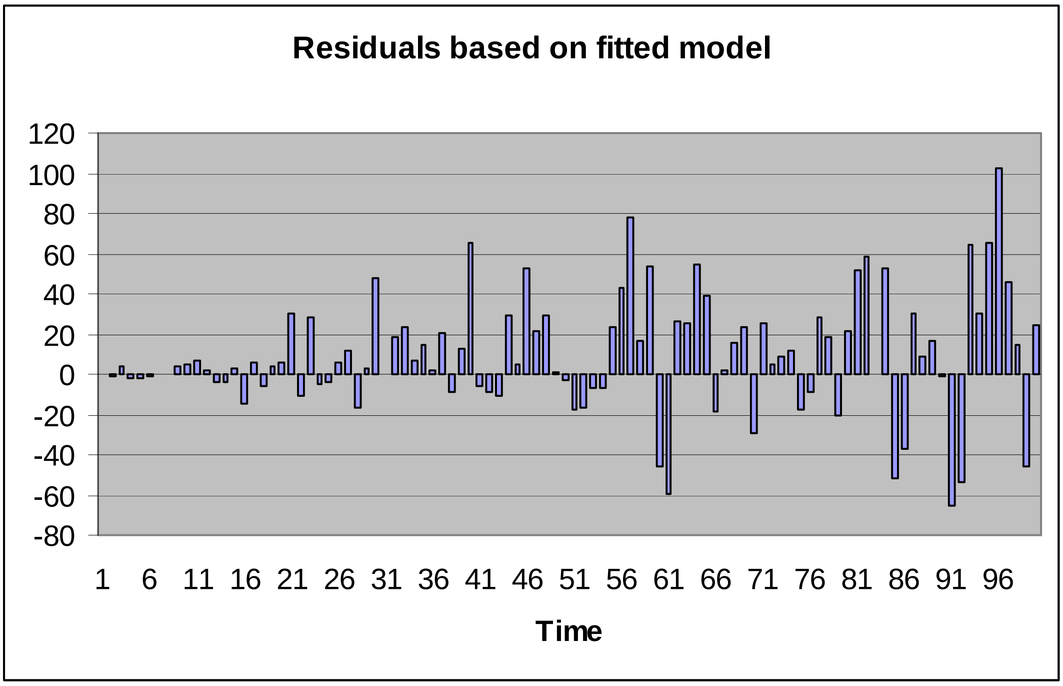
$$S = \begin{cases} \sum_{i=1}^{N} X_i & \text{if} \quad N > 0 \\ 0 & \text{if} \quad N = 0 \end{cases}$$

where N has a Poisson distribution with mean 2 and $X_1, X_2, \ldots, X_N$ is a sequence of independent and identically distributed random variables that are also independent of N. Their distribution is such that $P(X_i = 1) = 1/3$ and $P(X_i = 2) = 2/3$. An aggregate reinsurance contract has been arranged such that the amount paid by the reinsurer is $S - 3$ (if $S > 3$) and zero otherwise.

The aggregate claims paid by the direct insurer and the reinsurer are denoted by S_I and S_R, respectively.

Calculate $E(S_I)$ and $E(S_R)$. [8]

8 A modeller has attempted to fit an ARMA(p,q) model to a set of data using the Box-Jenkins methodology. The plot of residuals based on this proposed fit is shown below.



(i) Under the assumptions of the model, the residuals should form a white noise process.

(a) By inspection of the chart, suggest two reasons to suspect that the residuals do not form a white noise process.

(b) Define what is meant by a turning point.

(c) Perform a significance test on the number of turning points in the data above. (There are 100 points in the data and 59 turning points.)

[6]

(ii) On your suggestion, the original fitted model is discarded, and re-parameterised to:

$$X_{n+2} = 5 + 0.9(X_{n+1} - 5) + e_{n+2} + 0.5e_n .$$

Given the following observations:

$X_{99} = 2, \qquad X_{100} = 7$
$\hat{e}_{99} = -0.7, \qquad \hat{e}_{100} = 1.4$

Use the Box-Jenkins methodology to calculate the forward estimates $X_{100}(1), X_{100}(2)$ and $X_{100}(3)$.

[4]
[Total 10]

9 An insurer's NCD scale for motor policies has 3 levels of discount: 0%, 25% and 40%. The rules for moving between these levels are as follows:

- following a claim-free year, a policyholder moves to the next higher level of discount, or remains at 40% discount
- following a year of one or more claims, a policyholder at 40% discount moves to 25% discount while a policyholder at 25% or 0% moves to or stays at 0% discount

The full premium for each policyholder is £1,000. Following an accident, policyholders decide whether or not to claim by considering total outgoing over the next two years, assuming no further claims in this period and ignoring interest.

(i) Find the claim threshold for each level of discount. [3]

(ii) The probability of no accidents in any year for each policyholder is 0.88 and individual losses are assumed to have a lognormal distribution with $\mu = 6.0$ and $\sigma = 3.33$. Ignoring the possibility of more than 1 accident occurring in a year, calculate the transition matrix. [3]

(iii) Calculate the stationary distribution. [3]

(iv) Derive the stationary distribution under the alternative assumption that a policyholder always claims after a loss (regardless of the size of the claim). [3]

(v) Comment on the difference between the results of (iii) and (iv). [1]
[Total 13]

10 (i) The Gamma distribution with mean μ and variance μ^2/α has density function

$$f(y) = \frac{\alpha^\alpha}{\mu^\alpha \Gamma(\alpha)} y^{\alpha-1} e^{-\frac{y\alpha}{\mu}} \qquad (y > 0)$$

(a) Show that this may be written in the form of an exponential family.

(b) Use the properties of exponential families to confirm that the mean and variance of the distribution are μ and μ^2/α. [9]

(ii) Explain the difference between a continuous covariate and a factor. [3]

(iii) A company is analysing its claims data on a portfolio of motor policies, and uses a gamma distribution to model the claim severities. The company uses three rating factors:

policyholder age (as a continuous variable);
policyholder gender;
vehicle rating group (as a factor).

(a) Write down the form of the linear predictor when all rating factors are included as main effects.

(b) State how the linear predictor changes if an interaction between policyholder age and gender is included.

[4]
[Total 16]

11 The number, X, of claims on a given insurance policy over one year has probability distribution given by

$$P(X = k) = \theta^k (1-\theta) \quad k = 0, 1, 2, \ldots$$

where θ is an unknown parameter with $0 < \theta < 1$.

Independent observations $x_1, \ldots, x_n$ are available for the number of claims in the previous n years. Prior beliefs about θ are described by a distribution with density

$$f(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\alpha-1}$$

for some constant $\alpha > 0$.

(i) (a) Derive the maximum likelihood estimate, $\hat{\theta}$, of θ given the data $x_1, \ldots, x_n$.

(b) Derive the posterior distribution of θ given the data $x_1, \ldots, x_n$.

(c) Derive the Bayesian estimate of θ under quadratic loss and show that it takes the form of a credibility estimate

$$Z\hat{\theta} + (1 - Z)\mu$$

where μ is a quantity you should specify from the prior distribution of θ.

(d) Explain what happens to Z as the number of years of observed data increases.

[11]

(ii) (a) Determine the variance of the prior distribution of θ.

(b) Explain the implication for the quality of prior information of increasing the value of α. Give an interpretation of the prior distribution in the special case $\alpha = 1$.

[3]

(iii) Calculate the Bayesian estimate of θ under quadratic loss if $n = 3$, $x_1 = 3, x_2 = 3, x_3 = 5$ and

(a) $\alpha = 5$
(b) $\alpha = 2$

Comment on your results in the light of (ii) above. [4]

[Total 18]

END OF PAPER

EXAMINATION

April 2007

Subject CT6 — Statistical Methods
Core Technical

EXAMINERS' REPORT

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.

M A Stocker
Chairman of the Board of Examiners

June 2007

Comments

Comments on solutions presented to individual questions for this April 2007 paper are given below.

Question 1
Generally well answered.

Question 2
This bookwork question was poorly answered and only the best prepared candidates scored well.

Question 3
Relatively few candidates correctly identified that the deductible of 100 applied to all claims

Question 4
Most candidates scored full marks on this question.

Question 5
Candidates generally scored well on this straightforward question

Question 6
This was the most difficult question on the paper. Many answered the bookwork in part (i) but were unable to make a start on part (ii).

Question 7
Poorly answered. Most candidates correctly calculated E(S) but failed to work methodically through the calculation of $E(S_I)$.

Question 8
This question was a good differentiator — whilst weaker candidates struggled with (i)(c) and (ii),, Full credit was also given for those candidates who checked the probability that T>58.5 in (i)(c).

Question 9
This was well answered.

Question 10
Parts (i) and (ii) of this question were well answered on the whole. Only the best candidates answered part (iii) well.

Question 11
Part (i) of this question was well answered. Only the best candidates scored the second mark in (ii)(b). Although many candidates achieved full marks for the calculation of θ^* *relatively few were awarded both marks for their comments in (iii).*

1 (i) The policyholder must have an interest in the risk being insured, to distinguish between insurance and a wager.

The risk must be of a financial and reasonably quantifiable nature.

(ii) Pecuniary loss — protects against bad debts or other failure of a third party.

Fidelity guarantee — protects against losses caused by dishonest actions of employees.

Business interruption cover — protects against losses made as a result of not being able to conduct business.

2 (i) Two time series X, Y are cointegrated if X and Y are $I(1)$ random processes and there exists a non-zero vector (α, β) such that $\alpha X + \beta Y$ is stationary.

$I(1)$ means that ∇X and ∇Y are stationary
(α, β) is called a cointegrating vector

(ii) Examples:

One of the processes is driving the other
Both are being driven by the same underlying process

3 The claim amount, Z, is given by

$$Z = \begin{cases} 0 & Y < 100 \\ Y - 100 & Y \geq 100 \end{cases}$$

$$E[Z] = \int_{100}^{2000} (x - 100) f(x)\, dx + 1900 P(X \geq 2000)$$

$$= \int_{100}^{2000} x\lambda e^{-\lambda x} dx - 100\int_{100}^{2000} \lambda e^{-\lambda x} dx + 1900 e^{-2000\lambda}$$

where $\lambda = \dfrac{1}{1000}$

Hence $E[Z] = \left[-xe^{-\lambda x}\right]_{100}^{2000} + \int_{100}^{2000} e^{-\lambda x}dx - 100\left[-e^{-\lambda x}\right]_{100}^{2000} + 1900e^{-2000\lambda}$

$$= 100e^{-100\lambda} - 2000e^{-2000\lambda} + \left[-\frac{1}{\lambda}e^{-\lambda x}\right]_{100}^{2000} - 100\left(e^{-100\lambda} - e^{-2000\lambda}\right) + 1900e^{-2000\lambda}$$

$$= \frac{1}{\lambda}\left(e^{-100\lambda} - e^{-2000\lambda}\right) = 1000\left(e^{-0.1} - e^{-2}\right) = 769.5\,.$$

4 (i) **Revenue (£)**

	Cautious	*Best estimate*	*Optimistic*
Slots	840,000	1,200,000	1,380,000
Dice	1,680,000	2,400,000	2,760,000
Cards	2,240,000	3,200,000	3,680,000

Costs (£)

	Cautious	*Best estimate*	*Optimistic*
Slots	1,550,000	1,550,000	1,550,000
Dice	1,850,000	1,850,000	1,850,000
Cards	2,450,000	2,450,000	2,450,000

Profit (£)

	Cautious	*Best estimate*	*Optimistic*
Slots	-710,000	-350,000	-170,000
Dice	-170,000	550,000	910,000
Cards	-210,000	750,000	1,230,000

Slots is dominated by Dice and Cards, so could be left out from here on.

(ii)

	Maximum Cost
Slots	-710,000
Dice	-170,000
Cards	-210,000

The minimax decision is Dice.

(iii)

	Expected Profit
Slots	-404,000
Dice	442,000
Cards	606,000

The highest expected profit comes from Cards.

5 The cumulative cost of claims paid is (Figures in £000s):

	Development year		
Accident Year	*0*	*1*	*2*
2004	4,144	4,838	5,021
2005	4,767	5,599	
2006	5,903		

The number of accumulated settled claims is as follows:

	Development year			
Accident Year	*0*	*1*	*2*	*Ult*
2004	581	656	684	684
2005	626	697		727
2006	674			788

Grossing up factors for claim numbers

	Development year		
Accident Year	*0*	*1*	*2*
2004	0.849	0.959	1
2005	0.861	0.959	
2006	0.855		

Average cost per settled claim

	Development year			
Accident Year	*0*	*1*	*2*	*Ult*
2004	7.133	7.375	7.341	7.341
2005	7.615	8.033		7.996
2006	8.758			9.104

Grossing up factors for average claim amounts

	Development year		
Accident Year	*0*	*1*	*2*
2004	0.972	1.005	1.000
2005	0.952	1.005	
2006	0.962		

The total ultimate loss is therefore:

Accident Year	*ACPC*	*Claim Numbers*	*Projected Loss*
2004	7.341	684	5,021
2005	7.996	727	5,813
2006	9.104	788	7,174
			18,008

Claims paid to date	16,523
Outstanding claims	1,485

Assumptions:

Claims fully run-off by end of development year 3.
Projections based on simple average of grossing up factors.
Number of claims relating to each development year are a constant proportion of total claim numbers from the origin year.
Similarly for claim amounts i.e. same proportion of total claim amount for origin year.

6 (i) The requirement to calculate cos and sin is time consuming for a computer. The Polar Method avoids this by using the acceptance-rejection method.

(ii) (a) We must transform the values to a log-normal distribution with the appropriate mean and variance by calculating

$$X = e^{5+2Z}$$

The required value is $Y = \sqrt{X} = \sqrt{915.1} = 30.25$

(b) We require $n > \dfrac{z_{\alpha/2}^2 \tau^2}{\varepsilon^2} = \dfrac{1.645^2 \times 26.3}{1} = 71.17$

Hence, 72 simulations are required.

7

$$\begin{aligned} E(S) &= E(N) \times E(X_1) \\ &= 2 \times (1 \times 1/3 + 2 \times 2/3) \\ &= 10/3 \end{aligned}$$

and $S = S_I + S_R$.

We will calculate directly the distribution of S_I.

$$P(S_I = 0) = P(N = 0) = e^{-2} = 0.13534$$

$$P(S_I = 1) = P(N = 1)P(X_1 = 1) = e^{-2}\frac{2^1}{1!} \times \frac{1}{3} = 0.09022$$

$$\begin{aligned} P(S_I = 2) &= P(N = 1)P(X_1 = 2) + P(N = 2)P(X_1 = 1)P(X_2 = 1) \\ &= e^{-2}\frac{2^1}{1!} \times \frac{2}{3} + e^{-2}\frac{2^2}{2!} \times \frac{1}{3} \times \frac{1}{3} \\ &= 0.21052 \end{aligned}$$

$$P(S_I = 3) = 1 - 0.13534 - 0.09022 - 0.21052 = 0.56392$$

$$\begin{aligned} E(S_I) &= 0 \times 0.13534 + 1 \times 0.09022 + 2 \times 0.21052 + 3 \times 0.56392 \\ &= 2.20303 \end{aligned}$$

and hence

$$E(S_R) = E(S - S_I) = E(S) - E(S_I) = 10/3 - 2.20303 = 1.1303.$$

8 (i) (a) Magnitude of the residuals increases over time, suggesting that the variance is increasing over time.

More positive than negative residuals suggesting there is drift in the process.

(b) If y_i ($i = 1, 2, \ldots, n$) is a sequence of numbers, it has a turning point at time k if either $y_{k-1} < y_k$ and $y_k > y_{k+1}$, or $y_{k-1} > y_k$ and $y_k < y_{k+1}$.

(c) Let T represent the number of turning points, and let N=100 be the number of data points. Then

$$E(T) = 2/3(N-2) = 2/3(100-2) = 65.333$$
$$Var(T) = (16N-29)/90 = 17.45556 = 4.178^2$$
$$P(T \geq 59) \approx P(N(65.333, 4.178^2) > 59.5)$$
$$= P(N(0,1) > \frac{65.333-59.5}{4.178}) = P(N(0,1) > -1.396) = 0.919$$

This is a two-sided test so there is approximately a 16% chance of getting such an extreme number of turning points. This value is not significant at the 5% level, and so this test gives no significant evidence to suggest that the residuals are not a white noise process.

Full credit should also be given for calculating a confidence interval and checking if 59 is in this.

(ii) $X_{100}(1) = 5 + 0.9(X_{100} - 5) + 0 + 0.5e_{99} = 5 + 0.9(7-5) - 0.5 \times 0.7 = 6.45$

$X_{100}(2) = 5 + 0.9(X_{100}(1) - 5) + 0.5e_{100} = 5 + 0.9(6.45 - 5) + 0.5 \times 1.4 = 7.005$

$X_{100}(3) = 5 + 0.9(X_{100}(2) - 5) + 0.5e_{101} = 5 + 0.9(7.005 - 5) = 6.8045$

9 (i)

	£	Starting level (Year 0) 0%	25%	40%
Year 1	Premium if claim in year 0	1,000	1,000	750
	Premium if no claim in year 0	750	600	600
	Saving	250	400	150
Year 2	Premium if claim in year 0	750	750	600
	Premium if no claim in year 0	600	600	600
	Saving	150	150	0
	Claim threshold	400	550	150

(ii) $P(\mathrm{X} > 400) = P(\log X > \log 400) = P(\mathrm{Z} > (\log 400 - 6)/3.33) = P(\mathrm{Z} > -0.0026)$
$= 1 - 0.4990 = 0.5010$

$P(X > 550) = P(\log X > \log 550) = P(Z > (\log 550 - 6)/3.33)$
$= P(\mathrm{Z} > 0.0931) = 1 - 0.5371 = 0.4629$

$P(X > 150) = P(\log X > \log 150) = P(Z > (\log 150 - 6)/3.33)$
$= P(\mathrm{Z} > -0.2971) = 0.6168$

$0.12 \times 0.5010 = 0.0601$
$0.12 \times 0.4629 = 0.0556$
$0.12 \times 0.6168 = 0.0740$

Hence the transition matrix is

$$\begin{bmatrix} 0.0601 & 0.9399 & 0 \\ 0.0556 & 0 & 0.9444 \\ 0 & 0.0740 & 0.9260 \end{bmatrix}$$

(iii) $\underline{\pi} = \underline{\pi}\, P$

$\pi_0 = 0.0601\,\pi_0 + 0.0556\,\pi_{25}$
$\pi_{25} = 0.9399\,\pi_0 + 0.0740\,\pi_{40}$
$\pi_{40} = 0.9444\,\pi_{25} + 0.9260\,\pi_{40}$
$\pi_{25} = 16.905\,\pi_0$
$\pi_{40} = 215.740\,\pi_0$
$\pi_0 + \pi_{25} + \pi_{40} = 1$

Hence

$\underline{\pi} = (0.0043, 0.0724, 0.9234)$

(iv) Stationary distribution if policyholder claims after a loss

$$P = \begin{bmatrix} 0.12 & 0.88 & 0 \\ 0.12 & 0 & 0.88 \\ 0 & 0.12 & 0.88 \end{bmatrix}$$

$$\begin{aligned} \pi_0 &= 0.12\,\pi_0 + 0.12\,\pi_{25} \\ \pi_{25} &= 0.88\,\pi_0 + 0.12\,\pi_{40} \\ \pi_{40} &= 0.88\,\pi_{25} + 0.88\,\pi_{40} \end{aligned}$$

$$\begin{aligned} \pi_{25} &= 7.333\,\pi_0 \\ \pi_{40} &= 7.333\,\pi_{25} = 53.778\,\pi_0 \\ \pi_0 + \pi_{25} + \pi_{40} &= 1 \end{aligned}$$

Hence:

$\underline{\pi} = (0.0161, 0.1181, 0.8658)$

(v) Award 1 mark for any sensible comment on the reduction in the number of policyholders in the lower discount categories (or increase in the higher discount categories)

10 (i) (a) $$f(y) = \exp\left[-\frac{y\alpha}{\mu} - \alpha\log\mu + (\alpha-1)\log y + \alpha\log\alpha - \log\Gamma(\alpha)\right]$$

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

which is in the form of an exponential family.

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

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

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

(b) The mean and variance of the distribution are $b'(\theta)$ and $a(\varphi)b''(\theta)$

$$b'(\theta) = -\frac{1}{\theta} = \mu$$

which confirms that the mean is μ.

$$b''(\theta) = \frac{1}{\theta^2} = \mu^2$$

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

hence the variance is $\frac{1}{\alpha}\mu^2$, as required

(ii) A factor is categorical e.g. male/female.

For a continuous covariate, the value is included. For example, if x is a continuous covariate, a main effect would be $\alpha + \beta x$.

(iii) (a) The linear predictor has the form

$$\alpha_i + \beta_j + \gamma x$$

where α_i is the factor for policyholder gender (i = 1,2)
β_j is the factor for vehicle rating group
x is the policyholder age

($\alpha_1 = 0$, $\beta_1 = 0$)

(b) The linear predictor becomes

$$\alpha_i + \beta_j + \gamma_i x$$

11 (i) (a) The likelihood is given by

$$L \propto \theta^{x_1}(1-\theta)\cdots\theta^{x_n}(1-\theta)$$
$$= \theta^{x_1+\cdots x_n}(1-\theta)^n$$

and so the log-likelihood is given by

$$l = LogL = (x_1+\cdots+x_n)\log\theta + n\log(1-\theta)$$

and differentiating gives

$$\frac{dl}{d\theta} = \frac{x_1+\cdots+x_n}{\theta} - \frac{n}{1-\theta}$$

setting this expression to zero to find the maximum gives

$$\frac{x_1+\cdots+x_n}{\hat\theta} - \frac{n}{1-\hat\theta} = 0$$
$$(x_1+\cdots+x_n)(1-\hat\theta) = n\hat\theta$$
$$(x_1+\cdots+x_n) = (n+x_1+\cdots+x_n)\hat\theta$$
$$\hat\theta = \frac{x_1+\cdots+x_n}{n+x_1+\cdots+x_n}$$

to check this is a maximum, note that

$$\frac{d^2l}{d\theta^2} = -\frac{x_1+\cdots+x_n}{\theta^2} - \frac{n}{(1-\theta)^2} < 0$$

(b) The posterior distribution is given by

$$f(\theta|\underline{x}) \propto g(\underline{x}|\theta)f(\theta)$$
$$= \theta^{x_1+\cdots+x_n}(1-\theta)^n \times \theta^{\alpha-1}(1-\theta)^{\alpha-1}$$
$$= \theta^{\alpha+x_1+\cdots+x_n-1}(1-\theta)^{n+\alpha-1}$$

which is the pdf of a beta distribution with parameters $\alpha+x_1+\cdots+x_n$ and $\alpha+n$.

(c) First note that the prior distribution of θ is Beta with parameters α and α. Hence its mean is

$$\frac{\alpha}{\alpha+\alpha} = 1/2\,.$$

Under Bayesian loss, the estimator is given by the mean of the posterior distribution, which is

$$\theta^* = \frac{\alpha + \sum x_i}{2\alpha + \sum x_i + n} = \frac{\sum x_i}{\sum x_i + n} \times \frac{\sum x_i + n}{2\alpha + \sum x_i + n} + \frac{2\alpha}{2\alpha + \sum x_i + n} \times \frac{1}{2}$$
$$= \hat{\theta} \times Z + (1 - Z)\mu$$

Where $Z = \dfrac{\sum x_i + n}{2\alpha + \sum x_i + n}$ and $\mu = 1/2$ is the prior mean of θ.

(d) As *n* increases, *Z* tends towards 1, and the Bayes estimate approaches the maximum likelihood estimate, as more credibility is put on the data, and less on the prior estimate.

(ii) (a) The variance of the prior distribution is given by:

$$\frac{\alpha^2}{(2\alpha)^2(2\alpha + 1)} = \frac{1}{4(2\alpha + 1)}.$$

(b) Higher values of α result in a lower variance and hence imply greater certainty over the prior value of θ.

In the special case where $\alpha = 1$ the prior distribution is Uniform on [0,1] implying that we have no particular reason to believe that any prior value of θ is more or less likely than any other.

(iii) $\sum x_i = 3 + 3 + 5 = 11$

(a) $\theta^* = \dfrac{5 + 11}{2 \times 5 + 11 + 3} = \dfrac{16}{24} = \dfrac{2}{3} = 0.6667$

(b) $\theta^* = \dfrac{2 + 11}{2 \times 2 + 11 + 3} = \dfrac{13}{18} = 0.7222$

The first set of parameters has greater certainty attached to the prior estimate (i.e. a higher value of α), and therefore the posterior estimate is closer to the mean of the prior distribution (which is 0.5) than in the second case.

END OF MARKING SCHEDULE

EXAMINATION

2 October 2007 (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.

1 An investment actuary notices that the volatility of the price of a particular asset is much higher following a significant change in the price of the asset.

Define an ARCH model and explain what particular properties of the model would make it appropriate for modelling this asset. [5]

2 Two poachers can escape from either the farmyard or the fields of an estate. Two gamekeepers are employed by the estate to catch them. The gamekeepers have three options: both patrol the farmyard (action a_1); one of them patrols the farmyard and the other patrols the fields (action a_2); both gamekeepers patrol the fields (action a_3). The poachers also have three options: both try to escape through the farmyard (option θ_1); one of them tries to escape from the farmyard and the other from the fields (option θ_2); both try to escape over the fields (option θ_3). The number of birds they manage to poach under each combination of θ_i and a_j can be found in the following table:

	a_1	a_2	a_3
θ_1	0	90	120
θ_2	75	0	75
θ_3	120	90	0

Assume the goal of the gamekeepers and the poachers is to minimise and maximise the number of birds poached, respectively.

(i) Show that θ_1 does not dominate θ_2 and vice versa. [1]

(ii) Determine the minimax solution to this problem for the gamekeepers, and state the maximum number of birds that will be poached if they adopt this strategy. [2]

(iii) Given the prior distribution $P(\theta_1) = 0.25$, $P(\theta_2) = 0.35$ and $P(\theta_2) = 0.4$, determine the Bayes solution to the problem for the gamekeepers. [3]
[Total 6]

3 The number of claims, N, in a year on a portfolio of insurance policies has a Poisson distribution with parameter λ. Claims are either large (with probability p) or small (with probability $1 - p$) independently of one another.

Suppose we observe r large claims. Show that the conditional distribution of $N - r \mid r$ is Poisson and find its mean. [7]

4 The table below gives the cumulative incurred claims by year and earned premiums for a particular type of motor policy (Figures in £000s).

Claims paid to date total £15,000,000. The ultimate loss ratio is expected to be in line with the 2003 accident year.

	Development year				
Accident Year	*0*	*1*	*2*	*3*	*Earned Premiums*
2003	3,340	3,750	4,270	4,400	4,800
2004	3,670	4,080	4,590		4,900
2005	3,690	4,290			5,050
2006	4,150				5,200

Ignoring inflation, use the Bornhuetter-Ferguson method to calculate the total reserve required to meet the outstanding claims, assuming that the claims are fully developed by the end of development year 3. [8]

5 Aggregate annual claims on a portfolio of insurance policies have a compound Poisson distribution with parameter λ. Individual claim amounts have an exponential distribution with mean 1.

The insurer calculates premiums using a loading of α (so that the annual premium is $1+\alpha$ times the annual expected claims) and has initial surplus U.

(i) Show that if the first claim occurs at time t, the probability that this claim causes ruin is $e^{-U}e^{-(1+\alpha)\lambda t}$. [3]

(ii) Show that the probability of ruin on the first claim is $\dfrac{e^{-U}}{2+\alpha}$. [4]

(iii) Show that if the insurer wishes to set α such that the probability of ruin at the first claim is less than 1% then it must choose $\alpha > 100e^{-U} - 2$. [2]
[Total 9]

6 Claim sizes (in suitable units) for a portfolio of insurance policies come from a distribution with probability density function

$$f(x) = \begin{cases} axe^{-x^2} & 0 \le x \le 2 \quad a > 0 \\ 0 & \text{otherwise} \end{cases}$$

where $a > 0$ is a constant.

(i) Find a. [2]

(ii) Show that $|f(x)| \le 1$. [2]

(iii) Random numbers have been drawn from a $U(0,1)$ distribution, and are arranged in pairs. The first three pairs are:

0.7413 and 0.4601
0.3210 and 0.6316
0.5069 and 0.0392

Using the rectangle $\{(x, y) | 0 \le x \le 2, \quad 0 \le y \le 1\}$ and the pairs of random numbers in the order given above, use the acceptance-rejection method to generate a single observation from the claim size distribution.
[3]

(iv) Estimate the average number of $U(0,1)$ variables needed to generate a simulation using the approach in (iii). [1]

(v) The results of many thousands of simulations generated in this way are to be used to compare the impact of two possible reinsurance treaties on the insurer's profitability. Explain, with a reason, whether the modeller should use the same simulated claims for each treaty in his model, or whether a new sample for each treaty should be generated. [2]
[Total 10]

7 A no claims discount system has 4 levels. The premiums paid by policyholders in each level are as follows:

Level	*Premium*
0	100%
25	75%
40	60%
50	50%

The rules for moving between the levels are as follows:

- following a claim-free year, a policyholder moves to the next higher level of discount, or remains at 50% discount

- following a year of one or more claims, a policyholder moves to the next lower discount rate or remains at 0% discount

It is assumed that claims occur according to a Poisson process with rate λ per year per policyholder, and that the equilibrium distribution has been reached.

(i) Show that the average premium paid, if the premium paid by a policyholder in level 0 is £500, may be written as

$$500\left(\frac{1+0.75k+0.6k^2+0.5k^3}{1+k+k^2+k^3}\right)$$

where $k=\dfrac{e^{-\lambda}}{1-e^{-\lambda}}$ [5]

(ii) Calculate the average premium paid by policyholders whose claim rate per year is (a) 0.12, (b) 0.24, (c) 0.36. [3]

(iii) Comment on the results in (ii), in relation to the effectiveness of the no claims discount system discriminating between good and bad drivers. [2]
[Total 10]

8 The total claim amount, S, on a portfolio of insurance policies has a compound Poisson distribution with Poisson parameter 50. Individual loss amounts have an exponential distribution with mean 75. However, the terms of the policies mean that the maximum sum payable by the insurer in respect of a single claim is 100.

(i) Find $E(S)$ and Var (S). [7]

(ii) Use the method of moments to fit as an approximation to S:

(a) a normal distribution
(b) a log-normal distribution

[3]

(iii) For each fitted distribution, calculate $P(S > 3{,}000)$. [3]
[Total 13]

9 $y_1, y_2, \ldots, y_n$ are independent, identically distributed observations with probability function given by $f\left(y_i \mid \mu\right) = \dfrac{\mu^{y_i} e^{-\mu}}{y_i!}$.

(i) Show that the log-likelihood may be written as

$$\theta \sum_{i=1}^{n} y_i - nb(\theta) + \text{ terms not depending on } \theta$$

and identify the natural parameter, θ, and the function $b(\theta)$. [3]

(ii) The fitted value for observation y_i is denoted by $\hat{y}_i$.

(a) Write down the Pearson residual for y_i, in terms of y_i and $\hat{y}_i$.

(b) Explain why Pearson residuals are usually not suitable for model checking for the Poisson distribution.

[3]

(iii) Show that the conjugate prior density function for θ is proportional to $\exp\left\{\alpha\theta - \beta e^{\theta}\right\}$, and derive the posterior distribution for this prior. [4]

(iv) Use the identity $E\left[\dfrac{\partial \log f}{\partial \theta}\right] = 0$ (for any density function f) to show that

$E\left[b(\theta)\right] = \dfrac{\alpha}{\beta}$ and $E\left[b(\theta) \mid y_1, y_2, \ldots, y_n\right] = \dfrac{\alpha + \sum_{i=1}^{n} y_i}{\beta + n}$, and comment on these results. [5]
[Total 15]

10 The time series X_t is assumed to be stationary and to follow an ARMA (2,1) process defined by:

$$X_t = 1 + \frac{8}{15} X_{t-1} - \frac{1}{15} X_{t-2} + Z_t - \frac{1}{7} Z_{t-1}$$

where Z_t are independent $N(0,1)$ random variables.

(i) Determine the roots of the characteristic polynomial, and explain how their values relate to the stationarity of the process. [2]

(ii) (a) Find the autocorrelation function for lags 0, 1 and 2.

(b) Derive the autocorrelation at lag k in the form

$$\rho_k = \frac{A}{c^k} + \frac{B}{d^k}$$

[12]

(iii) Determine the mean and variance of X_t. [3]

[Total 17]

END OF PAPER

EXAMINATION

September 2007

Subject CT6 — Statistical Methods
Core Technical

EXAMINERS' REPORT

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.

M A Stocker
Chairman of the Board of Examiners

December 2007

Comments

Overall it is clear that candidates found this a tougher paper than in other recent sittings. In particular, candidates who did not have a firm grasp of the basic statistical material covered in subject CT3 struggled with some of the questions.

Comments on individual questions are as follows:

Q1 *The Examiners had intended this to be a straightforward bookwork question as it is taken largely verbatim from the core reading. However, very few candidates could recall this part of the core reading accurately and therefore most candidates struggled to score any marks.*

Q2 *Well answered. There was a minor typographical error in the question, but virtually all candidates seemed to have understood what was intended. The Examiners allowed any alternative interpretation provided it was clearly stated.*

Q3 *In general, this question was very poorly answered. Most candidates appeared unable to apply Bayes' Theorem to the situation given in the question.*

Q4 *Well answered.*

Q5 *Parts (i) and (iii) were generally well answered, a pleasing improvement relative to similar questions from recent sittings. Only the better candidates were able to complete part (ii).*

Q6 *This was the first time in a number of sittings that this material has been tested. Most of the better candidates coped reasonably well with parts (i), (ii) and (v). Weaker candidates struggled badly with (i) and (ii) which required only some calculus. Part (iii) was actually simpler than many candidates appear to have expected and as a result many solutions were over-complicated and scored poorly.*

Q7 *Well answered.*

Q8 *This question was well answered by the better candidates. Many candidates picked up significant follow through marks in parts (ii) and (iii) despite making numerical errors in part (i).*

Q9 *Part (i) was generally well answered. The remaining parts were tougher, with a number of candidates quoting the results in parts (iii) and (iv) rather than deriving them as instructed.*

Q10 *Parts (i) and (iii) were well answered. Most candidates were able to make some progress with (ii) (a) though a number struggled to derive three correct equations. Only the best candidates were able to answer (ii) (b).*

1 An ARCH(p) model is

$$X_t = \mu + e_t \sqrt{\alpha_0 + \sum_{k=1}^{p} \alpha_k (X_{t-k} - \mu)^2}$$

where e_t are independent $N(0,1)$.

Often, this is used to model $\ln(Z_t/Z_{t-1})$ where Z_t is the asset price.

It can be seen that a large departure in X_{t-k} from μ will result in X_t having a larger variance. This will then result in a large volatility for the asset price.

2 (i) θ_1 would dominate θ_2 (and vice versa) if the amount of birds poached under each of a_1, a_2 and a_3 is higher in each case for θ_1 (θ_2)

θ_1 provides a better outcome for a_2 (90 > 0) and a_3 (120 > 75)

θ_2 provides a better outcome for a_1 (0 < 75)

(ii) Minimax solution for gamekeepers. Worst loss under each option would be:

a_1 = max (0,75,120) = 120
a_2 = max (90,0,90) = 90
a_3 = max (120,75,0) = 120

Therefore, gamekeepers would choose a_2 and lose 90 birds.

(iii) $a_1 = 0.25 \times 0 + 0.35 \times 75 + 0.4 \times 120 = 74.25$
$a_2 = 0.25 \times 90 + 0.35 \times 0 + 0.4 \times 90 = 58.5$
$a_3 = 0.25 \times 120 + 0.35 \times 75 + 0.4 \times 0 = 56.25$

Hence the Bayes decision is a_3.

3

$$P(N-r=k \mid r \text{ big claims}) = P(N=r+k)\times P(of\ r+k \text{ claims } k \text{ are small}) / P(r \text{ big claims})$$

but

$$P(r \text{ big claims}) = \sum_{j=0}^{\infty} e^{-\lambda}\frac{\lambda^{r+j}}{(r+j)!}\times\frac{(r+j)!}{r!\,j!}p^r(1-p)^j$$

$$=\frac{\lambda^r}{r!}p^r\times\sum_{j=0}^{\infty}e^{-\lambda}\frac{\lambda^j}{j!}(1-p)^j$$

$$=\frac{\lambda^r}{r!}p^r e^{-\lambda}e^{\lambda(1-p)}$$

So

$$P(N-r=k \mid r \text{ big claims}) = P(N=r+k)\times P(of\ r+k \text{ claims } k \text{ are small}) / P(r \text{ big claims})$$

$$=\frac{e^{-\lambda}\frac{\lambda^{r+k}}{(r+k)!}\times\frac{(r+k)!}{r!k!}p^r(1-p)^k}{e^{-\lambda}e^{\lambda(1-p)}p^r\lambda^r/r!}$$

$$=e^{-\lambda(1-p)}\frac{\lambda^k(1-p)^k}{k!}$$

which is a probability from a Poisson distribution with parameter $\lambda(1-p)$. Hence conditional mean of $N - r$ is $\lambda(1-p)$.

4 The development factors are:

Accident Year	Development year 0	1	2	3	EP
2003	3,340	3,750	4,270	4,400	4,800
2004	3,670	4,080	4,590		4,900
2005	3,690	4,290			5,050
2006	4,150				5,200
		10,700	7,830	4,270	
		12,120	8,860	4,400	
Development factors		1.1327	1.1315	1.0304	
Cumulative DFs		1.3207	1.1660	1.0304	

Ultimate loss ratio 4,400/4,800 = 0.9167

Estimated ultimate loss for each accident year:

Accident Year	*ULR*	*EP*	*UL*	*Expected claims*	*Claims to be incurred*
2003	0.9167	4,800	4,400	4,400	0
2004	0.9167	4,900	4,492	4,359	133
2005	0.9167	5,050	4,629	3,970	659
2006	0.9167	5,200	4,767	3,609	1,158

Revised ultimate losses are:

Accident Year	*Claims incurred*	*Claims to be incurred*	*Revised UL*
2003	4,400	0	4,400
2004	4,590	133	4,723
2005	4,290	659	4,949
2006	4,150	1,158	5,308
			19,379
Claims to date			15,000
Reserve required			4,379

5 (i) Let X be the size of the first claim, so that X has an exponential distribution with parameter 1. Then for ruin to occur at time t we require $X > U + (1 + \alpha)\lambda t$.

$$P(X > U + (1+\alpha)\lambda t) = \int_{U+(1+\alpha)\lambda t}^{\infty} e^{-x} dx$$

$$= \left[-e^{-x}\right]_{U+(1+\alpha)\lambda t}^{\infty}$$

$$= e^{-U} e^{-(1+\alpha)\lambda t}.$$

[Note that it would be acceptable to quote the cumulative distribution function for the exponential distribution from the tables rather than calculate the integral]

(ii) Let T denote the time until the first claim. Then T has an exponential distribution with parameter λ and

$$P(\text{Ruin at first claim}) = \int_0^{\infty} P(\text{Ruin at first claim} \mid \text{first claim is at } t) \times f_T(t) dt$$

$$= \int_0^{\infty} e^{-U} e^{-(1+\alpha)\lambda t} \lambda e^{-\lambda t} dt$$

$$= \int_0^{\infty} e^{-U} \lambda e^{-(2+\alpha)\lambda t} dt$$

$$= \left[-e^{-U} \frac{\lambda}{(2+\alpha)\lambda} e^{-(2+\alpha)\lambda t}\right]_0^{\infty}$$

$$= \frac{e^{-U}}{2+\alpha}.$$

(iii) We require

$$\frac{e^{-U}}{2+\alpha} < 0.01$$

i.e. $e^{-U} < 0.01 \times (2 + \alpha)$
i.e. $100e^{-U} < 2 + \alpha$.
i.e. $100e^{-U} - 2 < \alpha$.

6 (i) We find a by solving:

$$\int_0^2 f(x)dx = 1$$

$$\int_0^2 axe^{-x^2}dx = 1$$

$$\left[-\frac{a}{2}e^{-x^2}\right]_0^2 = 1$$

$$\frac{-a}{2}(e^{-4} - 1) = 1$$

$$a = \frac{-2}{e^{-4} - 1} = 2.03731$$

(ii) We find the local maximum value of $f(x)$ by differentiation:

$$f'(x) = ae^{-x^2} - 2ax^2e^{-x^2} = ae^{-x^2}(1 - 2x^2)$$

and this derivative is zero when

$$2x^2 = 1$$
$$x = \pm 0.70711$$

and $f(0.70711) = 0.8738$.

Note that $f(0) = 0$ and $f(2) = 0.07463$ so that the maximum value on [0,2] is 0.8738 and so $|f(x)| < 1$ as required.

(iii) Take as our first point $(2\times 0.7413, 0.4601) = (1.4826, 0.4601)$

Now $f(1.4826) = 0.3353$ which is less than 0.4601 so we reject this point as it lies above the graph of $f(x)$.

Take as our second point $(2\times 0.3210, 0.6316) = (0.6420, 0.6316)$

Now $f(0.6420) = 0.86615 > 0.6316$ so this point lies below the graph of $f(x)$ and is therefore acceptable. Our random sample is therefore the x co-ordinate $= 0.6420$.

(iv) The box $0 < x < 2$, $0 < y < 1$ has area 2, and the area under the curve of $f(x)$ is 1 by definition. Therefore we expect half the points to be rejected as they lie above $f(x)$. Hence it will on average take 4 $U(0,1)$ simulations to determine one point using the acceptance-rejection method.

(v) It would be better to use the same simulated claims to evaluate the re-insurance arrangements

This avoids the possibility that the apparent superiority of one arrangement is in fact due to a favourable series of simulated claims

7 (i) Let $\theta = \exp(-\lambda)$

$\underline{\pi} = \underline{\pi}\, P$

$$P = \begin{bmatrix} 1-\theta & \theta & 0 & 0 \\ 1-\theta & 0 & \theta & 0 \\ 0 & 1-\theta & 0 & \theta \\ 0 & 0 & 1-\theta & \theta \end{bmatrix}$$

$$\begin{aligned}
\pi_0 &= (1-\theta)\,\pi_0 + (1-\theta)\,\pi_{25} \\
\pi_{25} &= \theta\,\pi_0 + (1-\theta)\,\pi_{40} \\
\pi_{40} &= \theta\,\pi_{25} + (1-\theta)\,\pi_{50} \\
\pi_{50} &= \theta\,\pi_{40} + \theta\,\pi_{50}
\end{aligned}$$

$$\pi_0 + \pi_{25} + \pi_{40} + \pi_{50} = 1$$

$$\begin{aligned}
\pi_{25} &= \frac{\theta}{1-\theta}\pi_0 &&= k\pi_0 \\
\pi_{40} &= \frac{\theta}{1-\theta}\pi_{25} &&= k^2\pi_0 \\
\pi_{50} &= \frac{\theta}{1-\theta}\pi_{40} &&= k^3\pi_0
\end{aligned}$$

and hence $\pi_0 + k\,\pi_0 + k^2\,\pi_0 + k^3\,\pi_0 = 1$

hence $\pi_0 = \dfrac{1}{1+k+k^2+k^3}$

hence the average premium paid is

$$500 \times \left(\frac{1+0.75k+0.6k^2+0.5k^3}{1+k+k^2+k^3} \right)$$

(ii) (a) $\theta/(1-\theta) = e^{-0.12}/(1-e^{-0.12}) = 7.8433$ Premium = £257.79
(b) $\theta/(1-\theta) = e^{-0.24}/(1-e^{-0.24}) = 3.6866$ Premium = £270.33
(c) $\theta/(1-\theta) = e^{-0.36}/(1-e^{-0.36}) = 2.3077$ Premium = £288.46

(iii) (a) to (b) λ increases by 100% but average premium paid increases only by 4.9%
(b) to (c) λ increases by 50% but average premium paid increases only by 6.7%

The no claims discount system is not effective at discriminating between good and bad drivers.

8 (i) Let the individual loss amounts have distribution *X*. Then

$$E(X) = \int_0^{100} 0.01333xe^{-0.01333x}dx + 100 \times P(X > 100)$$

$$= \left[-xe^{-0.01333x}\right]_0^{100} + \int_0^{100} e^{-0.01333x}dx + 100\int_{100}^{\infty} 0.01333e^{-0.01333x}dx$$

$$= -100e^{-1.333} + \left[-75e^{-0.01333x}\right]_0^{100} + 100\left[-e^{-0.01333x}\right]_{100}^{\infty}$$

$$= -100e^{-1.333} - 75e^{-1.333} + 75 + 100e^{-1.333}$$

$$= 55.2302$$

Hence $E(S) = 50 \times 55.2302 = 2761.5$

$$E(X^2) = \int_0^{100} 0.01333x^2 e^{-0.01333x} dx + 100^2 P(X > 100)$$

$$= \left[-x^2 e^{-0.01333x}\right]_0^{100} + \int_0^{100} 2xe^{-0.01333x} dx + 100^2 e^{-1.333}$$

$$= -100^2 e^{-1.333} + \left[-\frac{2x}{0.01333} e^{-0.01333x}\right]_0^{100} + \int_0^{100} \frac{2}{0.01333} e^{-0.01333x} dx + 100^2 e^{-1.333}$$

$$= -\frac{200}{0.01333} e^{-1.333} + \left[-\frac{2}{0.01333^2} e^{-0.01333x}\right]_0^{100}$$

$$= -\frac{200}{0.01333} e^{-1.333} - \frac{2}{0.01333^2} e^{-1.333} + \frac{2}{0.01333^2}$$

$$= 4330.6$$

and so

$$Var(S) = 50 \times 4330.6 = 216529 = (465.33)^2$$

(ii) (a) The normal distribution is $N(2761.5, 465.33^2)$

(b) The Log-Normal distribution has parameters μ and σ with

$$E(S) = e^{\mu + \sigma^2/2}$$
$$Var(S) = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1) = E(S)^2 \times (e^{\sigma^2} - 1)$$

So substituting gives

$$216529 = 2761.5^2 \times (e^{\sigma^2} - 1)$$

$$e^{\sigma^2} = \frac{216529}{2761.5^2} + 1 = 1.028394$$

$$\sigma^2 = \log(1.028394) = 0.027998$$

$$\sigma = 0.167327$$

And now we can substitute for σ to give

$$2761.5 = e^{\mu + 0.027998/2}$$

$$\mu = \log(2761.5) - 0.027998/2 = 7.90953$$

(iii) Using the Normal distribution:

$$P(N(2761.5, 465.33^2) > 3000) = P\left(N(0,1) > \frac{3000 - 2761.5}{465.33}\right)$$

$$= P(N(0,1) > 0.51) = 1 - 0.69497 = 0.30503$$

From tables.

Using the log-normal distribution,

$$P(\log N(7.90953, 0.167327^2) > 3000) = P(N(7.90953, 0.167327^2) > \log(3000))$$

$$= P(N(0,1) > \frac{\log 3000 - 7.90953}{0.167327}).$$

$$= P(N(0,1) > 0.58) = 1 - 0.71904 = 0.28096$$

9 (i) The likelihood is $\prod_{i=1}^{n} f\left(y_i \mid \mu\right) = \prod_{i=1}^{n} \frac{\mu^{y_i} e^{-\mu}}{y_i!}$

and hence the log-likelihood is

$$\log \mu \sum_{i=1}^{n} y_i - n\mu - \ldots \ = \theta \sum_{i=1}^{n} y_i - nb\left(\theta\right) + \text{ terms not depending on } \theta$$

where $\theta = \log\mu$

$b(\theta) = e^{\theta}$

(ii) (a) The Pearson residual is

$$\frac{y_i - \hat{y}_i}{\sqrt{\hat{y}_i}}$$

(b) The Pearson residuals are skewed.

This makes it difficult to assess the fit of the model by eye.

(iii) The conjugate prior has the same θ dependence as the likelihood, which is proportional to $\exp\left\{y\theta - e^{\theta}\right\}$. Hence the conjugate prior is $\exp\left\{\alpha\theta - \beta e^{\theta}\right\}$.

$$
\begin{aligned}
f(\theta \mid y_1, y_2, \ldots, y_n) &\propto f(y_1, y_2, \ldots, y_n \mid \theta)\, f(\theta) \\
&\propto \exp\left\{\theta \sum_{i=1}^{n} y_i - ne^{\theta}\right\} \exp\left\{\alpha\theta - \beta e^{\theta}\right\} \\
&\propto \exp\left\{\theta\left(\alpha + \sum_{i=1}^{n} y_i\right) - (\beta + n)e^{\theta}\right\}
\end{aligned}
$$

(iv) For the prior $\log f = \alpha\theta - \beta e^{\theta}$ and $\dfrac{\partial \log f}{\partial \theta} = \alpha - \beta e^{\theta}$. Hence

$$E\left[\frac{\partial \log f}{\partial \theta}\right] = E\left[\alpha - \beta e^{\theta}\right] = 0, \text{ and so } E\left[e^{\theta}\right] = \frac{\alpha}{\beta}.$$

For the posterior $E\left[\dfrac{\partial \log f}{\partial \theta}\right] = E\left[\alpha + \sum_{i=1}^{n} y_i - (\beta + n)e^{\theta}\right]$, and hence

$$E\left[e^{\theta} \mid y_1, y_2, \ldots, y_n\right] = \frac{\alpha + \sum_{i=1}^{n} y_i}{\beta + n}.$$

Note that $e^{\theta} = \mu$, and the posterior estimate can be written as

$$\frac{\beta}{\beta + n} \times \frac{\alpha}{\beta} + \frac{n}{\beta + n} \times \frac{\sum_{i=1}^{n} y_i}{n} = Z\frac{\alpha}{\beta} + (1 - Z)\frac{\sum_{i=1}^{n} y_i}{n}$$

ie a combination of the prior estimate and the estimate from the data.

10 (i) The characteristic equation is given by:

$$(1 - \frac{8}{15}\lambda + \frac{1}{15}\lambda^2) = (1 - \frac{1}{3}\lambda)(1 - \frac{1}{5}\lambda) = 0$$

which has roots = 3 and 5. They are both greater than 1. Hence, subject to the initial values having appropriate distributions, this implies (weak) stationarity.

(ii) (a)

Firstly, note that $Cov(X_t, Z_t) = 1$ and $Cov(X_t, Z_{t-1}) = 8/15 - 1/7 = 41/105$

We need to generate 3 distinct equations linking γ_0, γ_1 and γ_2

This can be done as follows:

(A)

$$\begin{aligned}\gamma_0 &= Cov(X_t, X_t) = Cov(1 + 8/15X_{t-1} - 1/15X_{t-2} + Z_t - 1/7Z_{t-1}, X_t)\\ &= 8/15\gamma_1 - 1/15\gamma_2 + 1 - 1/7 \times 41/105\\ &= 8/15\gamma_1 - 1/15\gamma_2 + 694/735\end{aligned}$$

(B)

$$\begin{aligned}\gamma_1 &= Cov(X_t, X_{t-1}) = Cov(1 + 8/15X_{t-1} - 1/15X_{t-2} + Z_t - 1/7Z_{t-1}, X_{t-1})\\ &= 8/15\gamma_0 - 1/15\gamma_1 - 1/7\end{aligned}$$

(C)

$$\begin{aligned}\gamma_2 &= Cov(X_t, X_{t-2}) = Cov(1 + 8/15X_{t-1} - 1/15X_{t-2} + Z_t - 1/7Z_{t-1}, X_{t-2})\\ &= 8/15\gamma_1 - 1/15\gamma_0\end{aligned}$$

Next stage is to solve these equations.

Substituting (C) into (A) gives

$$\gamma_0 = (8/15)\gamma_1 - (1/15)((8/15)\gamma_1 - (1/15)\gamma_0) + 694/735$$

so

$$(224/225)\gamma_0 = (112/225)\gamma_1 + 694/735$$

$$\gamma_0 = (1/2)\gamma_1 + 5205/5488$$

Now substituting into (B) gives

$\gamma_1 = 8/15((1/2)\gamma_1 + 5205/5488) - (1/15)\gamma_1 - 1/7$
so
$(4/5)\gamma_1 = 249/686$
$\gamma_1 = 1245/2744 = 0.4537$

And

$\gamma_0 = 1/2 \times 0.4537 + 5205/5488 = 1.1753$
$\gamma_2 = 8/15 \times 0.4537 - 1/15 \times 1.1753 = 0.1636$

Finally,

$$\rho_0 = 1,\ \rho_1 = \frac{\gamma_1}{\gamma_0} = 0.386,\ \rho_2 = \frac{\gamma_2}{\gamma_0} = 0.139$$

(b) $\rho_k = \frac{8}{15}\rho_{k-1} - \frac{1}{15}\rho_{k-2}$ for $k \geq 2$

We will show that the solution has the form:

$$\rho_k = A\left(\frac{1}{3}\right)^k + B\left(\frac{1}{5}\right)^k$$

Substituting the proposed solution into the recurrence relation gives

$$\frac{8}{15}\rho_{k-1} - \frac{1}{15}\rho_{k-2} = \frac{8}{15}\left[A\left(\frac{1}{2}\right)^{k-1} + B\left(\frac{1}{5}\right)^{k-1}\right] - \frac{1}{15}\left[A\left(\frac{1}{3}\right)^{k-2} + B\left(\frac{1}{5}\right)^{k-2}\right]$$

$$= A\left(\frac{1}{3}\right)^k\left(\frac{8}{15}\times 3 - \frac{1}{5}\times 9\right) + B\left(\frac{1}{5}\right)^k\left(\frac{8}{15}\times 5 - \frac{1}{15}\times 25\right)$$

$$= A\left(\frac{1}{3}\right)^k + B\left(\frac{1}{5}\right)^k$$

$$= \rho_k$$

So the solution does have this form.

The values of A and B are fixed by $\rho_0 = 1$, $\rho_1 = 0.386$

$$\therefore A + B = 1$$

$$\frac{1}{3} A + \frac{1}{5} B = 0.386$$

$$\rightarrow \frac{1}{3} A + \frac{1}{5}(1 - A) = 0.386$$

$$A = 1.395$$

$$B = -0.395$$

$$\therefore P_k = 1.395 \left(\frac{1}{3}\right)^k - 0.395 \left(\frac{1}{5}\right)^k$$

(iii) We require mean and variance of X_t which must be normally distributed since Z is normally distributed.

Variance is $\gamma_0 = 1.1753$ from (ii) (a)

$$E(X_t) = 1 + \frac{8}{15} E(X_t) - \frac{1}{15} E(X_t)$$

$$\therefore E(X_t) = \frac{15}{8}$$

END OF EXAMINERS' REPORT

EXAMINATION

8 April 2008 (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 Give two examples of the main types of liability insurance, stating for each example a typical insured peril. [4]

2 A claim amount distribution is normal with unknown mean μ and known standard deviation £50. Based on past experience a suitable prior distribution for μ is normal with mean £300 and standard deviation £20.

(i) Calculate the prior probability that μ, the mean of the claim amount distribution, is less than £270. [1]

(ii) A random sample of 10 current claims has a mean of £270.

(a) Determine the posterior distribution of μ.

(b) Calculate the posterior probability that μ is less than £270 and comment on your answer.

[5]
[Total 6]

3 (i) X and Y are independent Poisson random variables with mean λ. Derive the moment generating function of X, and hence show that $X + Y$ also has a Poisson distribution. [4]

(ii) An insurer has a portfolio of 100 policies. Annual premiums of 0.2 units per policy are payable annually in advance. Claims, which are paid at the end of the year, are for a fixed sum of 1 unit per claim. Annual claims numbers on each policy are Poisson distributed with mean 0.18.

Calculate how much initial capital is needed in order to ensure that the probability of ruin at the end of the year is less than 1%. [4]
[Total 8]

4 $Y_1, Y_2, \ldots, Y_n$ are independent observations from a normal distribution with $E[Y_i] = \mu_i$ and $\mathrm{Var}[Y_i] = \sigma^2$.

(i) Write the density of Y_i in the form of an exponential family of distributions. [2]

(ii) Identify the natural parameter and derive the variance function. [3]

(iii) Show that the Pearson residual is the same as the deviance residual. [4]
[Total 9]

5 The following table shows the claim payments for an insurance company in units of £5,000:

Accident year	Development year 0	1	2	3
2004	410	814	216	79
2005	575	940	281	
2006	814	1066		
2007	1142			

The inflation for a 12 month period to the middle of each year is given as follows:

2005	*2006*	*2007*
5%	5.5%	5.4%

The future inflation from 2007 is estimated to be 8% per annum.

Claims are fully run-off at the end of the development year 3.

Calculate the amount of outstanding claims arising from accidents in year 2007, using the inflation adjusted chain ladder method.

[9]

6 A portfolio of general insurance policies is made up of two types of policies. The policies are assumed to be independent, and claims are assumed to occur according to a Poisson process. The claim severities are assumed to have exponential distributions.

For the first type of policy, a total of 65 claims are expected each year and the expected size of each claim is £1,200.

For the second type of policy, a total of 20 claims are expected each year and the expected size of each claim is £4,500.

(i) Calculate the mean and variance of the total cost of annual claims, S, arising from this portfolio. [3]

The risk premium loading is denoted by θ, so that the annual premium on each policy is $(1+\theta)\times$expected annual claims on each policy. The initial reserve is denoted by u.

A normal approximation is used for the distribution of S, and the initial reserve is set by ensuring that

$$P(S < u + \text{annual premium income}) = 0.975.$$

(ii) (a) Derive an equation for u in terms of θ.

(b) Determine the annual premium required in order that no initial reserve is necessary. [7]

[Total 10]

7 Consider the following model applied to some quarterly data:

$$Y_t = e_t + \beta_1 e_{t-1} + \beta_4 e_{t-4} + \beta_1\beta_4 e_{t-5}$$

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

(i) Express in terms of β_1 and β_4 the roots of the characteristic polynomial of the MA part, and give conditions for invertibility of the model. [2]

(ii) Derive the autocorrelation function (ACF) for Y_t. [5]

For our particular data the sample ACF is:

Lag	*ACF*
1	0.73
2	0.14
3	0.37
4	0.59
5	0.24
6	0.12
7	0.07

(iii) Explain whether these results confirm the initial belief that the model could be appropriate for these data. [3]

[Total 10]

8 The NCD scale policy for an insurance company is:

Level 0	0%
Level 1	25%
Level 2	50%

The premium at the Level 0 is £800. The probability that a policyholder has an accident in a year is 0.2, and it is assumed that a policyholder does not have more than one accident each year.

In the event of a claim free year the policyholder moves to the next higher level of discount in the coming year or remains at Level 2.

In the event of a claim the policyholder moves to the next lower level of discount in the coming year or remains at Level 0.

Following an accident the policyholder decides whether or not to make a claim based on the claim size and the amount of premiums over the period of the next 2 policy years, assuming no more claims are made.

(i) For each discount level, find the minimum claim amount for which the policyholder will make a claim. [2]

(ii) Assuming that the cost of repair for each accident has an exponential distribution with mean £600, calculate the probability that a policyholder makes a claim at each level of discount. [5]

(iii) Write down the transition matrix and calculate the average premium payment for a year when the system has reached the equilibrium. [6]

[Total 13]

9 (i) Describe the difference between *strictly* stationary processes and *weakly* stationary processes. [2]

(ii) Explain why weakly stationary multivariate normal processes are also strictly stationary. [1]

(iii) Show that the following bivariate time series process, $(X_n, Y_n)^t$, is weakly stationary:

$$X_n = 0.5X_{n-1} + 0.3Y_{n-1} + e_n^x$$
$$Y_n = 0.1X_{n-1} + 0.8Y_{n-1} + e_n^y$$

where e_n^x and e_n^x are two independent white noise processes. [5]

(iv) Determine the positive values of c for which the process

$$X_n = (0.5 + c)\, X_{n-1} + 0.3Y_{n-1} + e_n^x$$
$$Y_n = 0.1X_{n-1} + (0.8 + c)\, Y_{n-1} + e_n^y$$

is stationary. [6]

[Total 14]

10 A bicycle wheel manufacturer claims that its products are virtually indestructible in accidents and therefore offers a guarantee to purchasers of pairs of its wheels. There are 250 bicycles covered, each of which has a probability p of being involved in an accident (independently). Despite the manufacturer's publicity, if a bicycle is involved in an accident, there is in fact a probability of 0.1 for each wheel (independently) that the wheel will need to be replaced at a cost of £100. Let S denote the total cost of replacement wheels in a year.

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

$$M_S(t) = \left[\frac{pe^{200t} + 18pe^{100t} + 81p}{100} + 1 - p \right]^{250}.$$ [4]

(ii) Show that $E(S) = 5{,}000p$ and $Var(S) = 550{,}000p - 100{,}000p^2$ [6]

Suppose instead that the manufacturer models the cost of replacement wheels as a random variable T based on a portfolio of 500 wheels, each of which (independently) has a probability of 0.1p of requiring replacement.

(iii) Derive expressions for $E(T)$ and Var(T) in terms of p. [2]

(iv) Suppose $p = 0.05$.

(a) Calculate the mean and variance of S and T.
(b) Calculate the probabilities that S and T exceed £500.
(b) Comment on the differences. [5]

[Total 17]

END OF PAPER

Subject CT6 — Statistical Methods
Core Technical

EXAMINERS' REPORT

April 2008

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.

M A Stocker
Chairman of the Board of Examiners

June 2008

Comments

Comments for individual questions are given after each of the solutions that follow.

1

Type	*Typical perils*
Employers' liability	• Accidents caused by employer negligence • Exposure to harmful substances • Exposure to harmful working conditions
Motor 3rd party liability	• Road traffic accidents
Public liability	• Will relate to the type of policy
Product liability	• Faulty design, manufacture or packaging of product • Incorrect or misleading instructions
Professional Indemnity	• Wrong medical diagnosis, error in medical operation etc.

Comment: Most of the candidates did very well here.

2 (i)

$$P(\mu < 270) = P(N(300, 20^2) < 270)$$
$$= P(N(0,1) < \frac{270-300}{20})$$
$$= P(N(0,1) < -1.5)$$
$$= 1 - 0.93319$$
$$= 0.06681$$

(ii)

(a) Using the result from page 28 of the tables, the posterior distribution of μ is normal, with mean

$$\mu_* = \left(\frac{10\times 270}{50^2} + \frac{300}{20^2}\right) \Big/ \left(\frac{10}{50^2} + \frac{1}{20^2}\right) = £281.54$$

and variance

$$\sigma_*^2 = \frac{1}{\frac{10}{50^2} + \frac{1}{20^2}} = 153.85 = 12.40^2$$

So the posterior distribution of μ is $N(281.54, 12.40^2)$.

(b) The posterior probability required is given by:

$$P(N(281.54, 12.40^2) < 270) = P(N(0,1) < \frac{270 - 281.54}{12.4})$$
$$= P(N(0,1) < -0.931)$$
$$= 1 - (0.9 \times 0.82381 + 0.1 \times 0.82639)$$
$$= 0.1759$$

Comment: The probability that the true mean is less than £270 has risen, as the sample evidence suggests that the true mean is less than the mean of the prior distribution. Nevertheless, the sample size is relatively small, and the variance of the prior distribution is also small, so that a reasonable weight is still given to the prior information.

Comment: Some candidates did not use tables for generating the posterior parameters in (ii)(a). A few gave the correct interpretation in (ii)(b).

3 (i)

$$M_X(t) = E(e^{tX})$$
$$= \sum_{k=0}^{\infty} e^{kt} e^{-\lambda} \frac{\lambda^k}{k!}$$
$$= \sum_{k=0}^{\infty} e^{-\lambda} \frac{(\lambda e^t)^k}{k!}$$
$$= e^{-\lambda} e^{\lambda e^t}$$
$$= e^{\lambda(e^t - 1)}$$

Hence

$$M_{X+Y}(t) = E(e^{t(X+Y)})$$
$$= E(e^{tX})E(e^{tY})$$
$$= M_X(t) M_Y(t)$$
$$= e^{\lambda(e^t-1)} e^{\lambda(e^t-1)}$$
$$= e^{2\lambda(e^t-1)}$$

which is the MGF of a Poisson distribution with parameter 2λ. Hence $X + Y$ is Poisson distributed.

(ii) Using the result above, aggregate claims on the portfolio over the year have a Poisson distribution with parameter 18.

Let the initial capital be U. At the end of the year, the surplus will be $U + 20 - N$ where N is the number of claims.

Now using the tables in the gold book, $P(N <= 28) = 0.9897$, and
$P(N < = 29) = 0.9941$.

So we need U to be large enough that ruin would only occur if there were 30 or more claims. So we need $U + 20 - 29 > 0$.

i.e. $U > 9$.

Comment: Many candidates struggled with (ii) here.

4 (i)

$$f(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_i - \mu_i)^2}{2\sigma^2}\right\}$$

$$\log f = -½\log(2\pi\sigma^2) - \frac{(y_i^2 - 2y_i\mu_i + \mu_i^2)}{2\sigma^2}$$

$$= \frac{y_i\mu_i - ½\mu_i^2}{\sigma^2} - ½\log(2\pi\sigma^2) - ½\frac{y_i^2}{\sigma^2}$$

which is the form of an exponential family of distributions.

(ii) The natural parameter is μ_i

$$\theta_i = \mu_i$$

$$b(\theta_i) = ½\mu_i^2 = ½\theta_i^2$$

$$b'(\theta_i) = \theta_i$$

$$b''(\theta_i) = 1$$

Hence $V(\mu_i) = 1$

(iii) The scaled deviance is

$$2\sum\left[-\frac{(y_i - y_i)^2}{2\sigma^2} - ½\log(2\pi\sigma^2) + \frac{(y_i - \hat{\mu}_i)^2}{2\sigma^2} + ½\log(2\pi\sigma^2)\right]$$

$$= \sum\frac{(y_i - \hat{\mu}_i)^2}{\sigma^2}$$

Hence the deviance residual is

$$\text{sign}(y_i - \hat{\mu})\sqrt{\frac{(y_i - \hat{\mu}_i)^2}{\sigma^2}} = \frac{y_i - \hat{\mu}_i}{\sigma}$$

The Pearson residual is $\dfrac{y_i - \hat{\mu}_i}{\sigma V(\hat{\mu}_i)} = \dfrac{y_i - \hat{\mu}}{\sigma}$

Comment: This was a relatively easy question and many scored full marks in (i) and (iii) here.

5 Multiply the claim payments with the corresponding inflation factors given below:

Development year

2004	1.16757	1.11197	1.05400	1.00000
2005	1.11197	1.05400	1.00000	
2006	1.05400	1.00000		
2007	1.00000			

The resulting table is:

Development year

2004	478.70	905.14	227.66	79.00
2005	639.38	990.76	281.00	
2006	857.96	1066.00		
2007	1142.00			

The inflation adjusted accumulated claim payments in mid 2007 are given below:

Development year

year	0	1	2	3
2004	478.70	1383.84	1611.50	1690.50
2005	639.38	1630.14	1911.14	**2004.83**
2006	857.96	1923.96	**2248.66**	**2358.90**
2007	1142.00	**2853.75**	**3335.38**	**3498.88**

Note only the values of the last row are needed for the answer.

The bolded values show the completed table using the basic chain ladder approach.

The development factors are 2.4989, 1.1688, 1.0490.

For the answer we only need to work with the projected values at the last row as:

$(2853.75 - 1142.00)*1.08+(3335.38 - 2853.75)*1.08^2+(3498.88 - 3335.38)*1.08^3$
$= 2616.43$

2616.43 * 5000 = £13,082,150

Comment: Many candidates scored full marks here. Some missed the conversion of the final figure from units to pounds (i.e. multiplying by £5000).

6 (i) $E[S] = 65 \times 1200 + 20 \times 4500 = 168000$

$\text{Var}[S] = 65 \times 2 \times 1200^2 + 20 \times 2 \times 4500^2$

$= 187{,}200{,}000 + 810{,}000{,}000$

$= 997{,}200{,}000$

(ii) (a) c = annual premium income = $(1 + \theta)\, E[S]$

$$P(S < u + c) \doteqdot \Phi\left(\frac{u+c-E[S]}{\sqrt{\text{Var}[S]}}\right)$$

$$\frac{u+c-E[S]}{\sqrt{\text{Var}[S]}} = 1.96$$

$$u + \theta E[S] = 1.96\sqrt{\text{Var}[S]}$$

$$u = 1.96\sqrt{\text{Var}[S]} - \theta E[S]$$

$$= 61893.8 - 168000\theta$$

(b) $61893.8 - 168000\theta = 0$

$\theta = 0.3684$

No initial reserve is required if $\theta \geq 0.3684$. i.e. when the premium $\geq$ £229,894.

Comment: There were some mixed answers for the second part of this question. Numerical figures varied in (b) due to the rounding of the square root of the Var(S).

7 (i) Using the back shift operator it can be seen that

$$Y_t = (1 + \beta_1 B)\,(1 + \beta_4 B^4)\, e_t.$$

The invertibility conditions are then $|\beta_1| < 1$ and $|\beta_4| < 1$.

(ii) Since $\mathbf{E}(Y_t) = 0$, $\gamma_0 = \mathbf{E}(y_t^2) = \sigma^2(1+\beta_1^2+\beta_4^2+\beta_1^2\beta_4^2) = \sigma^2(1+\beta_1^2)(1+\beta_4^2)$.
Similarly it can be shown that

$$\gamma_1 = \mathbf{E}(Y_t Y_{t-1}) = \sigma^2\beta_1(1+\beta_4^2)$$

$$\gamma_2 = 0$$

$$\gamma_3 = \sigma^2\beta_1\beta_4$$

$$\gamma_4 = \sigma^2\beta_4(1+\beta_1^2)$$

$$\gamma_5 = \gamma_3$$

$$\gamma_k = 0,\ k > 5$$

So the ACF is

$$\rho_1 = \frac{\beta_1}{1+\beta_1^2}$$

$$\rho_2 = 0$$

$$\rho_3 = \rho_5 = \frac{\beta_1\beta_4}{(1+\beta_1^2)(1+\beta_4^2)}$$

$$\rho_4 = \frac{\beta_4}{1+\beta_4^2}$$

$$\rho_k = 0,\ k > 5$$

(iii) Since in general the ratio $\left|\frac{u}{1+u^2}\right| \le 0.5$ then we see that for our model $|\rho_1| < 0.5, |\rho_3| < 0.25, |\rho_4| < 0.5$ and $|\rho_5| < 0.25$. These do not seem to be satisfied by the sample ACF. So the model is not appropriate for such data.

Other observations like those listed below can suffice here:

- $r(2)$ is not zero, and neither are $r(6)$ and $r(7)$.

- $r(3)$ is not close to $r(5)$.

- $r(1)r(4) = 0.43$. This should be similar in value to both $r(3)$ and $r(5)$. Whilst close to $r(3)$ it isn't close to $r(5)$.

Full marks for at least three correct statements.

Comment: There were some easy marks here. With the exception of part (iii), many candidates did well but some dropped many points when the concept of auto-correlation was not clear.

8 (i)

Level	*Prem. If. claim*	*Prem. No claim*	*Difference*
0	800 600	600 400	400
1	800 600	400 400	600
2	600 400	400 400	200

(ii) The claims are exponentially distributed with parameter $\lambda = 1/\mu = 1/600$ and so Pr(loss > u) = exp($-u\,\lambda$).

Since
$P(\text{claim}) = P(\text{accident})\,P(\text{claim}|\text{accident})$

We derive that for a policyholder at Level 0
$P(\text{claim at }0\%) = 0.2P(X > 400) = 0.2\exp(-400/600) = 0.102683$

for Level 1
$P(\text{claim at }25\%) = 0.2P(X > 600) = 0.2\exp(-600/600) = 0.07357589$

and for Level 2
$P(\text{claim at }50\%) = 0.2P(X > 200) = 0.2\exp(-200/600) = 0.1433063$

(iii) The transition matrix will be

$$\begin{pmatrix} 0.1027 & 0.8973 & 0.0000 \\ 0.0736 & 0.0000 & 0.9264 \\ 0.0000 & 0.1433 & 0.8567 \end{pmatrix}$$

$\pi P = \pi$ and hence

$$\pi_0 = 0.1027\pi_0 + 0.0736\pi_1$$
$$\pi_1 = 0.8973\pi_0 + 0.1433\pi_2$$
$$\pi_2 = 0.9264\pi_1 + 0.8567\pi_2$$

The first and the last equations imply

$$\pi_0 = 0.0736 / 0.8973\pi_1 = 0.0820\pi_1$$

and

$$\pi_2 = 0.9264 / 0.1433\pi_1 = 6.4748\pi_1$$

From $\pi_0 + \pi_1 + \pi_2 = 1$ we obtain $\pi_1 = 1/7.5568 = 0.1325$
with $\pi_0 = 0.0820\times0.1325 = 0.0109$ and $\pi_2 = 1 - 0.1325 - 0.0109 = 0.8566$.
The average premium is now:
$800\,\pi_0 + 600\,\pi_1 + 400\,\pi_2 = 800 \times 0.0109 + 600 \times 0.1325 + 400 \times 0.8566 =$ 430.85.

Comment: Generally, very good answers with many candidates scoring full marks.

9 (i) Strictly stationary processes have the property that the distribution of $(X_{t+1}, \dots X_{t+k})$ is the same as that of $(X_{t+s+1}, \dots X_{t+s+k})$ for each t, s and k. For the weakly stationary only the first two moments are needed to satisfy

$$\mathbf{E}(X_t) = \mu \quad \forall t$$

and

$$\text{cov}(X_t, X_{t+s}) = \gamma(s) \quad \forall t, s.$$

(ii) These two definitions coincide for the multivariate normal processes since the normal distribution is characterised by the first two moments only.

(iii) In order to confirm that we need to calculate the eigenvalues of the parameter matrix

$$A = \begin{pmatrix} 0.5 & 0.3 \\ 0.1 & 0.8 \end{pmatrix}.$$

So we need to solve $\det(A - \lambda I) = 0$ which implies the solution of

$$(0.5 - \lambda)(0.8 - \lambda) - 0.03 = 0$$

$$0.37 - 1.3\lambda + \lambda^2 = 0$$

We see that this equation is satisfied for $\lambda_1 = 0.8791288$ and $\lambda_2 = 0.4208712$. Since they are both smaller than 1, the process is stationary.

(iv) The parameter matrix here is $A^c = A + cI$, and the eigenvalues equation is now $\det(A + cI - \lambda I) = 0$ or $\det(A - (\lambda - c)I) = 0$.

So the eigenvalues of A^c are $\lambda_1 + c$ and $\lambda_2 + c$ where λ_i are those of A.
Since λ_i are positive then the required values for c are such that $\lambda_1 + c < 1$ and $\lambda_2 + c < 1$.
Hence $0 < c < 1 - \lambda_1 = 0.1208712$, since λ_1 is the largest of the two.

Comment: This was not the easiest question. Some struggled with (ii), (iii) and (iv). There were quite a few candidates who managed to avoid the calculation of the eigen values of the matrix A by explicitly expressing each X_n and Y_n series as stationary univariate AR(2) processes with some white noise terms.

10 (i) Let N denote the annual number of accidents. Then $N \sim B(250,p)$ and (from the tables) $M_N(t) = (pe^t + 1 - p)^{250}$

If there is an accident, then the total cost of replacement wheels, X, has the following distribution:

Number of wheels requiring replacement	0	1	2
Cost of replacement X	£0	£100	£200
Probability	0.81	0.18	0.01

And $M_X(t) = 0.01e^{200t} + 0.18e^{100t} + 0.81$.

So

$$\begin{aligned} M_S(t) &= M_N(\log M_X(t)) \\ &= (pe^{\log M_X(t)} + 1 - p)^{250} \\ &= (pM_X(t) + 1 - p)^{250} \\ &= (p(0.01e^{200t} + 0.18e^{100t} + 0.81) + 1 - p)^{250} \\ &= \left(\frac{pe^{200t} + 18pe^{100t} + 81p}{100} + 1 - p\right)^{250} \end{aligned}$$

(ii) $E(S) = M_S'(0)$

$$M_S'(t) = 250 \times \left(\frac{pe^{200t} + 18pe^{100t} + 81p}{100} + 1 - p\right)^{249} \times (2pe^{200t} + 18pe^{100t})$$

$$E(S) = M_S'(0) = 250 \times 1 \times 20p = 5000p$$

$$E(S^2) = M_S''(0)$$

$$M_S''(t) = 250 \times 249 \times \left(\frac{pe^{200t} + 18pe^{100t} + 81p}{100} + 1 - p \right)^{248} \times (2pe^{200t} + 18pe^{100t})^2$$

$$+250 \times \left(\frac{pe^{200t} + 18pe^{100t} + 81p}{100} + 1 - p \right)^{249} \times (400pe^{200t} + 1800pe^{100t})$$

$$M_S''(0) = 250 \times 249 \times (20p)^2 + 250 \times 1 \times 2200p = 24,900,000p^2 + 550,000p$$

$$Var(S) = E(S^2) - E(S)^2 = 24,900,000p^2 + 550,000p - (5000p)^2 = 550,000p - 100,000p^2.$$

Alternatively, we note that

$E(N) = 250p$ and $\text{Var}(N) = 250p\,(1 - p)$
$E(X) = 0.01 \times 200 + 0.18 \times 100 = 20$
$\text{Var}(X) = 40000 \times 0.01 + 10000 \times 0.18 - 20 \times 20 = 1800$

and

$E(S) = E(N)\,E(X) = 250p \times 20 = 5000p$
$\text{Var}(S) = E(N)\,\text{Var}(X) + \text{Var}(N) \times E(X) \times E(X)$
$\text{Var}(S) = 250p \times 1800 + 250p(1 - p) \times 20 \times 20$
$= 450,000p + 100,000p\,(1 - p) = 550,000p - 100,000\,p^2$.

(iii) Let W denote the total number of wheels needing replacement. Then $W \sim B(500, 0.1p)$ and $T = 100W$

Hence

$E(T) = 100E(W) = 100 \times 500 \times 0.1p = 5000p$

and

$$Var(T) = Var(100W) = 100^2 Var(W) = 100^2 \times 500 \times 0.1p \times (1 - 0.1p)$$
$$= 500,000p(1 - 0.1p)$$

(iv) (a) If $p = 0.05$ then

$E(S) = E(T) = 250.$

$Var(S) = 550{,}000 \times 0.05 - 100{,}000 \times 0.05 \times 0.05 = 27{,}250 = 165.08^2$
$Var(T) = 500{,}000 \times 0.05 \times 0.995 = 24{,}875 = 157.72^2$

(b) And so assuming both can be approximated by a normal distribution, and allowing for a continuity correction

$$P(S > 550) \approx P(N(0,1) > \frac{550-250}{165.08}) = P(N(0,1) > 1.817)$$
$$= 1-(0.7\times 0.96562 + 0.3\times 0.96485)$$
$$= 0.034611$$

$$P(T > 500) \approx P(N(0,1) > \frac{550-250}{157.72}) = P(N(0,1) > 1.902)$$
$$= 1-(0.8\times 0.97128 + 0.2\times 0.97193)$$
$$= 0.02859$$

(c) The two distributions have the same mean, but different variances – the variance of *S* is slightly higher than that of *T*. This leads to a higher probability for such a loss under *S* than under the approximation *T*. Though the probabilities are both small in absolute terms, that for *S* is 20% higher than that for *T*. Effectively, fewer accidents are needed under *S* to give a high loss, because each accident can lead to two wheels being replaced, whereas under *T* only one wheel can be damaged per accident.

Comment: This was a challenging question with many students scoring well here and some trying to fudge the answers for (i).

END OF EXAMINERS' REPORT

EXAMINATION

16 September 2008 (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 Claim amounts on a portfolio of insurance policies have an unknown mean μ. Prior beliefs about μ are described by a distribution with mean μ_0 and variance σ_0^2. Data are collected from n claims with mean claim amount $\bar{x}$ and variance s^2. A credibility estimate of μ is to be made, of the form

$$Z\bar{x} + (1 - Z)\mu_0.$$

Suggestions for the choice of Z are:

A $\dfrac{n\sigma_0^2}{n\sigma_0^2 + s^2}$

B $\dfrac{n\sigma_0^2}{n\sigma_0^2 + n}$

C $\dfrac{\sigma_0^2}{n + \sigma_0^2}$

Explain whether each suggestion is an appropriate choice for Z. [4]

2 Write down the general statistical model for the run-off triangle claim data and explain the terms used. [5]

3 An insurer is considering whether to outsource its advertising. If it decides not to outsource, it expects to spend £1,400,000 next year on its advertising, and believes that this would result in a portfolio of 100,000 policies. It has received quotations from two different companies for outsourcing its advertising.

Company A would cost £2,100,000 per year, and believes that this would result in the business expanding to 125,000 policies.

Company B would cost £3,000,000 per year, and believes that this would expand the business to 140,000 policies.

At present, each policy returns a profit to the company of £30 per year; but this is not guaranteed in the future. The company has assessed that it will stay at this level with probability 0.6, but could reduce to £20 per year with probability 0.25, or increase to £40 per year with probability 0.15.

(i) Explain which of the three options can be immediately discarded. [1]

(ii) Determine the Bayes solution to the problem of maximising the profit to the company over the coming year. [5]
[Total 6]

4 An insurance company provides warranties for a certain electrical gadget. At the start of 2006 there were 4,500 gadgets under warranty, each of which has a probability q of suffering complete failure in 2006 (independently between gadgets). The prior distribution of q is beta with mean 0.015 and standard deviation 0.005. Given that 58 gadgets suffer a complete failure in 2006, determine the posterior distribution of q. [6]

5 The table below shows the cumulative values (in units of £1,000) of incurred claims on a portfolio of an insurance company:

	Development year		
Underwriting year	*1*	*2*	*3*
2005	3,541	7,111	9,501
2006	2,949	6,850	
2007	3,894		

The estimated loss ratio for 2006 and 2007 is 87% and the respective premium income (also in units of £1,000) is:

	Premium income
2005	11,041
2006	11,314
2007	12,549

Given that the total of claims paid to date is £20,103,000, calculate the reserve for this portfolio using the Bornhuetter-Ferguson method. [7]

6 Consider the ARCH(1) process

$$X_t = \mu + e_t\sqrt{\alpha_0 + \alpha_1(X_{t-1} - \mu)^2}$$

where e_t are independent normal random variables with variance 1 and mean 0. Show that, for $s = 1, 2, \ldots, t\text{-}1$, X_t and X_{t-s} are:

(i) uncorrelated. [5]

(ii) not independent. [3]

[Total 8]

 PLEASE TURN OVER

7 (i) Let $U_1, U_2, \ldots, U_n$ be independent random numbers generated from a U(0,1) distribution. Write down the Monte Carlo estimator, $\hat{\theta}$, for the integral

$$\theta = \int_0^1 (e^x - 1)dx. \qquad [1]$$

(ii) Determine the variance of the estimator $\hat{\theta}$ in (i). [4]

(iii) Calculate the smallest value of n for which the estimator $\hat{\theta}$ has absolute error less than 0.1 with probability 90%. [4]

[Total 9]

8 An insurer has issued two five-year term assurance policies to two individuals involved in a dangerous sport. Premiums are payable annually in advance, and claims are paid at the end of the year of death.

Individual	*Annual Premium*	*Sum Assured*	*Annual Prob (death)*
A	100	1,700	0.05
B	50	400	0.1

Assume that the probability of death is constant over each of the five years of the policy. Suppose that the insurer has an initial surplus of U.

(i) Define what is meant by $\psi(U)$ and $\psi(U,t)$. [2]

(ii) Assuming U = 1,000

(a) Determine the distribution of $S(1)$, the surplus at the end of the first year, and hence calculate $\psi(U,1)$.

(b) Determine the possible values of $S(2)$ and hence calculate $\psi(U,2)$.

[8]

[Total 10]

9 A motor insurance company applies a NCD scale policy of discount levels

Level 0	0%
Level 1	20%
Level 2	50%

Following a claim-free year, the policyholder moves to the next higher level (or remains at Level 2). If one or more claims are made, the policyholder moves to the next lower level (or remains at Level 0).

The probability of a policyholder not making a claim in a policy year is $1 - p$.

The annual premium at Level 0 is £600.

(i) Derive, in terms of p, the expected proportions of policyholders at each discount level assuming that the system is in equilibrium . [7]

(ii) Calculate the average premium paid in the stable state for the particular values $p = 0.1$ and $p = 0.3$ and comment on the results. [3]
[Total 10]

10 From a sample of 50 consecutive observations from a stationary process, the table below gives values for the sample autocorrelation function (ACF) and the sample partial autocorrelation function (PACF):

Lag	*ACF*	*PACF*
1	0.854	0.854
2	0.820	0.371
3	0.762	0.085

The sample variance of the observations is 1.253.

(i) Suggest an appropriate model, based on this information, giving your reasoning. [2]

(ii) Consider the AR(1) model

$$Y_t = a_1 Y_{t-1} + e_t,$$

where e_t is a white noise error term with mean zero and variance σ^2.

Calculate method of moments (Yule-Walker) estimates for the parameters of a_1 and σ^2 on the basis of the observed sample. [4]

(iii) Consider the AR(2) model

$$Y_t = a_1 Y_{t-1} + a_2 Y_{t-2} + e_t,$$

where e_t is a white noise error term with mean zero and variance σ^2.

Calculate method of moments (Yule-Walker) estimates for the parameters of a_1 , a_2 and σ^2 on the basis of the observed sample. [7]

(iv) List two statistical tests that you should apply to the residuals after fitting a model to time series data. [2]

[Total 15]

11 Losses on a portfolio of insurance policies in 2006 are assumed to have an exponential distribution with parameter λ. In 2007 loss amounts have increased by a factor k (so that a loss incurred in 2007 is k times an equivalent loss incurred in 2006).

(i) Show that the distribution of loss amounts in 2007 is also exponential and determine the parameter of the distribution. [3]

Over the calendar years 2006 and 2007 the insurer had in place an individual excess-of-loss reinsurance arrangement with a retention of M. Claims paid by the insurer were:

2006: 4 amounts of M and 10 claims under M for a total of 13,500.
2007: 6 amounts of M and 12 claims under M for a total of 17,000.

(ii) Show that the maximum likelihood estimate of λ is:

$$\hat{\lambda} = \frac{22}{13{,}500 + \dfrac{17{,}000}{k} + 4M + \dfrac{6M}{k}}$$

[7]

(iii) The insurer is negotiating a new reinsurance arrangement for 2008. The retention was set at 1600 when the current arrangement was put in place in 2006. Loss inflation between 2006 and 2007 was 10% (i.e. k = 1.1) and further loss inflation of 5% is expected between 2007 and 2008.

(a) Use this information to calculate $\hat{\lambda}$.

(b) The insurer wishes to set the retention M' for 2008 such that the expected (net of re-insurance) payment per claim for 2008 is the same as the expected payment per claim for 2006. Calculate the value of M', using your estimate of λ from (iii)(a). [10]

[Total 20]

END OF PAPER

Faculty of Actuaries **Institute of Actuaries**

Subject CT6 — Statistical Methods
Core Technical

EXAMINERS' REPORT

September 2008

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

November 2008

Comments on individual questions

Q1 *There was a wide variety in the standard of answers to this question – it was generally well answered by the candidates who passed whilst weaker candidates struggled to relate the possible choices to the desirable characteristics of a credibility estimate.*

Q2 *This bookwork questions was generally well answered.*

Q3 *Well answered.*

Q4 *Many weaker candidates struggled to calculate* α *and* β. *Most were able to derive the posterior distribution.*

Q5 *Well answered.*

Q6 *Candidates found this (and Q7) the hardest on the paper, with very few able to make much headway. This was a little disappointing, since although the question was phrased around ARCH models it required only the standard definition on covariance and a little algebra.*

Q7 *Along with Q6, candidates found this the toughest question on the paper. Very few were able to write down the required estimator and therefore made no progress at all. Many candidates failed to recognise that the estimator had to be a function of the uniform random variables rather than a fixed number.*

Q8 *This question was a good differentiator – the candidates who passed were able to take a systematic approach to what was a fairly straightforward situation. Many weaker candidates failed to give an accurate definition of the probabilities in part (i) and a large number chose to make an approximation to the distribution in part(ii) when an exact calculation was simpler and quicker (and was what the question asked for).*

The notation used in this question differed slightly from that used in the core reading, which was unfortunate. Whilst this caused no problems for the majority of candidates, the examiners made allowance in those cases where the notation caused confusion.

Q9 *This question was generally well answered.*

Q10 *This question was well answered by the better candidates. Whilst most well prepared candidates were able to derive the relevant Yule-Walker equations, only the better candidates were able to move from these to estimates for the parameters.*

Q11 *Another good differentiator. Most candidates were able to derive the maximum likelihood estimate. Only the better candidates made much headway with part (iii).*

1 A This is an appropriate choice – the larger the value of n, the closer Z is to 1 and the more weight is placed on the data. Furthermore, high values of the variance of the prior (indicating uncertainty in the prior) lead to higher values of Z and so more weight on the sample data. Finally, high variance in the sample reduces the value of Z and places more reliance on the prior.

B This is not appropriate – the value is a constant independent of the size of the sample, whereas we would expect more weight on the sample the larger the sample.

C This is not appropriate – the greater the value of n, the lower the value of Z and the less weight is put on the sample. This is the reverse of what we would expect.

2 $C_{ij} = R_j S_i X_{i+j} + E_{ij}$

where

C_{ij} is the incremental claims from origin year i to j years ahead.
R_j is the development factor for year j independent of origin year i.
S_i is a parameter varying by origin year i, representing exposure (e.g. total claims incurred).
X_{i+j} is a parameter varying by calendar year, for example representing inflation.
E_{ij} is the error term.

3 (i) Company B can be discounted immediately because it is dominated by both of the other two options.

(ii) **Profit Table (in millions)**

	£20	£30	£40	*E[Profit]*
Current	0.6	1.6	2.6	1.5
Company A	0.4	1.65	2.9	1.525

The insurer should choose Company A.

4 Suppose that q has a $beta(\alpha,\beta)$ distribution as per the tables, and let X denote the number of failures in 2006 so that X has a $B(4500,q)$ distribution. Then

$$\begin{aligned} f(q|X) &\propto f(q)\times f(X|q) \\ &\propto q^{\alpha-1}(1-q)^{\beta-1}\times q^{x}(1-q)^{n-x} \\ &\propto q^{\alpha+x-1}(1-q)^{\beta+n-x-1} \end{aligned}$$

So the posterior distribution of q is beta with parameters $\alpha+x$ and $\beta+n-x$.

In our case, the parameters of the prior distribution are given by

$$\frac{\alpha}{\alpha+\beta}=0.015 \text{ and } \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}=0.005^2.$$

$$\alpha=0.015(\alpha+\beta)$$
$$0.985\alpha=0.015\beta$$
$$\alpha=\frac{3}{197}\beta$$

And substituting into the second equation gives

$$\frac{\frac{3}{197}\beta^2}{\left(\frac{200}{197}\beta\right)^2\left(1+\frac{200}{197}\beta\right)}=0.005^2$$

$$\frac{3}{197}=0.005^2\times\left(\frac{200}{197}\right)^2\times\left(1+\frac{200}{197}\beta\right)$$

$$591=1+\frac{200}{197}\beta$$

$$\beta=590\times\frac{197}{200}=\frac{11623}{20}=581.15$$
$$\alpha=\frac{3}{197}\times\frac{11623}{20}=8.85$$

So the revised parameters are given by:

$$\alpha_*=8.85+58=66.85$$
$$\beta_*=581.15+4500-58=5023.15$$

5 The main calculations are reported in the table:

Year	*Initial ultimate. Loss*	*r*	*f*	*1 - 1/f*	*Emerging Liability*
2006	9843.18	1.336099	1.336099	0.251552	2476.08
2007	10917.63	2.151156	2.874157=1.336099×2.151156	0.652072	7119.08

The second column (IUL) is obtained as 87% of Premium income values. The third column reports the development factors

7111/9501 = 1.336099 and (7111 + 6850) / (3541 + 2949) = 2.151156.

The emerging liability column is the product of IUL vales with those of 1 - 1/*f*.

The total emerged liability is now 2476.08 + 7119.08 = **9595.16** and the total reported liability is 9501 + 6850 + 3894 = 20245. Therefore the reserve value is 9595.16 + 20245 – 20103 = **9737.2**.

6 (i) Since e_t are independent from X_t, X_{t-1}, … and $\mathbf{E}(e_t) = 0$ we have that

$$\begin{aligned}
E(X_t) &= E(\mu + e_t\sqrt{\alpha_0 + \alpha_1(X_{t-1}-\mu)^2})\\
&= \mu + E(e_t\sqrt{\alpha_0 + \alpha_1(X_{t-1}-\mu)^2})\\
&= \mu + E(e_t)E(\sqrt{\alpha_0 + \alpha_1(X_{t-1}-\mu)^2}) \quad \text{since } e_t \text{ and } X_{t-1} \text{ are independent}\\
&= \mu + 0\times E(\sqrt{\alpha_0 + \alpha_1(X_{t-1}-\mu)^2})\\
&= \mu
\end{aligned}$$

The direct approach to showing that X_t and X_{t-s} are uncorrelated is shown below. The crucial steps involve noting that e_t is independent of X_{t-1} as above, and that e_t is independent of

$$e_{t-s}\sqrt{\alpha_0 + \alpha_1(X_{t-1}-\mu)^2}\sqrt{\alpha_0 + \alpha_1(X_{t-s-1}-\mu)^2}\,.$$

The algebra can be simplified by noting that adding a constant doesn't affect covariance, so the μ's can be ignored.

$$Cov(X_t, X_{t-s}) = E(X_t X_{t-s}) - E(X_t)E(X_{t-s})$$
$$= E\left((\mu + e_t\sqrt{\alpha_0 + \alpha_1(X_{t-1}-\mu)^2})(\mu + e_{t-s}\sqrt{\alpha_0 + \alpha_1(X_{t-s-1}-\mu)^2})\right) - \mu^2$$
$$= E(\mu^2 + \mu e_t\sqrt{\alpha_0 + \alpha_1(X_{t-1}-\mu)^2} + \mu e_{t-s}\sqrt{\alpha_0 + \alpha_1(X_{t-s-1}-\mu)^2}$$
$$+ e_t e_{t-s}\sqrt{\alpha_0 + \alpha_1(X_{t-s-1}-\mu)^2}\sqrt{\alpha_0 + \alpha_1(X_{t-1}-\mu)^2}) - \mu^2$$

$$= E(\mu^2) + \mu E(e_t)E(\sqrt{\alpha_0 + \alpha_1(X_{t-1}-\mu)^2}) + \mu E(e_{t-s})E(\sqrt{\alpha_0 + \alpha_1(X_{t-s-1}-\mu)^2})$$
$$+ E(e_t e_{t-s}\sqrt{\alpha_0 + \alpha_1(X_{t-s-1}-\mu)^2}\sqrt{\alpha_0 + \alpha_1(X_{t-1}-\mu)^2}) - \mu^2$$

$$= \mu^2 + \mu \times 0 \times E(\sqrt{\alpha_0 + \alpha_1(X_{t-1}-\mu)^2}) + \mu \times 0 \times E(\sqrt{\alpha_0 + \alpha_1(X_{t-s-1}-\mu)^2})$$
$$+ E(e_t)E(e_{t-s}\sqrt{\alpha_0 + \alpha_1(X_{t-s-1}-\mu)^2}\sqrt{\alpha_0 + \alpha_1(X_{t-1}-\mu)^2}) - \mu^2$$

$$= \mu^2 + 0 + 0 + 0 \times E(e_{t-s}\sqrt{\alpha_0 + \alpha_1(X_{t-s-1}-\mu)^2}\sqrt{\alpha_0 + \alpha_1(X_{t-1}-\mu)^2}) - \mu^2$$
$$= 0$$

(ii) The conditional variance of $X_t | X_{t-1}$ is

$\text{var}(X_t | X_{t-1}) = \text{var}(e_t)(\alpha_0 + \alpha_1(X_{t-1} - \mu)^2) = \alpha_0 + \alpha_1(X_{t-1} - \mu)^2$.

So the values of X_{t-1} are affecting the variance of X_t. If the same idea is applied recursively, it can be seen that the variance of X_t will be affected by the value of X_{t-s}. So X_t and X_{t-s} are not independent.

7 (i) $\hat{\theta} = \dfrac{1}{n}\sum_{t=1}^{n}(e^{U_t} - 1) = \left(\sum_{i=1}^{n}\dfrac{e^{U_t}}{n}\right) - 1$

(ii) We need to find the variance of $h(U) = e^U - 1$ where $U \sim U(0, 1)$

$$\mathbf{E}h(U) = \int_0^1 (e^x - 1)dx = e - 2$$

$$\mathbf{E}h(U)^2 = \int_0^1 (e^x - 1)^2 dx$$

$$= \int_0^1 (e^{2x} - 2e^x + 1)dx$$

$$= \frac{e^2 - 1}{2} - 2(e - 1) + 1$$

$$= \frac{e^2}{2} - 2e + 2.5$$

$\text{var}\,h(U) = \mathbf{E}h(U)^2 - (\mathbf{E}h(U))^2 = 0.242$ and $\text{var}\,\hat{\theta} = \dfrac{0.242}{n}$.

(iii) From the theory, the required n should satisfy

$$n \geq \frac{z_\alpha^2}{0.1^2} 0.242$$

where $P(|z| < z_\alpha) = 1 - \alpha$ with $z \sim N(0, 1)$. In our case $\alpha = 10\%$ and $z_\alpha = 1.64$ and so

$$n \geq \frac{1.64^2}{0.1^2}\ 0.242 = 65.09.$$

The answer is 66.

8 (i) Let $S(t)$ denote the insurer's surplus at time t. Then

$\psi(U) = \Pr(S(t) < 0$ for some value of $t)$ i.e. the probability of ruin at some time

$\psi(U, t) = \Pr(S(k) < 0)$ for some $k < t$ i.e. the probability that ruin occurs before time t.

(ii) (a) Immediately before the payment of any claims, the insurer has cash reserves of 1000 + 150 =1150.

The distribution of $S(1)$ is given by:

Deaths	*S(1)*	*Prob*
None	1150	$0.95 \times 0.9 = 0.855$
A only	$1150 - 1700 = -550$	$0.9 \times 0.05 = 0.045$
B only	$1150 - 400 = 750$	$0.95 \times 0.1 = 0.095$
A and B	$1150 - 2100 = -950$	$0.05 \times 0.1 = 0.005$

And the probability of ruin is given by 0.045 + 0.005 = 0.05.

(b) Assuming the surplus process ends if ruin occurs by time 1, then 2 possible values of $S(2)$ are -550 and -950.

If there are no deaths in year 1, possible values of $S(2)$ are

No deaths: 1150 + 150 = 1300

A only: 1150 + 150 – 1700 = -400
B only: 1150 + 150 – 400 = 900
A and B: 1150 + 150 – 1700 – 400 = -800

If B dies in year 1, the possible values of $S(2)$ are:

A lives: 750 + 100 = 850
A dies: 750 + 100 – 1700 = -850

The probability of ruin within 2 years is given by:

$0.05 + 0.855 \times (0.05 \times 0.9 + 0.05 \times 0.1) + 0.095 \times 0.05 = 0.0975$

Alternatively, note that ruin occurs within 2 years if and only if A dies during this time, the probability of which is $0.05 + 0.95 \times 0.05 = 0.0975$.

9 (i) The transition matrix is

$$\begin{pmatrix} p & 1-p & 0 \\ p & 0 & 1-p \\ 0 & p & 1-p \end{pmatrix}$$

The equilibrium probabilities (π_0, π_1, π_2) satisfy

$\pi_0 = p\pi_0 + p\pi_1$
$\pi_1 = \pi_0(1-p) + p\pi_2$
$\pi_2 = \pi_1(1-p) + (1-p)\pi_2$

From equations 1 and 3 above we obtain

$\pi_0 = p/(1-p)\,\pi_1$ and $\pi_2 = (1-p)/p\,\pi_1$.

Since $\pi_0 + \pi_1 + \pi_2 = 1$ we get $\pi_1 = \dfrac{p(1-p)}{p+(1-p)^2}$ together with $\pi_0 = \dfrac{p^2}{p+(1-p)^2}$

and $\pi_2 = \dfrac{(1-p)^2}{p+(1-p)^2}$

(ii) The average premium is now

£600$\left(\dfrac{p^2}{p+(1-p)^2} + 0.8\dfrac{p(1-p)}{p+(1-p)^2} + 0.5\dfrac{(1-p)^2}{p+(1-p)^2}\right)$ and for

$p = 0.1$ and $p = 0.3$ this quantity takes values £321.1 and £382 respectively.

The difference in average premiums is small given that the claim probability is three times higher, suggesting the NCD system does not discriminate well.

10 (i) ACF looks to decline exponentially suggesting an AR(p) process. While PACF becomes almost zero at lag 3 so AR(2) is a likely process here.

(ii) For p = 1 the equations are:

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

therefore $\hat{a}_1 = \hat{\rho}_1 = 0.854$ and $\hat{\sigma}^2 = \hat{\gamma}_0 - \hat{a}_1\hat{\gamma}_1 = \hat{\gamma}_0(1-\hat{\rho}_1^2) = 0.33917$.

(iii) For p = 2 the equations are:

$$\gamma_2 = a_1\gamma_1 + a_2\gamma_0$$
$$\gamma_1 = a_1\gamma_0 + a_2\gamma_1$$
$$\gamma_0 = a_1\gamma_1 + a_2\gamma_2 + \sigma^2$$

Solving the first two equations with respect to a_i after dividing both sizes by γ_0 we have that

$$a_2 = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2}$$
$$a_1 = \rho_1(1 - a_2)$$

and with the right substitution for $\hat{\rho}_1, \hat{\rho}_2$ we get $\hat{a}_1 = 0.5679$ and $\hat{a}_2 = 0.3350$.

Then the white noise variance can now be estimated as $\hat{\sigma}^2 = \hat{\gamma}_0 - \hat{a}_1\hat{\gamma}_1 - \hat{a}_2\hat{\gamma}_2 = 0.3011$.

(iv) Any 2 tests can be given, including:

- the turning point's test
- the "portmanteau" Ljung-Box χ^2 test
- the inspection of the values of the SACF values based on their 95% confidence intervals under the white noise null hypothesis

11 (i) Suppose X is exponentially distributed with parameter λ. Then we must show that $Y = kX$ is also exponentially distributed.

Now

$$P(Y < y) = P(kX < y)$$

$$= P(X < y/k)$$

$$= \int_0^{y/k} \lambda e^{-\lambda z} dz$$

$$= \int_0^{y} \lambda e^{-\frac{\lambda}{k}x} \frac{dx}{k} \text{ making the substitution } x = kz$$

$$= \int_0^{y} \frac{\lambda}{k} e^{-\frac{\lambda}{k}x} dx$$

Which is the distribution n function of an exponential distribution with parameter λ / k. So Y is exponentially distributed with parameter λ / k.

(ii) First note that the probability that a loss in 2006 is greater than M is given by $e^{-\lambda M}$ and likewise the probability that a loss in 2007 is greater than M is given by $e^{-\lambda M/k}$.

The likelihood of the data is given by:

$$L = C \times e^{-4\lambda M} \times \prod_i (\lambda e^{-\lambda x_i}) \times e^{-6\lambda M/k} \times \prod_j (\frac{\lambda}{k} e^{-\lambda y_i / k})$$

Where the x_i represent the claims in 2006 and y_j represent the claims in 2007.

The log-likelihood is given by

$$l = \log L = C' - 4\lambda M + 10\log\lambda - \lambda \sum x_i - 6\lambda M/k + 12\log\lambda - \lambda/k \sum y_j$$
$$= C' - 4\lambda M + 22\log\lambda - 13{,}500\lambda - 6\lambda M/k - 17{,}000\lambda/k$$

Differentiating gives

$$l' = -4M + \frac{22}{\lambda} - 13{,}500 - 6M/k - 17{,}000/k$$

And equating this to zero gives

$$-4M + 22/\hat{\lambda} - 13{,}500 - 6M/k - 17{,}000/k = 0$$

$$22/\hat{\lambda} = 13,500 + 17,000/k + 4M + 6M/k$$
$$\hat{\lambda} = \frac{22}{13,500 + 17,000/k + 4M + 6M/k}$$

We can check this is a maximum by noting that $l'' = -\frac{22}{\lambda^2} < 0$

(iii) (a) $\hat{\lambda} = \frac{22}{13,500 + 17,000/1.1 + 4\times 1,600 + 6\times 1,600/1.1} = 0.000499$

(b) Expected payment per claim for 2006 is given by:

$$\int_0^M x\lambda e^{-\lambda x}dx + M\int_M^\infty \lambda e^{-\lambda x}dx = \left[-xe^{-\lambda x}\right]_0^M + \int_0^M e^{-\lambda x}dx + M\left[-e^{-\lambda x}\right]_M^\infty$$
$$= -Me^{-\lambda M} + \left[-\frac{1}{\lambda}e^{-\lambda x}\right]_0^M + Me^{-\lambda M}$$
$$= \frac{1}{\lambda}(1 - e^{-\lambda M}) \quad (*)$$
$$= \frac{1}{0.000499}(1 - e^{-0.000499\times 1600})$$
$$= 1102.107$$

We know that claims for 2008 will have an exponential distribution with parameter $\mu = \lambda / 1.05\times 1.1$. We need to choose the retention R so that

$$1102.107 = \int_0^R x\mu e^{-\mu x}dx + R\int_R^\infty \mu e^{-\mu x}dx$$
$$= \frac{1}{\mu}(1 - e^{-\mu R}) \quad \text{using the result from (*)}$$
$$= \frac{1.05\times 1.1}{0.0004999}(1 - e^{-\frac{0.000499R}{1.05\times 1.1}})$$

And so

$$0.476148392 = 1 - e^{-0.00043203463R}$$

$$R = -\frac{1}{0.00043203462}\log(1 - 0.476148392)$$

=1496.52

END OF EXAMINERS' REPORT

EXAMINATION

28 April 2009 (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 Describe the essential characteristic of liability insurance. List four distinct examples of types of liability insurance. [4]

2 (i) Express the probability density function of the gamma distribution in the form of a member of the exponential family of distributions. Specify the natural and scale parameters. [3]

(ii) State the corresponding canonical link function for generalised linear modelling if the response variable has a gamma distribution. [1]
[Total 4]

3 An insurer's portfolio consists of three independent policies. Each policy can give rise to at most one claim per month, which occurs with probability θ independently from month to month. The prior distribution of θ is beta with parameters $\alpha = 2$ and $\beta = 4$. A total of 9 claims are observed on this portfolio over a 12 month period.

(i) Derive the posterior distribution of θ. [2]

(ii) Derive the Bayesian estimate of θ under all or nothing loss. [4]
[Total 6]

4 Individual claim amounts on a particular insurance policy can take the values 100, 150 or 200.

There is at most one claim in a year. Annual premiums are 60.

The insurer must choose between three reinsurance arrangements:

A no reinsurance
B individual excess of loss with retention 150 for a premium of 10
C proportional reinsurance of 25% for a premium of 20

(i) Complete the loss table for the insurer. [4]

	Reinsurance		
Claims	*A*	*B*	*C*
0			
100			
150			
200			

(ii) Determine whether any of the reinsurance arrangements is dominated from the viewpoint of the insurer. [2]

(iii) Determine the minimax solution for the insurer. [1]
[Total 7]

5 An insurance portfolio contains policies for three categories of policyholder: A, B and C. The number of claims in a year, N, on an individual policy follows a Poisson distribution with mean λ. Individual claim sizes are assumed to be exponentially distributed with mean 4 and are independent from claim to claim. The distribution of λ, depending on the category of the policyholder, is

Category	*Value of λ*	*Proportion of policyholders*
A	2	20%
B	3	60%
C	4	20%

Denote by S the total amount claimed by a policyholder in one year.

(i) Prove that $E(S) = E[E(S|\lambda)]$. [2]

(ii) Show that $E(S|\lambda) = 4\lambda$ and $\text{Var}(S|\lambda) = 32\lambda$. [2]

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

(iv) Calculate $\text{Var}(S)$. [2]

[Total 8]

6 The following information is available for a motor insurance portfolio:

The number of claims settled:

	Development Year		
Accident year	*0*	*1*	*2*
2006	442	151	50
2007	623	111	
2008	681		

The cost of settled claims during each year (in 000's):

	Development Year		
Accident year	*0*	*1*	*2*
2006	6321	1901	701
2007	7012	2237	
2008	7278		

Claims are fully run off after year 2. Calculate the outstanding claims reserve using the average cost per claim method with grossing up factors. Inflation can be ignored.

[10]

7 It is necessary to simulate samples from a distribution with density function $f(x) = 6x(1-x) \quad 0 < x < 1$.

(i) Use the acceptance-rejection technique to construct a complete algorithm for generating samples from *f* by first generating samples from the distribution with density $h(x) = 2(1-x)$. [5]

(ii) Calculate how many samples from *h* would on average be needed to generate one realisation from *f*. [1]

(iii) Explain whether the acceptance-rejection method in (i) would be more efficient if the uniform distribution were to be used instead. [4]

[Total 10]

8 An insurer has an initial surplus of *U*. Claims up to time *t* are denoted by *S*(*t*). Annual premium income is received continuously at a rate of *c* per unit time.

(i) Explain what is meant by the insurer's surplus process *U*(*t*). [2]

(ii) Define carefully each of the following probabilities:

(a) $\psi(U,t)$

(b) $\psi_h(U,t)$

[2]

(iii) Explain, for each of the following pairs of expressions, whether one of each pair is certainly greater than the other, or whether it is not possible to reach a conclusion.

(a) $\psi(10,2)$ and $\psi(20,1)$

(b) $\psi(10,2)$ and $\psi(5,1)$

(c) $\psi_{0.5}(10,2)$ and $\psi_{0.25}(10,2)$

[6]

[Total 10]

9 Individual claims under a certain type of insurance policy are for either 1 (with probability α) or 2 (with probability $1 - \alpha$).

The insurer is considering entering into an excess of loss reinsurance arrangement with retention $1 + k$ (where $k < 1$). Let X_i denote the amount paid by the insurer (net of reinsurance) on the ith claim.

(i) Calculate and simplify expressions for the mean and variance of X_i. [5]

Now assume that $\alpha = 0.2$. The number of claims in a year follows a Poisson distribution with mean 500. The insurer wishes to set the retention so that the probability that aggregate claims in a year will exceed 700 is less than 1%.

(ii) Show that setting $k = 0.334$ gives the desired result for the insurer. [5]

[Total 10]

10 Let Y_t be a stationary time series with autocovariance function $\gamma_Y(s)$.

(i) Show that the new series $X_t = a + bt + Y_t$ where a and b are fixed non-zero constants, is not stationary. [2]

(ii) Express the autocovariance function of $\Delta X_t = X_t - X_{t-1}$ in terms of $\gamma_Y(s)$ and show that this new series is stationary. [7]

(iii) Show that if Y_t is a moving average process of order 1, then the series ΔX_t is not invertible and has variance larger than that of Y_t. [6]

[Total 15]

11 A motor insurance company operates a No Claims Discount scheme with discount levels 0%, 25% and 50% of the annual premium of 1,000. The probability of having an accident during any year is 0.1 (ignore the possibility of more than one accident in a year). The policyholder moves one level up in the discount scheme (or stays at the 50% level) in the event of a claim free year and moves one level down (or stays at the 0% level) if the claim does not involve a criminal offence. If the claim involves a criminal offence then the policyholder automatically moves to the 0% discount level. One in ten accidents involves a criminal offence. The policyholder makes a claim only if the cost of repairs is higher than the aggregate additional premiums payable in the next two claim-free years.

(i) Calculate, for each level of discount, the cost of a repair below which the policyholder will not claim, distinguishing between claims that involve a criminal offence and claims that do not. [4]

(ii) Calculate the probability of a claim for each level of discount given that an accident has occurred, given that the repair cost following an accident has an exponential distribution with mean 400. [5]

(iii) Calculate the stationary distribution of the proportion of policyholders at each discount level. [7]

[Total 16]

END OF PAPER

Subject CT6 — Statistical Methods
Core Technical

EXAMINERS' REPORT

April 2009

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

June 2009

Comments

Comments on solutions presented to individual questions for this April 2009 paper are given below.

Q1 *This was generally well answered with various examples for liability insurance mentioned.*

Q2 *This is a standard theoretical question. Many candidates scored full marks here.*

Q3 *Many candidates misspecified the parameters of the Beta distribution in (i)*

Q4 *Some candidates ignored the premium of 60 so all their figures were out by this amount. This however did not affect the answers in (ii) and (iii).*

Q5 *Part (i) was generally not well answered despite the fact that it is a standard result about conditional expectations. Alternative correct solutions were also presented. The following parts of the question were straightforward.*

Q6 *This question was well answered with many candidates scoring full marks. The final figures could differ from those shown in the model solution due to differences in rounding in the intermediate steps. Full credit was given for these provided a consistent approach was taken to rounding.*

Q7 *Many candidates were able to give a correct general idea, but very few were able to specify a complete algorithm so that not many scored full marks.*

Q8 *This was a straightforward question where many candidates scored well.*

Q9 *Many marks were dropped in the last stages of the solution.*

Q10 *This was one of the least well answered questions and many marks were dropped particularly in parts (ii) and (iii).*

Q11 *Many candidates scored well here despite the lengthy calculations required. Some however dropped marks at the calculation of P(claim/accident) at the 50% discount level because they did not distinguish between the cases where the accident was due to a criminal offences and those where it was not.*

1 The essential characteristic of liability insurance is to provide indemnity where the insured, owing to some form of negligence, is legally liable to pay compensation to a third party.

Examples

- Employer's liability
- Motor 3rd party liability
- Public liability
- Product liability
- Professional indemnity

2 (i) We need to express the distribution function of the Gamma distribution in the form:

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

Suppose Y has a Gamma distribution with parameters α and λ. Then

$$f_Y(y) = \frac{\lambda^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\lambda y}$$

And substituting $\lambda = \frac{\alpha}{\mu}$ we can write the density as

$$f_Y(y;\theta,\varphi) = \frac{\alpha^\alpha}{\mu^\alpha \Gamma(\alpha)} y^{\alpha-1} e^{-\frac{y\alpha}{\mu}}$$

$$f_Y(y;\theta,\varphi) = \exp\left[\left(-\frac{y}{\mu} - \log\mu\right)\alpha + (\alpha-1)\log y + \alpha\log\alpha - \log\Gamma(\alpha)\right]$$

Which is in the right form with $\theta = -\frac{1}{\mu}$; $\varphi = \alpha$; $a(\varphi) = \frac{1}{\varphi}$; $b(\theta) = -\log(-\theta)$ and $c(y,\varphi) = (\varphi-1)\log y + \varphi\log\varphi - \log\Gamma(\varphi)$.

Thus the natural parameter is $\frac{1}{\mu}$, ignoring the minus sign, and the scale parameter is α.

(ii) The corresponding link function is $\frac{1}{\mu}$.

3 (i)

$$\begin{aligned} f(\theta|x) &\propto f(x|\theta)f(\theta) \\ &\propto \theta^9(1-\theta)^{27}\theta^{\alpha-1}(1-\theta)^{\beta-1} \\ &\propto \theta^9(1-\theta)^{27}\theta^1(1-\theta)^3 \\ &\propto \theta^{10}(1-\theta)^{30} \end{aligned}$$

Which is the pdf of a Beta(11,31) distribution.

(ii) Under all or nothing loss, the Bayes estimate is the value that maximises the pdf of the posterior.

$$\begin{aligned} f(\theta) &= C\times\theta^{10}(1-\theta)^{30} \\ f'(\theta) &= C\left(10\theta^9(1-\theta)^{30}+\theta^{10}\times 30(1-\theta)^{29}\times -1\right) \\ &= C\theta^9(1-\theta)^{29}\left(10(1-\theta)-30\theta\right) \end{aligned}$$

And $f'(\theta)=0$ when

$$10(1-\theta)-30\theta=0$$
$$40\theta=10$$
$$\theta=1/4$$

We can check this is a maximum by observing that $f(0.25) > 0$ whilst $f(0) = f(1) = 0$. Since the maximum on [0,1] must occur either at the end-points of the interval, or at turning point we can see that we do have a maximum.

4 (i) The loss table is as follows:

Claims	*Impact of reinsurance*			*Insurer's Loss*			
	A	*B*	*C*	*A*	*B*	*C*	
0	0	–10	–20	–60	–50	–40	
100	0	–10	5	40	50	35	
150	0	–10	17.5	90	100	72.5	
200	0	40	30	140	100	110	

(ii) If there are no claims, A gives the best result.
If claims are 100 C gives the best result.
If claims are 200 then B gives the best result.

So each strategy can be best under certain circumstances, and so no approach is dominated.

(iii) The maximum losses are:

A	140
B	100
C	110

The lowest is for B, so approach B is the minimax solution.

5 (i) Let $f(s)$ denote the marginal probability density for S and let $f(s|\lambda)$ denote the conditional probability density for $S|\lambda$. Then

$$E[E(S|\lambda)] = \sum_{i=1}^{3} p(\lambda_i)\int_0^\infty sf(s|\lambda)ds$$
$$= \int_0^\infty s\sum_{i=1}^{3} p(\lambda_i) f(s|\lambda)ds$$

But $\sum_{i=1}^{3} p(\lambda_i) f(s|\lambda) = f(s)$ by definition, so

$$E\left[E\left(S|\lambda\right)\right] = \int_0^\infty sf(s)ds = E(S)$$

(ii) Using the results for compound distributions, we have:

$$E(S|\lambda) = E(N|\lambda)E(X|\lambda) = E(N|\lambda)E(X) = 4\lambda$$

$$Var(S|\lambda) = E(N|\lambda)Var(X) + Var(N|\lambda)E(X)^2$$
$$= \lambda\times 16 + \lambda\times 4^2$$
$$= 32\lambda$$

(iii) $E(S) = E[E(S|\lambda)] = E(4\lambda) = 4E(\lambda) = 12$

(iv) First note that $E(\lambda) = 3$ and

$$Var(\lambda) = 0.2 \times 2^2 + 0.6 \times 3^2 + 0.2 \times 4^2 - 9 = 0.4$$

$$\begin{aligned}
Var(S) &= Var[E(S|\lambda)] + E[Var(S|\lambda)] \\
&= Var(4\lambda) + E(32\lambda) \\
&= 16 \times Var(\lambda) + 32E(\lambda) \\
&= 16 \times 0.4 + 32 \times 3 \\
&= 102.4
\end{aligned}$$

6 The accumulated claims are:

	Development Year			
Accident year	*0*	*1*	*2*	*Ult*
2006	442	593	643	***643***
2007	623	734		***796***
2008	681			***927***

The accumulated claim costs are:

	Development Year		
Accident year	*0*	*1*	*2*
2006	6321	8222	8923
2007	7012	9249	
2008	7278		

The average costs per claim are:

	Development Year			
Accident year	*0*	*1*	*2*	*Ult*
2006	14.301	13.865	13.877	***13.877***
2007	11.255	12.601		***12.612***
2008	10.687			***11.115***

The grossing-up factors for the claim numbers:

Accident year	*0*	*1*	*2*
2006	0.687	0.922	1
2007	0.783	0.922	
2008	0.735		

The grossing-up factors for the average cost per settled claim:

Accident year	*0*	*1*	*2*
2006	1.031	0.999	1
2007	0.892	0.999	
2008	0.961		

Projected loss: 643 × 13.877 + 796 × 12.612 + 927 × 11.115 = 29265.67
Claims paid to date: 8923 + 9249 + 7278 = 25450
Outstanding Claims are therefore: 3815.7

The calculations are based on rounding the intermediate calculations to the accuracy shown. Different rounding will give slightly different answers, and is acceptable. Carrying all the calculations through in full accuracy gives a solution of 3808.0.

7 (i) We must find C where

$$C = Max\frac{f(x)}{h(x)} = \max\frac{6x(1-x)}{2(1-x)} = \max 3x = 3$$

We need to be able to generate a random variable from the distribution with density $h(x)$. We can do this as follows.

First note that the cdf of $h(x)$ is

$$H(x) = \int_0^x g(t)dt = \int_0^x 2(1-t)dt = \left[2t - t^2\right]_0^x = 2x - x^2$$

Given a random sample z from $U(0,1)$ we can use the inverse transform method to sample from h by setting:

$$2x - x^2 = z$$
$$x^2 - 2x + z = 0$$
$$x = \frac{2 \pm \sqrt{4-4z}}{2} = 1 \pm \sqrt{1-z}$$

And we can see that the solution we want is $x = 1 - \sqrt{1-z}$

The algorithm to generate a sample y from f is then:

- Sample z from $U(0,1)$
- Generate x from $h(x)$ via $x = 1 - \sqrt{1-z}$
- Generate u from $U(0,1)$
- If $u < \dfrac{f(x)}{3h(x)} = \dfrac{6x(1-x)}{6(1-x)} = x$ then generate $y = x$ otherwise begin again.

(ii) On average, we expect to use C = 3 realisations from h to generate one sample from f.

(iii) In this case, we must find the maximum value of $f(x)$.

$$f'(x) = 6 - 12x$$

And $f'(x)$ = 0 when x = 0.5

Since $f(0) = f(1) = 0$ and $f(0.5) = 6/4 = 1.5$ we can see that this is the maximum.

Since C is lower for $g(x) = 1$, using the constant function would be more efficient.

8 (i) $U(t)$ represents the insurers surplus at time t. It represents the initial surplus plus cumulative premiums received less claims incurred:

$$U(t) = U + ct - S(t)$$

Where U is the initial surplus and c is the annual premium income (assumed payable continuously).

(ii) (a) $\psi(U,t) = P(U(s) < 0$ for some $0 \le s \le t$ given that $U(0) = U)$

(b) $\psi_h(U,t) = P(U(s) < 0$ for some $s, s = h, 2h, \ldots, t-h, t$ given that $U(0) = U)$

(iii) (a) We can say that $\psi(10,2) > \psi(10,1) > \psi(20,1)$

The first inequality holds because the longer the period considered when checking, the more likely that ruin will occur. The second inequality holds because a larger initial surplus reduces the probability of ruin (higher claims are required to cause ruin).

(b) We can reach no definite conclusion here. The first term has a longer term suggesting a higher probability, but a higher initial surplus suggesting a lower probability. We can't say anything about the size of the two effects, so we can't reach a definite conclusion.

(c) We can say that $\psi_{0.5}(10,2) \le \psi_{0.25}(10,2)$.

This is because the second term checks for ruin at the same times as the first term, as well as at some additional times. So the probability of ruin when checking over the larger set of times must be higher.

9 (i)

$$E(X_i) = \alpha + (1+k)(1-\alpha)$$
$$= 1 + k(1-\alpha)$$

$$Var(X_i) = E(X_i^2) - E(X_i)^2$$
$$= \alpha + (1+k)^2(1-\alpha) - (1+k(1-\alpha))^2$$
$$= \alpha + (1-\alpha) + 2k(1-\alpha) + k^2(1-\alpha) - 1 - 2k(1-\alpha) - k^2(1-\alpha)^2$$
$$= k^2(1-\alpha)(1-(1-\alpha))$$
$$= k^2\alpha(1-\alpha)$$

(ii) Let Y denote the aggregate claims in a year. Then Y has a compound Poisson distribution, and using the standard results from the tables:

$$E(Y) = 500 \times E(X_i) = 500 + 500k(1-\alpha) = 500 + 400k = 633.60$$

And

$$Var(Y) = 500 \times E(X_i^2)$$
$$= 500\left(\alpha + (1+k)^2(1-\alpha)\right)$$
$$= 500\left(\alpha + (1-\alpha) + 2k(1-\alpha) + k^2(1-\alpha)\right)$$
$$= 500 + (1000k + 500k^2)(1-\alpha)$$
$$= 500 + 800k + 400k^2$$
$$= 811.82$$

Using a normal approximation, we find

$$P(Y > 700) = P(Z > \frac{700 - 633.6}{\sqrt{811.82}})$$
$$= P(Z > 2.33)$$
$$= 0.01$$

10 (i) Since Y_t is stationary, we know there is a constant μ_Y such that $E(Y_t) = \mu_Y$ for all values of t.

But then $E(X_t) = E(a + bt + Y_t) = a + bt + \mu_Y$ which depends on t since b is non-zero.

Hence X_t is not stationary.

(ii) First note that

$$
\begin{aligned}
&E(\Delta X_t) = E(X_t - X_{t-1}) \\
&= E(X_t) - E(X_{t-1}) \\
&= E(a + bt + Y_t) - E(a + b(t-1) + Y_{t-1}) \\
&= a + bt + \mu_Y - a - b(t-1) - \mu_Y \\
&= b
\end{aligned}
$$

i.e. the mean is a constant independent of *t*.

Secondly,

$$
\begin{aligned}
&Cov(\Delta X_t, \Delta X_{t-s}) = Cov(X_t - X_{t-1}, X_{t-s} - X_{t-s-1}) \\
&= Cov(b + Y_t - Y_{t-1}, b + Y_{t-s} - Y_{t-s-1}) \\
&= Cov(Y_t - Y_{t-1}, Y_{t-s} - Y_{t-s-1}) \\
&= Cov(Y_t, Y_{t-s}) - Cov(Y_{t-1}, Y_{t-s}) - Cov(Y_t, Y_{t-s-1}) + Cov(Y_{t-1}, Y_{t-s-1}) \\
&= \gamma_Y(s) - \gamma_Y(s-1) - \gamma_Y(s+1) + \gamma_Y(s) \\
&= 2\gamma_Y(s) - \gamma_Y(s-1) - \gamma_Y(s+1)
\end{aligned}
$$

Since the autocovariance depends only on the lag *s*, and the mean is constant, the new series is stationary.

(iii) Suppose that $Y_t = e_t + \beta e_{t-1}$ where e_t is a white noise process with variance σ^2.

Then

$$
\begin{aligned}
&\Delta X_t = a + bt + e_t + \beta e_{t-1} - a - b(t-1) - e_{t-1} - \beta e_{t-2} \\
&= b + e_t + (\beta - 1)e_{t-1} - \beta e_{t-2} \\
&= b + (1 + (\beta - 1)B - \beta B^2)e_t \\
&= b + (1 - B)(1 + \beta B)e_t
\end{aligned}
$$

So the lag polynomial has a unit root, and hence the time series is not invertible.

Now

$$
\begin{aligned}
&Var(Y_t) = Var(e_t + \beta e_{t-1}) \\
&= Var(e_t) + \beta^2 Var(e_{t-1}) \\
&= (1 + \beta^2)\sigma^2
\end{aligned}
$$

So $\gamma_Y(0) = (1 + \beta^2)\sigma^2$

Also $\gamma_Y(1) = \gamma_Y(-1) = Cov(e_t + \beta e_{t-1}, e_{t-1} + \beta e_{t-2}) = \beta\sigma^2$

So using the result from part (ii)

$$Var(\Delta X_t) = 2\gamma_Y(0) - \gamma_Y(1) - \gamma_Y(-1)$$
$$= 2(1+\beta^2)\sigma^2 - 2\beta\sigma^2$$
$$= 2(1-\beta+\beta^2)\sigma^2$$

And finally

$$Var(\Delta X_t) - Var(Y_t) = (2 - 2\beta + 2\beta^2)\sigma^2 - (1+\beta^2)\sigma^2$$
$$= (1 - 2\beta + \beta^2)\sigma^2$$
$$= (1-\beta)^2\sigma^2 > 0$$

11 (i) Premiums at the three levels are £1,000, £750 and £500.

0% level

Claims involving criminal offences make no difference here.
Claim: 1000 + 750 = 1750
No Claim: 750 + 500 = 1250

No claim if the accident cost is less than 500.

25% level

Claims involving criminal offences make no difference here.
Claim: 1000 + 750 = 1750
No Claim: 500 + 500 = 1000

No claim if the accident cost is less than 750.

50% level

Criminal offence claims make a difference here so we distinguish two scenarios:

No criminal offence

Claim: 750 + 500 = 1250
Not claim: 500 + 500 = 1000

No claim if the accident cost is less than 250.

Criminal offence

Claim: 1000 + 750 = 1750
Not claim: 500 + 500 = 1000

No claim if the accident cost is less than 750.

(ii) Let X be the repair cost for each accident then $X \sim Exp\left(\frac{1}{400}\right)$

i.e. $P(X > x) = e^{-\frac{x}{400}}$

For 0% level: $P(\text{claim}|\text{accident}) = P(X > 500) = e^{-\frac{5}{4}} = 0.2865$

For 25% level: $P(\text{claim}|\text{accident}) = P(X > 750) = e^{-\frac{75}{40}} = 0.1534$

For 50% level: $P(\text{claim}|\text{accident}) = P(\text{claim}|\text{criminal offence})$
$+ P(\text{claim}|\text{ no criminal offence})$

$= P(X > 750) \times 0.1 + P(X > 250) \times 0.9$

$= 0.1e^{-\frac{75}{40}} + 0.9e^{-\frac{25}{40}} = 0.4971$

(iii) Note that $P(\text{Accident}) = 0.1$ and $P(\text{criminal offence}|\text{ accident}) = 0.1$.

Hence at 0% level:

$P_{11} = P(\text{current level}) = P(\text{make a claim})$
$= P(\text{claim}|\text{accident})\, P(\text{accident}) = 0.02865$

$P_{12} = P(\text{move to 25\% level}) = P(\text{not claim}) = 1 - 0.02865 = 0.97135$

$P_{13} = P(\text{move to 50\% level}) = 0$

At 25% level:

$P_{21} = P(\text{move down to 0\% level}) = P(\text{make a claim})$
$= P(\text{claim}|\text{accident})\, P(\text{accident}) = 0.015335$

$P_{22} = 0$ since it can never remain at this level for more than a year at a time

$P_{23} = 1 - P_{21} = 1 - 0.015335 = 0.984665$

At the 50% levels:

P_{31} = P(make a claim and criminal offence involved)
= P(accident) P(criminal offence|accident) P(claim|criminal offence)

$= \times\ 0.1P(X > 750) = 0.01 \times e^{-\frac{75}{40}} = 0.0015335$

P_{32} = P(make a claim and no criminal offence involved) P(accident)
P(no criminal offence|accident) P(claim|no criminal offence)

$= 0.1 \times 0.9 \times P(X > 250) = 0.09 \times e^{-\frac{25}{40}} = 0.0481735$

P_{33} = P(no claim) = 1 − P(claim)

$= 1 - 0.01 \times e^{-\frac{75}{40}} - 0.09 \times e^{-\frac{25}{40}}$

= 0.950293 (rounding errors possible here)

Therefore the transition matrix is now

$$P = \begin{pmatrix} 0.02865 & 0.97135 & 0 \\ 0.015335 & 0 & 0.984665 \\ 0.0015335 & 0.0481735 & 0.950293 \end{pmatrix}$$

For the stationary distribution we need to find $\pi = (\pi_0, \pi_1, \pi_2)$ such that $\pi P = \pi$:

$\pi_0 0.02865 + \pi_1 0.015335 + \pi_2 0.0015335 = \pi_0$
$\pi_0 0.97135 + 0 + \pi_2 0.481735 = \pi_1$
$0 + \pi_1 0.984665 + \pi_2 0.950293 = \pi_2$
$\pi_0 + \pi_1 + \pi_2 = 1$

Further calculations show that $\pi_0 = 0.002256424$, $\pi_1 = 0.047946812$ and $\pi_2 = 0.949796764$.

END OF EXAMINERS' REPORT

EXAMINATION

7 October 2009 (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 Consider the stationary autoregressive process of order 1 given by

$$Y_t = 2\alpha Y_{t-1} + Z_t, \qquad |\alpha| < 0.5$$

where Z_t denotes white noise with mean zero and variance σ^2.

Express Y_t in the form $Y_t = \sum_{j=0}^{\infty} a_j Z_{t-j}$ and hence or otherwise find an expression for the variance of Y_t in terms of α and σ. [4]

2 An insurance company has a portfolio of two-year policies. Aggregate annual claims from the portfolio follow an exponential distribution with mean 10 (independently from year to year). Annual premiums of 15 are payable at the start of each year. The insurer checks for ruin only at the end of each year. The insurer starts with no capital. Calculate the probability that the insurer is not ruined by the end of the second year. [5]

3 The loss function under a decision problem is given by:

	θ_1	θ_2	θ_3
d_1	10	15	5
d_2	8	20	15
d_3	12	15	10
d_4	5	23	8

where d_1, d_2, d_3 and d_4 are the possible decisions and θ_1, θ_2 and θ_3 are the possible states of nature.

(i) State which decision can be discounted immediately. [1]

(ii) Determine the minimax solution to the problem. [2]

(iii) Determine the Bayes criterion solution to the problem given that $P(\theta_1) = 0.4$, $P(\theta_2) = 0.25$ and $P(\theta_3) = 0.35$. [2]

[Total 5]

4 A portfolio consists of k independent travel insurance policies. Each policy covers the policyholder's trips over one year. For policy i, the number of claims in the jth month of the covered year, Y_{ij}, is assumed to have a distribution given by

$$P(Y_{ij} = y) = \theta_{ij}(1-\theta_{ij})^y \quad \text{for } y = 0,1,2,\ldots$$

where θ_{ij} are unknown constants between 0 and 1.

(i) Write down the likelihood function and obtain the maximum likelihood estimate for the parameters θ_{ij}. [3]

(ii) Show that $P(Y_{ij} = y)$ can be written in exponential family form and suggest its natural parameter. [2]

(iii) Suppose that θ_{ij} depends on the temperature x_j recorded in the jth month. Explain why it is not appropriate to set $\theta_{ij} = \alpha + \beta x_j$. Suggest another relationship between θ_{ij} and $\alpha + \beta x_j$ that might be used. [3]

[Total 8]

5 The following claim amounts are believed to come from a lognormal distribution with unknown parameters μ and σ^2:

50, 87, 103, 119, 126, 154, 183, 203

Estimate the parameters μ and σ^2 using:

(i) the method of moments; [5]

(ii) the method of percentiles, using the upper and lower quartiles. [5]

[Total 10]

6 The following data is observed from $n = 500$ realisations from a time series:

$$\sum_{i=1}^{n} x_i = 13153.32 \ , \ \sum_{i=1}^{n} (x_i - \overline{x})^2 = 3153.67 \text{ and } \sum_{i=1}^{n-1} (x_i - \overline{x})(x_{i+1} - \overline{x}) = 2176.03 \, .$$

(i) Estimate, using the data above, the parameters μ, a_1 and σ from the model

$$X_t - \mu = a_1(X_{t-1} - \mu) + \varepsilon_t$$

where ε_t is a white noise process with variance σ^2. [7]

(ii) After fitting the model with the parameters found in (i), it was calculated that the number of turning points of the residuals series $\hat{\varepsilon}_t$ is 280.

Perform a statistical test to check whether there is evidence that $\hat{\varepsilon}_t$ is not generated from a white noise process. [3]
[Total 10]

7 The transition rules for moving between the three levels 0%, 35% and 50% of a No Claims Discount system are as follows:

If no claim is made in a year, the policyholder moves to the next higher level of discount, or remains at the 50% level. When at the 0% or 35% level, the policyholder moves to (or remains at) the 0% level when one or more claims is made during the year. When at the 50% level of discount, the policyholder moves to the 35% level if exactly one claim is made during the year, or moves to the 0% level if two or more claims are made during the year.

It is assumed that the number of claims X made each year has a geometric distribution with parameter q such that

$$P(X = x) = q^x(1-q), \ x = 0,1,2,\ldots$$

The full premium is 350.

(i) (a) Write down the transition matrix.

(b) Verify that the equilibrium distribution (in increasing order of discount) is of the form:

$$(kq^2(2-q), kq(1-q), k(1-q)^2)$$

for some constant k. Express k in terms of q. [8]

(ii) The value of the expected premium in the stationary state paid by "low risk" policyholders (with $q = 0.05$) is 178.51.

(a) Calculate the corresponding figure paid by "high risk" policyholders (with $q = 0.1$).

(b) Comment on the effectiveness of the No Claims Discount system.

[4]
[Total 12]

8 The cumulative incurred claims for an insurance company for the last four accident years are given in the following table:

	Development year			
Accident year	*0*	*1*	*2*	*3*
2005	96	136	140	168
2006	100	156	160	
2007	120	130		
2008	136			

It can be assumed that claims are fully run off after three years. The premiums received for each year from 2005 to 2008 are 175, 181, 190 and 196 respectively.

Calculate the reserve at the end of year 2008 using:

(a) The basic chain ladder method.

(b) The Bornhuetter-Ferguson method.

[12]

9 A certain proportion p of electrical gadgets produced by a factory is defective. Prior beliefs about p are represented by a Beta distribution with parameters α and β. A sample of n gadgets is inspected, and k are found to be defective.

(i) Explain what is meant by a conjugate prior distribution. [1]

(ii) Derive the posterior distribution for beliefs about p. [3]

(iii) Show that if $X \sim \text{Beta}(\alpha, \beta)$ with $\alpha > 1$ then $E\left(\frac{1}{X}\right) = \frac{\alpha+\beta-1}{\alpha-1}$. [3]

(iv) It is required to make an estimate d of p. The loss function is given by $L(d,p) = \frac{(d-p)^2}{p}$.

Determine the Bayes estimate d^* of p. [4]

(v) Determine a parameter Z such that d^* can be written as

$$d^* = Z \times \frac{k}{n} + (1-Z) \times \frac{1}{\mu}$$

where μ is the prior expectation of $1/p$. [2]

(vi) Under quadratic loss, the Bayes estimate would have been $\frac{\alpha+k}{\alpha+\beta+n}$.

Comment on the difference in the two Bayes' estimates in the specific case where $\alpha = \beta = 3$, $k = 2$ and $n = 10$. [2]

[Total 15]

10 The total number of claims N on a portfolio of insurance policies has a Poisson distribution with mean λ. Individual claim amounts are independent of N and each other, and follow a distribution X with mean μ and variance σ^2. The total aggregate claims in the year is denoted by S. The random variable S therefore has a compound Poisson distribution.

(i) Derive an expression for the moment generating function of S in terms of the moment generating function of X. [4]

(ii) Derive expressions for the mean and variance of S in terms of λ, μ and σ. [6]

For a particular type of policy, individual losses are exponentially distributed with mean 100. For losses above 200 the insurer incurs an additional expense of 50 per claim.

(iii) Calculate the mean and variance of S for a portfolio of such policies with $\lambda = 500$. [9]

[Total 19]

END OF PAPER

Subject CT6 — Statistical Methods.
Core Technical

September 2009 examinations

EXAMINERS' REPORT

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

December 2009

Comments for individual questions are given with the solutions that follow.

1

$$
\begin{aligned}
Y_t &= 2\alpha Y_{t-1} + Z_t \\
Y_t - 2\alpha Y_{t-1} &= Z_t \\
(1-2\alpha B)Y_t &= Z_t \\
Y_t &= (1-2\alpha B)^{-1} Z_t \\
&= (1 + 2\alpha B + (2\alpha B)^2 + (2\alpha B)^3 + \cdots)Z_t \\
&= \sum_{j=0}^{\infty} (2\alpha)^j Z_{t-j}
\end{aligned}
$$

So $\quad \mathrm{Var}(Y_t) = \mathrm{Var}\left(\sum_{j=0}^{\infty} (2\alpha)^j Z_{t-j} \right)$

$$
\begin{aligned}
&= \sum_{j=0}^{\infty} (4\alpha^2)^j \sigma^2 \\
&= \frac{\sigma^2}{(1-4\alpha^2)}
\end{aligned}
$$

Other valid approaches to deriving the variance were given full credit.

This is a nice, short question involving time series. It requires a little knowledge about series expansions, some algebraic manipulation and the formula for a geometric progression. The question was answered reasonably well, with most candidates recalling the series expansion for $(1-X)^{-1}$. Strong candidates spotted that the condition $|\alpha|<0.5$ was needed to use the formula for an infinite geometric series.

2

Let X_1 denote aggregate claims in year 1, and let X_2 denote aggregate claims in year 2. Then to avoid ruin after the first year, we require X_1<15 and to avoid ruin after 2 years we require $X_1 + X_2$<30.

$$
\begin{aligned}
P(X_1 < 15 \text{ and } X_1 + X_2 < 30) &= \int_0^{15} f_{X_1}(x) P(X_2 < 30 - x)dx \\
&= \int_0^{15} 0.1e^{-0.1x}(1 - e^{-0.1(30-x)})dx \\
&= \int_0^{15} 0.1e^{-0.1x} - 0.1e^{-3} dx
\end{aligned}
$$

$$= \left[-e^{-0.1x} - 0.1xe^{-3} \right]_0^{15}$$

$$= -e^{-1.5} - 1.5e^{-3} + 1$$

$$= 0.702189$$

This was a question involving the concept of ruin in discrete time and requiring candidates to calculate a probability by integrating the pdf of the exponential distribution. This question was not answered well. Many candidates did not recognise the condition for ruin at t=2, ie $X_1+X_2<30$.

3

(i) Decision d_3 is dominated by d_1 and can be discounted immediately.

(ii) Maximum losses are:

d_1	15
d_2	20
d_4	23

So the minimax solution is to choose d_1

(iii) Expected losses are given by:

$$E(L(d_1)) = 0.4\times 10 + 0.25\times 15 + 0.35\times 5 = 9.5$$
$$E(L(d_2)) = 0.4\times 8 + 0.25\times 20 + 0.35\times 15 = 13.45$$
$$E(L(d_4)) = 0.4\times 5 + 0.25\times 23 + 0.35\times 8 = 10.55$$

So the Bayes solution is also to choose d_1.

A straightforward question involving outcomes of 3 decision functions and requiring the candidates to derive the minimax solution and the Bayes criterion solution. This question was answered very well by most candidates.

4

(i) The likelihood is

$$L = \prod_{i=1}^{k} \prod_{j=1}^{12} \theta_{ij}(1-\theta_{ij})^{y_{ij}}$$

Where y_{ij} is the number of claims on the ith policy in the jth month.

Taking the logarithm of L we have

$$\log L = \sum_{i=1}^{k}\sum_{j=1}^{12}\left[\log\theta_{ij} + y_{ij}\log(1-\theta_{ij})\right]$$

and so

$$\frac{\partial \log L}{\partial \theta_{ij}} = \frac{1}{\theta_{ij}} - \frac{y_{ij}}{1-\theta_{ij}}$$

And setting the derivative to zero we find $1-\hat{\theta}_{ij} = y_{ij}\hat{\theta}_{ij}$ so that

$$\hat{\theta}_{ij} = \frac{1}{1+y_{ij}}$$

(ii) $P(Y_{ij} = y) = \theta_{ij}(1-\theta_{ij})^y = e^{\log\theta_{ij} + y\log(1-\theta_{ij})}$

The natural parameter is $\log(1-\theta_{ij})$.

(iii) The range of $\alpha + \beta x_j$ is $(-\infty, +\infty)$ which means it is not suitable for modelling parameters $\theta_{ij} \in [0,1]$.

A possible relationship to consider is $\log\dfrac{\theta_{ij}}{1-\theta_{ij}} = \alpha + \beta x_j$.

Other sensible alternatives should be given credit.

A question testing derivation of the m.l.e. of a non-standard p.d.f, with an example application of the theory in the last part. The question was not answered well, particularly part iii).

5

(i) For the given sample

$$\sum_{i=1}^{8} \frac{x_i}{8} = 128.125$$

$$\sum_{i=1}^{8} \frac{x_i^2}{8} = 18{,}641.125$$

From the tables:

$$E(X) = e^{\mu + \frac{\sigma^2}{2}}$$

$$E(X^2) = \ e^{\sigma^2} - 1\ E(X)^2 + E(X)^2 = E(X)^2 e^{\sigma^2}$$

Substituting into the second of these, we have:

$$18{,}641.125 = 128.125^2 \times e^{\sigma^2}$$

$$\sigma^2 = \log\left(\frac{18{,}641.125}{128.125^2}\right) = 0.12711274$$

And substituting back into the first expression

$$128.125 = e^{\mu + \frac{\sigma^2}{2}}$$

$$\mu + \frac{\sigma^2}{2} = \log 128.125$$

$$\mu = \log 128.125 - 0.5 \times 0.12711274 = 4.78945$$

(ii) The lower and upper quartile points in the data set are 95 and 168.5

We need to solve:

$$e^{\mu + 0.6745\sigma} = 168.5 \quad \text{and} \quad e^{\mu - 0.6745\sigma} = 95$$

Dividing the first by the second gives:

$$e^{2 \times 0.6745\sigma} = 1.773684211$$

$$\sigma = \frac{\log 1.773684211}{2 \times 0.6745} = 0.424802711$$

$$\sigma^2 = 0.180457343$$

And substituting back into the first equation:

$$e^{\mu + 0.6745\sigma} = 168.5$$

$$\mu = \log 168.5 - 0.6745 \times 0.424802711$$

$$\mu = 4.840406$$

It is possible to use other definitions of upper and lower quartile. Other sensible choices were given full credit provided the subsequent calculations followed through correctly

A numerical question that tested the theory of fitting a distribution using two different methods to sample data. This question was answered reasonably well, with some candidates scoring very highly indeed. This question was a good differentiator with strong candidates showing they had learnt the theory of distribution fitting thoroughly and accurately calculating the answers. NB Both percentile definitions as per CT3 were given credit.

6

(i) Clearly $\hat{\mu} = \frac{13153.32}{500} = 26.30644.$

Using the known expression of the auto covariance function for AR(1) processes: $\rho_k = a_1^k$, we see that

$$a_1^k, = \hat{\rho}_1 = \frac{\sum_{i=1}^{499}(x_i - \bar{x})(x_{i+1} - \bar{x})}{\sum_{i=1}^{500}(x_i - \bar{x})^2} = \frac{2176.03}{3153.67} = 0.6899993$$

Taking the variance of both sides of

$X_t - \mu = a_1(X_{t-1} - \mu) + \varepsilon_t$

and using the fact that $\gamma_0 = \text{var}(X_t - \mu) = \text{var}(X_{t-1} - \mu)$

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

Hence $\hat{\sigma}^2 = \hat{\gamma}_0(1 - a(hat)_1^2) = \dfrac{3153.67}{500}(1 - 0.6899993^2) = 3.304416,$

i.e. $\hat{\sigma} = \sqrt{3.304416} = 1.817805$

(ii) Using the fact that under the white noise assumptions the mean and variance of the number of change points are

$\dfrac{2(N-2)}{3} = 332$ and $\dfrac{(16N-29)}{90} = 88.56667$

respectively where $N = 500$. Therefore since the 95% confidence interval is

$(332 - 1.96 \times \sqrt{88.56667}, 332 + 1.96 \times \sqrt{88.56667}) = (313.6, 350.4)$

which does not contain the observed number 280, there is a strong evidence that the errors are not close to those of a white noise.

This was a slightly more complicated parameter fitting question for a time-series model, and with a chi-square significance test to finish. The question was answered poorly, with many candidates finding this one tough. Candidates scoring poorly for part i) usually did not get ii) out as well.

7

(i)

a. Note first that

$$P(X = 0) = 1 - q$$
$$P(X = 1) = q(1-q)$$
$$P(X \geq 2) = 1 - (1-q) - q(1-q) = q^2$$
$$P(X \geq 1) = 1 - (1-q) = q$$

The transition matrix is

$$P = \begin{pmatrix} q & 1-q & 0 \\ q & 0 & 1-q \\ q^2 & q(1-q) & 1-q \end{pmatrix}$$

b. $(kq^2(2-q), kq(1-q), k(1-q)^2)P = (\pi_1, \pi_2, \pi_3)$

Where

$$\pi_1 = kq^3(2-q) + kq^2(1-q) + kq^2(1-q)^2$$
$$= kq^2 \;\; q(2-q) + (1-q) + (1-q)^2$$
$$= kq^2 \;\; 2q - q^2 + 1 - q + 1 - 2q + q^2$$
$$= kq^2(2-q)$$

$$\pi_2 = kq^2(2-q)(1-q) + k(1-q)^2 q(1-q)$$
$$= kq(1-q) \;\; q(2-q) + (1-q)^2$$
$$= kq(1-q)(2q - q^2 + 1 - 2q + q^2)$$
$$= kq(1-q)$$

$$\pi_3 = kq(1-q)^2 + k(1-q)^3$$
$$= k(1-q)^2(q + 1 - q)$$
$$= k(1-q)^2$$

Since the proportions must sum to 1, we have

$$k = \frac{1}{q^2(1-q) + q(1-q) + (1-q)^2} = \frac{1}{2q^2 - q^3 - q + 1}$$

(ii)

a. Average premium is:

$$\text{L}\,350 \times \frac{1}{2\times 0.1^2 - 0.1^3 - 0.1 + 1} \times \;\; 0.1^2 \times 1.9 + 0.65 \times 0.1 \times 0.9 + 0.5 \times 0.9^2$$

= 183.76

b. Policyholders are twice as likely to claim, but the premium increases only by 3%! Suggests that the NCD system is not effective.

A 3x3 NCD problem with generic P(claim) = q, and P(not claim) = 1-q, and then a numerical application. Despite some fiddly algebra, this question was based on standard NCD theory and answered very well.

8

a. The development factors are given by

R_1 = (136 + 156 + 130) / (96 + 100 + 120) = 1.335443

R_2 = (140 + 160) / (136 + 156) = 1.027397

R_3 = 168 / 140 = 1.2

The fully developed table using the chain ladder is below:

Incident year	*0*	*1*	*2*	*3*
2005	96	136	140	168
2006	100	156	160	**192**
2007	120	130	**133.56**	**160.28**
2008	136	**181.62**	**186.60**	**223.92**
R	1.335443	1.027397	1.2	1
f	1.646436	1.232876	1.2	1

Reserve = (168 + 192 + 160.28 + 223.92) – (168 + 160 + 130 + 136) = 150.2

b. B-F method

Estimated loss ratio: 168/175 = 0.96

	2008	*2007*	*2006*	*2005*
F	1.646436	1.232876	1.2	1
1 – 1/f	0.392627	0.188888	0.1666667	0
IUL	188.16	182.4	173.76	168
Emerging liab.IUL(1 – 1/f)	73.87678	34.45325	28.96	0

Reserve is now = 73.87678 + 34.45325 + 28.96 = 137.29

A standard chainladder / Bornhuetter-Ferguson question which candidates answered very well.

9

(i) A distribution is a conjugate prior for an unknown parameter if when used as a prior distribution for that parameter it leads to a posterior distribution which is from the same family.

(ii) $f(p|k) \propto f(k|p) \times f(p)$

$$\propto p^k (1-p)^{n-k} p^{\alpha-1}(1-p)^{\beta-1}$$

$$\propto p^{\alpha+k-1}(1-p)^{\beta+n-k-1}$$

Which is the pdf of a Beta($\alpha+k,\beta+n-k$) distribution.

(iii)

$$E\left(\frac{1}{X}\right) = \int_0^1 \frac{1}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}dx$$

$$= \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-2}(1-x)^{\beta-1}dx$$

$$= \frac{\Gamma(\alpha-1)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\alpha+\beta-1)}\int_0^1 \frac{\Gamma(\alpha+\beta-1)}{\Gamma(\alpha-1)\Gamma(\beta)}x^{\alpha-2}(1-x)^{\beta-1}dx$$

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

$$= \frac{\alpha+\beta-1}{\alpha-1}.$$

Other derivations are acceptable.

(iv)Let $h(d) = E(L(d,p))$

$$h(d) = E\left(\frac{(d-p)^2}{p}\right) = E\left(\frac{d^2-2dp+p^2}{p}\right)$$

$$= d^2E(1/p) - 2d + E(p)$$

$$h'(d) = 2dE(1/p) - 2$$

And $h'(d) = 0$

$$2d^*E(1/p) = 2$$

When $d^* = 1/E(1/p) = \dfrac{\alpha+k-1}{\alpha+\beta+n-1}$

Using the result from (iii) applied to the posterior distribution for *p*.

(v)

$$d^* = \frac{\alpha+k-1}{\alpha+\beta+n-1} = \frac{\alpha-1}{\alpha+\beta-1}\times\frac{\alpha+\beta-1}{\alpha+\beta+n-1} + \frac{k}{n}\times\frac{n}{\alpha+\beta+n-1}$$

$$= \mu(1-Z) + Z\frac{x}{n}$$

Where $Z = \dfrac{n}{\alpha+\beta+n-1}$ and μ is the prior expectation of 1/*p*.D

(vi) The estimates are:

Using the given loss function the estimate is

(3 + 2 – 1) / (3 + 3 + 10 – 1) = 4/15 = 0.266666

Using Bayesian loss, we have (3 + 2) / (3 + 3 + 10) = 5/16 = 0.3125.

The mean of the prior is 0.5 and the observed sample mean is 0.2. The loss function in (iv) penalises mis-estimates particularly when the true value of p is lower. This means that the estimate in (iv) is lower than would result from straight quadratic loss.

A longer Bayes question with derivation of a posterior Beta distribution, a credibility factor Z and consideration of a non-standard loss function (given) and the quadratic loss. There was a wide range of quality of answers for this 6 part question. Generally i) and ii) were answered well, iii) to v) proved trickier, in particular deriving d in part v).*

10

(i) $M_S(t) = E(e^{tS})$

$= E(E(e^{t(X_1+X_2+\cdots+X_N)} | N)$

$= E(M_X(t)^N)$ since the X_i are independent and identically distributed

$= E(e^{N \log M_X(t)})$

$= M_N(\log M_X(t))$

$= \exp(\lambda(\exp(\log(M_X(t)-1))))$

$= \exp(\lambda(M_X(t)-1))$

(ii) $M'_S(t) = M_S(t) \times \lambda M'_X(t)$

$E(S) = M'_S(0) = M_S(0) \times \lambda \times M'_X(0)$

$= 1 \times \lambda \times \mu$

$= \lambda\mu$

$M''_S(t) = M'_S(t) \times \lambda \times M'_X(t) + M_S(t) \times \lambda \times M''_X(t)$

$E(S^2) = M''_S(0) = M'_S(0) \times \lambda \times M'_X(0) + M_S(0) \times \lambda \times M''_X(0)$

$= \lambda\mu \times \lambda \times \mu + 1 \times \lambda \times (\sigma^2 + \mu^2)$

$= \lambda^2\mu^2 + \lambda\mu^2 + \lambda\sigma^2$

And so

$$\begin{aligned}\mathrm{Var}(S) &= E(S^2) - E(S)^2\\ &= \lambda^2\mu^2 + \lambda\mu^2 + \lambda\sigma^2 - \lambda^2\mu^2\\ &= \lambda(\mu^2 + \sigma^2)\end{aligned}$$

(iii) First, we must calculate the mean and variance of a single claim, say Y. Let us denote by X the underlying loss. Then

$$\begin{aligned}
E(Y) &= \int_0^{200} 0.01xe^{-0.01x}dx + \int_{200}^{\infty} (x+50)\times 0.01\times e^{-0.01x}dx \\
&= \int_0^{\infty} 0.01xe^{-0.01x}dx + 50\int_{200}^{\infty} 0.01e^{-0.01x}dx \\
&= E(X) + 50\times P(X > 200) \\
&= 100 + 50\times e^{-200\times 0.01} \\
&= 100 + 6.76676 \\
&= 106.76676
\end{aligned}$$

$$\begin{aligned}
E(Y^2) &= \int_0^{200} 0.01x^2e^{-0.01x}dx + \int_{200}^{\infty} (x+50)^2\times 0.01e^{-0.01x}dx \\
&= \int_0^{\infty} 0.01x^2e^{-0.01x}dx + \int_{200}^{\infty} xe^{-0.01x}dx + 50^2\int_{200}^{\infty} 0.01e^{-0.01x}dx \\
&= E(X^2) + \left[-100xe^{-0.01x}\right]_{200}^{\infty} + \int_{200}^{\infty} 100e^{-0.01x}d + 50^2 P(X > 200) \\
&= 100^2 + 100^2 + 20,000e^{-2} + \left[-100^2e^{-0.01x}\right]_{200}^{\infty} + 2,500e^{-2} \\
&= 20,000 + 20,000e^{-2} + 10,000e^{-2} + 2,500e^{-2} \\
&= 20,000 + 32,500e^{-2} \\
&= 24,398.39671
\end{aligned}$$

And finally, using the results from part (ii)

$$\begin{aligned}
E(S) &= 500E(X) = 500\times 106.76676 \\
&= 53,383.38
\end{aligned}$$

and

$$\text{Var}(S) = 500E(X^2) = 500\times 24,398.39671 = 12,199,198.36$$

A long question about deriving mgf of a compound Poisson distribution. Mixture of bookwork and proof, and a numerical application to finish. This question was answered reasonably well, in particular part i) and part ii).

END OF EXAMINERS' REPORT