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 $\mathrm{Var}(\lambda) = \alpha/\beta^2$. Now

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$\therefore$ The autocorrelation function is

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

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

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

$$\begin{aligned} P(\text{Ruin}) &= P(N(40{,}000, (8{,}602.33)^2) > 50{,}000) \\ &= P\left(N(0,1) > \frac{50{,}000 - 40{,}000}{8{,}602.33} \right) \\ &= 1 - \Phi(1.162) \\ &= 0.123. \end{aligned}$$

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 $\theta = -\dfrac{1}{\mu}$.

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

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

$\hat{y}_i = 14.2$

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

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

$= 0.4$

Hence the deviance residual is $-\sqrt{0.4} = -0.633$.

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 $\dfrac{s}{\alpha - 1}$. 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 $\dfrac{s}{\alpha - 1} = 3$ 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