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 $\Pr(\text{loss} > u) = \exp(-u\,\lambda)$.

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 $\pi_0 = 0.0820 \times 0.1325 = 0.0109$ 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 $(X_{t+1}, \ldots X_{t+k})$ is the same as that of $(X_{t+s+1}, \ldots X_{t+s+k})$ 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

$$\operatorname{cov}(X_t, X_{t+s}) = \gamma(s) \quad \forall t, s.$$

(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 $\mathrm{Var}(N) = 250p\,(1 - p)$
$E(X) = 0.01 \times 200 + 0.18 \times 100 = 20$
$\mathrm{Var}(X) = 40000 \times 0.01 + 10000 \times 0.18 - 20 \times 20 = 1800$

and

$E(S) = E(N)\,E(X) = 250p \times 20 = 5000p$
$\mathrm{Var}(S) = E(N)\,\mathrm{Var}(X) + \mathrm{Var}(N) \times E(X) \times E(X)$
$\mathrm{Var}(S) = 250p \times 1800 + 250p(1 - p) \times 20 \times 20$
$= 450{,}000p + 100{,}000p\,(1 - p) = 550{,}000p - 100{,}000\,p^2$.

(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

$$Var(T) = Var(100W) = 100^2 Var(W) = 100^2 \times 500 \times 0.1p \times (1 - 0.1p)$$
$$= 500{,}000\,p(1 - 0.1p)$$

(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