# 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 $\alpha = 2$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

(vi) 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