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 *i*th root of the *i*th autocorrelation:

i	*i*th 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:

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

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)

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

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

$$= E(E(e^{X_1 t}e^{X_2 t}\ldots e^{X_N t}|N))$$

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

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

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

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

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

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

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

$k = \left(\dfrac{n_1 y_1 (1+r)^2 + n_2 y_2 (1+r)}{n_1 + n_2} \right)$ 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 $$\frac{u + c - E(S)}{\sqrt{\text{Var}(S)}} = 2.326$$

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

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

$$\therefore u' = \alpha(u + 0.1E[S]) - 1.1E[S] + (1 + \xi)(1 - \alpha)\, E[S] + \alpha E[S]$$
$$= \alpha u + (0.1\alpha - 1.1 + 1 - \alpha + \xi(1 - \alpha) + \alpha)\, E(S)$$
$$= \alpha u + (1 - \alpha)\, (\xi - 0.1)\, E(S)$$

(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