ASI SUBJECT CT6 – STATISTICAL MODELS
June 2005 Examinations - solutions

Solution 1.

(i) We obtain the following table of premiums for the next three years:

Current Level	**Claim**			**No Claim**			**Smallest loss for which the claim will be made**
	Year 1	**Year 2**	**Year 3**	**Year 1**	**Year 2**	**Year 3**	
0%	100	85	70	85	70	50	50
15%	100	85	70	70	50	50	85
30%	85	70	50	50	50	50	55
50%	85	70	50	50	50	50	55

[2]

(ii) 0% level: $P(\text{Cost}>50) = e^{-50/500} = 0.905$

[1]

15% level: $P(\text{Cost}>85) = e^{-85/500} = 0.844$

[1]

30% and 50% level: $P(\text{Cost}>55) = e^{-55/500} = 0.896$

[1]
[Total 5]

Solution 2.

(i) Let x_1, x_2, x_3 be the observed claims. The Bayesian estimate under quadratic loss is the posterior mean. We first find the posterior distribution of q. The posterior density of q is:

$$f(q|x) \propto \exp\left(\frac{-1}{2s_1^2}\sum_{j=1}^{n}\left(x_j - q\right)^2 - \frac{1}{2s_2^2}\left(q - m\right)^2\right)$$

$$\propto \exp\left(\frac{-1}{2}\left\{\left(\frac{n}{s_1^2} + \frac{1}{s_2^2}\right)q^2 - 2q\left(\frac{m}{s_2^2} + \frac{n\bar{x}}{s_1^2}\right)\right\}\right)$$

[2]

The posterior mean is thus:

$$\frac{\dfrac{n\bar{x}}{s_1^2} + \dfrac{m}{s_2^2}}{\dfrac{n}{s_1^2} + \dfrac{1}{s_2^2}} = \frac{\dfrac{n}{s_1^2}}{\dfrac{n}{s_1^2} + \dfrac{1}{s_2^2}}\bar{x} + \frac{\dfrac{1}{s_2^2}}{\dfrac{n}{s_1^2} + \dfrac{1}{s_2^2}}m$$

[1]

Using the values given:

$$n = 3, \quad s_1^2 = 16, \quad s_2^2 = 49, \quad m = 100$$

$we \quad obtain:$

$$\frac{\dfrac{3}{16}}{\dfrac{3}{16} + \dfrac{1}{49}}\bar{x} + \frac{\dfrac{1}{49}}{\dfrac{3}{16} + \dfrac{1}{49}}100$$

This is of the form $Z\bar{x} + (1 - Z)100$ where 100 is the prior mean for q.

[1]

(ii) The credibility factor Z is $\dfrac{\dfrac{3}{16}}{\dfrac{3}{16} + \dfrac{1}{49}} = 0.9018.$

If $\bar{x} = 110,$ then the credibility premium is 109.02.

[1]

(iii) If the variance of 16 is decreased, then the value of Z would increase, and the credibility estimate would move closer to the past data. This makes sense, since decreasing this variance means that the claim amounts within each risk are less variable, and so we should put relatively more weight on past data.

[1]
[Total 6]

Solution 3.

Without policy excess, the mean and variance of the aggregate claims, S, are:

$$E(S) = 200 \times 500 = 1{,}00{,}000$$

$$E(X) = \frac{l}{a-1}$$

$$V(X) = \frac{a\,l^2}{(a-1)^2(a-2)}$$

$$\frac{V(X}{\left(E(X)\right)^2} = \frac{a}{a-2} = \frac{(1500)^2}{(500)^2} = 9$$

$$a = 2.25$$

$$l = (2.25-1) \times 500 = 625$$

[2]

$$X' = \begin{cases} 0 & if \quad X <= 100 \\ X - 100 & otherwise \end{cases}$$

$$\therefore \; E(X') = \int_{100}^{\infty} (x-100)\tfrac{a\,l^a}{(l+x)^{a+1}}\,dx$$

$$= \frac{l^a}{(a-1)(l+100)^{a-1}} = 415.33404$$

[2]

$$E(S') = 415.33 \times 200 = 83{,}066.81$$

$$\% \quad reduction = \frac{100{,}000 - 83{,}066.81}{100{,}000} = 16.93\%$$

[1]

[Total 5]

Solution 4.

(i) The assumptions for the inflation adjusted chain ladder method are:

- The first accident year is fully run-off.
- For each accident year, the amount of claims paid, in real terms, in each development year is a constant proportion of the total claims, in real terms, from that accident year.
- Explicit allowance for past inflation
- Explicit allowance for future inflation

[1]

(ii) We can derive the incremental claims from the cumulative claims as under:

	Development year			
Accident year	**0**	**1**	**2**	**3**
2001	2,047	815	355	268
2002	2,471	1,257	190	
2003	2,388	1,438		
2004	2,580			

[1]

The inflation adjusted claims are:

	Development year			
Accident year	**0**	**1**	**2**	**3**
2001	2,724.56	986.15	390.50	268.00
2002	2,989.91	1,382.70	190.00	
2003	2,626.80	1,438.00		
2004	2,580.00			

[2]

The inflation adjusted cumulative claims are:

	Development year			
Accident year	**0**	**1**	**2**	**3**
2001	2,724.56	3,710.71	4,101.21	4,369.21
2002	2,989.91	4,372.61	4,562.61	
2003	2,626.80	4,064.80		
2004	2,580.00			

[1]

The development factors are:

Development year	**1**	**2**	**3**
Development Factor	**1.456388**	**1.071815**	**1.065347**

[1]

The estimated outstanding claims for 2004 are

= 2,580*1.456388*1.071815*1.065347-2,580

= **Rs1,710.49.**

[2]
[Total 8]

Solution 5.

(i) The total claims from a portfolio are given by:

$$S = X_1 + X_2 + \dots + X_n$$

where n is the total number of fixed policies and X_i is the total claim amount from the ith policy. We assume that the X_i are independent but not necessarily identically distributed.

[2]

(ii) Since N has a negative binomial distribution with parameters *k=3* and *p=0.9* we have:

$$P(N=n) = \binom{n+2}{n}(0.9)^3(0.1)^n$$

and thus

$$\frac{P(N=n)}{P(N=n-1)} = 0.1 \times \binom{n+2}{n} / \binom{n+1}{n-1} = 0.1\frac{(n+2)((n+1)}{(n+1)n} = 0.1 + \frac{0.2}{n}$$

Thus the relationship holds with *a=0.1 and b=0.2.* **[2]**

(iii) ***By Recursive method***

The recursive formula states that:

$$P(S=s) = \sum_{x=1}^{s}\left(a + \frac{bx}{s}\right)P(X=x)P(S=s-x), \quad s \geq 1$$

and **[1]**

$$P(S=0) = P(N=0)$$

Working in units of 500 we have:

$$P(S=0) = P(N=0) = 0.9^3 = 0.729$$

$$P(S=1) = 0.1\left(1 + \frac{2\times 1}{1}\right)P(X=1)P(S=0) = 0.1\times 3\times 0.5\times 0.729$$

[1]

$$= 0.10935$$

$$P(S=2)=0.1\left(1+\frac{2\times 1}{2}\right)P(X=1)P(S=1)+0.1\left(1+\frac{2\times 2}{2}\right)P(X=2)P(S=0)$$

$$=0.06561$$

$$P(S=3)=0.1\left(1+\frac{2\times 1}{3}\right)P(X=1)P(S=2)+0.1\left(1+\frac{2\times 2}{3}\right)P(X=2)P(S=1)$$

$$=0.01185 \quad (\sin ce \quad P(X=3)=0)$$

$$P(S=4)=0.1\left(1+\frac{2\times 1}{4}\right)P(X=1)P(S=3)+0.1\left(1+\frac{2\times 2}{4}\right)P(X=2)P(S=2)$$

$$+0.1\left(1+\frac{2\times 4}{4}\right)P(X=4)P(S=0)$$

$$=0.05884 \quad (\sin ce \quad P(X=3)=0)$$

Hence the probability that the aggregate claim amount is less than or equal to Rs2,000 is:

$$P(S\le 4)=0.729+0.10935+0.06561+0.01185+0.05884=0.97465$$

[4]

[Total 10]

Solution 6.

Let X be the gross claim amount.

For 8 claims, X < 1,000

For 19 claims, $\sum_{i=1}^{19} x_i = 66{,}666 + 19 \times 1{,}000 = 85{,}666$

For 13 claims, X > 21,000

[2]

$$P(X < 1000) = 1 - e^{-1{,}000\ l}$$
$$P(X > 21000) = e^{-21{,}000\ l}$$

Likelihood function is:

$$\left(\prod_{i=1}^{19} l e^{-l x_i}\right)\left(1 - e^{-1{,}000\,l}\right)^8 \left(e^{-21{,}000\,l}\right)^{13}$$

[2]

The logLikelihood function is therefore:

$$19 \log\ l - l \sum_{i=1}^{19} x_i + 8\log(1 - e^{-1{,}000 l}) - 13 \times 21{,}000\,l$$
$$= 19 \log\ l - (85{,}666 + 273{,}000)\,l + 8\log(1 - e^{-1{,}000\,l})$$
$$= 19 \log\ l - 358{,}666\,l + 8\log(1 - e^{-1{,}000\,l})$$

[2]
[Total 6]

7 (i) Let the first claim amount be D_1. $E(D_1) = (5{,}000+15{,}000)/2 = 10{,}000$.

Time-to first claim, T_1, has the exponential distribution with mean 1/0.4. [1]

The surplus amount at time T_1 is $10{,}000 + 1.25\,.(0.4)\,T_1\,.10{,}000 - D_1$. [1]

Probability of ruin at first claim

$$= P(10{,}000 + 12{,}500(0.4)T_1 - D_1 < 0)$$

$$= P(10{,}000 + 12{,}500(0.4)T_1 - 15{,}000 < 0).P(D_1 = 15{,}000)$$

$$+ P(10{,}000 + 12{,}500(0.4)T_1 - 5{,}000 < 0).P(D_1 = 5{,}000)$$

$$= P(10{,}000 + 12{,}500(0.4)T_1 - 15{,}000 < 0)/2 + 0$$

$$= P((0.4)T_1 < 0.4)/2 = (1 - e^{-0.4}) / 2 = 0.16484.$$ [2]

(ii) Let the second claim amount be D_2. $E(D_2) = E(D_1) = 10{,}000$.

Time-to second claim, T_2, has the gamma distribution with scale parameter λ and shape parameter 2. [1]

The surplus amount at time T_2 is $10{,}000 + 1.25\,.(0.4)\,T_2\,.10{,}000 - D_1 - D_2$. [1]

Probability of ruin at second claim

$$= P(10{,}000 + 1.25\,.(0.4)\,T_2\,.10{,}000 - D_1 - D_2 < 0)$$

$$= P(10{,}000 + 1.25\,.(0.4)\,T_2\,.10{,}000 - 10{,}000 < 0).P(D_1 = D_2 = 5{,}000)$$

$$+ P(10{,}000 + 1.25\,.(0.4)\,T_2\,.10{,}000 - 30{,}000 < 0).P(D_1 = D_2 = 15{,}000)$$

$$+ P(10{,}000 + 1.25\,.(0.4)\,T_2\,.10{,}000 - 20{,}000 < 0).P(D_1 = 5{,}000, D_2 = 15{,}000)$$

$$+ P(10{,}000 + 1.25\,.(0.4)\,T_2\,.10{,}000 - 20{,}000 < 0).P(D_1 = 15{,}000, D_2 = 5{,}000)$$

$$= 0 + P((0.4)\,T_2 < 1.6)/4 + P((0.4)\,T_2 < 0.8)/4 + P((0.4)\,T_2 < 0.8)/4$$

$$= (1 - 2.6e^{-1.6})/4 + (1 - 1.8e^{-0.8})/2 = 0.21437.$$ [2]

(iii) It is clear from part (i) that even if there is one claim of size Rs. 15,000 before one year, ruin will take place. Therefore, all claims occurring before one year must be of size Rs. 5,000. [1]

If there are two or fewer claims (of size Rs. 5,000) before one year, surplus will still be positive. However, if there are three or more claims occurring before one year, then ruin will occur even if all the claims have size Rs. 5,000. This is because the initial surplus will be spent on paying out the first two claims while there would not be sufficient accumulation of premium to pay out the third. [1]

Therefore, the probability that ruin does not occur before one year is

P(No claims) + P(1 claim of size Rs. 5,000) + P(2 claims of size Rs. 5,000)

$$= e^{-0.4} [1 + (0.4)(1/2) + (0.4^2/2)(1/4)] = 0.81779. \quad [2]$$

8 (i)

d_1 100 policies premium Rs.850 per annum

d_2 150 policies premium Rs.810 per annum

d_3 200 policies premium Rs.790 per annum

θ_1 Intensity I_1 claim costs Rs.400 per policy per annum

θ_2 Intensity I_2 claim costs Rs.450 per policy per annum

θ_3 Intensity I_3 claim costs Rs.570 per policy per annum

θ_4 Intensity I_4 claim costs Rs.600 per policy per annum

Figures in Rs.Lakhs

Strategy	d_1	d_2	d_3
Total premiums	850	1,215	1,580
Fixed expenses	150	150	150
Per policy expenses	180	270	360
Premium less expenses	520	795	1,070

[2]

Hence annual profits (Rs.lakhs):

	θ_1	θ_2	θ_3	θ_4
d_1	120	70	−50	−80
d_2	195	120	−60	−105
d_3	270	170	−70	−130

[3]

Minimax = minimise maximum loss

d_1	80	← choose d_1, set premiums at Rs.850 per annum
d_2	105	
d_3	130	

[2]

(ii) Bayes criterion

$d_1 = 0.1 \times 120 + 0.4 \times 70 - 0.3 \times 50 - 0.2 \times 80 = 9$

$d_2 = 0.1 \times 195 + 0.4 \times 120 - 0.3 \times 60 - 0.2 \times 105 = 28.5$

$d_3 = 0.1 \times 270 + 0.4 \times 170 - 0.3 \times 70 - 0.2 \times 130 = 48 \leftarrow$ choose d_3

Choose d_3 , premiums of Rs.790 per annum [2]

9 $X \sim N(\mu, 100^2), \mu \sim N(500, 20^2)$

(i) $P(\mu > 535) = P\left(Z > \frac{535 - 500}{20} = 1.75\right)$

$= 1 - 0.960 = 0.040$ [1]

(ii) $n = 10$, $\bar{x} = 535$

$\mu \mid \bar{x} \sim N(\mu_*, \sigma_*^2)$

$$\mu_* = \frac{\frac{10(535)}{100^2} + \frac{500}{20^2}}{\frac{10}{100^2} + \frac{1}{20^2}} = \frac{0.001(535) + 0.0025(500)}{0.0035}$$

$= 510$ [1]

$$\sigma_*^2 = \frac{1}{0.0035} = 16.9^2$$

$\therefore \mu \mid \bar{x} \sim N(510, 16.9^2)$ [1]

$\boxed{\text{(iii)}}$ $P(\mu > 535 \mid \bar{x}) = P\left(Z > \frac{535 - 510}{16.9} = 1.48\right)$

$= 1 - 0.934 = 0.07$ [1]

Since $\bar{x} >$ prior mean, the posterior probability in (ii) is larger than the prior one in (i). [2]

10 (i) If Y has a Poisson distribution with mean μ, then

$$f(y, \mu) = e^{-\mu}\mu^{y} / y! = \exp\left(\frac{y\log\mu - \mu}{1} - \log y!\right),$$

which is of exponential family form. [1]

The link function is $g(\mu) = \log(\mu)$. [1]

The linear predictor is $\eta = \alpha_i$.

So this is a generalised linear model. [1]

(ii) The likelihood is

$$\prod_{i=1}^{3}\prod_{j=1}^{m}\frac{e^{-\mu_{ij}}\mu_{ij}^{y_{ij}}}{y_{ij}!},$$

so the log-likelihood is

$$\sum_{i=1}^{3}\sum_{j=1}^{m}(-\mu_{ij} + y_{ij}\log(\mu_{ij}) - \log(y_{ij}!))$$

i.e., in terms of α_i's, writing y_{i+} for the sum of the observations in the ith group, the log-likelihood is

$$l(\alpha_1, \alpha_2, \alpha_3) = -\sum_{i=1}^{3} me^{\alpha_i} + \sum_{i=1}^{3} y_{i+}\alpha_i + \text{constant.}$$ [2]

Differentiating,

$$\frac{\partial l}{\partial \alpha_i} = -me^{\alpha_i} + y_{i+},$$

so the maximum likelihood estimator of α_i is

$$\hat{\alpha}_i = \log(y_{i+}/\mathrm{m}). \qquad [2]$$

(iii) *Comparing models 2 and 3*:

There are 60 observations altogether.

Model 3 has one parameter estimate, and so has degrees of freedom 59.

Model 2 has degrees of freedom 58.

The drop in deviance in going from model 3 to model 2 is 72.53 − 61.64 = 10.89.

The corresponding drop in degrees of freedom is 59 −58 = 1.

So to test for a significant improvement, compare 10.89 to a χ_1^2.

The upper 5% point of χ_1^2 is 3.841, the upper 1% point is 6.635, this is a significant improvement. We prefer model 2 to model 3. [3]

Comparing models 2 and 1:

Model 1 has degrees of freedom 57.

The drop in deviance is 61.64 − 60.40 = 1.24, and this should be compared to χ_1^2.

It is not significant; do not prefer model 1 to model 2. [2]

(iv) *Interpretation of models*:

Model 3 says that there is no difference in the average number of claims for the three age groups.

Model 2 says that there is no difference in the average number of claims between age groups 1 and 2, but that the third age group may be different.

Model 1 gives the possibility of different average number of claims for each age group. [3]

11. (i) In terms of the backwards shift operator we have

$$(1 + 2\alpha B - \alpha^2 B^2)Y = Z.$$

We must find the values of α such that the roots of the polynomial $1 + 2\alpha x - \alpha^2 x^2$ lie outside the unit circle.

The roots are $\frac{1}{\alpha}\left(1 \pm \sqrt{2}\right)$, so we require that $\frac{\sqrt{2}+1}{|\alpha|} > 1$ and $\frac{\sqrt{2}-1}{|\alpha|} > 1$, in other words that $|\alpha| < \sqrt{2} - 1$. [3]

(ii) $Y_t = -2\alpha Y_{t-1} + \alpha^2 Y_{t-2} + Z_t$

$\mathrm{Cov}[Y_t, Y_t] = \gamma_0 = -2\alpha\gamma_1 + \alpha^2\gamma_2 + \sigma^2$ (1)

$\mathrm{Cov}[Y_t, Y_{t-1}] = \gamma_1 = -2\alpha\gamma_0 + \alpha^2\gamma_1$ (2)

$\mathrm{Cov}[Y_t, Y_{t-2}] = \gamma_2 = -2\alpha\gamma_1 + \alpha^2\gamma_0$ (3) [2]

From (2); $\gamma_1 = -\dfrac{2\alpha\gamma_0}{1-\alpha^2}$ (4)

Substitute for γ_1 from (4) into (3)

$$\gamma_2 = 2\alpha \cdot \frac{2\alpha\gamma_0}{1-\alpha^2} + \alpha^2\gamma_0 = \gamma_0 \cdot \left(\frac{5\alpha^2 - \alpha^4}{1-\alpha^2}\right) \quad (5)$$ [2]

substitute for γ_1 from (4) and γ_2 from (5) into (1)

$$\Rightarrow \quad \gamma_0 = \frac{\sigma^2\left(1-\alpha^2\right)}{\left(1+\alpha^2\right)\left(1-6\alpha^2+\alpha^4\right)} \quad (6)$$

substitute for γ_0 from (6) into (4) and (5) to find γ_1 and γ_2

$$\Rightarrow \quad \gamma_1 = \frac{-2\alpha\sigma^2}{\left(1+\alpha^2\right)\left(1-6\alpha^2+\alpha^4\right)}$$

and

$$\gamma_2 = \frac{\left(5\alpha^2-\alpha^4\right).\sigma^2}{\left(1+\alpha^2\right)\left(1-6\alpha^2+\alpha^4\right)}$$

$\boxed{[2]}$

12 (i) (a) First the u_k need to be transformed so that their distribution is something suitable for the white noise sequence of a time series, since at the very least the mean of the sequence needs to be zero. $N(0, \sigma_e^2)$ is the standard choice: one method of achieving this is to define, for each integer t,

$$e_{2t} = \sigma_e \sqrt{-2 \log u_{2t}} \sin(2\pi u_{2t+1})$$

$$e_{2t+1} = \sigma_e \sqrt{-2 \log u_{2t}} \cos(2\pi u_{2t+1}),$$

but there are others, such as the polar method, inverse transform method or acceptance-rejection sampling.

The values of the e_t can now be fed into the formula to give the values of the X_t, whichever model is in use. [3]

(b) The ability to re-use a pseudo-random number sequence is important when comparing the ability of different mechanisms to control a process which is affected by randomness: in order to ensure fair comparison of the mechanisms, the must be subjected to the same degree of "random" input. [1]

(ii) The models do not possess the correct correlation structure. [1]

(iii) (a) $\rho_1 = \text{Corr}(X_t, X_{t-1}) = \alpha_1 \text{Corr}(X_{t-1}, X_{t-1}) + \alpha_2 \text{Corr}(X_{t-2}, X_{t-1}) = \alpha_1 + \alpha_2 \rho_1$.

Hence $\rho_1 = \alpha_1 / (1 - \alpha_2)$

$\rho_2 = \text{Corr}(X_t, X_{t-2}) = \alpha_1 \text{Corr}(X_{t-1}, X_{t-2}) + \alpha_2 \text{Corr}(X_{t-2}, X_{t-2}) = \alpha_1 \rho_1 + \alpha_2$.

[2]

(b) We have $0.7 = \rho_1 = \alpha_1 / (1 - \alpha_2)$

and $0.5 = \rho_2 = \alpha_2 + \alpha_1^2 / (1 - \alpha_2) = \alpha_2 + 0.7\alpha_1$. Two equations in two unknowns. Solution: $\alpha_1 = \dfrac{35}{51}$, $\alpha_2 = \dfrac{1}{51}$.

[2]

(2 marks for the observation that $\alpha_1 = 0.7$ and $\alpha_2 = 0$ is very close to giving the right answer, as it gives $\rho_2 = 0.49$.)