Institute of Actuaries of India

Subject CT6-Stastatical Models

May 2008 Examination

INDICATIVE SOLUTION

Introduction

The indicative solution has been written by the Examiners with the aim of helping candidates. The solutions given are only indicative. It is realized that there could be other points as valid answers and examiner have given credit for any alternative approach or interpretation which they consider to be reasonable.

Question 1

(i) A Saddle point is the name given to :

- an entry of a payoff matrix
- that is the largest in its column
- and smallest in its row

(ii) If there is no saddle point, a randomized strategy can be adopted to enable a player to minimize his/her maximum expected loss.

This means that the player will vary his or her choice of strategy in a random fashion but in accordance with some fixed set of probabilities.

(iii) The largest number in the second column is 7.

But this number is not the smallest in its row.

The largest number in the third column could be 8 or y (if y > 8)

But 8 is not the smallest in its row,

And if y > 8 it cannot be the smallest in its row either

A saddle point can only occur in the first column.

If x < 4, then the largest number in the first column is 4, which is the smallest in its row, and hence is a saddle point.

If $4 < x \le 5$, then the largest number in the first column is x, which is the smallest in is row, and hence is a saddle point.

If x=4, then both x and 4 are two saddle points.

If x > 5, then the largest number in the first column is x, which is not the smallest in is row, and hence there is no saddle point.

[8]

Question 2

(i) (a) Consider the model $Y_t = a + bt + X_t$. If X_t is stationary with mean *m*, then mean of Y_t is (*m*+*a*)+*bt*, which is linear in *t*. The time plot of Y_t would exhibit a linear trend.

(b) Consider the model $Y_t = \sum_{s=1}^{t} X_s$. If X_t is stationary with mean *m*, then mean of Y_t is *mt*, which is linear in *t*. The time plot of Y_t would exhibit a linear trend.

(ii) (a) Assume model (a), fit linear regression and subtract fitted straight line, to obtain an approximation of the stationary time series X_t.

(b) Assume model (b), and difference the time series to obtain the stationary time series X_t.

[4]

Question 3

(i) Coefficient of variation, $\sqrt{\exp(\sigma^2)-1} = \frac{1000}{2000} = 0.5.$

It follows that $\sigma^2 = \ln(1.25) = 0.2231.$

We have $\exp(\mu + \sigma^2/2) = 2000.$ Hence, μ = 7.4893.

Proportion of claims involving re-insurer

P(claim > 3000) = P(log(claim) > log(3000))

$= 1 - \Phi\left(\frac{\ln(2500)-\mu}{\sigma}\right)$ = 1- Φ(0.7086)

= 0.2393 = 23.93%.

(ii) The mean of claim amounts paid by re-insurer is

$$\frac{1}{P(claim > 2500)} \int_{2500}^{\infty} y.\frac{1}{y}\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{\log y - \mu}{\sigma}\right)^2\right]dy - 2500$$

$$= \frac{1}{0.2393}\int_{[\log(2500)-\mu]/\sigma}^{\infty} \exp(\mu+\sigma z)\frac{1}{\sqrt{2\pi}}\exp\left[-\tfrac{1}{2}z^2\right]dz - 2500$$

$$= \frac{\exp\left(\mu+\tfrac{1}{2}\sigma^2\right)}{0.2393}\int_{[\log(2500)-\mu]/\sigma}^{\infty} \frac{1}{\sqrt{2\pi}}\exp\left[-\tfrac{1}{2}(z-\sigma)^2\right]dz - 2500$$

$$= \frac{2000}{0.2393}\int_{[\log(2500)-\mu]/\sigma-\sigma}^{\infty} \frac{1}{\sqrt{2\pi}}\exp\left[-\tfrac{1}{2}u^2\right]du - 2500$$

$$= \frac{2000}{0.2393}\left[1-\Phi\left(\frac{\log(2500)-\mu}{\sigma}-\sigma\right)\right] = \frac{2000}{0.2393}\times 0.4066 - 2500 = Rs.899.$$

(iii) Let the total number of claims be *N*, and the number of claims experienced by the re-insurer be *M*. Given *N*=*n*, the random variable *M* has the binomial distribution with parameters *N* and 0.2393. Thus,

$$P(M = m \mid N = n) = \binom{n}{m}(0.2393)^m(1-0.2393)^{n-m}.$$

(iv) The range of possible values of *N* is *m* to infinity.

(v) The unconditional probability that the re-insurer is involved in *m* claims is

$$
\begin{aligned}
P(M = m) &= \sum_{n=m}^{\infty} P(M = m \mid N = n) P(N = n) \\
&= \sum_{n=m}^{\infty} \binom{n}{m} (0.2393)^m (1-0.2393)^{n-m} e^{-100} \frac{100^n}{n!} \\
&= \sum_{l=0}^{\infty} \binom{l+m}{m} (0.2393)^m (1-0.2393)^l e^{-100} \frac{100^{l+m}}{(l+m)!} \\
&= \frac{(0.2393)^m \times (100)^m}{m!} \sum_{l=0}^{\infty} (1-0.2393)^l e^{-100} \frac{100^l}{l!} \\
&= \frac{(0.2393)^m \times (100)^m}{m!} e^{-100} \times e^{100(1-0.2393)} = e^{-23.93} \frac{(23.93)^m}{m!}.
\end{aligned}
$$

Thus, *M* has the Poisson distribution with mean 23.93.

(vi) Aggregate claim amount paid by the re-insurer has a compound Poisson distribution. The expected value of the aggregate claims is

E(individual claim size) x E(claim number) = 899 x 23.93 = Rs. 21500.

[12]

Question 4

(i) The distribution of *Y* is geometric with probability parameter *p*, i.e.,

$$P(Y = y) = (1-p)^{y-1} p, \qquad y = 1,2,3,\ldots$$

(ii) $$E(Y) = \sum_{y=1}^{\infty} yp(1-p)^{y-1} = 1/p.$$

(iii) $$P(Y = y) = \exp[y \log(1-p) + \log[p/(1-p)]].$$

Comparing this expression with the general form of an exponential family, $\exp\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\varphi)\right]$, one can set $\theta = \log(1-p)$, $\phi = 1$, $a(\phi) = 1$, and $b(\theta) = -\log[p/(1-p)] = \log(1-p) - \log p = \theta - \log(1-e^{\theta})$.

The natural parameter is $\theta = \log(1-p)$.

The mean function is $b'(\theta) = 1 + e^{\theta}/(1-e^{\theta}) = 1/(1-e^{\theta})$, which is the same as $1/p$.

The variance function is $b''(\theta) = e^{\theta} /(1-e^{\theta})^2$, which is equivalent to $(1-p)/p^2$.

(iv) Since the mean function is $E(Y) = b'(\theta) = 1/(1-e^{\theta})$, the canonical link function is $\theta = \log[1-1/E(Y)]$.

A generalized linear regression model based on the canonical link function is

$$\log[1-1/E(Y \mid X)] = \beta_0 + \beta_1 X \ .$$

(v) Note that $\beta_0 + \beta_1 X = \log[1-p]$, i.e., $p = 1 - e^{\beta_0 + \beta_1 X}$. The likelihood function is

$$\prod_{i=1}^{n}\left[p_i(1-p_i)^{Y_i-1}\right] = \prod_{i=1}^{n}\left[\left(1-e^{\beta_0+\beta_1 X_i}\right)\left(e^{\beta_0+\beta_1 X_i}\right)^{Y_i-1}\right].$$

[10]

Question 5

(i) Let X be the size of the first claim, so that X has an exponential distribution with parameter 1. Then for ruin to occur at time t we need $X > U + (1+\alpha)\lambda t$.

$$P[X > U + (1+\alpha)\lambda t] = \int_{U+(1+\alpha)\lambda t}^{\infty} \exp(-x)dx$$
$$= \exp[-U - (1+\alpha)\lambda t]$$

(ii) Let T denote the time until the first claim. Then T has an exponential distribution with parameter λ, and

P(Ruin at first claim)

$= \int_0^\infty$ P(Ruin at first claim| first claim occurs at t) * $f_T(t)\,dt$ $= \int_0^\infty$ exp(-U)*exp(-(1+α)λt)* λexp(-λt) dt

$= \int_0^\infty$ exp(-U)* λ*exp(-(2+ α) λt) dt

= [- exp(-U) * {λ / ((2+ α) λ)}* exp(-(2+ α) λt)$]_0^\infty$

= exp(-U) / (2+ α)

(iii) We need $\exp(-U) / (2+ \alpha) < 0.05$

i.e., $\exp(-U) < 0.05 * (2+ \alpha)$

i.e., $20 \exp(-U) < 2+ \alpha$

i.e., $\alpha > 20 \exp(-U) – 2$

[10]

Question 6

Let G be the event that the defendant is guilty, and E be the event that the DNA test would find a match between his blood and blood sample found at the crime scene.

$P(G) = 0.10$

$P(E|G) = 0.99$

$P(E|\text{not } G) = 0.01$

Revised probability the defendant is guilty, given the DNA evidence

$$P(G \mid E) = \frac{P(E \mid G)P(G)}{P(E \mid G)P(G) + P(E \mid \text{not } G)P(\text{not } G)}$$

= (0.10 * 0.99)/[(0.10 * 0.99) + (0.90 * 0.01)]

= 0.099/(0.099 + 0.009)

= 0.9167.

[5]

Question 7

The cumulative distribution function of X is

$F_x(X \le x)$ = 0.00 for $x < 2$
= 0.15 for $2 \le x < 3$
= 0.35 for $3 \le x < 5$
= 0.60 for $5 \le x < 7$
= 0.80 for $7 \le x < 11$
= 1.00 for $11 \le x$

If the uniform random variate u is such that $F(x_{i-1}) < u \le F(x_i)$ then return x_i as the random variate from P_x.

Using the above method following is the table of the random variates from Px corresponding to the given Uniform variates.

Unifrom Variate	0.011	0.757	0.438	0.258	0.981	0.518	0.4	0.351	0.672
Random Variate from Px	2	7	5	3	11	5	5	5	7

[4]

Question 8

(i) Level 1 & 2 P(0 Claims) = exp(–0.60) = 0.5488

P(>0 Claims) = 1 – 0.5488 = 0.4512

Level 3 & P P(0 Claims) = exp(-0.40) = 0.6703

P(1 Claim) = 0.4*exp(-0.40) = 0.2681

P(>1 Claim) = 1 – 0.6703 – 0.2681 = 0.0616

Probability of moving to Level P:

From Level 2 = 10% * 0.5488 = 0.05488

From Level 3 = 10% * 0.6703 = 0.06703

From Level P = 25%*(1 – 0.0616) = 0.23461

Probability of moving to Level 3:

From Level 2 = 90% * 0.5488 = 0.4939

From Level 3 = 90% * 0.6703 = 0.6033

From Level P = 75% * (1- 0.0616) = 0.7038

Therefore the transition matrix is

$$\begin{bmatrix} 0.4512 & 0.5488 & 0 & 0 \\ 0.4512 & 0 & 0.4939 & 0.0549 \\ 0.0616 & 0.2681 & 0.6033 & 0.0670 \\ 0.0616 & 0 & 0.7038 & 0.2346 \end{bmatrix}$$

(ii) The vector of state probabilities Π, in the steady state, satisfies the equation

$$\Pi \begin{bmatrix} 0.4512 & 0.5488 & 0 & 0 \\ 0.4512 & 0 & 0.4939 & 0.0549 \\ 0.0616 & 0.2681 & 0.6033 & 0.0670 \\ 0.0616 & 0 & 0.7038 & 0.2346 \end{bmatrix} = \Pi$$

$$
\begin{array}{lllll}
0.4512\,\pi_1 & + 0.4512\,\pi_2 & + 0.0616\,\pi_3 & + 0.0616\,\pi_p & = \pi_1 \\
0.5488\,\pi_1 & & + 0.2681\,\pi_3 & & = \pi_2 \\
 & 0.4939\,\pi_2 & + 0.6033\,\pi_3 & + 0.7038\,\pi_p & = \pi_3 \\
 & 0.0549\,\pi_2 & + 0.0670\,\pi_3 & + 0.2346\,\pi_p & = \pi_p \\
\pi_1 & + \pi_2 & + \pi_3 & + \pi_p & = 1
\end{array}
$$

Hence:

$\pi_p = 0.0717\,\pi_2 + 0.0875\,\pi_3$

$\pi_3 = 1.245\,\pi_2 + 1.774\,\pi_p$

$\pi_p = 0.0717\,\pi_2 + 0.1089\,\pi_2 + 0.1552\,\pi_p$

$\pi_3 = 1.245\,\pi_2 + 0.1272\,\pi_2 + 0.1552\,\pi_3$

$\pi_p = 0.2138\,\pi_2$

$\pi_2 = 0.6156\,\pi_3$

$0.5488\,\pi_1 + 0.2681\,\pi_3 = 0.6156\,\pi_3$

$\pi_3 = 1.5793\,\pi_1$

$\pi_2 = 0.9722\,\pi_1$

$\pi_p = 0.2078\,\pi_1$

Solving the equations, we get:

$\pi_1 = 0.266$,

$\pi_2 = 0.2586$,

$\pi_3 = 0.42$,

$\pi_p = 0.0553$.

(iii) The premium at the different levels are:

Level 1 = Rs. 500,
Level 2 = Rs. 375,
Level 3 = Rs. 250,
Level P = Rs. 250 + Rs. 50 = Rs. 300.

Average premium per policy

= 0.266*500 + 0.2586*375 + 250*0.42 + 300*0.0553

= Rs. 351.57.

[15]

Question 9

Adjust Individual claim amounts to mid-2007 prices:

Figures in Rs 000's

	Development Year			
Policy Year	0	1	2	3
2004	1778	1440	950	350
2005	2271	1662	1200	
2006	2532	1900		
2007	3000			

Cumulative Claim amounts:

Figures in Rs 000's

	Development Year			
Policy Year	0	1	2	3
2004	1778	3218	4168	4518
2005	2271	3933	5133	
2006	2532	4432		
2007	3000			
Column sum		11583	9301	4518
Column sum - last entry	6518	7151	4168	

Development factors for each year:

$D_3 = 4518/4168 = 1.08397$

$D_2 = 9301/7151 = 1.30066$

$D_1 = 11583/6518 = 1.76007$

Ultimate Claim amount for 2007 policies = 0.83 * Rs 55,00,000

= Rs 45,65,000.

Hence, outstanding amounts in Rs 000's arising from 2007 policies:

2007, 1 = $4{,}565 * \left(1 - \dfrac{1}{1.08397}\right)$ = 354

2007, 2 = $4{,}565 * \left(\dfrac{1}{1.08397} - \dfrac{1}{1.08397 * 1.30066}\right)$ = 973

2007, 3 = 4,565 * $\left(\frac{1}{1.08397*1.30066} - \frac{1}{1.08397*1.30066*1.76007}\right)$ = 1398

Adjust for future inflation = 354 * 1.05^3 + 973 *1.05^2 + 1398 *1.05

= 2950 = Rs 29,50,000.

[13]

Question 10

(i) If the prior distribution of a parameter θ is such that, given a sample from a distribution parametrized by θ, the posterior distribution of θ belongs to the same family as the prior distribution, then the prior is called a conjugate prior.

(ii) Suppose that the prior distribution of λ is Gamma(α,*s*). If X_1, X_2, …, X_n are samples from the exponential distribution with parameter λ, then the posterior density satisfies:

f(λ/ *X*) ∞ *f*(*X*/λ)*f*(λ)

$$= \left[\prod_{i-1}^{n}\left[\lambda\exp(-\lambda X_i)\right]\right]\frac{s^{\alpha}\lambda^{\alpha-1}e^{-s\lambda}}{\Gamma(\alpha)}$$

$$\propto \quad \lambda^{\alpha+n-1}\exp\left[-\lambda\left(s+\sum_{i=1}^{n}X_i\right)\right]$$

$\propto$ pdf of $\Gamma\left(\alpha+n, s+\sum_{i=1}^{n}X_i\right)$.

This means that the posterior distribution of also follows a Gamma distribution and therefore the Gamma distribution satisfies the definition of a conjugate prior.

(iii) (a) We know that λ ~ $\Gamma(\alpha, s)$. So

$$E(1/\lambda) = \int_0^\infty \frac{f(\lambda)}{\lambda}\,d\lambda$$

$$= \int_0^\infty \frac{s^{\alpha}\lambda^{\alpha-1}e^{-s\lambda}}{\lambda\Gamma(\alpha)}\,d\lambda$$

$$= \int_0^\infty \frac{s^{\alpha}\lambda^{\alpha-2}e^{-s\lambda}}{\Gamma(\alpha)}\,d\lambda$$

$$= \frac{s}{\alpha-1} \int_0^\infty \frac{s^{\alpha-1} \lambda^{\alpha-2} e^{-s\lambda}}{\Gamma(\alpha-1)} \, d\lambda$$

$$= \frac{s}{\alpha-1}$$

(b) Posterior mean is E(1/λ) where λ ~ $\Gamma\left(\alpha+n, s+\sum_{i=1}^{n} Xi\right)$. The result of part (a) implies that the posterior mean is given by

$$\frac{s+\sum_{i=1}^{n} Xi}{\alpha+n-1} = \frac{s}{\alpha+n-1} + \frac{\sum_{i=1}^{n} Xi}{\alpha+n-1}$$

$$= \frac{\alpha-1}{\alpha+n-1} * \frac{s}{\alpha-1} + \frac{n}{\alpha+n-1} * \frac{\sum_{i=1}^{n} Xi}{n}$$

$$= (1-Z) * \frac{s}{\alpha-1} + Z * \frac{\sum_{i=1}^{n} Xi}{n},$$

where the credibility factor, Z, is equal to $\frac{n}{\alpha+n-1}$.

[11]

Question 11

(i) A stochastic process X_t is said to be a Markov Process if it satisfies the property

$$P[X_t \in A \mid X_{S_1} = x_1, X_{S_2} = x_2, \ldots X_{S_n} = x_n, X_S = x] = P[X_t \in A \mid X_S = x]$$

for all times $s_1 < s_2 < \ldots < s_n < s < t$, all states $x_1, x_2, \ldots, x_n$ and all subsets A of the state space S.

(ii) 1.

(iii) $$\begin{bmatrix} X_t \\ X_{t-1} \end{bmatrix} = \begin{bmatrix} 0.5 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} X_{t-1} \\ X_{t-2} \end{bmatrix} + \begin{bmatrix} e_t \\ 0 \end{bmatrix}.$$

[5]

Question 12

The number of turning points T approximately follows the normal distribution with mean $E(T) = 2(N-2)/3 = 65.33$ and variance $Var(T) = (16N-29)/90 = 17.455$ (using $N = 100$).

The test statistic is $[T- E(T)]/S.D.(T)$. As $T = 43$, the test statistics is -5.345. This is very much outside the range of +/- 1.96.

So we reject the hypothesis that the residuals are independent.

[3]

****