Institute of Actuaries of India

Subject CT6 – Statistical Methods

September 2018 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.

Solution 1:

i) Exp(μ) model can be written as

$f(y) = \mu \exp(-\mu y)$
$= \exp(\log\mu - \mu y)$; $E(y) = 1/\mu$

$\theta = -\mu$, $b(\theta) = -\log(-\theta)$, $\phi = 1$, $a(\phi) = \phi$
and $c(y, \phi) = 0$

[2]

ii) Maximum likelihood estimation
$\log L(\mu 1, \mu 2) = \sum_{i=1}^{n} \log\mu i - \mu i y i$

$\log L(\alpha, \beta) = \sum_{i=1}^{15} \log\mu i + \sum_{i=16}^{30} \log\mu i - \sum_{i=1}^{15} \mu i y i - \sum_{i=16}^{30} \mu i y i$

Substituting the model form in above equation we get,

$\log L(\alpha, \beta) = \sum_{i=1}^{15} -(\frac{1}{2\alpha}) + \sum_{i=16}^{30} \beta - \sum_{i=1}^{15} e\hat{}(-\frac{1}{2\alpha}) y i - \sum_{i=16}^{30} e\hat{}\beta y i$

partially differentiating wrt to α and setting = 0 we get

$\frac{\partial \log L}{\partial \alpha}$ $= 15/2\alpha^2 - (e^{(-1/2\alpha)} / 2\alpha^2) * \sum_{i=1}^{15} yi = 0$

$\Rightarrow e^{-(1/2\alpha)} = 15/\sum_{i=1}^{15} yi = 15/166$

Therefore,

$-1/2\alpha = -2.404$

$\alpha = 0.21$

Similarly setting the partial derivative wrt $\hat{\beta} = 0$ we

$\frac{\partial \log L}{\partial \beta} = 15 - \sum_{i=16}^{30} yi e^{\beta}$

$\Rightarrow e^{\beta} = 15/157$

$= \beta = -2.348$

above are the Mle's of α and β. [6]

iii) Scaled deviance for the model m:

$SD = 2(\log Ls - \log Lm)$

Where logLs is the value of the likelihood function for the saturated model and logLm is the value of the likelihood function for model m.

For the saturated model $\mu i = 1/yi$

Ls = $\prod_{i=1}^{30} \mu i \, e^{-\mu i y i}$

= $\prod_{i=1}^{30} \mu i \, e^{-1}$

logLs= $-\sum_{1}^{30} logyi$ -30

= -84.94

Similarly with the estimate of α and β calculated above log likelihood of the model I is

= $-15/2\ \alpha - e^{-1/2\alpha} \sum_{1}^{15} yi + 15\ \beta - e^{\beta} \sum_{16}^{30} yi$

= -35.71 -15.34 -35.22 -15

= -101.27

Scaled deviance = 2(-84.94 +101.27)

= 33.52 [4]

[12 Marks]

Solution 2:

i) The formula for the total claims from the portfolio is:

S = *X*1 + *X*2 +….+ *Xn*

where *Xi* is the claim amount from the *i* th member (which may be zero).

The assumptions underlying this model are:

- there are a fixed number of risks (*ie* members), *n*
- claims occur independently for each member
- the number of claims for each member is either 0 or 1. [2]

ii) Let *X* = *bI* , where *I* is an indicator random variable denoting whether or not a claim is paid, *ie* *P*(*I* = 0) = 1- *q* , *P*(*I* = 1) = *q* and *b* is the fixed benefit amount.

Therefore *E*(*I*) = *q* , var(*I*) = *q*(1- *q*) .

So:

E(*X*) = *bE*(*I*) = *bq*

and:

var(*X*) = b^2var(*I*) = $b^2q(1- q)$

[3]

iii) If *S* is the total claim amount, then:

E(*S*) = 1,250 * 50,000 * 0.008 + 250 * 20,000 * 0.012 = 560,000

& var(S) = 1250* $50{,}000^2$ * 0.008 * 0.992 + 250 *20000 * 0.012 * 0.988

= 2.59856 *10^{10} [2]

iv) Using a normal approximation for *S* , we have:
S ~ N (560000, 2.59856 * 10^{10})
So the probability is given by:
P(S > 1,000,000) = P(Z > (1,000,000 - 560,000)/Sqrt(2.59856 *10^{10}))
= 1 – P(Z <2.59856) , Where Z is a standard normal variate.
= 0.00317 [3]

v) A Normal approximation provides correct value if the distribution symmetric. However, according to law of large numbers, a distribution with sufficiently large number of observations may be approximated by a normal distribution. Here, considering large number of data points of 1500, we may conclude that calculated probability is close to accurate. [1]

[11 Marks]

<u>Solution 3:</u>

i) Number of claims follow type 2 negative binomial distribution with parameter 2 and p.
Hence pdf can be written as:

$$p(x) = \frac{\Gamma(2+x)}{\Gamma(x+1)} p^2 (1-p)^x, x = 0,1,2,....$$

$$p(x) = (x+1) p^2 (1-p)^x, x = 0,1,2,....$$

The likelihood function for n sample observed claims can be written as

$$L(p) = (x_1 + 1) p^2 (1-p)^{x_1} \times ......\times (x_n + 1) p^2 (1-p)^{x_n}$$

$$L(p) = p^{2n} (1-p)^{\sum x_i} \prod (x_i + 1)$$

Taking log both sides

$$\log L(p) = 2n \log p + \sum x_i \log(1-p) \quad + C$$

Differentiating w.r.t p and setting equal to 0 we get

$$\frac{\partial}{\partial p} \log L(p) = \frac{2n}{p} - \frac{\sum x_i}{1-p} = 0$$

$$\hat{p} = \frac{2n}{2n + \sum x_i}$$

Differenting second time we get

$\frac{\partial^2}{\partial p^2} \log L(p) = -\frac{2n}{p^2} - \frac{\sum x_i}{(1-p)^2}$, which is negative hence the derived value is maximum.

Hence mle of p is $\frac{2n}{2n+\sum x_i}$ [3]

ii) The prior distribution of p follows Beta distribution with parameters α and β.

$$prior(p) \propto p^{\alpha-1}(1-p)^{\beta-1}, 0 < p < 1$$

The likelihood function is given by

$$L(p) \propto p^{2n}(1-p)^{\sum x_i}$$

Hence the posterior distribution will follow

$$post(p) \propto p^{2n+\alpha-1}(1-p)^{\sum x_i+\beta-1}$$

The above follows another beta distribution with parameters ($2n+\alpha$, $\sum x_i + \beta$).

The posterior estimate under squared error loss is the mean of the posterior distribution. The mean of the above beta distribution is given by

$$Mean = \frac{2n+\alpha}{2n+\alpha+\sum x_i + \beta}$$

The above can be written as

$$Mean = \frac{2n}{2n+\alpha+\sum x_i + \beta} + \frac{\alpha}{2n+\alpha+\sum x_i + \beta}$$

$$= \frac{2n}{2n+\sum x_i}\frac{2n+\sum x_i}{2n+\alpha+\sum x_i + \beta} + \frac{\alpha}{\alpha+\beta}\frac{\alpha+\beta}{2n+\alpha+\sum x_i + \beta}$$

Lets $Z = \frac{2n+\sum x_i}{2n+\alpha+\sum x_i + \beta}$

Then the above can be written as

$$\hat{p} = MLE(p)Z + \frac{\alpha}{\alpha+\beta}(1-Z)$$

Which is the required the credibility estimate form. [5]

iii) Under squared error loss, the posterior estimate is given by mean of the distribution.

Here, n = 20, $\sum x_i$ =150, α = 5,β=3

$$Mean = \frac{2n+\alpha}{2n+\alpha+\sum x_i + \beta} = \frac{2\times 20+5}{2*20+5+150+3} = 0.2273$$

Hence the posterior estimate of p under squared error loss is given by 0.2273

The posterior estimate under all-or-nothing is mode of the distribution.

Mode of beta distribution (α,β) is given by (α-1)/(α+β-2)

The same can be derived as

$$\log f(p) = (\alpha - 1)\log p + (\beta - 1)\log(1 - p) + C$$

Differentiating and setting equal to zero we get

$$\frac{\partial}{\partial p}\log f(p) = \frac{\alpha - 1}{p} - \frac{\beta - 1}{1 - p} = 0$$

$$p = \frac{\alpha - 1}{\alpha + \beta - 2}$$

$$\frac{\partial^2}{\partial p^2}\log f(p) = -\frac{\alpha - 1}{p^2} - \frac{\beta - 1}{(1 - p)^2}$$, which is negative for α,β>1

Hence the mode of posterior distribution = (2*20+5-1)/(2*20+5+150+3-2)= 0.2245
Hence the required posterior estimate under all-or-nothing is 0.2245.
The posterior estimate under absolute error loss is median of the distribution.
Median of beta distribution (α,β) is given by (α-1/3)/(α+β-2/3) (approx.)
Hence the median of the posterior distribution = (2*20+5-1/3)/(2*20+5+150+3-2/3)= 0.2264
Hence the required posterior estimate under absolute error loss is 0.2264. [5]

[13 Marks]

Solution 4:

i) Lack of stationarity:

- Lack of stationarity may be caused by the presence of deterministic effects in the quanitity being observed. For ex. Deterministic trend or cycle such as seasonal effects.
- If the process observed is the integrated version of a more fundamental process.
- For ex. A company which sells greeting cards will find that the sales in some months will be much higher than in others. [3]

ii)

a) Any of the below is a form of ARIMA(p,d,q) process
$(1-B)^d$ ø(B) (Xt-μ) = θ(B)e_t where ø(B) = 1-$\sum_1^p B^i \alpha i$ and θ(B) = $\sum_1^q B^j \beta j$

$(\nabla^d Xt - \mu)$ = $\sum_1^p \alpha i(\nabla^d Xt - 1 - \mu)$ +e_t + $\sum_1^q B^j et - j$ [2]

b) Main steps involved in Box-Jenkins methodology:

- Tentative identification of a model from the ARIMA class
- Estimation of the parameter in the identified model
- Diagnostic checks [2]

c) If the sample auto correlation coefficients decay slowly from 1 then this indicates that further differencing is required. This is not the case for d=1, which means that the differencing of the original series is required once hence d=1.
Further the correct value of d minimise the sample variance. This also indicates that d=1. [2]

d) Classify the time series as ARIMA(p,d,q)

1. $X_t = 0.8e_{t-1} + e_t$

This is a MA(1) process and hence it is stationary. Therefore we can calssify it as ARIMA(0,0,1) process.

2. $Xt=2X_{t-2}+ e_t +0.5\ e_{t-3}$

This is an ARMA(2,3) process. This process can not be differenced so we can classify it as ARIMA (2,0,3), we also must see if I(0) is stationary.

$=(1-2B^2)X_t = e_t+0,5e_{t-3}$, the characeteristic equation of AR terms is
$= ø(λ)=1-2λ^2=0$
$λ = +- 1/√2$, both the roots of the characteristic equation are less than 1. We can not classify the process as ARIMA(2,0,3). Hence its a non stationary ARMA(2,3) process.

3. $Xt=1.5X_{t-1}+.5X_{t-2}+ e_t + e_{t-1}$

This is an ARMA(2,1) process.
The process can be differenced as follows
$X_t-1.5X_{t-1}+.5X_{t-2} = e_t + e_{t-1}$
$= (X_t-X_{t-1})-.5(X_{t-1}-X_{t-2}) = e_t + e_{t-1}$
$=\nabla Xt - .5\nabla Xt - 1 = e_t + e_{t-1}$

This process can not be differenced further hence we take d=1.
To classify the process as ARIMA(1,1,1) we need to check if the differenced process is stationary.
The characteristic equation is (1-.5λ)=0 which implies λ=2
The differenced process is stationary hence we can classify it as ARIMA(1,1,1) process.

[6]

[15 Marks]

Solution 5:

i) assumption under B-F method

- Payment from each origin year will develop in the same way
- Past inflation will be repeated in future
- The first year is fully run-off
- Loss ratios are appropriate [2]

ii) first lets calculate the development factors

$F_{1.2} = \frac{426+460+540}{400+380+508} = 1.107$

$F_{2,3=} = \frac{469+490}{426+460} = 1.082$

$F_{3,4} = \frac{525}{469} = 1.119$

For year 2013, the estimated loss ratio is = 525/675 = 77.78%

For accident year 2014 the initial estimate of the loss is =77.78%*700=544.44
Expected claims incurred by end of development year 3 is 544.44/$f_{3,4}$ = 486.37
Actual claims incurred by end of DY3 is 490 which is 3.62 higher than expected.
Revised estimate of the ultimate loss is 544.44+3.62=548.07

For accident year 2015
Initial estimate of ultimate loss is =77.78%*890 = 692.2
Expected claims incurred by end of year 2 is = 692.2/$f_{3,4}$*$f_{2,3}$
=692.2/(1.119*1.082)=571.31
Actual claims incurred by end of DY2 is 540 which is 31.31 lower than expected.
Revised estimate of the ultimate loss is 692-31.31=660.9

For accident year 2016
Initial estimate of ultimate loss is =77.78%*870= 676.67
Expected claims incurred by end of year 1 is = 676.67/$f_{2,3}$*$f_{1,2}$*$f_{3,4}$
=676.67/(1.119*1.082*1.107)=504.43
Actual claims incurred by end of DY1 is 490 which is 14.42 lower than expected.
Revised estimate of the ultimate loss is 676.67-14.42= 662.23
Hence the estimate of the total amount of calims incurred is 525+ 548.07+660.9+662.23=2396.22
and the outstanding claims reserve is 2396.22-1900=496.22 [7]

iii) We need to adjust the non-cumulative claims adjusted for past Inflation, claims at mid 2016 prices

Claims incurred ('000s)	Development Year				
		1	2	3	4
Accident year	2013	420.00	28.67	49.78	68.07
	2014	399.00	88.20	34.73	
	2015	533.40	35.28		
	2016	514.50			

we get following table for the cumulative claims;

Claims incurred ('000s)	Development Year				
Accident year		1	2	3	4
	2013	420.00	448.67	498.44	566.51
	2014	399.00	487.20	521.93	
	2015	533.40	568.68		
	2016	514.50			

Calculating the development factors
$F_{1,2}$ =1.113
$F_{2,3}$= 1.090
$F_{3,4}$=1.137

Expected cumulative claims adjusted for past and future inflation

Claims incurred ('000s)	Development Year				
Accident year		1	2	3	4
	2013	420.00	448.67	498.44	566.51
	2014	399.00	487.20	521.93	622.86
	2015	533.40	568.68	651.03	776.94
	2016	514.50	601.00	688.03	821.09

The expected claim amount is 2787.40 and the outstanding claims reserve after adjusting for inflation is 2787.40-1900=887.40 [7]

[16 Marks]

Solution 6:

i) Linear Congruential Generators

To obtain the desired sequence of random numbers u1,u2,...,uN , first generate a sequence of integers x1, x2,..., xN in the range {0,1,2,...,m -1} starting from an initial value x0 , known as the seed. The sequence of integers is generated using the recursive rule:
xn = (axn-1 + c)(mod m), n = 1, 2,..., N
where a is multiplier, c is increment, m is modulus

After that, un is set equal to xn/ m. This method generates random numbers on the interval [0,1) rather than [0,1]. In fact, it is clear that it can only give numbers of the form i /m where i = 0,1,2,....,m-1. [2]

ii) Mean of Beta distribution is given by = 3/(3+2) = 0.6
Variance of Beta distribution is given by = 3*2/((3+2) ^{2}(3+2+1))= 0.04

Box Muller Method
Using Box-Muller formula we get

$$z_1 = \sqrt{-2\ln 0.346} \times \cos(2\pi \times 0.762) = 0.1097$$
$$z_2 = \sqrt{-2\ln 0.346} \times \sin(2\pi \times 0.762) = -1.4528$$

Now Z follows N(0,1)
And let Y follows beta distribution and approximating the same is normal distribution we get
$Y \approx N(0.6, 0.04)$
We can use the z = (y-mean)/standard deviation to generate numbers from beta distribution.

Hence, y_1 = 0.6+0.2*z_1= 0.6+0.2*0.1097 = 0.6219
And y_2 = 0.6+0.2*z_2= 0.6+0.2*(-1.4528) = 0.3094

Polar Method

We need randome variates from V ~ U(-1,1)
Now, F(v) = (v+1)/2 ; v = 2u-1 using the inverse transform method

Hence, v_1 = - 0.3080, v_2 = 0.5240, and s = v_1^2+ v_2^2= 0.3694

Standard normal variates can be calculated using the below formula:

$$z_1 = v_1\sqrt{\frac{-2\ln s}{s}} = -0.7151$$
$$z_2 = v_2\sqrt{\frac{-2\ln s}{s}} = 1.2166$$

Using normal approximation of beta distribution we get,
y_1 = 0.6+0.2*z_1= 0.6+0.2*(-0.7151) = 0.4570
y_2 = 0.6+0.2*z_2= 0.6+0.2*1.2166 = 0.8433

[4]

iii) Under absolute error method, the minimum number of simulations required follow below equation:

$$n > \frac{z_{\alpha/2}^2 \hat{\tau}^2}{\varepsilon^2}$$

Here, $z_{\alpha/2}$=2.17, $\hat{\tau}$ = 0.2, ε =0.5

Hence n> 18.84

The minimum number of simulation required under absolute error method is 19.

Under relative error method, the minimum number of simulations required follow below equation:

$$n > \frac{z_{\alpha/2}^2 \hat{\tau}^2}{\varepsilon^2 \theta^2}$$

Here, $z_{\alpha/2}$=2.17, $\hat{\tau}$ = 0.2, ε=0.5, θ= 0.6
Hence n>52.33

The minimum number of simulation required under relative error method is 53. [3]

[9 Marks]

Solution 7:

i) Individual claims (X) follow Gamma distribution with parameter 20 and 0.4.

Hence, Mean = E(X) = 20/0.4 = 50
Variance = V(X) = 20/ (0.4^2) = 125
Now, $E(X^2) = V(X) + (E(X))^2 = 2625$

Expenses (Y) incurred which follows pareto distribution with parameter 5 and 0.2 and independent of the claim amount

Hence, Mean = E(Y) = 100/(5-1) = 25
Variance = V(Y) = 1041.67
Now, $E(Y^2) = V(Y) + (E(Y))^2 = 1666.67$

Net Outgo for the Company for the ith claim is given by $O_i = X_i+Y_i$

Hence the aggregate claim $S = O_1+ ... + O_n$

S follows compound poisson distribution with parameter 0.5n

Now, E(S) = λE(O) = 0.5*1000*[E(X)+E(Y)] = 37,500
And $Var(S) = \lambda E(O^2) = 0.5*1000*[E((X+Y)^2)]=0.5*1000*E[X^2+Y^2+2XY]$
$=0.5*1000*(E(X^2)+E(Y^2)+2E(X)E(Y))$ (as X and Y are independent)
= 33,95,833

Mean and variance of aggregate net outgo are 37,500 and 33,95,833 respectively. [5]

ii) Using normal approximation for aggregate net outgo we get,
$S \approx N(37500, 3395833)$
Let p is the annual for each policy.

To generate profit, S < 1000p

$$P(S < 1000p) = P(Z < \frac{1000p - 37500}{\sqrt{3395833}})$$

The probability will be atleast 95% if

$$\frac{1000p - 37500}{\sqrt{3395833}} \geq 1.6449$$

Or p ≥ 40.53

Hence the required minimum annual premium should be charged is 40.53. [2]

[7 Marks]

Solution 8:

i) Let Y denotes the net claim payable by the insurer.

Hence $Y = \begin{cases} X & ,X < M \\ M + 0.2(X - M) & ,X \geq M \end{cases}$

Now the net premium income by the insurer is given by,

$$C_{net} = \lambda\,(1 + 0.25)\mathrm{E(X)} - \lambda(1 + 0.6)\mathrm{E(Z)}$$

$$= 1.25\lambda.\frac{1}{\beta} - 1.6\lambda \int_M^\infty 0.8(x - M)\beta e^{-\beta x} dx$$

$$= 1.25\lambda.\frac{1}{\beta} - 1.28\lambda[\frac{e^{-\beta M}}{\beta}]$$

$$= \frac{\lambda}{\beta}[1.25 - 1.28e^{-\beta M}]$$

Now the net claim payable by the insurer is = $\lambda\, E(Y)$

$$E(Y) = \int_0^M xf(x)dx + \int_M^\infty [M + 0.2(x - M)]f(x)dx$$

$$= \int_0^M x\beta e^{-\beta x} dx + 0.8M \int_M^\infty \beta e^{-\beta x} dx + 0.2 \int_M^\infty x\beta e^{-\beta x} dx$$

$$= [-xe^{-\beta x}]_0^M - [\frac{1}{\beta} e^{-\beta x}]_0^M + 0.8Me^{-\beta M} + 0.2[\,[-xe^{-\beta x}]_M^\infty - [\frac{1}{\beta} e^{-\beta x}]_M^\infty]$$

$$= -Me^{-\beta M} - \frac{1}{\beta}e^{-\beta M} + \frac{1}{\beta} + 0.8Me^{-\beta M} + 0.2[Me^{-\beta M} + \frac{1}{\beta}e^{-\beta M}]$$

$$= \frac{1}{\beta}(1 - 0.8e^{-\beta M})$$

The insurer will not make any loss if

$$\lambda\frac{1}{\beta}\left[1.25 - 1.28\, e^{-\beta M}\right] \geq \lambda\frac{1}{\beta}(1 - 0.8e^{-\beta M})$$

Or $0.48\, e^{-\beta M} \leq 0.25$

So, $M \geq 28.33$ [5]

ii) $M_Y(R) = E(e^{RY})$

$$= \int_0^M e^{Rx}\beta e^{-\beta x}dx + \int_M^\infty e^{R(0.8M+0.2x)}\beta e^{-\beta x}dx$$

$$= \frac{\beta - \beta e^{-M(\beta-R)}}{\beta-R} + \beta e^{08M}\int_M^\infty e^{-(\beta-0.2R)x}dx$$

$$= \frac{\beta - \beta e^{-M(\beta-R)}}{\beta-R} + \frac{\beta e^{-M(\beta-R)}}{\beta-0.2R}$$

$$= \frac{\beta}{\beta-R}[1 - \frac{0.8R}{\beta-0.2R}e^{-M(\beta-R)}]$$ [3]

iii) Equation of adjustment coefficient is

$\lambda + C_{net}R = \lambda M_Y(R)$

Or $\lambda + \frac{\lambda}{\beta}\left[1.25 - 1.28\, e^{-\beta M}\right]R = \frac{\lambda\beta}{\beta-R}[1 - \frac{0.8R}{\beta-0.2R}e^{-M(\beta-R)}]$

Or $-\beta R + (\beta - R)\left[1.25 - 1.28\, e^{-\beta M}\right]R = -\frac{0.8R\beta^2}{\beta-0.2R}e^{-M(\beta-R)}]$

Or $-\beta(\beta - 0.2R) + (\beta - 0.2R)(\beta - R)\left[1.25 - 1.28\, e^{-\beta M}\right] + 0.8e^{MR}e^{-M\beta}\beta^2 = 0$

Or $0.8e^{MR}\beta^2 - \beta e^{M\beta}(\beta - 0.2R) + (\beta - 0.2R)(\beta - R)\left[1.25e^{M\beta} - 1.28\right] = 0$ [3]

iv) Using M = 50 and β = 0.01 and $e^x = 1 + x + \frac{x^2}{2}$ we get

$$1 + 50\,R + 1250\,R^2 - 206.09(0.01 - 0.2R) + (0.0001 - 1.2\beta R + 0.2R^2)[9761.27] = 0$$

Solving for R and taking smallest +ve root we get R = 0.01509

[2]

v) Here M = 0, then we get the revised equation as

$$0.8\beta^2 - \beta^2 + 0.2\beta R - 0.03\,(\beta^2 - 1.2\beta R + 0.2R^2) = 0$$

Solving we get, R = 0.01, taking smallest +ve root [2]

vi) Initial surplus U = 100

Using Lundberg's Inequality $\psi(U) \leq e^{-RU}$

Hence, maximum ruin probability under (d) = exp(-0.01509*100) = 22.11%

And maximum ruin probability under (e) = exp(-0.01 * 100) = 36.78%

We can see that if the retention level reduces then the ruin probability increases. [2]

[17 Marks]

****