Institute of Actuaries of India

Subject CT6 – Statistical Methods

April 2016 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)

$$Y_t^2 - \beta_1 e_t^2 Y_{t-1}^2 = 2(Y_t - \beta_1 e_t^2 Y_{t-1})\mu - (1-\beta_1 e_t^2)\mu^2 + \beta_0 e_t^2$$

Or, $Y_t^2 - 2Y_t\mu + \mu^2 = e_t^2(\beta_0 + \beta_1 Y_{t-1}^2 - 2\beta_1 Y_{t-1}\mu + \beta_1\mu^2)$

Or, $(Y_t - \mu)^2 = e_t^2(\beta_0 + \beta_1(Y_{t-1} - \mu)^2)$

Or, $Y_t = \mu + e_t(\beta_0 + \beta_1(Y_{t-1} - \mu)^2)^{0.5}$

Now,

$$E(Y_t) = E(\mu) + E(e_t(\beta_0 + \beta_1(Y_{t-1} - \mu)^2)^{0.5})$$

e_t and Y_{t-1} are independent.

$$E(Y_t) = \mu + E(e_t)E((\beta_0 + \beta_1(Y_{t-1} - \mu)^2)^{0.5})$$

Or, $E(Y_t) = \mu + 0 \times E((\beta_0 + \beta_1(Y_{t-1} - \mu)^2)^{0.5})$

Hence, $E(Y_t) = \mu$

Now,

$$Cov(Y_t, Y_{t-s}) = E(Y_t Y_{t-s}) - E(Y_t)E(Y_{t-s})$$

Or, $Cov(Y_t, Y_{t-s}) = E((\mu + e_t(\beta_0 + \beta_1(Y_{t-1} - \mu)^2)^{0.5})(\mu + e_{t-s}(\beta_0 + \beta_1(Y_{t-s-1} - \mu)^2)^{0.5})) - \mu^2$

Or, $Cov(Y_t, Y_{t-s}) = E(\mu^2 + \mu e_t(\beta_0 + \beta_1(Y_{t-1} - \mu)^2)^{0.5} + \mu e_{t-s}(\beta_0 + \beta_1(Y_{t-s-1} - \mu)^2)^{0.5} + e_t(\beta_0 + \beta_1(Y_{t-1} - \mu)^2)^{0.5} e_{t-s}(\beta_0 + \beta_1(Y_{t-s-1} - \mu)^2)^{0.5}) - \mu^2$

Or, $Cov(Y_t, Y_{t-s}) = \mu^2 + \mu E(e_t)E((\beta_0 + \beta_1(Y_{t-1} - \mu)^2)^{0.5}) + \mu E(e_{t-s})E((\beta_0 + \beta_1(Y_{t-s-1} - \mu)^2)^{0.5}) + E(e_t)E(e_{t-s})E((\beta_0 + \beta_1(Y_{t-1} - \mu)^2)^{0.5}(\beta_0 + \beta_1(Y_{t-s-1} - \mu)^2)^{0.5}) - \mu^2$

Or, $Cov(Y_t, Y_{t-s}) = \mu^2 + 0 + 0 + 0 \times E((\beta_0 + \beta_1(Y_{t-1} - \mu)^2)^{0.5}(\beta_0 + \beta_1(Y_{t-s-1} - \mu)^2)^{0.5}) - \mu^2$

Or, $Cov(Y_t, Y_{t-s}) = 0$ [6]

ii)

Now,

$$Var(Y_t \mid Y_{t-1}) = Var(e_t)Var(\beta_0 + \beta_1(Y_{t-1} - \mu)^2) = \beta_0 + \beta_1(Y_{t-1} - \mu)^2$$

From the above equation we can see that variance of Y_t depends on Y_{t-1}. Similarly recursively we can see that variance of Yt will depend on Y_{t-s}. Hence Y_t and Y_{t-s} are dependent. [2]

iii)

The first difference of Xt can be written as given below:

$$\Delta X_t = X_t - X_{t-1}$$

Now,

$$E(\Delta X_t) = E(X_t) - E(X_{t-1})$$

Or, $E(\Delta X_t) = E(0.5Y_t + 0.3t + 0.1) - E(0.5Y_{t-1} + 0.3(t-1) + 0.1)$

Or, $E(\Delta X_t) = 0.5\mu + 0.3t + 0.1 - 0.5\mu - 0.3(t-1) - 0.1 = 0.3$

The mean is independent of t and hence constant.

Now,

$$Cov(\Delta X_t, \Delta X_{t-s}) = Cov(X_t - X_{t-1}, X_{t-s} - X_{t-s-1})$$

Or, $Cov(\Delta X_t, \Delta X_{t-s}) = Cov(0.3 + Y_t - Y_{t-1}, 0.3 + Y_{t-s} - Y_{t-s-1})$

Or, $Cov(\Delta X_t, \Delta X_{t-s}) = Cov(Y_t - Y_{t-1}, Y_{t-s} - Y_{t-s-1})$

Or, $Cov(\Delta X_t, \Delta X_{t-s}) = Cov(Y_t, Y_{t-s}) - Cov(Y_t, Y_{t-s-1}) - Cov(Y_{t-1}, Y_{t-s}) + Cov(Y_{t-1}, Y_{t-s-1})$

Or, $Cov(\Delta X_t, \Delta X_{t-s}) = 0 - 0 - 0 + 0 = 0$

The auto covariance function is constant hence the first difference of X_t is stationary.

[5]

[13 Marks]

Solution 2:

i)

One advantage that the negative binomial distribution has over the Poisson distribution is that its variance exceeds its mean. Mean and variance are equal for the Poisson distribution. Thus, the negative binomial distribution may give a better fit to a data set which has a sample variance in excess of the sample mean. This is often the case in practice.

[1]

ii)

Let S denote the aggregate claims for first insurance company.

$$S = X_1 + X_2 + \ldots + X_N$$

Where N is the total number of claims for first insurer

Now the moment generating function of S is given by

$$M_S(t) = E(e^{St}) = E(E(e^{X_1 t} e^{X_1 t} \ldots e^{X_1 t} \mid N)) = E(M_X(t)^N) = E(e^{N \log M_x(t)}) = M_N(\log M_X(t))$$

Now

$$M_N(t) = \left(\frac{p}{1 - qe^t}\right)^k$$

And

$$M_X(t) = \frac{\lambda}{\lambda - t}$$

Hence, $M_S(t) = \left(\frac{p}{1 - qM_X(t)}\right)^k = \left(\frac{p(\lambda - t)}{\lambda - t - q\lambda}\right)^k = \left(\frac{p(\lambda - t)}{p\lambda - t}\right)^k$ ------(1)

Now from Negative Binomial distribution we get,

$$Mean = \frac{kq}{p} = 100$$

$$Variance = \frac{kq}{p^2} = 150$$

So, p = 1/1.5 = 0.67, k = 200

Now from Exponential distribution we get,

Mean = 1/ λ = 100, hence λ = 0.01

Hence MGF of aggregate claim is given by $M_S(t) = (\frac{0.67(0.01-t)}{0.0067-t})^{200}$

Mean of aggregate claim = E(N)E(X)=100*100=10,000

Variance of aggregate claim = E(N)Var(X)+Var(N)(E(X))^2= 25,00,000

[7]

iii)

Let T denotes aggregate claim for second insurer.

$T = Y_1 + Y_2 + ... + Y_M$

From Binomial distribution we get

Mean = np = 40 and Variance = npq = 16, p = 0.6, n =66.67

The individual claims follow exponential distribution with parameter μ.

Hence E(Y)=1/ μ and Var(Y) =$(1/\mu)^2$

Now, MGF of binomial distribution with n and p parameters are given by,

$$M_R(t) = (q + pe^t)^n$$

And $M_Y(t) = \frac{\mu}{\mu - t}$

Now MGF of aggregate claim for second insurer is given by

$$M_S(t) = (q + p\frac{\mu}{\mu - t})^n = (\frac{q\mu - qt + p\mu}{\mu - t})^n = (\frac{\mu - qt}{\mu - t})^n \quad ....(2)$$

Hence, mean of aggregate claim = E(M)E(Y) = 40 /μ

And variance of aggregate claim = E(M)Var(Y)+Var(M)(E(Y))^2 = 56/(μ^2)

Comparing the above mean and variance with that of part (b), we get

40 /μ = 10,000 therefore μ = 0.004

56/(μ^2) = 25,00,000 therefore μ = 0.0047 (taking positive root) [5]

[13 Marks]

Solution 3:

i)

The earned premium in units of Rs 5000 for each of the accident year is as given below

AY	Earned premium(in units rs 5000)
2011	17500000/5000=3500
2012	19250000/5000=3850
2013	1850000/5000=370
2014	2050000/5000=410

Since claims are fully run off for Accident year 2011, we can compute the loss ratio as

= (1572+820+425+325)/3500 = 89.77%

The cumulative claims data is

Accident year	Development year			
	0	1	2	3
2011	1572	2392	2817	3142
2012	1600	2350	2800	
2013	1823	2723		
2014	1700			

The development factor for development year (0, 1) = (2392+2350+2350)/ (1572+1600+1823) = 1.494

Similarly the d. f for development year (1, 2) and development year (2, 3) are 1.185 and 1.115 respectively.

Initial ultimate liability = Earned Premium (for each AY) * loss ratio

Thus

AY	2011	2012	2013	2014
Initial ultimate Liability	3142.00	3456.20	332.154	368.063

Thus the emerging liability for each of the accident years is

	Emerging Liability (In rs 5000)	Emerging Liability "Column (2) *5000 "
AY		
2011	0.00	0.00
2012	=3456.20 * (1-1/1.115)= 357.50	1787500.00
2013	=332.154 * (1-1/1.321)= 80.747	403735.599
2014	=368.063 * (1-1/1.974)=181.654	908272.314

Thus the total emerging liability is 3099507.913 [5]

ii)

AY (t)	2011	2012	2013	2014
Inflation Index	400	429	465	516
Annual Inflation rate (t-1,t)		7.25%	8.39%	10.97%

We need to project the non-cumulative data after allowing for inflation. Thus we have

Non-Cumulative Claims data (after allowance for past inflation)

Accident year	Development year			
	0	1	2	3
2011	=1572*1.2900= 2027.88	=820*1.2028= 986.29	=425*1.1097= 471.61	325.00
2012	=1600*1.2028= 1924.48	=750*1.1097= 832.26	450.00	
2013	=1823*1.1097= 2022.94	900.00		
2014	1700.00			

Where 1.2900 = 1.1097*1.0839*1.0725

1.2028 = 1.1097*1.0839

The cumulative claims data is thus

Accident year	Development year			
	0	1	2	3
2011	2027.88	3014.17	3485.79	3810.79
2012	1924.48	2756.73	3206.73	
2013	2022.94	2922.94		
2014	1700.00			

The revised development factors based on inflation adjusted cumulative claims data is

DY		1	2	3
Development factor		1.4550	1.1597	1.0932

Projected Cumulative claims data after allowance for past inflation

AY	DY			
	0	1	2	3
2011	2027.88	3014.17	3485.79	3810.79
2012	1924.48	2756.73	3206.73	3505.72
2013	2022.94	2922.94	3389.74	3705.78
2014	1700.00	2473.44	2868.45	3135.89

Projected Non-cumulative claims

Accident year	Development year			
	0	1	2	3
2011				
2012				298.98
2013			466.79	316.04
2014		773.44	395.01	267.44

Projected Non-cumulative claims (after allowing for future inflation)

Accident year	Development year			
	0	1	2	3
2011				
2012				=298.98*1.10= 328.88
2013			=466.79*1.10= 513.47	=316.04*1.10^2= 382.41
2014		=773.44*1.10= 850.78	=395.01*1.10^2= 477.96	=267.44*1.10^3= 355.97

Thus the total reserve under Inflation adjusted basic chain ladder method is

AY	Reserve
2012	=(328.88)*5000=1644402.522
2013	=(382.41+513.47)*5000=4479432.986
2014	=(355.97+477.96+850.78)=8423548.999
Total	14547384.51

[9]

[14 Marks]

Solution 4:

i)

a) E (S) = E (A) + E (B)

= 3000 * 20 + 2000 * 10

= 80000 [2]

b) Variance (S) = Variance (A) + Variance (B)

= 20 * (3000^2 + 3000^2) + 10 * 2000^2

= 400000000

= 4 * 10^8 [2]

c) We need u such that P (u + c < S) = 0.025

Thus we have

P ((S – E (S))/ $\sqrt{Var(S)}$) > (u + c – E(S))/ $\sqrt{Var(S)}$) = 0.025

⇨ (u + c – E(S))/ $\sqrt{Var(S)}$) = 1.96

- ⇨ U = - c + E (S) +1.96 *$\sqrt{Var(S)}$)
- ⇨ U = -1.25 E (S) + E(S) + 1.96 * $\sqrt{Var(S)}$)
- ⇨ U = -.25 * 80000 + 1.96 * 20000
- ⇨ U = 19200

Thus the initial capital required to ensure that the probability of ruin at the end of first year is less than 2.5 % is 19200. [3]

ii)

a) The loss table for the reinsurer is as given below

Claims	Impact of reinsurance		
	Type 1	Type 2	No reinsurance
0	-600	-300	0
2000	-100	-300	0
3000	150	-300	0
4000	400	-300	0
5000	650	700	0

Where each cell computes the loss to the reinsurer as

Benefit paid by reinsurer less premium received by reinsurer

On this basis the loss table for the insurer is built which is as given below

Claims	Impact of reinsurance		
	Type 1	Type 2	No reinsurance
0	-900	-1200	-1500
2000	600	800	500
3000	1350	1800	1500
4000	2100	2800	2500
5000	2850	2800	3500

Where each cell computes the loss to the reinsurer as

Benefit paid by insurer less premium received by insurer less benefit received from reinsurer plus premium paid to reinsurer

[6]

b) The maximum losses under each of the three categories are

Type 1 = 2850
Type 2 = 2800
No reinsurance = 3500

The minimum among these is 2800. Hence the minimax solution to this problem is to choose excess of loss reinsurance (Type 2). [2]

[15 Marks]

Solution 5:

"p" has a beta (α,β) distribution and let X denote the no. of failures, so X has a Binomial (9000, p) distribution

The posterior distribution of p is

$f(p|X) \propto f(X|p) * f(p) \propto p^{\alpha-1}.(1-p)^{\beta-1}.\ p^{x}.(1-p)^{n-x}$

$\propto p^{\alpha+x-1}.(1-p)^{\beta+n-1}$

Thus the posterior distribution is Beta with parameters

$\alpha_x = \alpha + x$

$\beta_x = \beta + n - x$

From the prior distribution of p we have

$\alpha/(\alpha+\beta) = 0.013$ ………………….(i)

$\alpha\beta/((\alpha+\beta)^2.\ (\alpha+\beta+1)) = 0.004^2$ ……………………….(ii)

From equation (i) we have

$\alpha = (13/987)*\beta$ ………(iii)

Substituting (iii) in (ii) we get

$(13/987)\ \beta^2 / ((1000\ \beta/987)^2(1+1000\ \beta/987)) = 0.004^2$

- ⇨ $(13/987) = 0.004^2*(1000\ /987)^2*(1+1000\ \beta/987)$
- ⇨ $801.93 = 1+ 1000\ \beta/987$
- ⇨ $800.93 = 1000\ \beta/987$
- ⇨ $\beta = 790.52$

Therefore

$\alpha = (13/987)*\beta = 10.41$

Thus the parameters of the posterior distribution are

$\alpha_x = \alpha + x = 10.41 + 92 = 102.41$

$\beta_x = \beta + n - x = 790.52 + 9000 - 92 = 9698.52$

[6 Marks]

Solution 6:

i)

θ follows normal distribution with mean 5 and variance 10.

Hence, E (θ) = 5

And $Var(\theta) = E(\theta^2) - (E(\theta))^2 = 10$

$E(\theta^2) = 35$

Now $E(X1) = E(E(X1|\theta)) = E(10+\theta) = 10 + E(\theta) = 15$

Similarly, $E(X2) = E(E(X2|\theta)) = E(5+3\theta) = 5 + 3E(\theta) = 20$

Now $Var(X1) = Var(E(X1|\theta)) + E(Var(X1|\theta))$

Or, $Var(X1) = Var(10+\theta) + E(2+3\theta^2) = Var(\theta) + 2 + 3E(\theta^2) = 10 + 2 + 3*35 = 117$

Similarly, $Var(X2) = Var(E(X2|\theta)) + E(Var(X2|\theta))$

Or $Var(X2) = Var(5+3\theta) + E(3+5\theta^2) = 9Var(\theta) + 3 + 5E(\theta^2) = 90 + 3 + 5*35 = 268$ [5]

ii)

$E(X1X2) = E(E(X1X2|\theta)) = E(E(X1|\theta)E(X2|\theta))$, as X1 and X2 are conditionally independent given θ

Now, $E(X1X2) = E((10+\theta)(5+3\theta)) = E(50 + 35\theta + 3\theta^2) = 50 + 35*5 + 3*35 = 330$

Now if X1 and X2 are unconditionally independent then,

$E(X1X2) = E(X1)E(X2)$

Here, $E(X1)E(X2) = 15*20 = 300 \neq E(X1X2)$

Hence X1 and X2 are not unconditionally independent. [3]

[8 Marks]

Solution 7:

i)

Let each day t number of units are produced.

Hence, no of sales in each day (till 9 PM) = d , if d ≤ t

= t , if d > t

Hence, profit (p) = yd + z(t-d) – xt ,if d ≤ t

= yt – xt , if d > t

Hence, expected profit

$$E(p) = \sum_{d=0}^{t}(yd + z(t-d) - xt)p(d) + \sum_{d=t+1}^{\infty}(yt - xt)p(d)$$

Or, $E(p) = (y-z)\sum_{d=0}^{t} d\,p(d) + (zt - xt)\sum_{d=0}^{t} p(d) + (yt - xt)\sum_{d=t+1}^{\infty} p(d)$

Or, $E(p) = (y-z)\sum_{d=0}^{t} d\,p(d) + zt\sum_{d=0}^{t} p(d) + yt(1 - \sum_{d=0}^{t} p(d)) - xt$

Or, $E(p) = (y-z)\sum_{d=0}^{t} d\,p(d) - (yt - zt)\sum_{d=0}^{t} p(d) + yt - xt$

Or, $E(p) = (y-z)\sum_{d=0}^{t} (d-t)\,p(d) + (y-x)t$

Now if we produce (t-1) units then the expected profit is given by,

$$E(p-1) = (y-z)\sum_{d=0}^{t-1}(d - t + 1)\,p(d) + (y-x)(t-1)$$

Or, $E(p-1) = (y-z)\sum_{d=0}^{t-1}(d-t)\,p(d) + (y-z)\sum_{d=0}^{t-1} p(d) + (y-x)t - (y-x)$

Or, $E(p-1) = (y-z)\sum_{d=0}^{t}(d-t)\,p(d) - (y-z)(t-t)\,p(t) + (y-z)\sum_{d=0}^{t-1} p(d) + (y-x)t - (y-x)$

Or, $E(p-1) = (y-z)\sum_{d=0}^{t}(d-t)\,p(d) + (y-x)t + [(y-z)\sum_{d=0}^{t-1} p(d) - (y-x)]$

Or, $E(p-1) = E(p) + [(y-z)\Pr(D < t) - (y-x)]$

Or, $E(p) - E(p-1) = (y-x) - (y-z)\Pr(D < t)$

Hence incremental production is positive (ie t is better than t-1) if the below follows:

$(y-x) - (y-z)\Pr(D < t) > 0$

Or, $\Pr(D < t) < \dfrac{y-x}{y-z}$ [8]

ii) We can use the above solution, $\Pr(D < t) < \frac{y-x}{y-z}$

Here, $\frac{y-x}{y-z} = \frac{70-35}{70-25} = \frac{35}{45} = 0.78$

Hence the optimum quantity = 500. [1]

iii)

The expected profit can be calculated using the below equation.

$$E(p) = (y-z)\sum_{d=0}^{t}(d - t)\,p(d) + (y-x)t$$

$$= 45[(100-500)*0.1+(200-500)*0.3+(300-500)*0.2+(400-500)*0.15+(500-500)*0.05]+35*500$$

=9,175 [2]

[11 Marks]

<u>Solution 8</u>:

The likelihood function is given by L(c) = $\prod 2cx_i \exp(-cx_i^2)*(P(X > 1000000))^2$, the product is taken for i = 1,2,….,8.

P(x > 1000000) = $\int 2cx \exp(-cx^2)dx$, where the integration is taken from 1000 to α, using the distribution function in INR thousands. = $[-\exp(-cx^2)]$ over 1000 to ∞ = exp(-c 1000^2).

The log likelihood function is LogL $=\sum(\log c + \log 2 + \log x_i - cx_i^2 - 2c*1000^2)$

Differentiating with respect to *c* we obtain:

$\partial \log L/\partial c = 8/c - \sum x_i^2 - 2.1000^2$, the summation is taken from 1 to 8.

Setting this derivative equal to 0, it becomes $c = 8 / (\sum x_i^2 + 2.1000^2) = 1.71*10^{-6}$

The second order derivative is $-8/c^2$, which is less than 0. Thus this value of c is the maximum likelihood estimate of c.

Alternatively, the answer is 1.71*10^-12 if actual claim payments (rather than units of 1000) are used.

[7 Marks]

<u>Solution 9</u>:

i)

The key components of a Generalised Linear Model are given below:

1. a distribution for the data (Poisson, exponential, gamma, normal or binomial etc)
2. a linear predictor (a function of the covariates that is linear in the parameters)
3. a link function (that links the mean of the response variable to the linear predictor)

[1]

(ii)

There are two types of covariate: variables and factors.

A variable is a type of covariate whose real numerical value enters the linear predictor directly, such as age (x).

The other main type of covariate is a factor, which takes a categorical value. For example, the sex of the policyholder is either male or female, which constitutes a factor with 2 categories (or levels).

[1]

iii)

a) Using the linear predictor $\eta_i = a + bx_i$, we have $E(Yi) = \mu_i = exp(\eta_i)$. so $\eta_i = g(\mu_i) = \ln(\mu_i)$ is the natural link function. [2]

b) Assuming that $Yi \sim Exp(\lambda i)$, we have the following likelihood function:

$L = \prod f(y_i) = \prod \lambda i exp(-\lambda i y_i)$, where i is taken from 1 to 249.

$\ln(L) = \sum \ln(\lambda i) - \sum(\lambda i y_i)$, sum is taken over 1 to 249.

$= \sum \ln(1/\mu_i) - \sum(y_i/\mu_i)$, as $\mu_i = E(Yi) = 1/\lambda i$,

$= -\sum(a + bx_i) - \sum y_i \exp(-(a + bx_i))$,

Differentiating partially with respect to a and b and equating those to 0, we get

$\sum y_i \exp(-(a + bx_i)) = 249$ ….(1) &

$\sum x_i y_i \exp(-(a + bx_i)) = \sum x_i$, where the summations are taken from 1 to 249. [4]

c) An approximate 95% confidence interval for a is:

MLE of a ± 1.96 *s.e.*(MLE of a) = 8.477 ± (1.96 *1.655) = 8.477 ± 3.244 = (5.233,11.721)

Since this confidence interval does not contain zero we are 95% confident that the

parameter is non-zero and should be kept. [2]

d) Test $H0{:}b = 0$ against $H1{:}b \neq 0$. The test statistic is:

$\Delta dev = 226.282 - 219.457 = 6.826$

Comparing with chi square distribution with 1 degree of freedom, we find that the value of the test statistic exceeds the upper 1%point (6.635) of this distribution. We therefore reject the null hypothesis and conclude that Model B significantly reduces the scaled deviance (*i.e.* it is significantly better fit to the data) so survival time (in years) of the pensioners is dependent on last drawn annual salary.

[3]

[13 Marks]

****