Institute of Actuaries of India

Subject CT3 – Probability & Mathematical Statistics

May 2015 Examinations

Indicative Solutions

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 approaches leading to a valid answer and examiners have given credit for any alternative approach or interpretation which they consider to be reasonable.

Solution 1:

(i)

P [A ∪ B] = P [A] + P [B] – P [A ∩ B] … I

P [A ∪ B^C] = P [A] + P [B^C] - P [A ∩ B^C] … II

Adding I and II, we have

P [A ∪ B] + P [A ∪ B^C] = {P [A] + P [B]-P [A ∩ B]} + {P [A] + P [B^C]-P [A ∩ B^C]}

2 P [A] + {P [B] + P [B^C]} – {P [A ∩ B] + P [A ∩ B^C]}= 0.6 + 0.8

2 P [A] + 1 – P [A] =1.4

Thus P [A] = 0.4 **[3]**

(ii)

Let X be number of people with 6 fingers in the sample of 400 persons

Assuming X ~ Binomial (400, 0.02)

E (X) = 400(0.02) = 8; Var (X) = 400 (0.02) (0.98) = 7.84

Using continuity correction:

P (X ≥ 8) = P ((X - E(X)) / SD(X)) > (7.5 – 8)/2.8)

= 1 - Φ (-0.18) = Φ (0.18) = 0.57142

Alternatively, assuming X ~ Poisson (400(0.02)) = Poisson (8)

E (X) = 8; Var (X) = 8

Using continuity correction:

P (X ≥ 8) = P ((X - E(X)) / SD(X))> (7.5 – 8)/2.828)

= 1 - Φ (-0.18) = Φ (0.18) = 0.57142

[3]

(iii)

Simulation from exponential distribution:

Here λ = (1/5) = 0.2;

F(x) =1- $e^{-\lambda x}$= U => x = (-1/λ) Ln (1-U),

where U is given as 0.0923, 0.8657 and 0.3494.

x_1=-5Ln (1-0.0923) = 0.4842

x_2=-5Ln (1-0.8657) = 10.038

x_3=-5Ln (1-0.3494) = 2.1493 **[3]**

(iv)

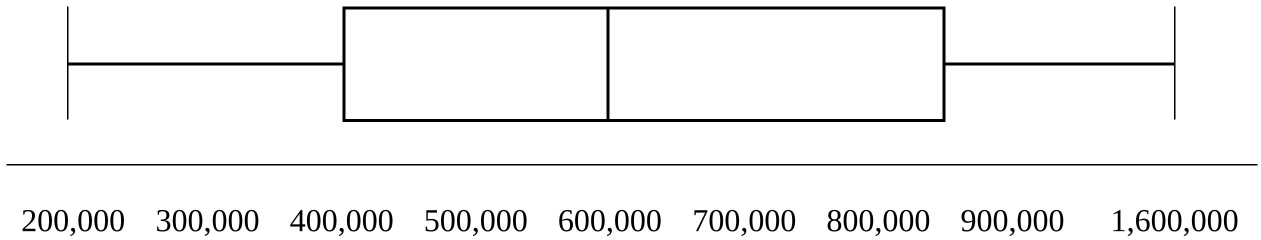
Data ordered (Rupees in Lakhs): 2,3,5,5,6,7,8,9,16

The median is the 5th observation = 6,00,000

Q1 = 2.5th observation = 4,00,000

Q3 = 7.5th observation = 8,50,000

The box plot is:



Here, Mean = 6,77,778;
Median = 6,00,000;
Mode = 5,00,000
Here mean > median > mode and hence the distribution is positively skewed.

[3]
[12 Marks]

Solution 2:

(i)

Let T denote lifetime of the mobiles

$$f(t)=\frac{1}{2}e^{-t}/2 , \quad 0\leq t\leq\infty$$

$$P\ [T\leq 1] = \int_0^1 \frac{1}{2}e^{-t}/2 = 1- e^{-1/2} = 0.393$$

$$P\ [1\leq T\leq 2] = \int_1^2 \frac{1}{2}e^{-t}/2 = e^{-1/2}-e^{-1} = 0.239$$

$$P\ [T>2] = \int_2^\infty \frac{1}{2}e^{-t}/2 = e^{-1}- 0 = 0.368$$

[3]

(ii) $X_{i:}$ Refund for the ith mobile i = 1, 2, 3…500 where $X_{i's}$ are assumed to be independent identically distributed.

X_i = 10,000 with probability 0.393
X_i = 5000 with probability 0.239
X_i = 0 with probability 0.368

$E(X_i)$ = 10,000(0.393) + 5,000(0.239) + 0(0.368) = Rs 5,125

Therefore expected refunds equal to $\sum_1^{500} E(X_i)$ = 500(5,125) = Rs 25,62,500

[3]
[6 Marks]

Solution 3:

F: Event that laptops are faulty

Bi (i =1, 2, 3 and 4): Event that the laptops come from manufacturer A if i=1; manufacturer B if i=2; manufacturer C if i=3 and manufacturer D if i=4.

We have

$$P(B_2|F) = \frac{P(F|B2)\ P(B2)}{P(F|B1)\ P(B1)+P(F|B2)\ P(B2)+P(F|B3)\ P(B3)+P(F|B4)\ P(B4)}$$

$$= \frac{10\%(30\%)}{15\%(40\%)+10\%(30\%)*+5\%(20\%)+3\%(10\%)}$$

= 0.03/0.103 = 0.29126

[4 Marks]

Solution 4:

(i)

$P_{n+1} = 0.6\, P_n$; n ≥ 0

$P_1 = 0.6\, P_0$

$P_2 = 0.6\, P_1 = 0.6^2\, P_0$

…

$P_k = 0.6^k\, P_0$; where k ≥ 1

We know that $P_0 + P_1 + P_2 + \ldots = 1$

$P_0 + \sum_1^\infty P_k = 1$

$P_0 + P_0(0.6^1 + 0.6^2 + 0.6^3 + \ldots) = 1$

$P_0 + P_0\,(0.6/(1-0.6)) = P_0 + P_0\,(3/2) = 1$

$P_0 = 2/5 = 0.4$

$P_1 = 0.6\, P_0$

$P_1 = 0.6\,(0.4) = 0.24$

$P_2 + P_3 + P_4 + \ldots\ldots = (1 - 0.4 - 0.24) = 0.36$

Number of Claims	0	1	2 and more
Probability	0.40	0.24	0.36
Claim Amount	0	5000	2500

[5]

(ii)

E(X) = 0.4(0) + 0.24(5000) + 0.36(2500) = 2100

$V(X) = E(X^2) - [E(X)]^2$

$E(X^2)$ = 0.4(0) (0) + 0.24(5000) (5000) + 0.36(2500) (2500) = 82,50,000

$V(X) = 8250000 - 2100^2 = 8250000 - 4410000 = 3840000$

$SD(X) = (3840000)^{0.5} = 1959.59$ **[3]**

[8 Marks]

Solution 5:

(i) $M_X(t) = E(e^{tx}) = \sum_{x=0}^{\infty} P(X=x)\, e^{tx}$

$= (2/3) \sum \left(\frac{e^t}{3}\right)^x$ for $e^t<3$

$= (2/3)\,(1/(1-(\frac{e^t}{3}))) = (2)\,(3-e^t)^{-1}$ for $e^t<3$ **[3]**

(ii)

$M'(t) = (2)(-1)(3-e^t)^{-2})(-e^t) = (2)e^t(3-e^t)^{-2}$

$M''(t) = (2)\{e^t(-2)(3-e^t)^{-3}(-e^t) + e^t(3-e^t)^{-2}\} = (2)\{2e^{2t}(3-e^t)^{-3} + e^t(3-e^t)^{-2}\}$

Mean = $E(X) = M'(0) = (2)e^0(3-e^0)^{-2} = (2)(1)(2)^{-2}$

$= (2/4) = 0.5$

Variance $(X) = E(X^2) - (E(X))^2 = M''(0) - (M'(0))^2$

$M''(0) = (2)\{2/8+1/4\} = (2)(4/8) = 1$

Variance $(X) = 1 - (0.5)^2$

Variance(X) = 0.75 **[3]**

[6 Marks]

Solution 6:

(i)

Probability distribution of S_N:

Here N = 0, 1, and 2

Number of Claims	Claim Amount (in Rs)
0	0
1	1000 or 2000
2	(1000,1000) or (1000,2000) or (2000,1000) or (2000,2000)

$P(S_N=0) = 0.4$

$P(S_N=1000)$ = P (N=1) (P (One claim of Rs 1000 given one claim))

= (1/2) (0.7) = 0.35

$P(S_N=2000)$ = P (N=1) (P (One claim of Rs 2000 given one claim)) + P (N=2) (P (Both claims of Rs1000 each given 2 claims))

= (1/2) (0.3) + (1/10) (0.7) (0.7 = 0.199

$P(S_N=3000)$ = P(N=2) {P (First claim of Rs 1000 and second claim of Rs 2000 given two claims) + P (First claim of Rs 2000 and second claim of Rs 1000 given two claims)}

= (1/10) [{0.7) (0.3) + (0.3) (0.7)] = 0.042

$P(S_N=4000)$ = P (N=2) (P (Both claims of Rs 2000 each given two claims))

= (1/10) (0.3) (0.3) = 0.009

It can be checked that the total probability of the compound distribution is 1.

$P(S_N=0) + P(S_N=1000) + P(S_N=2000) + P(S_N=3000) + P(S_N=4000)$

$= 0.40 + 0.35 + 0.199 + 0.042 + 0.009 = 1$ **[5]**

(ii) $E(S_N)$ using the probability distribution of S_N:

$E(S_N) = (0)P(S_N=0) + (1000)P(S_N=1000) + (2000)P(S_N=2000) + (3000)P(S_N=3000) + (4000)P(S_N=4000)$

= (0) 0.40 + (1000) 0.35 + (2000) 0.199 + (3000) 0.042 + (4000) 0.009
= Rs 910 **[2]**

(iii)
E (N) = (0) 2/5 + (1) 1/2 + (2) 1/10 = 0 + 0.5 + 0.2 = 0.7
[0.75]
E (X) = (1000) 0.7 + (2000) 0.3 = 700 + 600 = 1300
[0.75]
Hence, E (S_N) = E (N) E (X) = (0.7) (1300) = 910 [Verified (ii)].
[0.5] **[2]**

[9 Marks]

Solution 7:

(i) Sample mean: $\bar{X} = \frac{\sum Xi}{n}$

E [$\sum$ Xi] = $\sum$ E[Xi] = $\sum \mu = n\mu$; since they are identically distributed

Var [$\sum$ Xi] = $\sum$ Var[Xi] = $n\sigma^2$; since they are iid

E [$\bar{X}$] = μ.

Var [$\bar{X}$] = $\frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$ **[2]**

(ii) Sample variance : $S^2 = \frac{1}{n-1}[\sum Xi^2 - n\bar{X}^2]$

$E[S^2] = \frac{1}{n-1}(\sum E[Xi^2] - n\, E[\bar{X}^2])$

$= \frac{1}{n-1}[\sum(\sigma^2 + \mu^2) - n(\frac{\sigma^2}{n} + \mu^2)]$

$= \frac{1}{n-1}[n(\sigma^2 + \mu^2) - \sigma^2 - n\mu^2] = \frac{1}{n-1}[(n-1)\sigma^2] = \sigma^2$ **[2]**

(iii) The sampling distribution of $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ when sampling from a normal population, with mean μ and variance σ^2

The variance of χ^2_k is 2k.

Hence, Var $[\frac{(n-1)S^2}{\sigma^2}] = 2(n-1) \Rightarrow Var[S^2] = \frac{\sigma^4}{(n-1)^2} 2(n-1) = \frac{2\sigma^4}{n-1}$ **[1]**

(iv) It is given that σ^2 = 100 and n = 10.

We need to compute P ($50 < S^2 < 150$)

We know that $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ and in this case it is χ^2_9

Let Y = 9 S^2 / 100. Then, for S^2= 50: Y = 9 (50)/100 = 4.5;
and for S^2= 150: Y = 9 (150)/100 = 13.5

Thus, P (50<S^2<150) = P (4.5< Y<13.5)
= P (Y<13.5) - P (Y<4.5) [from page 165 of Tables]
= 0.8587 – 0.1245 = 0.7342 **[3]**

[8 Marks]

Solution 8:

(i)

From the Formulae and Tables for Actuarial Examinations this pdf corresponds to two parameter version of the distribution given by

$$f(x) = \frac{\alpha \lambda^\alpha}{(\lambda + x)^{\alpha+1}} \; ; x > 0 , \lambda > 0 \text{ and } \alpha > 0$$

E[X] = λ / (α – 1) and Var [X] = α λ^2 / ((α – 1)^2 (α – 2)); α > 2

Thus the pdf of X, $f(x) = \frac{C \beta^3}{(x + \beta)^4}$; x > 0 and β >0 can be identified with c=3 (α known) λ= β (unknown).

E[X] = β/2

[0.5]

Var [X] = 3/4 β^2

[0.5]

If the students obtain the results using the first principles full credit to be given

$$f(x) = \frac{\alpha \lambda^\alpha}{(\lambda + x)^{\alpha+1}} \; ; x > 0 , \lambda > 0 \text{ and } \alpha > 0$$

$\int_0^\infty \frac{C \beta^3}{(x + \beta)^4}$ dx =1implies $\left[- \frac{C}{3} \frac{\beta^3}{(x + \beta)^3}\right]_0^\infty$=1 giving c = 3 **[3]**

Mean and Variance

(ii)

The method of moments estimator:

Equating sample mean to E[X] from (i) gives:

$\bar{X}_n$ = β / 2 => The moments estimator $\hat{\beta}$ = 2 $\bar{X}_n$

The mean square error of $\hat{\beta}$:

MSE [$\hat{\beta}$] = Var [$\hat{\beta}$] + (Bias [$\hat{\beta}$])^2

E [$\hat{\beta}$] = E [2 $\bar{X}_n$] = 2 E [$\bar{X}_n$] = 2 E[X] = 2 (β / 2) = β and Bias = E [$\hat{\beta}$] – β =0

–This estimator is unbiased.

Using the fact that the individual values of *Xi* are independent:
Var $[\hat{\beta}]$ = Var $[2\,\bar{X}_n]$ = 4 Var $[\bar{X}_n]$ = 4 Var[X] / n = 4 (3/4 β^2) / n= 3 β^2 / n

Hence, MSE $[\hat{\beta}]$ = 3 β^2/ n + 0 = 3 β^2/ n
This estimator is consistent since MSE tends to zero as n → ∞ **[4]**

(iii)
MSE $[b\,\bar{X}_n]$ = Var $[b\,\bar{X}_n]$ + (Bias $[b\,\bar{X}_n]$)^2

$$= b^2 \frac{3}{4}\frac{\beta^2}{n} + (E[b\,\bar{X}_n] - \beta)^2$$

$$= b^2 \frac{3}{4}\frac{\beta^2}{n} + \left(\beta\left(\frac{b}{2} - 1\right)\right)^2$$

$$= \frac{\beta^2}{n}\left(\frac{3}{4}b^2 + n\left(\frac{b}{2} - 1\right)^2\right)$$

Differentiating the MSE respect to b:
d (MSE)/db = $\frac{\beta^2}{n}\left(\frac{3}{2}b + n\left(\frac{b}{2}-1\right)\right)$
Setting this equal to 0 gives b = 2n / (n + 3) = 2 / (1 + $\frac{3}{n}$)

Differentiating the MSE a second time with respect to b, we obtain:
d^2(MSE)/db^2 = $\frac{\beta^2}{n}\left(\frac{3}{2} + \frac{n}{2}\right)$ which is positive
=> a minimum value for the MSE is at b = 2 / (1 + 3/n). **[5]**

(iv)
It is seen in (ii) that $\hat{\beta} = 2\,\bar{X}_n$ is an unbiased estimator. The estimator $b\bar{X}_n$ in (iii) when b = 2 / (1 + $\frac{3}{n}$) is negatively biased estimator as b < 2 for n ≥ 1.

As n → ∞, b tends to 2. So, the estimator in (iii) is also consistent, in view of (ii) **[2]**
[14 Marks]

Solution 9:

(i) The likelihood function is:

$$L(p) = C\,[(1-p)^4]^{86}\,[4p(1-p)^3]^{75}\,[6p^2(1-p)^2]^{16}\,[4p^3\,(1-p)]^2\,[p^4]^1$$

$$= C\,[(1-p)^{344}]\,[p^{75}(1-p)^{225}]\,[p^{32}(1-p)^{32}]\,[p^6(1-p)^2]\,[p^4]$$

$$= C\,(1-p)^{603}\,p^{117}$$

C is a constant.

Taking logs and differentiating with respect to p and setting equal to zero gives:

dLn L/dp = - 603/(1-p) + 117/p = 0
=> p = 117/ (603+117) = 117/720 = 0.1625
Checking for maximum:
d^2Ln L/dp^2 = - 603/ $(1-p)^2$ - 117/p^2 < 0 **[4]**
=> Maximum value for Ln L is at p = 0.1625

(ii)

Goodness of fit test:

We are testing the following hypotheses using a χ2 goodness of fit test:
H_0: the probabilities conform to a Bin (4, p) distribution
H_1: the probabilities do not conform to a Bin (4, p) distribution

Using $\hat{p}$ = 0.1625 from part (a), the probabilities for this binomial distribution are:

P (X = 0) = $(1-p)^4$ = 0.49197
P (X = 1) = 4p $(1-p)^3$ = 0.38183
P (X = 2) = 6p^2 $(1-p)^2$ = 0.11113
P (X = 3) = 4p^3 (1-p) = 0.01437
P (X = 4) = p^4 = 0.00070

The expected values are 88.55, 68.73, 20.00, 2.59 and 0.13.
Combining the expected values less than 5 to third group, we get

No of Claims	0	1	2 or more
Observed O_i	86	75	19
Expected E_i	88.55	68.73	22.72

The degrees of freedom = 3-1-1=1.

$\chi 2 = \sum (O_i - E_i)^2 / E_i$

= $(86-88.55)^2$ / 88.55 + $(75-68.73)^2$ / 68.73 + $(19-22.72)^2$ / 22.72
= 0.074 + 0.572 + 0.608 = 1.254

This is less than the 5% critical value of 3.841.We have insufficient evidence at 5% level to reject Ho. Hence, the model is a good fit. **[4]**

[8 Marks]

Solution 10:

(i)

Probability of a Type I error is probability of rejecting the null hypothesis when it is true.
= P (X > 3 |$\mu = 1$)
= 1 - P (X ≤ 3 |$\mu = 1$)
= 1 – 0.98101 *(from tables page 175)* = 0.01899 => approx 2% **[2]**

(ii)

The power of a test is 1 minus the probability of a Type II error, i.e. the probability of rejecting the null hypothesis when it is false.

= P (X > 3 $|\mu = 2$)

= 1 - P (X $\leq$ 3 $|\mu = 2$)

= 1 - 0.85712 *(from tables page 175)* = 0.14288 => approx 14% **[2]**

[4 Marks]

Solution 11:

(i)

We know that $\hat{\beta} = \frac{S_{xy}}{S_{xx}}$ = 9818.500 / 10,804.875 = 0.9087

And $\hat{\alpha} = \bar{Y} - \hat{\beta}\ \bar{X}$ = 126.5 – 0.9087 (75.125) = 58.2331

Hence the prediction equation is ŷ = 58.2331 + 0.9087x **[2]**

(ii)

Total sum of squares = $\sum(Y - \bar{Y})^2 = S_{yy}$ = 11,250.000

Residual sum of squares = $\sum(y - \hat{y})^2 = (S_{yy} - \frac{(S_{xy})^2}{S_{xx}})$

= (11,250.000 – $(9818.500)^2$/10,804.875) = 2,327.829

Reg. sum of squares = Total s.s– Residual s.s = 11,250 - 2,327.829 = 8,922.171

Source of variation	Degrees of Freedom	Sum of squares	Mean squares (MSS)
Regression	1	8,922.171	8,922.1710
Residuals	6	2,327.829	387.9715
Total	7	11,250.000	

The variance ratio is F = 8,922.1710 / 387.9715 = 22.997
For testing Ho: β = 0 at 5% level, we have the critical value F (1, 6) at 5.987.
Hence, we reject Ho. **[4]**

(iii)

95% confidence limits for β

$$\hat{\beta} \pm t_{0.025,6}\sqrt{\frac{MSS\ Residual}{S_{xx}}}$$

$$= 0.9087 \pm 2.447\sqrt{\frac{387.9715}{10,804.875}}$$

$$= 0.9087 \pm 0.4637 \Rightarrow \beta \in [0.445, 1.372]$$ **[2]**

(iv)

H_0: ρ = 0 Vs H_1 : ρ ≠ 0

Sample correlation coefficient, $r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{9818.500}{\sqrt{(10804.875)\ (11250.000)}} = 0.8906$

$$t_{n-2} = \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}} \Rightarrow t_6 = \frac{|0.8906|\sqrt{6}}{\sqrt{1-0.8906^2}} = 4.7955$$

Table value of t_6 at 1% level is 3.707. Hence, reject Ho. **[3]**

(v)

Coefficient of Determination

R^2 is estimated by sample correlation coefficient,

$r^2 = 0.8906^2 = 79.31\%$

Coefficient of determination (R^2) is used to measure the goodness of fit of a linear regression model. It is, usually quoted as a percentage, the proportion of total variability of the responses “explained” by the model. **[2]**

[13 Marks]

Solution 12:

(i)

Ho: Each agent has the same mean number of policies per month

H_1: There are differences among the mean numbers of policies generated by different agents.

The given summary measures are: $y_{1.}$ = 38; $y_{2.}$ = 27; $y_{3.}$ = 31; $y_{4.}$ = 21; $y_{..}$ = 117

$\sum\sum y_{ij}^2$= 677

SS T = 677 – 117^2/26 = 150.50

SS B = (38^2/8 + 27^2/7 + 31^2/6 + 21^2/5) – 117^2/26 = 6.51

SS R = 150.50 – 6.51 = 143.99

Source of variation	Degrees of Freedom	Sum of squares	Mean squares
Between treatments	3	6.51	2.170

Residuals	22	143.99	6.545
Total	25	150.50	

The variance ratio is F = 2.170 / 6.545 = 0.3315

Under Ho, this has an F (3, 22) distribution. The 5% critical point is 3.049, so we cannot reject Ho, and we conclude that the average number of policies generated doesn't differ between agents.

The assumptions are: The underlying population distribution is normal with common variance. It is also assumed that the samples have been drawn randomly and independently of each other. **[5]**

(ii) Sample sizes: $n_1 = 8$; $n_2 = 7$; $n_3 = 6$; $n_4 = 5$

Sample totals: $y_{1.} = 38$; $y_{2.} = 27$; $y_{3.} = 31$; $y_{4.} = 21$

Sample means: $\bar{y}_{1.} = 4.75$, $\bar{y}_{2.} = 3.86$, $\bar{y}_{3.} = 5.17$, $\bar{y}_{4.} = 4.20$

For agents with highest and lowest mean numbers, least significant difference at 5% level is

$$t_{(0.025, n-k)}\ \hat{\sigma}\ \left(\frac{1}{n_3} + \frac{1}{n_2}\right)^{0.5}$$

$$= (2.074)\ (6.545)^{0.5}\ (1/6 + 1/7)^{0.5} = 2.952$$

The least significant difference of 2.952 is greater than $\bar{y}_{3.} - \bar{y}_{2.} = 1.310$.
Hence, the difference is not significant. **[3]**

[8 Marks]

****