Institute of Actuaries of India

Subject CT3 – Probability & Mathematical Statistics

September 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 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) Let S be the salary (constant) of each of the 99 employees.

Hence, average salary of 99 employees = (S+S+…+S)/99 = S
The variance of the salary of 99 employees was 0 (as they are getting the constant salary)

[1]

(ii) S+1000 is the salary of the 100th employee.

So, the average salary of 100 employees = [(S+S+…+S) +(S+1000)] /100
= (100S+1000)/100 = S+10

With the addition of the 100th employee, the average salary of 100 employees has increased by Rs 10.

Now, the variance of the salary of 100 employees

$$\frac{1}{100}[\{(-10)^2+(-10)^2+\ldots+(-10)^2\}+990^2]$$

$$= \frac{1}{100}[9900+980100] = \frac{1}{100}(990000) = 9900$$

The standard deviation = $\sqrt{9900}$ = Rs 99.50, which is positive

[3]
[4 Marks]

Solution 2:

Given that

Number	3	5
Frequency	f_1	f_2

Mean : $\frac{3f_1+5f_2}{(f_1+f_2)}$

Variance : $\frac{1}{(f_1+f_2)}[(9f_1+25f_2)-(f_1+f_2)\left(\frac{3f_1+5f_2}{(f_1+f_2)}\right)^2]$

$$= \frac{(9f_1{}^2+9f_1f_2+25f_1f_2+25f_2{}^2)-(9f_1{}^2+30f_1f_2+25f_2{}^2)}{(f_1+f_2)^2}$$

$$= \frac{4\,f_1f2}{(f_1+f_2)^2}$$

[3 Marks]

Solution 3:

(i) Nine people can sit in a row with 9! ways.

The three persons A, B and C sit together in a particular order (ABC) say and remaining 6 people can sit in any order.

Hence, there are 7! ways of sitting: namely, (ABC) sit together in a particular order with remaining 6 persons in any order.

Thus, probability of A, B and C sit together in a particular order is 7! /9! = 1/72. **[2]**

(ii) For A, B, C sitting together in any order implies that A, B, C can sit in 3! Ways.

Three persons A,B,C sit together in the any order {(A,B,C),(A,C,B), (B,A,C), (B,C,A), (C,A,B),(C,B,A) } say and remaining 6 person can sit in any order.

Hence, there are 3! (7!) ways of sitting (A, B, C) sit together in any order along with other 6 persons.

Thus, probability of A, B and C sit together in any order is 3!(7!) /9! =1/12. **[2]**

(iii) If A and C occupy the end seats then remaining people can be seated in 7! ways. However A and C can be seated first and last and vice versa.

Hence, probability A and C occupy the end seats is 2 (7!)/9! = 1/36. **[2]**

(iv) If B occupies the middle seat, then remaining 8 people can sit in 8! different ways.

Hence, probability B always occupies the middle seat in the row is 8!/9! = 1/9 **[2]**

[8 Marks]

Solution 4:

(i) Let *T*: Accident with tyre burst

C: Accident due to collision with the road divider and

D: Death casualty in a car accident

Given that: *P(T)* = 0.6, *P (C)* =0.4, *P(D|T)* = 0.3 and *P(D|C)* =0.5.

Given accidental death casualty, the probability of tyre burst is:

$$P(T|D) = \frac{P(T)P(D\,|T\,)}{P(T)\,P(D\mid T) + P(C)\,P(D\,|C\,)}$$

$$= \frac{0.6\,(0.3)}{(0.6\,(0.3) + 0.4\,(0.5))} = \frac{18}{38} = 0.4737.$$

[4]

(ii) To find the most probable cause of death casualty due to accidents, we need to compare given accidental death casualty, the probability of tyre burst and probability of collision with the road divisor; i.e *P(T| D)* and *P(C|D)*

Now, $P(C|\,D) = \frac{P(C)P(D\mid C\,)}{P(T)\,P(D\mid T) + P(C)\,P(D\mid C\,)}$

$$= \frac{0.4\,(0.5)}{(0.6\,(0.3)+ 0.4\,(0.5))} = \frac{20}{38} = 0.5263.$$

[Alternately, $P(C/D) = 1- P(T/D) = \frac{20}{38}$]

Since, *P(T|D) < P(C|D),* collision with the road divider is the most probable cause of accidental death casualty.

[2]
[6 Marks]

Solution 5:

(i) P(obtaining correct password in the third try)
= P(obtaining incorrect password in the first two attempts and obtaining correct password in the third attempt)
= (1- 1/100) (1- 1/99) (1/98)
=1/100. **[2]**

(ii) File can be accessed if the password is correct in the first attempt or second attempt or third attempt.

That is: P (Password is correct in the first attempt) + P(password is incorrect in the first attempt and correct in the second attempt) + P(password is incorrect in the first two attempts and correct in the third attempt)

= 1/100 + (1- 1/100)(1/99) + (1 - 1/100) (1 - 1/99) (1/98)
=3/100. **[2]**

(iii) P (Correct password is found on the 10th try)
= [P (Incorrect password in the first 9 attempts)][P(correct password on the 10th attempt)]
[1]
= [$(1 - 1/100)^9$][1/100] =0.009135. **[2]**

[6 Marks]

Solution 6:

(i) The Joint density of *(X,Y):*
We know that *f(y|x)* = $\frac{f(x,\ y)}{f(x)}$.
Hence, *f(x,y)= f(y|x)f(x)*

$$=\frac{1}{x}\,(8x)$$

Thus, *f(x,y)* = $\begin{cases} 8 & if\ 0 < x < \frac{1}{2}\ and\ 0 < y < x \\ 0 & Otherwise. \end{cases}$ **[1]**

(ii) The marginal density of *Y:*

f(y)=$\int f(x,y)dx = \int_y^{0.5} 8\,dx = 4(1-2y)$ *for* $0<y<\frac{1}{2}$.

[2]

(iii) The Mean and Variance of *X* and *Y*:

$E(Y)=\int_0^{0.5}(y)\,4\,(1-2y)\,dy = \frac{1}{2}-\frac{1}{3} = \frac{1}{6}$.

$V(Y) = E(Y^2)-(E(Y))^2$.

$E(Y^2)=\int_0^{0.5} y^2\,4(1-2y)\,dy = \frac{1}{6} - \frac{1}{8} = \frac{1}{24}$.

$V(Y)=\frac{1}{24} - \left(\frac{1}{6}\right)^2 = \frac{1}{72}$. **[4]**

[7 Marks]

Alternate Solution

[Full credit to be given if the candidate provides alternate solution]

E(Y) = E(E(Y|X))
E(Y|X=x) = $\int_0^x y\,\frac{1}{x}\,dy = \frac{x}{2}$
Hence *E(Y|X)* = $\frac{X}{2}$
E(Y) = E(E(Y|X)) = $\int_0^{0.5} \frac{x}{2}\,(8x)dx = \frac{1}{6}$

V(Y) =E (Var(Y|X)) +Var (E(Y|X))

Now *Var(Y|X)* = $E(Y^2 \mid X) - (E(Y|X))^2$

$E(Y^2|X=x) = \int_0^x y^2 \frac{1}{x} dy = \frac{x^2}{3}$
Hence, *Var(Y|X)* = $\frac{x^2}{3} - (\frac{x}{2})^2 = \frac{x^2}{12}$
E (Var(Y|X)) = $E(\frac{X^2}{12}) = \int_0^{0.5} \frac{x^2}{12}\,(8x)dx = \frac{1}{96}$ (I)

Now *Var (E(Y|X))* = $Var\left(\frac{X}{2}\right) = \frac{(E(X^2)-(E(X))^2)}{4}$
And $E(X^2) = \int_0^{0.5} x^2\,(8x)dx = \frac{1}{8}$
And $E(X) = \int_0^{0.5} x\,(8x)dx = \frac{1}{3}$
Hence *Var (E(Y|X))* = $Var\left(\frac{X}{2}\right) = \frac{(E(X^2)-E(X)^2)}{4} = \frac{1}{4}\left(\frac{1}{8} - \frac{1}{9}\right) = \frac{1}{4\,(8)(9)}$ (II)

Adding (I) and (II) we get

V(Y) =E (Var(Y|X)) +Var (E(Y|X)) $= \frac{1}{72}$.

[7 Marks]

Solution 7:

$$P[(\bar{X}-1) > S] = P\left[\frac{(\bar{X}-1)}{S} > 1\right] = P\left[\frac{4(\bar{X}-1)}{S} > 4\right] = P\left[\frac{(\bar{X}-1)}{S/\sqrt{16}} > 4\right] = P[t_{15} > 4]$$

From actuarial table page 163 the probability is between 0.001 and 0.0005.

[3 Marks]

Solution 8:

(i) Mean of sample mean cost is $E(\bar{X}) = \mu$ that is 20,000. **[1]**

(ii) Standard error of sample mean: SE $(\bar{X}) = \frac{\sigma}{\sqrt{n}} = \frac{3000}{\sqrt{100}} = \frac{3000}{10} = 300$ **[1]**

(iii) $P(\bar{X} < 20{,}300) = \left(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{20{,}300 - 20{,}000}{300}\right)$

$= P(Z<1) = 0.84134.$ **[2]**

[4 Marks]

Solution 9:

It is given that the population random variable X *as* pdf : $f(x) = \theta e^{-\theta x}\,; x, \theta > 0.$

$H_0: \theta = 1$ and $H_1: \theta = \frac{1}{2}$; Sample size 1, Critical region $> -log\alpha$, $0 < \alpha < 1$

(i) $P(Type\ I\ error) = P_{H_0}(x > -\log\alpha) = \int_{-\log\alpha}^{\infty} e^{-x}\,dx = e^{-(-Log\,\alpha)} = \alpha$

[2]

(ii) $Power\ of\ the\ test = P_{H_1}(x > -\log\alpha) = \int_{-\log\alpha}^{\infty} \frac{1}{2} e^{-\frac{x}{2}}\,dx = \int_{\frac{-\log\alpha}{2}}^{\infty} e^{-y} dy$

$= e^{\frac{\log\alpha}{2}} = \sqrt{\alpha}$ **[3]**

[5 Marks]

Solution 10:

(i)

- An estimator is a rule, as a function of the random sample, often expressed as a formula that tells how to calculate the value of an estimate.

- If we have a random sample $\underline{X} = (X_1, X_2, \ldots, X_n)$ from a distribution with an unknown parameter ϑ and $g(\underline{X})$ is an estimator of ϑ, it seems desirable that $E[g(\underline{X})] = \vartheta$.

This is the property of unbiasedness.

- The MSE of an estimator $g(\underline{X})$ for θ is defined by: $MSE[g(\underline{X})] = E\left((g(\underline{X}) - \theta)^2\right)$
  Alternative expression, $MSE = Variance + (bias)^2$

- It is also desirable that an estimator gets better as the sample size increases i.e. it is desirable that *MSE* → *0* as n → ∞. This property is known as consistency.

[4]

(ii)

- Method of moments
- Method of maximum likelihood
- Method of Percentiles **[1]**

(iii)

Steps for finding the maximum likelihood estimator in two parameter cases are:

- Write down the likelihood function, *L*
- Find log *L* and simplify the resulting expression
- Partially differentiate log L with respect to each parameter to be estimated
- Set the derivatives equal to zero
- Solve these equations simultaneously.
- Check the condition that the Hessian matrix i.e. the matrix of second derivatives, is negative definite. **[2]**

(iv)

The CRLB provides a lower bound for the variance of an unbiased estimator as a function of the true parameter value. It can be used to compare the efficiency of different estimators.

It also provides an approximate value for the variance of the MLE of a parameter when the sample size is large. Hence, it may be used to obtain approximate confidence intervals. **[2]**

(v)

Let *X* denote the gross claim amount following Pareto (α, 60,000) distribution

with pdf $f(x_i) = \frac{\alpha(60{,}000)^{\alpha}}{(60{,}000 + x_i)^{(\alpha+1)}}$; $\mathrm{x_i} > 0$

Of the 10 claims, 2 claims are right censored.

The likelihood function is: $L(\alpha) = \prod_{i=1}^{8} f(x_i)\,[P(X > 30{,}000)]^2$

Using the distribution function of the Pareto distribution:

$$P\ (X > 30{,}000) = 1 - F(30{,}000) = \left(\frac{60{,}000}{60{,}000+30{,}000}\right)^{\alpha} = \left(\frac{2}{3}\right)^{\alpha}$$

Hence: $L\ (\alpha) = \prod_{i=1}^{8} \frac{\alpha(60{,}000)^{\alpha}}{(60{,}000+\ x_i)^{(\alpha+1)}} \left(\frac{2}{3}\right)^{2\alpha}$

Taking log gives:

$$logL = \sum_{i=1}^{8}[log\,\alpha + \alpha\, log\, 60{,}000 - (\alpha+1)\, log(60{,}000 +\ x_i)] + 2\alpha log\left(\frac{2}{3}\right)$$

$$8\, log\,\alpha + 8\alpha\, log\, 60{,}000 - (\alpha+1)\sum_{i=1}^{8} log(60{,}000 +\ x_i) + 2\alpha log\left(\frac{2}{3}\right)$$

Now differentiate with respect to α and

Setting this expression equal to 0 we find that:

$$\frac{d\, log\, L}{d\,\alpha} = \frac{8}{\alpha} + 8\, log\, 60{,}000 - \sum_{i=1}^{8} log(60{,}000 +\ x_i) + 2log\left(\frac{2}{3}\right)$$

$$\hat{\alpha} = \frac{8}{\sum_{i=1}^{8}\log(60{,}000+\mathrm{x_i}) - 8\log 60{,}000 - 2\log\left(\frac{2}{3}\right)}$$

Note also that:

$\frac{d^2\, log\, L}{d\,\alpha^2} = -\frac{8}{\alpha^2} < 0$So the turning point is a maximum.

Using the data given in the question we find that the ML estimate of α is:

$$\hat{\alpha} = \frac{8}{90.32117 - 88.01680 + 0.81093} = \frac{8}{3.115302} = 2.567969$$

[7]
[16 Marks]

Solution 11:

(i) Frequency of rare events like earthquakes can be modeled as Poisson distribution.

X ~ Poi (μ).

We require:

$P(X \geq 1) = 0.025\ underpoi\ (\mu_1)$

$P\ (X \leq 1) = 0.025\ underpoi\ (\mu_2)$

From the first equation:

$0.975 = P(X = 0) = e^{-\mu_1}$

Solving this we obtain μ_1 = 0.0253

From the second equation:
$0.025 = P(X = 0) + P(X = 1) = e^{-\mu_2} + \mu_2 e^{-\mu_2}$

We can solve this numerically or using interpolation.

From Tables page 177:
$P\ (X \leq 1) = 0.02656$ for μ = 5.50 and;
$P\ (X \leq 1) = 0.02441$ for μ = 5.60

By interpolation, we obtain μ_2 = 5.5725

So the confidence interval is 0.0253 < μ<5.5725 **[4]**

(ii) Since n is large enough to use a normal approximation for Poisson distribution, the pivotal quantity is:

$\frac{\sum X - n\mu}{\sqrt{n\hat{\mu}}} \sim N(0,1)$ or $\frac{\hat{\mu}-\mu}{\sqrt{\frac{\hat{\mu}}{n}}} \sim N\ (0,1)$; where $\hat{\mu} = \bar{X}$.

Hence, a 95% confidence interval can be obtained for μ from:

$\frac{\sum X \pm 1.96\sqrt{n\hat{\mu}}}{n}$ or $\hat{\mu} \pm 1.96\sqrt{\frac{\hat{\mu}}{n}}$

Substituting n = 36; $\sum X_i = 36$and $\hat{\mu} = \bar{X} = 1$gives:

$\frac{36 \pm 1.96\sqrt{36\ (1)}}{36}$ or $1 \pm 1.96\sqrt{\frac{1}{36}}$

The confidence interval μ is (0.6733, 1.3267).

[2]

(iii) The large sample gives a much narrower confidence interval.

With a large sample we can predict the value of μ (average earthquakes p.a.) with greater certainty.

[2]

[8 Marks]

Solution 12:

We are interested in testing:

H_0:Exponential is a suitable distribution for salaries; against
H_1:Exponential is not a suitable distribution for salaries

We first need to estimate the value of λ using the method of moments. The mean of the salaries (Exponential) distribution is $\frac{1}{\lambda}$.

Sample mean = $\frac{158\ (1.00)+\ 24\ (3.50)+\ 12\ (7.50)+\ 6\ (55.00)}{200} = 3.31$

Setting mean of the salaries distribution equal to the sample mean gives a value of 0.3021148$\left(\text{i. e.} \frac{1}{3.31}\right)$for λ.

The probability, P that an exponential random variable lies between a and b is:
$P = F(b) - F(a) = \ e^{-\lambda a} - e^{-\lambda b}$

Expected number of Employees (E) is the probability multiplied by total employees (200)

Salary Bands	X < 2	2 ≤ X < 5	5 ≤ X < 10	X≥ 10
Observed Employees (O)	158	24	12	6
Probabilities (P)	0.4535	0.3257	0.1720	0.0487
Expected Employees (E)	90.70	65.14	34.41	9.75

$$\sum \frac{(O-E)^2}{E} = \frac{(158-90.70)^2}{90.70} + \frac{(24-65.14)^2}{65.14} + \frac{(12-34.41)^2}{34.41} + \frac{(6-9.75)^2}{9.75} = 91.95$$

The underlying distribution is χ^2 with 4 – 1 – 1 = 2 degrees of freedom (since we have set the total and estimated the mean from the data).

The critical value of the χ^2_2 distribution is 9.210, so we have evidence to reject H_0 at the 1% level and conclude that the exponential is not an appropriate distribution for salaries.

[6 Marks]

Solution 13:

(i)

We know that $\hat{\beta} = \frac{S_{XY}}{S_{XX}} = \frac{28,500,000}{5,680,000} = 5.0176$

And $\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} = 13,050 - 5.0176(1,400) = 6,025.35$

Hence the prediction equation is $\hat{y} = 6,025.35 + 5.0176\,x$

[2]

(ii)

99% confidence interval for β, the slope parameter, is: $\hat{\beta} \pm t_{0.005,8}\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}$

$$\hat{\sigma}^2 = \frac{1}{n-2}\left\{S_{YY} - \frac{(S_{XY})^2}{S_{XX}}\right\} = \frac{1}{10-2}\left\{156,225,000 - \frac{(28,500,000)^2}{5,680,000}\right\} = 1,652,904.93$$

99% confidence interval for β is: $\hat{\beta} \pm 3.355\sqrt{\frac{1,652,904.93}{5,680,000}}$

= (3.2078, 6.8275) **[3]**

(iii)

Ho: β = 0; Vs. H_1: β ≠ 0

Because 0, the hypothesized value of β is not included in the confidence interval calculated in above part, we can reject Ho and conclude that significant statistical relationship exists between the student size and daily revenue.

[2]

(iv)

For student population of 1,000; estimate of mean daily revenue is given by:

$\hat{\mu} = \hat{\alpha} + \hat{\beta}X_i = 6,025.35 + 5.0176(1000) = 11,042.96$

Standard error of the estimate Se $(\hat{\mu})$ = $\sqrt{\left\{\frac{1}{n} + \frac{(X_i - \bar{X})^2}{S_{XX}}\right\}\hat{\sigma}^2}$

Se $(\hat{\mu})$ = $\sqrt{\left\{\frac{1}{10} + \frac{(1,000-1,400)^2}{5,680,000}\right\}1,652,904.93}$ = 460.27

95% confidence interval for $\widehat{\mu}$ is $\hat{\mu} \pm t_{0.025,n-2}Se\,(\hat{\mu})$

Substituting values of $\hat{\mu} = 11,042.96$; $t_{0.025,n-2} = 2.306$ and Se $(\hat{\mu})$ = 460.27

95% confidence limits for mean daily revenue in Rs = (9,981.57, 12,104.35)

[3]

(v)

The standard error of mean daily revenue estimate,

se of $(\hat{\mu})$ = $\sqrt{\left\{\frac{1}{n} + \frac{(X_i - \bar{X})^2}{S_{XX}}\right\}\hat{\sigma}^2}$ will be minimum at $X_i = \bar{X}$ since X_i term is squared and minimum can be zero at $X_i = \bar{X}$

For any confidence interval, the shortest length is attained at minimum se $(\hat{\hat{\mu}})$. Hence for a college with 1,400 students, the 95% confidence interval for the mean daily revenue will be the shortest.

[2]

(vi)

Standard error of the estimate Se $(\hat{y})$ = $\sqrt{\left\{1 + \frac{1}{n} + \frac{(X_i - \bar{X})^2}{S_{XX}}\right\}\hat{\sigma}^2}$

Se $(\hat{y})$ = $\sqrt{\left\{1 + \frac{1}{10} + \frac{(1{,}000 - 1{,}400)^2}{5{,}680{,}000}\right\} 1{,}652{,}904.93}$ = 1,365.56

95% confidence interval for $\hat{y}$ is $\hat{\mu} \pm t_{0.025,n-2}\ Se\ (\hat{y})$

Substituting values of $\hat{\mu}$=$11{,}042.96$; $t_{0.025,n-2} = 2.306$ and Se $(\hat{y})$ = 1,365.56

95% CI for individual daily revenue in Rs = (7,893.97, 14,191.94)

[2]

(vii)

CI for the daily revenue of a particular college with 1,000 students is wider than CI for mean daily revenue with 1,000 students.

The difference reflects the fact that we are able to estimate mean value of y more precisely than individual value of y.

The resulting interval for an individual response is wider than the corresponding interval for the mean response because the uncertainty associated with individual estimator is more than the relatively more stable mean response.

[1]

[15 Marks]

Solution 14:

(i)

Least squares estimates:

$\hat{\mu} = \overline{Y_{..}} = \frac{1}{n}\sum_i \sum_j Y_{ij}$

⇨ $\hat{\mu} = \frac{1}{16}(70) = 4.375$

$\hat{\tau_i} = \overline{Y_{i.}} - \overline{Y_{..}} = \frac{1}{n_i}\sum_j Y_{ij} - \frac{1}{n}\sum_i \sum_j Y_{ij}$

⇨ $\hat{\tau_1} = \frac{1}{5}(22) - 4.375 = 0.025$

⇨ $\hat{\tau_2} = \frac{1}{4}(18) - 4.375 = 0.125$

⇨ $\hat{\tau_3} = \frac{1}{7}(30) - 4.375 = -0.089$

[4]

(ii)

For the given data, summary measures are:

$y_{1.} = 22$; $y_{2.} = 18$; $y_{3.} = 30$; $y_{..} = 70$; $\sum\sum y_{ij}^2 = 324$

SS T = $324 - 70^2/16 = 324 - 306.25 = 17.75$
SS B = $(22^2/5 + 18^2/4 + 30^2/7) - 70^2/16 = 0.12$
SS R = $17.75 - 0.12 = 17.63$

$\hat{\sigma}^2 = \frac{SS\ R}{n-k} = \frac{17.63}{16-3} = 1.356$

Considering highest and lowest sample means:
$\bar{y}_{2.} = 4.500$, $\bar{y}_{3.} = 4.286$

For the second and third companies, least significant difference at 5% level is

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

$$= (2.160)\ (1.356)^{0.5}\left(\frac{1}{4} + \frac{1}{7}\right)^{0.5} = 1.58$$

The least significant difference of 1.58 is more than the difference between the average salaries of Health and Life insurance companies: 0.214 (4.500 – 4.286).

Hence, based on the analysis of the data a fresher is
not able to distinguish the three companies for applying for a job. **[5]**

[9 Marks]

****