Actuarial Society of India

EXAMINATIONS

June 2005

CT3 – probability and Mathematical Statistics

Indicative Solution

Q.1 **a)** Stem and leaf display of the given data

i)

4	8							
5	0	1	7					
6	1	3	6	6	7			
7	1	2	2	3	6	8	8	9
8	2	4	4	5	7	9		
9	3	4	9					
10	0							

[2]

ii) Range = 100 – 48 = 52 **[1]**

Interquartile range = 85 – 66 = 19 **[1]**

(Quartiles by alternative approach accepted)

b) $\bar{x} = \dfrac{1104}{20} = 55.2$ **[1]**

$s^2 = \dfrac{\sum x^2 - n\bar{x}^2}{n-1}$

$= \dfrac{61{,}226 - 20(55.2)^2}{19}$

$= 285.2$

$s = 3.874$ **[1]**

Interval $\bar{x} \pm s \sim (51.326, 59.074)$

All the items except the following 7 fall within the interval –

49 50 51 49 61 60 62

required percentage = $\dfrac{13}{20} \times 100 = 65\%$ **[1]**

Total [7]

Q.2 **a)** Statement **[1]**

Proof **[2]**

b) **i)** Since P(X = 2, Y = 3) = 0 $\neq P(X = 2)P(Y = 3)$ X and Y are dependent. **[2]**

ii) Adjust the cell probabilities so that

$P[X = x, Y = y] = P(X = x)P(Y = y)$ **[1]**

This leads to the following table

V\ U	1	2	3	
2	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{3}$
3	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{3}$
4	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{3}$
	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

[2]

Total [8]

Q.3 **a)** Given P(A) = 0.2, P(B) = 0.25, P(C) = 0.4

$P(\bar{A}) = 0.8$, $P(\bar{B}) = 0.75$, $P(\bar{C}) = 0.6$ **[1]**

$$\text{Expected no} = x_1P(\bar{A}) + x_2(\bar{B}) + x_3(\bar{C})$$
$$= 10(0.8) + 16(0.75) + 20(0.6)$$
$$= 8 + 12 + 12$$
$$= 32$$ **[2]**

b) i)

$$M_2(t) = \int_0^1 ze^{tz}dz + \int_1^2 e^{tz}(2-z)\,dz$$ **[1]**

Integrating by parts

$$\int_0^1 ze^{tz}\,dz = \left[\frac{ze^{tz}}{t}\right]_0^1 - \int_0^1 \frac{e^{tz}}{t}dz$$

$$= \frac{e^{tz}}{t} - \frac{1}{t^2}\left(e^t - 1\right)$$ **[1]**

Similarly

$$\int_1^2 e^{tz}(2-z)dz = -\frac{e^t}{t} + \frac{1}{t^2}\left(e^{2t} - e^t\right)$$

Addition gives $M_2(t) = \frac{1}{t^2}\left(e^t - 1\right)^2$ **[1]**

ii)

$$M_{X_1}(t) = \frac{e^t - 1}{t}$$

$$M_{X_2}(t) = \frac{e^t - 1}{t}$$ **[1]**

where X_1 and X_2 are independent uniform random variables on (0,1)

$$M_{X_1+X_2}(t) = \left(\frac{e^t - 1}{t}\right)^2$$ **[1]**

Comment : g(z) is the pdf of the sum of two independent uniform random variables.
That is the sum of two independent uniform random variables follows triangular distribution. **[1]**

Total [9]

Q.4 a) Expected number of accidents in a group of 1000 policyholders, is given by

$1000 \times \frac{1}{10000} = 0.1,$ which is taken as the parameter of the Poisson distribution **[1]**

Let $p(x) = e^{-0.1}\frac{(0.1)^x}{x!}$ x = 0,1,2,...........

Required Probability = P(0) + P(1) + P(2)

$$= e^{-0.1}\left(1 + 0.1 + \frac{0.01}{2}\right)$$ **[1]**

$$= 0.9048\left(1.105\right) = 0.9998$$ **[1]**

b) Define Poisson process X(t) as a discrete process, stating all the postulates namely

i) $P\left[\text{1 occurrence in (t, t+}\Delta\text{t)}\right] = l\Delta t + 0(\Delta t)$

ii) $P\left[\text{no occurrence in (t, t+}\Delta\text{t)}\right] = 1 - l\Delta t + 0(\Delta t)$

iii) $P\left[\text{2 or more occurrence in (t, t+}\Delta\text{t)}\right] = 0(\Delta t)$

iv) X(t) is independent of the number of occurrences of the event in any interval prior and after the interval (0,t) **[2]**

Let E_i takes placeat time t_i and T be the interval between the occurrence of E_i and E_{i+1}. T is a cts random variable.

Now $P[T>t] = P[E_{i+1}$ does not occur in $(t_i, t_i + t)]$

$= P$[no occurrence in the interval of length t]

$= P[X(t) = 0] = e^{-lt}$

$F(t) = P(T < t) = 1 - e^{-lt}$ and hence

$f(t) = F'(t) = l e^{-lt} \quad (t \geq 0)$

which is an exponential distribution **[2]**

c)

$$P(X = r) = P(Y = r) = pq^{r-1}, \quad r = 1,2,..........$$

$$P[X = r \mid X + Y = k] = \frac{P[X = r, X + Y = k]}{P[X + Y = k]}$$

$$= \frac{P[X = r].P[Y = k - r]}{\sum P[X = r].P[Y = k - r]}$$ **[2]**

$$= \frac{pq^{r-1}.pq^{k-r-1}}{\sum_{r=1}^{k-1} pq^{r-1}.pq^{k-r-1}}$$

$$= \frac{q^{k-2}}{\sum_{r=1}^{k-1} q^{k-2}} = \frac{1}{k-1}$$

$r = 1,2,.......(k-1)$

which is a discrete uniform distribution. **[2]**

Total [11]

Q.5 a) Statement of central limit theorem for i.c.r.v.'s **[1]**

Let Z be standard Normal r.v.

$\bar{X} \sim N(167, 4.5)$

$$P[163 < \bar{X} < 171] = P\left[\frac{163-167}{4.5} < \frac{\bar{X}-167}{4.5} < \frac{171-167}{4.5}\right]$$

$$= P\left[-0.8889 < \frac{\bar{X} - m}{4.5} < 0.8889\right]$$

$$\simeq 2P[Z < 0.89] = 0.626$$ **[2]**

b)

$$V = \frac{4}{3} p X^3$$

clearly $V \le \frac{4}{3} p$ and for $0 \le V_0 \le \frac{4}{3} p$

$$P(V > V_0) = P\left(\frac{4}{3} p X^3 > V_0\right)$$

$$= P\left(X^3 > \frac{3}{4}\frac{V_0}{p}\right)$$

$$= P\left[X > \left(\frac{3}{4}\frac{V_0}{p}\right)^{\frac{1}{3}}\right]$$

$$= 1 - F\left(\frac{3}{4}\frac{V_0}{p}\right)^{\frac{1}{3}}$$ **[2]**

$$= 1 - 3\left(\frac{3}{4}\frac{V_0}{p}\right)^{\frac{2}{3}} + \frac{3}{2}\frac{V_0}{p}$$ **[1]**

The above probability is zero if $V_0 > \frac{4}{3} p$ since X lies between 0 and 1. **[1]**

Total [7]

Q.6 i)

Likelihood $L(l) = \frac{1}{l^n} e^{-\sum x_i / l}$ **[1]**

$$\log L(l) = -n \log l - \frac{\sum x_i}{l}$$

$\frac{\partial \log L}{0 l} = 0$ gives

$$\hat{l} = \frac{\sum x_i}{n}$$ **[1]**

$$\frac{\partial^2 \log L}{\partial l^2} = \frac{n}{l^2} - 2\frac{\sum x_i}{l^3} < 0 \text{ at } \hat{l} = \bar{x}$$ **[1]**

ii)

$$E(\hat{l}) = E\left(\frac{\sum x_i}{n}\right) = l$$ **[1]**

iii)

$$-E\left(\frac{\partial^2}{\partial l^2} \log L(l)\right) = \frac{2}{l^3} n l - \frac{n}{l^2}$$

$$= \frac{n}{l^2}$$ **[1]**

$$CRLB = \frac{1}{\left(\frac{n}{l^2}\right)} = \frac{l^2}{n}$$ **[2]**

iv)

$$Var(\hat{l}) = V\left[\frac{\sum x_i}{n}\right] = \frac{1}{n^2} n l^2 = \frac{l^2}{n}$$ **[1]**

v)

$$\hat{l} \sim N\left(l, \frac{l^2}{n}\right)$$

$$\frac{\bar{x} - l}{l/\sqrt{n}} \sim N(0,1)$$

[1]

95% confidence interval for l is obtained as follows

$$P\left[-1.96 < \frac{\bar{x} - l}{\left(l/\sqrt{n}\right)} < 1.96\right] = 0.95$$

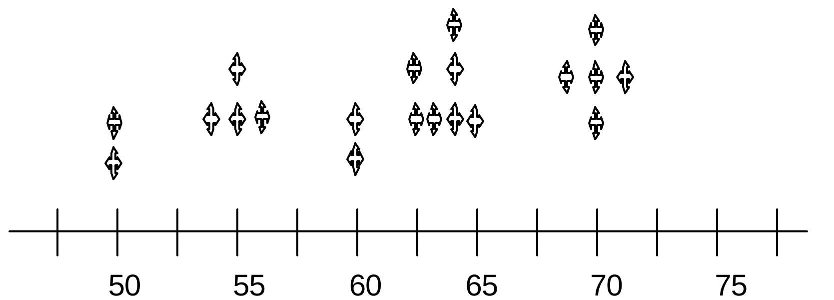
This gives the interval as

$$\frac{\bar{x}}{1 + 1.96/\sqrt{n}}, \frac{\bar{x}}{1 - 1.96/\sqrt{n}}$$

[1]

Total [10]

Q.7 a)



seems to be approximately normal. **[1]**

b)

$$\bar{x} = \frac{\sum x}{n}$$

$$= \frac{1235}{20}$$

$$= 61.75$$

$$S^2 = \frac{\sum x^2 - n\bar{x}^2}{n-1}$$

$$= \frac{77117 - \frac{1235^2}{20}}{19}$$

$$= 45.04$$

[2]

95% CI is

$$\bar{x} \pm t_{\alpha/2}\sqrt{\frac{S^2}{n}}$$

$$61.75 \pm 2.093\sqrt{\frac{45.04}{20}}$$

[1]

Or 58.61 and 64.89 **[2]**

c)

Here we apply the paired t-test

Differences in pair d (before and after treatment)

d: 3 2 2 0 4 4 -5 0 2 -1

2 0 -1 1 4 2 1 3 -5 -1

$\sum d = 17 \qquad \sum d^2 = 141$

If m be the true mean of the difference we test $H_0 : m = 0$ (treatment ineffective) versus $H_1 : m > 0$ (treatment effective) **[1]**

$\bar{d} = 0.85 \qquad S^2 = 6.66$ **[1]**

Under H_0,

$$t = \frac{\bar{d} - 0}{S}\sqrt{20}$$

$$= \frac{0.85\sqrt{20}}{\sqrt{6.66}}$$

$= 1.47$ with 19 d.f **[1]**

5% value of t (one-tailed) for 19 df is 1.729

obs t < critical value - H_0 holds

Treatment not effective. **[2]**

Total [11]

Q.8

We apply the c^2 - test

Under H_0 that the park is visited in the same proportion by birds of the six categories, the expected frequency (E) for each category is

$$\frac{1}{6} * 54 = 9$$

$$c^2 = \sum \frac{(0 - E)^2}{E}$$ over the six cells **[1]**

Calculated $c^2 = 7.778$ with 5 d.f **[1]**

$c^2_{0.05}(5) = 11.07$

Obs. Value of c^2 not significant H_0 holds. **[2]**

Total [4]

Q.9

a) Scatter plot – Y against X – shows linear trend **[1]**

b)

$$\bar{X} = \frac{\sum X}{n}$$

$$= \frac{50}{5}$$

$= 10$

$$\bar{Y} = \frac{\sum Y}{n}$$

$$= \frac{10}{5}$$

$= 2$

$\sum X^2 = 510 \quad \sum Y^2 = 30 \quad \sum XY = 109$

$$S_{XX} = \sum X^2 - \frac{\left(\sum X\right)^2}{n}$$

$= 10$ **[2]**

Similarly, $S_{YY} = 10 \quad S_{XY} = 9$

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = 0.9$$

[1]

c)

$$\hat{b}_1 = b_1 = \frac{S_{xy}}{S_{xx}} = 0.9$$

[2]

$\hat{b}_0 = b_0 = \overline{Y} - b_1\overline{X} = -7$ **[2]**

line $\overline{\text{Y}} = -7 + 0.9X$

d) Standard error of estimate Se is given by

$$S_e^2 = \frac{1}{n-s}\sum\left(Y_i - Y_i\right)^2$$

$$= \frac{1}{n-2}\left(S_{YY} - \frac{S_{XY}^2}{S_{XX}}\right)$$

[1]

= 0.6333

Se = 0.7958 **[1]**

e) 95% CI

$$b_1 \pm t_{a/2, n-2}\sqrt{\frac{S_e^2}{S_{XX}}}$$

[1]

$$0.9 \pm (3.182)\sqrt{\frac{19}{30*10}}$$

0.9 ± 0.8009 **[1]**

Or 0.0991 and 1.701

If e_i denotes the error term,

$e_i^{'}s$ are IN$(0, s^2)$ **[1]**

f) The above interval does not include the value zero. **[2]**

Hence Ho is rejected.

Total [15]

Q.10 We test the hypothesis Ho that the three methods have the same mean level of productivity against H_1 that means are not all equal. **[1]**

$n_1 = n_2 = n_3 = 5$

$y_1 = 147 \quad y_2 = 178 \quad y_3 = 127$

$$CF = \frac{452^2}{15} = 13620.3$$

$SST = 14326 - 13620.3 = 705.7$

$$SSB = \frac{147^2 + 173^2 + 127^2}{5} - CF = 264.1$$

$SSE = SST - SSB = 441.6$ **[2]**

[2]

ANOVA Table

Source	SS	Df	MSS	F
Between methods	264.1	2	132.05	3.59
Error	441.6	12	36.8	
Total	705.7	14		

Table value $F_{2,12}$ (0.05) = 3.885

No evidence against Ho. **[1]**

Total [6]

Q.11 a)

$$f_X(x) = \int_0^\infty x e^{-x(1+y)} dy = e^{-x}$$

$$f_Y(y) = \int_0^\infty x e^{-x(1+y)} dx = \frac{1}{(1+y)^2}, y > 0$$

As $\int_0^\infty \frac{y}{(1+y)^2}$ does not converge, E(Y) does not exist **[2]**

$$f(y \mid x) = \frac{f(x,y)}{f(x)}$$

$$= xe^{-xy}, y > 0$$

$$E(Y \mid x) = \int_0^\infty yxe^{-xy} dy = \frac{1}{x} \quad x>0$$

[2]

b) i) Theory **[2]**

ii) $m_N = 3 \qquad s_N^2 = 3$

$$m_X = 15 * \frac{1}{3} = 5$$

$$s_X^2 = 15 * \frac{1}{3} * \frac{2}{3} = \frac{10}{3}$$ **[1]**

$E(S) = m_N m_X = 15$ **[2]**

Var(S) = $m_N s_x^2 + s_N^2 m_x^2$ **[1]**

= 85 **[2]**

Total [12]

****