The Institute of Actuaries of India

Subject CT3 – Probability & Mathematical Statistics

15th May 2007

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.

Arpan Thanawala
Chairperson, Examination Committee

1. a) Steam and leaf plot

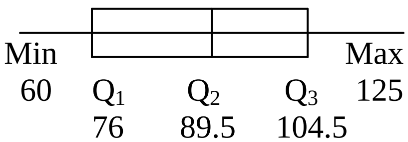
6	0,5,5,8,9
7	2,4,4,5,7,8
8	2,3,3,5,7,8,9
9	0,0,1,4,4,5,7
10	0,2,7,8
11	0,2,4,5
12	2,4,5

b) $Q_1 : \left(\frac{n+2}{4}\right)\downarrow :76$

$Q_2 : \left(\frac{n+1}{2}\right)\downarrow :89.5$

$Q_3 : \left(\frac{n+2}{4}\right)\uparrow :104.5$

c) Box Plot



[6]

2.a) Total No. of outcomes = $2^3 = 8$

$A = \{(HTH),(HTT),(THH),(THT)\}$

Similarly B and C have 4 outcomes each

$\therefore \ P(A) = P(B) = P(C) = \frac{1}{2}$

Now $A \cap B = A = \{(HTH),(THT)\}$

Similarly there are two outcomes in $B \cap C$ and $C \cap A$, so

$P(A \cap B) = P(B \cap C) = P(C \cap A) = \frac{2}{8} = \frac{1}{4}$

$$\left.\begin{aligned} P(A/B) &= \frac{P(A\cap B)}{P(B)} = \frac{1}{2} = P(A) \\ P(A/C) &= \frac{P(A\cap C)}{P(C)} = \frac{1}{2} = P(A) \end{aligned}\right\}$$

b) Since $P(A\cap B) = P(A).P(B)$

$P(C\cap B) = P(C).P(B)$

$P(A\cap C) = P(C).P(A)$

$\Rightarrow A, B$ and C are pairwise independent.

[4]

3.a) $\int_0^{50} k(50-x)dx = 1 \Rightarrow k = \frac{1}{1250}$

b) $\int_0^{50} \frac{1}{1250} x(50-x)dx = \frac{1}{1250}\left[\frac{50x^2}{2} - \frac{x^3}{3}\right]_0^{50}$

$$= \frac{1}{1250}\left[\frac{50\times(50)^2}{2} - \frac{50^3}{3}\right]$$

$= 16.67$

c) $\int_{25}^{50} \frac{1}{1250}(50-x)dx = \frac{1}{1250}\left[50x - \frac{x^2}{2}\right]_{25}^{50}$

$= 0.25$

d) $P\left(X < 30 / X > 25\right) = \frac{P(25 < X < 30)}{P(X > 25)}$

$$= \frac{\int_{25}^{30} f(x)dx}{\int_{25}^{50} f(x)dx}$$

$$= \frac{0.09}{0.25} = 0.36$$

[7]

4. Let $A=\{H\}$ - Head on first toss

$B = \{T,H\}$ - Tail on first toss and Head on second toss

$C = \{T,T\}$ - Tail on both tosses

Let X denote his random earnings

$EX = P(A)E(X/A) + P(B)E(X/B) + P(C)E(X/C)$

$$\left.\begin{array}{l} E[X/A] = 1, E[X/B] = -1+1 = 0 \quad E(X/C] = -1-1 = -2 \\ P(A) = 1/2 \\ P(B) = 1/4 \\ P(C) = 1/4 \end{array}\right\}$$

$EX = \frac{1}{2}.1 + \frac{1}{4}.0 + \frac{1}{4}(-2)$

$= 0$

[4]

5. N_i $i = 1,2,3,4$ denotes number of hospitalization in September, October, November and December respectively

From the given information

$N_1 \sim Poi(1),\ N_2 \sim Poi(1)\ N_3 \sim Poi(2),\ N_4 \sim Poi(3)$

Let $X = N_1+N_2+N_3+N_4$

$N_j^{'s}$ are assumed to be mutually independent

$X \sim Poi(1+1+2+3)=Poi(7)$

The desired probability is $P[X<5]$

$$\therefore P[X < 5] = \sum_{x=0}^{4} \frac{7^x e^{-7}}{x!}$$

$= 0.0009 + 0.0064 + 0.0223 + 0.0521 + 0.0912$

$= 0.1729$

[4]

6. a) $f(x) = \dfrac{\lambda^\alpha e^{-\lambda x} x^{\alpha-1}}{\Gamma\alpha} = k \quad e^{-\lambda x} x^{\alpha-1}$

where $k = \dfrac{\lambda^\alpha}{\Gamma\alpha}$

Hence, $\log f(x)= \log k-\lambda x+(\alpha-1)\log x$. Differentiating

$$\frac{f'(x)}{f(x)} = -\lambda + \frac{(\alpha-1)}{x}$$

$$f'(x) = k(\alpha-1)e^{-\lambda x} x^{\alpha-1} - \lambda k\, e^{-\lambda x} x^{\alpha-1}$$

$f'(x) = 0$ gives

$$\left(\frac{\alpha-1}{\lambda}\right) = x$$

$$\left.\begin{array}{l} \text{Hence, the mode is } \dfrac{\alpha-1}{\lambda}\ if\ \alpha > 1 \\ \qquad (or) \\ \qquad \text{mode} = 0\ if\ \alpha < 1 \end{array}\right\}$$

b) From the data, the average demand $= \dfrac{\alpha}{\lambda} = a \quad \rightarrow \qquad (1)$

(For Gamma, Mean = α/λ)

Most likely demand = b = Mode

$$\text{i.e. } \frac{\alpha-1}{\lambda} = b \quad \rightarrow \qquad (2)\int$$

From (1) and (2) $\quad (a-b) = 1/\lambda$

For Gamma distribution,

Var X = $\alpha/\lambda^2 = \left(\dfrac{\alpha}{\lambda}\right)\left(\dfrac{1}{\lambda}\right)$

$= a(a-b)$ from (1) and (2)

c) Skewness = $\dfrac{Mean - Mode}{S.D.}$

$$= \frac{\left(\frac{\alpha}{\lambda}\right) - \left(\frac{\alpha - 1}{\lambda}\right)}{\left(\frac{\alpha}{\lambda^2}\right)}$$

$= \lambda/\alpha$

So, when $\alpha = \lambda = 2$, Skewness = 1

(The computation of skewness can be done using any other formula also)

[10]

7. Let X denote the number of policies in the block for which there is atleast one claim in the coming period

Here X ~ B(250, 0.10)

We have to evaluate $P[X > 30] = \sum_{x=31}^{250} \binom{250}{x} (0.10)^x (0.90)^{250-x}$

To apply normal approximation

EX = 250(0.10) = 25

$Var\ X$ = 250(0.10) (0.90) = 22.5

When n is large, $X \sim N\left(25, \sqrt{22.5}\right)$

$P(X > 30) = P[X \geq 30.5]$

$= \left[\frac{X - 25}{\sqrt{22.5}} \geq \frac{30.5 - 25}{\sqrt{22.5}}\right]$

$= P\ [Z \geq 1.1595]$

$= 1 - \Phi(1.1595)$

$= 1 - 0.877$

$= 0.123$

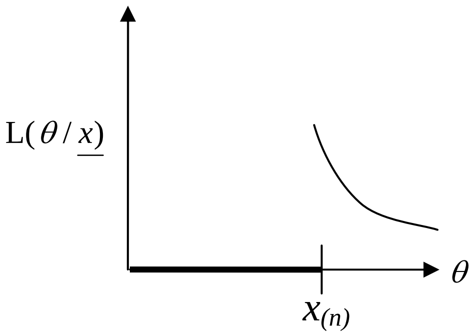
[4]

8. a) Likelihood :

$$L(\theta / \underline{x}) = \begin{cases} \dfrac{2^n \prod x_i}{\theta^{2n}} & if\ 0 < x_1, ..., x_n < \theta \\ 0 & otherwise \end{cases}$$

$$= \begin{cases} \dfrac{2^n \prod x_i}{\theta^{2n}} & if\ 0 < x_{(n)} < \theta \\ 0 & otherwise \end{cases}$$

Sketch



Since $L(\theta/\underline{x})$ is decreasing function of θ whenever $\theta > x_{(n)}$, the MLE is $X_{(n)}$.

b) Since $X_{(n)} < \theta$ for all θ implies $EX_{(n)} < \theta$ for all θ. Hence, $X_{(n)}$ is not unbiased for θ.

[6]

9. a) From the data $n = 1034$, $X = 848$

$$\hat{p} = \frac{X}{n} = \frac{848}{1034} = 0.82 \text{ (app)}$$

$H_0 : p = 0.5\ (p_0)$

$H_1 : p \neq 0.5$

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.82 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{1034}}}$$

= 20.64 (on rounding the denominator as 0.0155)

Critical value (5% level) = 1.96

Cal Z > critical value

Reject H_0 .

b) 90% confidence interval for p is

$$\hat{p} \pm Z_{\alpha/2} \frac{\hat{p}(1-\hat{p})}{n}$$

i.e. $0.82 \pm 1.645\sqrt{\frac{0.82(1-0.82)}{1034}}$

0.82 ± 0.0197

Hence, 90% confidence interval for p is (0.8003, 0.8397)

[5]

10. a) Choice of Policy

		Whole Life	Endowment with Profit	Endowment without profit	
Nativity	Rural	14	8	8	30
	Urban	10	18	17	45

0	E	$(0\text{-}E)^2$	$(0\text{-}E)^2/E$
14	9.6	19.36	2.0167
8	10.4	5.76	0.5538
8	10.	4.00	0.4000
10	14.4	19.36	1.3440
18	15.6	5.76	0.3690
17	15.0	4.00	0.2667
			4.9502

Critical value of χ^2 for 2d.f at 5% level is 5.99. Don't reject H_0.

b) Consider

	Whole life	Endowment
Rural	14	16
Urban	10	35

$$\chi^2 = \frac{(330)^2 \times 75}{51 \times 24 \times 30 \times 45} = 4.94$$

Reject H_0 , since the critical value of χ^2 for 1d.f. at 5% level is 3.84.

[6]

11. a)

Before Training (X)	42	35	37	46	53	38	44	40	43
After Training (Y)	47	28	26	54	42	17	44	31	44
Diference (d)	-5	7	11	-8	11	21	0	9	-1

$\bar{x}_d = 5$, $s_d = 9.206$, $n = 9$

$H_0 = \mu_d = 0$; $H_1 : \mu_d \neq 0$

Test statistic $t = \dfrac{\bar{x}_d}{s_d / \sqrt{n}} = 1.63$

Critical value of t at 5% level for 8d.f. is = 2.306. Do not reject H_0

b) 95% confidence interval for the mean change in ability of trainees

$$(\bar{x}_d - t_{\frac{\alpha}{2}} s_d / \sqrt{n},\ \bar{x}_d + t_{\frac{\alpha}{2}} s_d / \sqrt{n})$$

(-2.088,12.088)

The above confidence interval is for $(\mu_X - \mu_Y)$. However, the confidence interval for $(\mu_Y - \mu_X)$ is (-12.088, 2.088).

c) Computation of correlation coefficient :

$\Sigma X = 378$, $\Sigma Y = 333$, $\Sigma X^2 = 16112$, $\Sigma Y^2 = 13471$, $\Sigma XY = 14340$

$$r = \frac{\frac{1}{n}\sum XY - \overline{X}\,\overline{Y}}{\sqrt{\frac{1}{n}\Sigma X^2 - (\overline{X})^2}\ \sqrt{\frac{1}{n}\Sigma Y^2 - (\overline{Y})^2}}$$

$= 0.6796$

Test the significance of ρ :

$H_0 : \rho = 0$ *vs.* H_1 $\rho \neq 0$

Test statistic $t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}}$

$$= \frac{0.6796 X \sqrt{7}}{0.7336} = 2.451$$

Critical value of t at 5% level for 8 d.f.is 2.306, Reject H_0

[13]

Q11(c) The "Indicative solutions" has a minor error in the computation of test statistic t.

The correct value is $\dfrac{0.6796 X \sqrt{7}}{0.7336} = 2.451$ (approx)

12. a) Assumptions of Regression model : The errors ($\in$'s) are independent and have normal distribution with mean zero and a common (unknown) variance σ^2.

b) $\Sigma x_i = 2{,}000$, $\Sigma x_i^2 = 5{,}32{,}000$

$\Sigma y_i = 8.35$, $\Sigma y_i^2 = 9.1097$

$n = 10$

$\Sigma x_i y_i = 2175.40$

$$S_{xx} = 5{,}32{,}000 - \frac{(2000)^2}{10} = 1{,}32{,}000$$

$$S_{xy} = 2175.40 - \frac{2000(8.35)}{10}$$

$= 505.40$

$$S_{yy} = 9.1097 - \frac{(8.35)^2}{10}$$

$= 2.13745$

$$\hat{\beta} = b = \frac{S_{xy}}{S_{xx}} = 0.00383$$

$$a = \bar{y} - b\bar{x} = 0.069$$

Hence $\hat{y} = 0.069 + 0.00383x$

c) $s_e^2 = \dfrac{S_{yy} - S_{xy}^2 / S_{xx}}{n-2}$

$= 0.0253$

d) 95% confidence interval for α is $a \pm t_{\alpha/2} s_e \sqrt{\dfrac{1}{n} + \dfrac{(\bar{x})^2}{S_{xx}}}$

$t_{\alpha/2}(n-2) = 2.306$ (from tables)

Hence 95% confidence interval is (–0.164 , 0.302)

e) $H_0 : \beta = 0$ *vs.* H_1 $\beta \neq 0$

Test statistic $t = \dfrac{b - \beta}{s_e}\sqrt{S_{xx}}$

$= 8.75$

For the two sided test t value for 8 d.f. = 2.306

Reject H_0

f) The formula is $(a + bx_0) \pm t_{\alpha/2}\ s_e \sqrt{\dfrac{1}{n} + \dfrac{(x_0 - \bar{x})^2}{S_{xx}}}$

where $x_0 = 190$

(0.68, 0.93) is the 95% confidence interval for the mean evaporation coefficient $\alpha + 190\beta$

[14]

13. $H_0 : \mu_A = \mu_B = \mu_C$

H_1 : atleast one pair is not equal.

where μ_A , μ_B , and μ_C are mean scores awarded to logos *A*,*B*,*C*.

ANOVA TABLE

Source	SS	df	MSS	*F*
Between logos	31	2	15.50	3.02
Error	107.63	21	5.13	
Total	138.63	23		

Critical value of $F_{0.05}(2,21) = 3.47$

Do not Reject H_0

[6]

14. $f(x,y) \begin{array}{l} = 2e^{-(x+y)} \quad if \quad 0 \le y \le x < \infty \\ = \quad 0 \qquad otherwise \end{array}$

$$f(y) = \int_y^\infty 2e^{-(x+y)} dx$$

$$= 2e^{-y} \int_y^\infty e^{-x} dx = 2e^{-2y}; y > 0$$

Hence,

$$f(x/y) = \frac{f(x,y)}{f(y)} = \frac{2e^{-(x+y)}}{2e^{-2y}}$$

$$= e^{y-x} \quad ; \quad x > y$$

$$\therefore E(X/Y = y) = \int_y^\infty x e^{y-x} dx$$

$$= e^y \int_y^\infty x e^{-x} dx$$

$$= e^y e^{-y}(y+1) = y+1$$

[4]

15. $F_C(c) = P[C \le c]$

$$= \sum_{n=0}^\infty P\left[C \le c / N = n\right].P[N = n]$$

$$= \sum_{n=0}^\infty \left[W_1 + W_2 + ... + W_N \le c / N = n\right] P[N = n]$$

$$= \sum_{n=0}^\infty \left[W_1 + W_2 + ... + W_n \le c\right] P[N = n] \quad (\because W_j \ \ s \text{ are indep of } N)$$

$$= \sum_{n=0}^\infty F_{W*}(c) \ \ P_N(n)$$

where W^* is the n fold convolution of W.

Let $M_C(t)$ be the mgf of C

$M_C(t) = E[e^{tc}]$

$$= E_N\left[E(e^{ct} / N)\right]$$

$$= E_N\left[E\{e^{(W_1+W_2+...+W_N)t} / N\}\right]$$

$$= E_N\left[E(e^{W_1 t})E(e^{W_2 t})...E(e^{W_N t})\right] \quad (\because W_j \ \ s \text{ are indep of } N)$$

$$= E_N\left[E(e^{Wt})^N\right]$$

$$= E_N\left[M_W(t)^N\right]$$

$= E_N\left[e^{N \log M_W(t)}\right]$

$= M_N\left[\log M_W(t)\right]$

Hence, $\log M_C(t) = \psi_C(t) = \psi_N\left(\psi_W(t)\right)$ which is the cumulant generating function of C.

[7]

* * * * * * * * * * * * *