Institute of Actuaries of India

CT3: Probability and Mathematical Statistics

Indicative Solution

November 2008

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

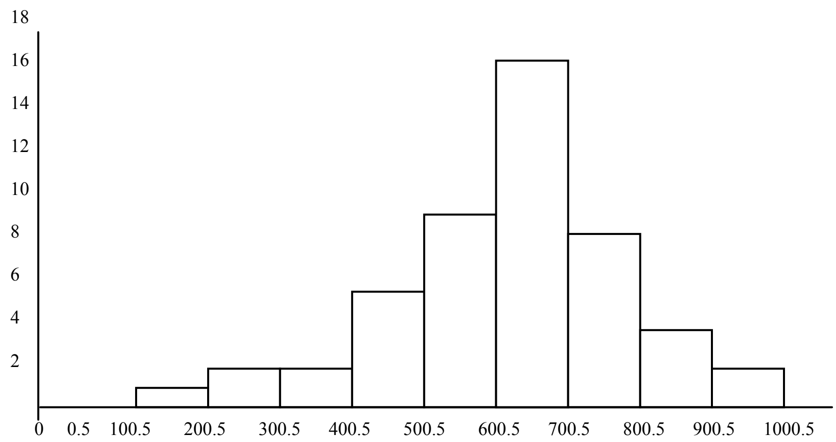
1.a) Stem and leaf diagram

1	06
2	06 58
3	29 69
4	12 28 65 71 89 95
5	09 23 28 34 56 67 71 79 88
6	08 19 28 28 38 39 39 54 56 59 71 73 74 79 79 88 93
7	08 28 38 39 39 59 79
8	08 38 39 59
9	08 98

b)Frequency Distribution

Class interval (in '000s)		Frequency	Class interval (in '000s)		Frequency
1	100	0	501	600	9
101	200	1	601	700	17
201	300	2	701	800	7
301	400	2	801	900	4
401	500	6	901	1000	2

Histogram



c) Negatively Skewed

d) If the observation 998 is omitted, there are only 49 observations. The median position is $\frac{49+1}{2} = 25$. The median is 638. As the median position is 25, there are 24 each to its right and left. Thus Q_1 is $\frac{24+1}{2}$ th position. $Q_1 = \frac{509+523}{2} = 516$

(Total 10 Marks)

2. $y = 1 - e^{-\lambda x} \Rightarrow x = -\frac{1}{\lambda}\log(1-y)$

Hence, $\frac{dx}{dy} = \frac{1}{\lambda(1-y)}$. Therefore $f(y) = f(x)\left|\frac{dx}{dy}\right|$; x in terms of y

$$= \lambda e^{-\lambda\left[-\frac{1}{\lambda}\log(1-y)\right]} \frac{1}{\lambda(1-y)} = 1 \quad ; \quad 0 < y < 1$$

Marks may be awarded if the candidate gives the df

(Total 3 Marks)

3 The joint *pdf* of (X,Y) is

$$f(x,y) = x\exp[-x(1+y)]\,;\; x > 0, y > 0$$

c) Marginal of X: $g(x) = \int_0^\infty x\exp[-x(1+y)\,dy$

$$= e^{-x} \quad ; \quad x > 0$$

Marginal of Y: $h(y) = \int_0^\infty x e^{-x}\; e^{-xy}\,dx$

$$= \frac{1}{(1+y)^2} \;;\; y > 0$$

b) Clearly $E[Y] = \int_0^\infty y\frac{1}{(1+y)^2}\,dy = \infty$

$E[Y]$ does not exist

c) $E[Y/x] = \int_0^\infty yxe^{-xy}\,dy = \frac{1}{x}\int_0^\infty te^{-t}\,dt = \frac{1}{x}$

Comment: $E[Y]$ does not exist even though *E[Y/x]* exists.

(Total 7 Marks)

4 a) *Mgf* of exponential random variable is $\frac{\lambda}{\lambda - t}$, $t < \lambda$

Mgf of $Y = X_1 + ... + X_n$ $is\left(\frac{\lambda}{\lambda - t}\right)^n$

b) *Cgf* of Y is $C_Y(t) = n\log\left(\frac{\lambda}{\lambda - t}\right)$

$$E[Y] = C_Y'(0) = n\left(\frac{\lambda - t}{\lambda}\right)\frac{\lambda}{(\lambda - t)^2}\Bigg|_{t=0} = \frac{n}{\lambda}$$

$$V[Y] = C_Y''(0) = \frac{n}{\lambda^2}$$

(Total 6 Marks)

5. Let A_1,A_2,A_3 and A_4 denote the four blood types *A,B,AB* and *O* respectively.

Let *S* denote the soldiers blood type classified as type *A* (*ie.*A_1).

It is given that i) $P(A_1)$=0.41, $P(A_2)$ = 0.09, $P(A_3)$=0.04 and $P(A_4)$=0.46

ii) $P(S/A_1)$=0.88, $P(S/A_2)$=0.04, $P(S/A_3)$=0.10 and $P(S/A_4)$=0.04

Hence, $P(A_1/S) = \dfrac{P(S/A_1)P(A_1)}{\sum_{L=1}^{4} P(S/A_i)P(A_i)}$

$$= \frac{0.88 \times 0.41}{(0.88 \times 0.41) + (0.04)(0.09) + (0.10)(0.04) + (0.04)(0.46)}$$

$$= 0.93$$

(Total 6 Marks)

6. $P(X_3 = 1) = P(X_1 = 1, X_2 = 1) + P(X_1 = -1, X_2 = -1)$

$$= P(X_1 = 1).P(X_2 = 1) + P(X_1 = -1).P(X_2 = -1)$$

$$= \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

$P(X_3 = -1) = 1/2$

$P(X_1 = 1, X_3 = 1) = P(X_1 = 1).P(X_3 = 1/X_1 = 1)$

$$= P(X_1 = 1).\frac{P(X_3 = 1, X_1 = 1)}{P(X_1 = 1)}$$

$$= P(X_1 = 1).\frac{P(X_2 = 17.P(X_1 = 1)}{P(X_1 = 1)} = \frac{1}{2}.\frac{1}{2} = \frac{1}{4}$$

Similarly for other choices

$P(X_1 = 1, X_3 = -1) = P(X_1 = 1)P(X_3 = -1)$

$P(X_1 = -1, X_3 = 1) = P(X_1 = -1)P(X_3 = 1)$

$P(X_1 = -1, X_3 = -1) = P(X_1 = -1).P(X_3 = -1)$

But $P(X_1 = 1, X_2 = 1, X_3 = -1) \neq P(X_1 = 1).P(X_2 = 1).P(X_3 = -1)$

(Total 3 Marks)

7. Let X denote the number of children with birth defects whose mothers smoke while pregnant.

If there is no relationship between maternal smoking and birth defects, then X follows binomial with $n = 20, p = 0.4$

$P(X \geq 12) = P\left(\dfrac{X - np}{\sqrt{npq}} > \dfrac{11.5 - 8}{\sqrt{4.8}}\right]$,on applying normal approximation and correction or continuity.

$$= P(Z > 1.59)$$

$$= 0.0559$$

(Total 3 Marks)

8.a) We have

$$E[X_k^2] = VAR[X_k] + \left(E[X_k]\right)^2 = 1 + 0^2 = 1$$

$$E[Y] = \sum_{k=1}^{50} E[X_k^2] = 50$$

b) We can write that Y follows $N(50, \sigma_Y^2)$ (approximately), where

$$\sigma_Y^2 = E[Y^2] - (E[Y])^2 = 2600 - (50)^2 = 100$$

Then, $P[Y < 60] \approx P\left[N(0,1) < \dfrac{60-50}{10}\right] = 0.8413$

Alternatively: X_i follows N (0,1); X_i^2 follows Chi-square with 1 df.

$\sum X_i^2$ follows Chi square dist with 50 df (assuming X_i^2 are independent).

which is same as Gamma distribution with parameters α = 50/2 = 25 and λ =1/2

E[Y] = α/λ = 25/0.5 = 50;Var[Y] = α/ λ^2 = 100

(Total 5 Marks)

9. The joint *pdf* of (*X*,λ) is

$$f_{X,\lambda}(x,\lambda) = \frac{r^p}{\Gamma p\, x!}\lambda^{x+p-1}\ e^{-\lambda(r+1)}$$

Hence, $f_x(x) = \int\limits_0^\infty f_{x,\lambda}(x,\lambda)d\lambda$

$$= \frac{r^p\Gamma(x+p)}{x!\Gamma p(+1)^{x+p}}$$

$$= \left(\frac{r}{r+1}\right)^p\left(\frac{1}{(r+1)^x}\right)\frac{(x+p-1)!}{(p-1)!x!}$$

$$= \binom{x+p-1}{x}\left(\frac{1}{r+1}\right)^x\left(\frac{r}{r+1}\right)^p$$

which is a negative binomial with parameters p and $\left(\dfrac{r}{r+1}\right)$

Hence $E[X] = \dfrac{p}{r}$ and $V[X] = \dfrac{p(r+1)}{r^2}$

(Total 5 Marks)

10. Let *W* be the time taken for the first occurrence of the event and *F(w)* be its distribution function.
The first occurrence o the event will take place after *w* only, if no occurrence takes place in [0,*w*]
Let *X* denote the number occurrences of the event.
Hence, $P(W > w) = P(X = 0) = e^{-\lambda w}$ (by assumption)
Hence *W* has exponential distribution with mean $\dfrac{1}{\lambda}$

(Total 3 Marks)

11. a) $X \sim N(12, 0.5)$, so $\overline{X} \sim N(12, 0.25)$

Probability of type I error : $P_{H_0}(\overline{X} < 11.5) = P\left(\frac{\overline{X}-12}{0.25} < \frac{11.5-12}{0.025}\right)$

$$= P\left(Z < \frac{-.5}{0.25}\right) = P(Z < -2) = 0.0228$$

b)Power at 11.25 : $P_{H_1}(\overline{X} < 11.5) = P\left(Z < \frac{11.5-11.25}{0.25}\right)$

$$= P(Z < 1) = 0.8413$$

(Total 4 Marks)

12. a) $n = 400$ $\qquad p = \frac{64}{400} = 0.16$ $\qquad q = 0.84$

Standard error (se) : $\sqrt{pq/n} = \sqrt{\frac{0.16 \times 0.84}{400}} = 0.01839$

b)99% confidence interval is

$$(p - 2.58\ se,\quad p + 2.58\ se) = (0.16 - 2.58 \times 0.0183,\ 0.16 + 2.58 \times 0.0183)$$
$$= (0.1128, 0.2072)$$

(Total 3 Marks)

13. a) $H_0 : \sigma_1^2 = \sigma_2^2$ $\qquad H_1 : \sigma_1^2 \neq \sigma_2^2$, $\qquad n_1 = n_2 = 16$

$$F = \frac{s_1^2}{s_2^2} = \frac{0.040}{0.028} = 1.43 \text{ with (15,15) df}$$

Table $F_{(15,15)}$ at 0.2 level is 1.55

Do not reject H_0. Pooling of variances is appropriate

b) 95% confidence interval for difference in means

$$\left[(\bar{x}_1 - \bar{x}_2)\right] - t_{\alpha/2}\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}, (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2}\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$= (0.94 - 0.62) - 2.042\sqrt{0.034\left(\frac{1}{16} + \frac{1}{16}\right)},$$

$$= (0.19, 0.45)$$

where $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$

c) The mean exposure in 1973 was higher than in 1979 since, the interval does not contain zero and is positive valued.

(Total 7 Marks)

14. a) Likelihood: $L(\alpha,\beta)=\begin{cases}\prod_{i=1}^{n}\left(\frac{\alpha}{\beta}\right)\left(\frac{x_i}{\beta}\right)^{\alpha-1} & if \quad 0<x_1,...,x_n<\beta \\ 0 & otherwise\end{cases}$

$$=\begin{cases}\left(\frac{\alpha}{\beta}\right)^{n}\frac{1}{\beta^{n(\alpha-1)}}\prod_{i=1}^{n}x_i^{(\alpha-1)} & if \quad 0<x_{(n)}<\beta \\ 0 & otherwise\end{cases}$$

MLE of β, say $\hat{\beta}$ is the largest observation , $X_{(n)}=t$

$$L(\alpha)=\left(\frac{\alpha}{t}\right)^{n}\frac{1}{t^{n(\alpha-1)}}\prod_{i=1}^{n}x_i^{\alpha-1}$$

log $L(\alpha)= n\ log\ \alpha - n\ log\ t - n\ (\alpha\text{-}1)\ log\ t + \Sigma(\alpha\text{-}1)\ log\ x_i$

$$\frac{\partial \log L}{\partial \alpha}=\frac{n}{\alpha}-n\log\ t\ +\sum\ \log x_i=0 \qquad \Rightarrow$$

$$\hat{\alpha}=n/(n\log\ t\ -\ \sum\log\ x_i)$$

b) $\hat{\beta}=25$

$$\hat{\alpha}=\frac{14}{[14\log 25-(\log 22+\log 23.9+\log 20.9+\log 23.8+\log 25+\log 24+\log 21.7 +\log 23.8+\log 22.8+2\log 23.1+\log 23.5+\log 23.0)]}$$

$$=\frac{14}{1.1115}= \ 12.595$$

(Total 10 Marks)

15. a) The expected frequencies

Quality	Mon	Tue	Wed	Thur	Fri
			Days		
Excellent	45.6	74.4	78.6	71.4	30
Good	15.2	24.8	26.2	23.8	10
Fair	13.2	21.6	22.8	20.7	8.7
Poor	1.97	3.2	3.4	3.1	1.13

20% of all frequencies is 4.
The number of cell frequencies less than 5 and less than 1 are 5 and 0 respectively.
Hence, the criterion is not satisfied.

b) After combining we have the following table.

	Days	
	157	243
Quality	43	57

The calculated χ^2 value 0.47. The table value of χ^2 at 5% level for 1 *df* is 3.841.
Do not reject. The quality and the production days are not independent.

(Total 4 Marks)

16. $H_0 : \mu_A = \mu_B = \mu_C$ H_1 : at least one of them is not equal

Correction factor (CF): $(17.09)^2/15=19.4712$

Total sum of squares(TSS): $1.69^2+0.64^2+\ldots+0.62^2$- CF=1.679

Variety sum of squares: $\{[(5.65)^2+(6.82)^2+(4.62)^2]/5\}$- CF=0.485

Error sum of squares (ESS):1.194

ANOVA

Source	Sum of squares	df	MSS	F
Between varieties Error	0.485 1.194	2 12	.242 .100	2.435
Total	1.679	14		

Table value of $F_{2,12,.05}$=3.88. Do not Reject H_0

(Total 4 marks)

17 a) The estimated correlation coefficient

$$S_{yy} = \left[n\Sigma y^2 - (\Sigma y)^2\right]/n = 63.89; \quad S_{xy} = \left[n\Sigma xy - (\Sigma x)(\Sigma y)\right]/n = -572.439$$
$$S_{xx} = \left[n\Sigma x^2 - (\Sigma x)^2\right]/n = 7150.05$$
$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$
$$= -0.847$$

For testing the hypothesis $H_0 : \rho = 0$ against $H_1 : \rho \neq 0$, the test statistic is

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = -7.635$$

The table value of $t_{23,0.05}$ is 1.714. Reject H_0

b) The random variables $e_i^{'s}$ are independently and normally distributed

The mean of $(Y/X=x_i)$ is $\beta_0 + \beta_1 x_i$

The variance of $e_i^{'s}$ are σ^2

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$
$$= -0.08$$
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}\bar{x} = 13.64$$

c) The unbiased estimate of σ^2 is $s^2 = \frac{1}{n-2}[S_{yy} - \frac{S_{xy}^{\ 2}}{S_{xx}}]$

$$= \frac{18.09}{(25-2)} = 0.79$$

d) $t = \frac{\hat{\beta}_1}{s/\sqrt{S_{xx}}} \sim t_{23}(0.05)$

= -7.62

The table value of $t_{23}(0.05)$ is 1.714 .

Hence, Reject H_0

(The candidate can choose other levels also and give their decisions)

e) The 90% confidence interval for β_0 is

$$\left[\hat{\beta}_0 - t_{0.05} s\sqrt{\Sigma x^2} / \sqrt{nS_{xx}}, \hat{\beta}_0 + t_{0.05} s\sqrt{\Sigma x^2} / \sqrt{nS_{xx}}\right]$$

= [12.64, 14.64)]

The 99% confidence interval for β_1 is

$$\left[\hat{\beta}_1 - t_{0.005}\, s\sqrt{S_{xx}}, \hat{\beta}_1 + t_{0.005}\, s\sqrt{S_{xx}}\right]$$

= [-0.109, - 0.051]

f) $R^2 = \frac{S_{xy}^{\ 2}}{S_{xx}S_{yy}}$

$= 0.717$

(Total 17 Marks)

[Total 100 Marks]

*********//END//*********