Institute of Actuaries of India

Subject CT3-Probability and Mathematical Statistics

October/Nov 2007 Examination

INDICATIVE SOLUTION

1. The *mgf* of Bernoulli distribution is

   $M_X(t) = E[e^{tX}]$

   $= \sum_{x=0,1} e^{tx} p^x q^{1-x}$; $0<p<1$

   $= (q+pe^t)$

   Hence, $M_{X_1+X_2+...+X_n}(t) = M_{X_1}(t) M_{X2_1}(t) ... M_{Xn}(t)$ (Since X_i are *iid*)

   $= (q+pe^t)^n$, which is the *mgf* of Binomial distribution.

   The sum of *n iid* Bernoulli random variables follows Binomial distribution.

   (Proving by other methods also is to be given marks)

   Total [2]

2. $P\left(A\overline{B} \cup \overline{A}B\right) = P\left(A\overline{B}\right) + P\left(\overline{A}B\right)$,Since they are disjoint

   $= P(A)P\left(\overline{B}\right) + P\left(\overline{A}\right)P(B)$, Since A and B are independent

   $= 1/2 \; 1/2 + 1/2 \; 1/2$

   $= 1/2$

   Total [3]

3. Let X_i denote the weight of i th container and let $S = X_1+X_2+\ldots+X_{25}$

   Given that $X \sim N(150,15^2)$

   Then, $E(S) = 25E(X_i)$ (since $X_i^{'}s$ are *iid*)

   = 25 x 150 = 3750 Kgs.

   $Var(S)$ = 25Var(X_1) = 25 x 15^2 = 5625

   Let $Z = \dfrac{S - E(S)}{\sigma_s} = \dfrac{S - 3750}{75} \Rightarrow S = 3750 + 75Z$

   If p is the probability of overloading the truck, then

   $p = P(S > 4000) = P(3750+75Z > 4000)$

   $= P(Z > 3.333)$

   $= 0.5 - P(0 < Z < 3.333)$

   $= 0.5 - 0.4996$

   $= 0.0004$

   On an average, the truck shall be overloaded 4 times in 10,000 occasions.

   Total [3]

4. Given that

$P(A) = 4/9$, $P(B) = 2/9$, $P(C) = 3/9$.

Let E denote the event of introducing new insurance schemes.

It is also given that

$$\left.\begin{aligned} P(E/A) &= 0.3 \\ P(E/B) &= 0.5 \\ P(E/C) &= 0.8 \end{aligned}\right\}$$

Hence, $P(E) = P(A)P(E/A) + P(B)P(E/B) + P(C)\,P(E/C)$

$$= \frac{4}{9}.\frac{3}{10} + \frac{2}{9}.\frac{5}{10} + \frac{3}{9}.\frac{8}{10} = \frac{46}{90}$$

Using Baye's theorem

$$P(A/E) = \frac{P(A).P(E/A)}{P(E)}$$

$$= \frac{\frac{4}{9} \times \frac{3}{10}}{\frac{46}{90}} = \frac{12}{46} = 0.26086$$

Total [4]

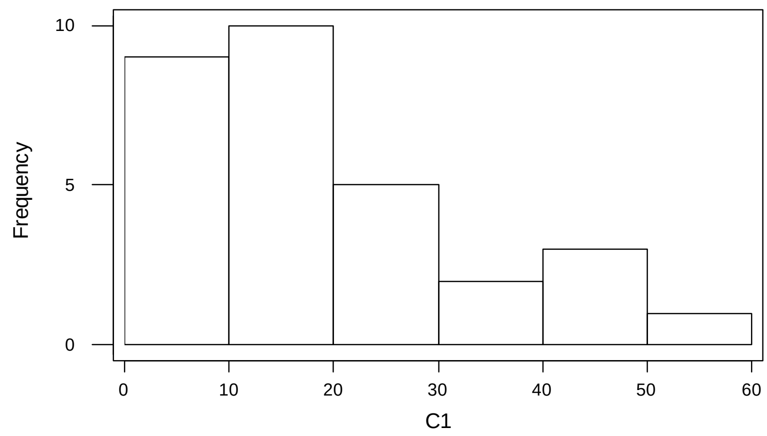
5. a)

Stem	Leaf									
0	1	2	3	4	6	7	7	8	8	
1	0	0	2	3	3	3	4	4	6	9
2	1	4	5	6		8				
3		4	6							
4		1	5	7						
5			0							
6										
7										
8										
9										

b)

CI	*frequency*
$0 < x \le 10$	9
$10 < x \le 20$	10
$20 < x \le 30$	5
$30 < x \le 40$	2
$40 < x \le 50$	3
$50 < x \le 60$	1

	N = 30



It is a positively skewed distribution.

Total [6]

6.a) $f(x) = kxe^{-x/2} \quad ; \quad x > 0$

$$\int_0^\infty f(x)dx = 1 \Rightarrow \int_0^\infty kxe^{-x/2}\,dx = 1$$

$$\Rightarrow k\int_0^\infty 2ue^{-u}\,2du = 1 \qquad \text{if } x/2 = u$$

$$\Rightarrow 4k = 1 \;\text{ since } \int_0^\infty e^{-u}udu = 1$$

$$\Rightarrow k = 1/4$$

b) $M_x(t) = 1/4\int_0^\infty xe^{tx}e^{-x/2}dx$

$$= \frac{1}{4}\int_0^\infty xe^{-x(1/2-t)}dx$$

$$= \frac{1}{4}\int_0^\infty \frac{u}{(1/2-t)^2}e^{-u}du \qquad \textit{if } x(1/2\text{-}t)=u$$

$$= \frac{1}{(1-2t)^2}$$

$$C_x(t) = \log M_x(t) = -2\log(1-2t)$$

c) $EX = C_x'(t)|_{t=0} = 4$

$V(X) = C_x''(t)|_{t=0} = 8$

Total [7]

7. Given that $Y = \log X \sim N(10,4)$ (Logarithm to base e)

a) $f_X(x) = \frac{1}{2x\sqrt{2\pi}} e^{\frac{1}{2}\left[\frac{\log X - 10}{2}\right]^2} \quad ; x > 0$

b) $E(X) = e^{\mu_y + \frac{1}{2}\sigma_Y^2}$

$= e^{10+\frac{1}{2}2(4)} = e^{12} \underset{\sim}{} 162.754$

$V(X) = e^{2\mu_Y + \sigma_Y^2}\left(e^{\sigma_Y^2} - 1\right)$

$= \mu_X^2\left(e^{\sigma_Y^2} - 1\right)$

$= e^{24}\left(e^4 - 1\right) \underset{\sim}{} 53.598e^{24}$

c) $P[X \le 1000] = P[\log X \le \log 1000]$

$= P[Y \le \log 1000]$

$= P\left[Z \le \frac{\log 1000 - 10}{2}\right] ; Z \sim N(0,1)$ (or) $1.419\, \chi_{10}^{.12}$

$= P(Z \le -1.55)$

$= 0.0611$

Total [7]

8.a) H_0 : Payment being good or delinquent is independent of person's income.
From the contingency table, the expected frequencies are

$$\left.\begin{array}{ll} E(45) = 40 & E(50) = 56 \\ E(65) = 64 & E(5) = 10 \\ E(20) = 14 & E(15) = 16 \end{array}\right\}$$

[1]

As $\chi^2 = \sum \frac{(0 - E)^2}{E}$

$= 6.417$

[1]

Critical value of $\chi^2_{0.05}(2) = 5.99$

Reject H_0.

b) H_0 Payment type is independent on Income level.
The contingency table is

	Income level		Total
	Not High	High	
Good	95	65	160
Delinquent	25	15	40

Proceeding as above the calculated value of $\chi^2 = 0.130$
[1]

Critical value of χ^2 at 0.05 level is 3.84 for 1 df.

Reject H_0.

Total [7]

9. a) A counting process *{N(t) ; t ≥ 0}* is said to be a Poisson process if the following conditions are satisfied

i) *N(t)* is independent of the number of occurrences in an interval prior to the interval *(0,t)*

ii) $P_n(t)$ depends only on the length of the interval and is independent of where this interval is situated.

iii) In the interval of infinitesimal length *h*, the probability of exactly one occurrence is $\tau h + o(h)$ (τ is constant) and that of more than occurrence is of *o(h)*.

b) Measuring time *t* in hours from 9.00 a.m. it is asked to determine $P[X(1/2)=1,\ X(5/2)=5]$

Using the independence of *X(*5/2) – *X*(1/2) and *X*(1/2), the question is reformulated as:

$P[X(1/2)=1,\ X(5/2)=5] = P[X(1/2)=1, X(51/2)-X(1/4)=4]$

$$= \left\{\frac{e^{-4(1/2)}4(1/2)}{1!}\right\}\left\{\frac{e^{-4(2)}(4(2))^4}{4!}\right\}$$

$$= \left(2e^{-2}\right)\quad\left(e^{-8}8^4/4!\right)$$

$$= \left(2e^{-2}\right)\quad\left(\frac{512}{3}e^{-8}\right)$$

$$= 0.0154965$$

Total [7]

10. a) $H_0 : \mu_1 = \mu_2 = \mu_3$

H_1 : At least $\mu_i^{'s}$ are not equal for one *i*.

From the data, the ANOVA table is

	SS	Df	MSS	*F*
Treatment	516	2	258.0	8.990
Error	430	15	28.7	
Total	946	17	-	

Calculated value of *F* value at (2,15) df is 8.990
Critical value *F* at (2,15) df at 0.05 level is 3.6823
Reject H_0.

b) Confidence interval is $\bar{x} \pm t_{\alpha/2}\ s/\sqrt{n}$

s being the estimate of population s.d. σ. In this ANOVA table, the best estimate of σ is square root of MSE which is equal to $\sqrt{28.67} = 5.354$

For plant at *X,* the mean test score = 79

Hence, 95% confidence interval is

$$= 79 \pm 2.131\left(\frac{5.354}{\sqrt{6}}\right)$$
$$79 \pm .4.66$$
$$= (74.34,\ 83.66)$$

Note that $t_{\alpha/2} = 2.131$ for 15 df. and $\frac{\alpha}{2} = 0.025$

Total [8]

11. a) The marginal density of X_1 is

$$= f_1(x_1) = \int_0^2 \left(x_1^2 + \frac{x_1 x_2}{3}\right) dx_2$$

$$= \begin{cases} 2x_1^2 + \frac{2}{3}x_1, & ;\ 0 < x_1 \le 1 \\ 0 \ \ elsewhere \end{cases}$$

The conditional density of X_2 given X_1

$$f_{X_2/X_1}(x_2) = \frac{\left(x_1^2 + \frac{x_1 x_2}{3}\right)}{\left(2x_1^2 + \frac{2}{3}x_1\right)}$$

$$= \begin{cases} \frac{1}{2}\left(\frac{3x_1 + x_2}{3x_1 + 1}\right) & ;\ 0 < x_1 \le 1\ ; 0 \le x_2 \le 2 \\ 0 \ \ elsewhere \end{cases}$$

b) Now $E(X_2 / X_1 = x_1) = \int_0^2 x_2 \frac{1}{2}\left(\frac{3x_1 + x_2}{3x_1 + 1}\right) dx_2$

$$= \frac{9x_1 + 4}{9x_1 + 3}$$

c) $E[E(X_2 / X_1 = x_1)] = \int_0^1 \left(\frac{9x_1 + 4}{9x_1 + 3}\right)\left(2x_1^2 + \frac{2}{3}x_1\right) dx_1 = \frac{10}{9}$

The marginal density of X_2 is

$$f_2(x_2) = \int_0^1 \left(x_1 + \frac{x_1 x_2}{3}\right) dx_1 = \frac{1}{3} + \frac{x_2}{6}\ ;\ 0 < x_2 \le 2$$

0 *elsewhere*

Hence, $E(X_2) = \int_0^2 x_2 \left(\frac{1}{3} + \frac{x_2}{6}\right) dx_2 = \frac{10}{9}$

Total [7]

12. a) $E\left(\frac{1}{X}\right) = \int_0^\infty \frac{1}{x} \frac{\theta^m x^{m-1} e^{-\theta x}}{(m-1)!} dx$

$= \frac{\theta^m}{(m-1)!} \int_0^\infty x^{m-2} e^{-\theta x} dx$

$= \frac{\theta^m}{(m-1)!} \int_0^\infty \left(\frac{u}{\theta}\right)^{m-2} e^{-u} \frac{du}{\theta} \quad if \;\; \theta x = u$

$= \frac{\theta}{m-1}$

[2]

Hence $E\left(\frac{m-1}{X}\right) = \theta$

$\Rightarrow \frac{m-1}{X}$ is an unbiased estimate of θ.

(Marks may be awarded if the candidate adopts alternative methods for the answer)

b) Given that $f(x,\theta) = (1+\theta)\, x^{\theta} \;\; ; \;\; 0 < x < 1$

$L(\theta; \underline{x}) = L = (1+\theta)^n \Pi x_i^{\theta}$

$Log\, L = n \log (1+\theta) + \theta (\Sigma \log x_i)$

$\frac{\partial \log L}{\partial \theta} = \frac{n}{1+\theta} + \Sigma \log x_i$

$\Rightarrow \hat{\theta} = \frac{-n}{\Sigma \log x_i} - 1$

Further,

$\frac{\partial^2 \log L}{\partial \theta^2} = -\frac{n}{(1+\theta)^2} < 0$ for all θ and hence at $\hat{\theta} = \frac{-n}{\Sigma \log x_i} - 1$

Hence $\hat{\theta}$ is the m.l.e. of θ

c) $f(x,\theta) = \theta e^{-\theta x}; 0 < x < \infty$

$\frac{\partial}{\partial \theta} \log f(x,\theta) = \frac{\partial}{\partial \theta}\left[\log \theta - \theta x\right]$

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

and so $E_\theta \left[\frac{\partial}{\partial \theta} \log f(x,\theta)\right]^2 = E_\theta \left(\frac{1}{\theta} - X\right)^2$

$= \text{Var X} = \frac{1}{\theta^2}$

Hence, the CR lower bound for the variance of any unbiased estimator of θ is $\frac{\theta^2}{n}$

Total [10]

13.a) $H_0 : \sigma_1^2 = \sigma_2^2 \quad v_s \quad H_1 : \sigma_1^2 \neq \sigma_2^2$

The test statistic is

$F = \frac{s_2^2}{s_1^2} = \frac{0.9604}{0.7225} = 1.329$

Critical value of F for (14,11) df at 5% level = 2.7316

Do not Reject H_0.

b) $H_0 : \mu_1 = \mu_2$; $H_1 : \mu_1 \neq \mu_2$

The pooled estimate of common variance σ^2 is $s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1+n_2-2)}$

$$= \frac{11(0.85)^2 + 14(0.98)^2}{12+15-2} = 0.8557$$

Hence, $s_p = 0.925$

The test statistic is

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

$$= \frac{|24.6-22.1|}{0.925\sqrt{\frac{1}{12}+\frac{1}{15}}} = \frac{2.5}{0.925\sqrt{0.0833+0.0667}}$$

$$= \frac{2.5}{0.925\sqrt{0.15}} = \frac{2.5}{0.925(0.3873)} = \frac{2.5}{0.3583} = 1.70874$$

The table value of t at 5% level for 25 df is : 2.06

Reject H_0.

c) The 95% confidence interval for μ_1-μ_2 is $\left(\bar{x}_1 - \bar{x}_2\right) \pm t_{\alpha/2, n_1+n_2-2} s_s \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$

$$= \left[\left(24.6-22.1\right) \pm 2.06(0.925)\sqrt{\frac{1}{12}+\frac{1}{15}}\right]$$

=(-3.24,-1.76) or (1,76,3.24) for (μ_2-μ_1)

d) 95% confidence interval for σ_1^2 / σ_2^2

$$\frac{s_1^2}{s_2^2} F_{0.025}(14,11) \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{s_1^2}{s_2^2} F_{0.975}(14,11)$$

Substituting the values of s_1^2, s_2^2 and the table values, $F_{0.025}(14,11)=0.3205$ and $F_{0.975}(14,11)=3.43$ we have the 95% CI as $0.3205(1/1.329) \leq \frac{\sigma_1^2}{\sigma_2^2} \leq 3.43(1/1.329)$ which results in $\{ 0.2411 \leq \frac{\sigma_1^2}{\sigma_2^2} \leq 2.581 \}$.

(use the fact that $F_{0.025}(14,11) = \dfrac{1}{F_{0.975}(11,14)} = 0.3205$)

Total [11]

14.a)

```
n       90+                                                 *
o         -
          -                                            *
o         -
f         -                                       *
        75+                                  *
s         -
e         -                             *
t         -                        *
t         -
l       60+                   *
e         -
m         -              *
e         -         *
n         -
t       45+    *
s         -
           ------+---------+---------+---------+---------+---------+
                100       120       140       160       180       200
                              no. of claims
```

The scatter diagram indicates that there is a strong relationship between the number of claims and the number of settlements and assumption of the straight line model $Y=\beta_0+\beta_1 X+\epsilon$ appears to be reasonable.

b) From the data, we have

$n = 10, \quad \Sigma X_i = 1450$

$\Sigma Y_i = 673, \ \overline{X} = 145, \ \overline{Y} = 67.3$

$\Sigma X_i^2 = 218500 \qquad \Sigma Y_i^2 = 47225$

$\Sigma XY = 101570$

$$S_{XX} = \Sigma X_i - \frac{1}{10}(\Sigma X_i)^2$$

$$= 218500 - \frac{(1450)^2}{10} = 8250$$

$$S_{YY} = \Sigma Y_i^2 - n\overline{Y}^2$$

$$= 47225 - 10\,(67.3)^2$$

$$= 1932.10$$

$$S_{XY} = \Sigma X_i Y_i - \frac{1}{10}(\Sigma X_i)(\Sigma Y_i)$$

$$= 101570 - \frac{(1450)(673)}{10} = 3985$$

Therefore $\hat{\beta} = \frac{S_{XY}}{S_{XX}} = \frac{3985}{8250} = 0.483$

$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$

$= 67.3 – (0.483)145$

$= - 2.739$

c) $\hat{\sigma}^2 = \frac{SSE}{n-2}$

$SSE = S_{YY} - S_2^{XY} / S_{XX}$

$$=1932.10 - \frac{3985^2}{8250}$$

$= 1932.10 – 1924.87$

$= 7.23$

Hence $\sigma^2 = \frac{SSE}{n-2} = \frac{7.23}{8} = 0.90$

d) $H_0 : \beta = 0$; $H_1 : \beta \neq 0$. The following ANOVA table is needed.

Source of variation	S.S.	df	MSS	F
Regression Error	1924.9 7.2	1 8	1924.9 0.9	2131.57
Total	1932.10	9		

Table *F(1,8)* at 0.05 level of significance is 5.32
at 0.01 level of significance is 11.30

Reject H_0

NOTE: This test can also be carried out using a t-test. The calculated value for t is the square root of F which is 46.17, at 8 degrees of freedom. The table value is 2.306. We reject H_0.

e) 95% confidence for β is

$$\hat{\beta} + t_{0.025}(8)\sqrt{\frac{MSE}{S_{XX}}} \leq \beta \leq \hat{\beta} + t_{0.975}(8)\sqrt{\frac{MSE}{S_{XX}}}$$

$$= 0.483 - 2.306\sqrt{\frac{0.90}{8250}} \leq \beta \leq 0.483 + 2.306\sqrt{\frac{0.90}{8250}}$$

$= 0.459 \leq \beta \leq 0.507.$

f) $H_0 : \rho = 0$; $H_1 : \rho \neq 0$

$$r(x,y) = \frac{S_{XY}}{(S_{XX}S_{YY})^{1/2}} = \frac{3985}{\sqrt{8250}\sqrt{1932.10}} = 0.99814$$

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

$$= \frac{0.99814 \times \sqrt{8}}{\sqrt{0.00372}} = 46.288$$

Calculated value of t = 46.288

Table value of *t at 5% level* with 8 degrees of freedom is 2.306.

Reject H_0.

Total [17]

NOTE: There is a printing error in Q14(f) which might have been rectified later. The hypothesis to be tested is $H_0 : \rho = 0$ against $H_1 : \rho \neq 0$. However, if any candidate has tested for $H_0 : \rho = 0.75$ against $H_1 : \rho \neq 0.75$ using Z-transformation, the marks can be awarded according to the correctness of the computation.

****