Institute of Actuaries of India

Subject CS1-Actuarial Statistics
(Paper A)

November 2019 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 points as valid answers and examiner have given credit for any alternative approach or interpretation which they consider to be reasonable.

Solution 1:

i. MGF of X:

$M_X(t) = E[e^{tX}] = \sum_{x=1}^{\infty} e^{tx} P(X = x) = \sum_{x=1}^{\infty} e^{tx} p(1-\mathrm{p})^{\mathrm{x}-1}$ [1]

$M_X(t) = p\mathrm{e}^{\mathrm{t}} + p(1-\mathrm{p})\mathrm{e}^{2\mathrm{t}} + p(1-\mathrm{p})^3\mathrm{e}^{3\mathrm{t}} + \cdots$ [1]

This is infinite Geometric Series with summation given as:

$M_X(t) = {p\mathrm{e}^{\mathrm{t}}}/{(1-(1-p)\mathrm{e}^{\mathrm{t}})}$ [1]

CGF is hence given by:

$C_X(t) = \ln(M_X(t)) = ln\left({p\mathrm{e}^{\mathrm{t}}}/{(1-(1-p)\mathrm{e}^{\mathrm{t}})}\right)$ [1]

[4]

ii. Determining E(X)

Alternative solution 1: $M'_X t \ at \ t = 0$	Alternative solution 2: $C'_X t \ at \ t = 0$
$M'_X(t) = {p\mathrm{e}^{\mathrm{t}}}/{(1-(1-p)\mathrm{e}^{\mathrm{t}})} + {p\mathrm{e}^{\mathrm{t}}.(1-p)\mathrm{e}^{\mathrm{t}}}/{(1-(1-p)\mathrm{e}^{\mathrm{t}})^2}$ At t=0, $M'_X(t) = {p}/{(p)} + {p.(1-p)}/{(p)^2} = 1/p$	$C_X(t) = \ln(M_X t) = ln\left({p\mathrm{e}^{\mathrm{t}}}/{(1-(1-p)\mathrm{e}^{\mathrm{t}})}\right)$ $C_X(t) = \ln(p) + t - \ln(1-(1-p)\mathrm{e}^{\mathrm{t}})$ $C'_X(t) = 1 - {(-1)(1-p)\mathrm{e}^{\mathrm{t}}}/{(1-(1-p)\mathrm{e}^{\mathrm{t}})}$ At t=0 $C'_X(t) = 1 + {1-p}/{p} = 1/p$

[2]

[6 marks]

Solution 2:

i) Variable t_k is defined as $t_k \Rightarrow \frac{N(0,1)}{\sqrt{\chi_k^2/k}}$ [1]

where k denotes the degrees of freedom [0.5]
and the two random variables $N(0,1)$ and χ_k^2 are independent. [0.5]

[2]

ii) Mean and variance of t_k for k>2 are 0 [0.5]
and k/(k-2) respectively. [0.5]

[1]

iii)

a) We know that for a sample from a normal population,

$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$ [0.5]

For the given confidence level $t_9 = 3.25$, [0.5]

Confidence interval for μ is thus $\left(50 - 3.25 \times \sqrt{\frac{48.667}{10}}, 50 + 3.25 \times \sqrt{\frac{48.667}{10}}\right)$

i.e. $(42.83\ ,\ 57.17)$ [1]

b)

From (ii) above, we know that $t_{n-1} \sim N\left(0, \sqrt{\frac{n-1}{n-3}}\right) \sim N(0\ , \sqrt{\frac{9}{7}})$ [1]

i.e.

$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim N(0\ , \sqrt{\frac{9}{7}})$ [0.5]

$\frac{\bar{X}-\mu}{S/\sqrt{7n/9}} \sim N(0\ ,1)$ [0.5]

Critical value for given level of confidence is 2.58 [0.5]

Confidence interval for μ is thus $\left(50 - 2.58 \times \sqrt{\frac{48.667}{10} \times \frac{9}{7}}, 50 + 2.58 \times \sqrt{\frac{48.667}{10} \times \frac{9}{7}}\right)$

[0.5]

i.e. $(43.54\ ,\ 56.45)$ [1]

[6]

[9 Marks]

Solution 3:

i) The likelihood is given by

$L(\theta) = \frac{\theta^{x_1}}{x_1!} e^{-\theta} \times \frac{\theta^{x_2}}{x_2!} e^{-\theta} \ \dots \times \frac{\theta^{n}}{n!} e^{-\theta}$ [1]

$\propto \theta^{\sum x_i} . e^{-n\theta}$ [0.5]

The prior distribution $\theta \sim Gamma(\alpha\ , \lambda)$ is given by

$f(\theta) = \frac{\lambda^{\alpha}}{\Gamma(\alpha\ )} \theta^{(\alpha\ -1)} e^{-\lambda\theta} \ \propto \theta^{(\alpha\ -1)} e^{-\lambda\theta}$ [1]

The posterior distribution is given by:

$posterior\ \propto prior\ \times likelihood$ [0.5]

$= \theta^{(\alpha\ -1)} e^{-\lambda\theta} \ \times \theta^{\sum x_i} . e^{-n\theta}$ [0.5]

$= \theta^{\sum x_i + (\alpha\ -1)} e^{-(\lambda+n)\theta}$ [0.5]

[4]

ii) Bayesian estimate of θ under quadratic loss is the mean of the posterior distribution.

[1]

The posterior distribution in part (i) is of the form

$Gamma(\alpha + \sum x_i\ , \lambda + n)$ [1]

The mean of above distribution is given by $\frac{(\alpha+\sum x_i)}{(\lambda+n)}$

i.e. $\tilde{\theta} = \frac{(\alpha+\sum x_i)}{(\lambda+n)}$ [1]

[3]

iii) From part (ii) we have $\tilde{\theta} = \frac{(\alpha+\sum x_i)}{(\lambda+n)}$

$= \frac{(\alpha)}{(\lambda+n)} + \frac{(\sum x_i)}{(\lambda+n)}$ [1]

$= \left(\frac{\alpha}{\lambda}\right)\left(\frac{\lambda}{\lambda+n}\right) + \left(\frac{\sum x_i}{n}\right)\left(\frac{n}{\lambda+n}\right)$ [1]

$= (mean\ of\ prior\ distribution) \times (1 - Z) + (sample\ mean) \times (Z)$

Where $Z = \frac{n}{\lambda+n}$ $hence\ (1 - Z) = \frac{\lambda}{\lambda+n}$ [1]

[3]

[10 Marks]

Solution 4:

i) The prior probability is the probability assessed by the investment department, 80% [1]

ii) The conditional probability = P (CS = H| FS = H) = 0.75 [2]

iii) The posterior probability can be computed as:

P (FS = H| CS = H) = P(FS=H ∩ CS=H)$P(FS = H|CS = H) =$

$$\frac{P(FS=\ \ )*P(CS=H|FS=H)}{P(FS=L)*P(CS = H|FS = L)+P(FS=H)*P(CS=H|FS=H)} = \frac{0.8*0.75}{0.2*0.4+0.8*0.75} = 0.88$$ [3]

[6 Marks]

Solution 5:

i) *Note: Alternate solution is possible depending on the assumption whether you get the head on m^{th} flip or $m+1^{st}$ flip.*

Let X be the number of times the coin needs to be flipped before getting heads for the first time. Then X|p is a Type 1 Geometric distribution with parameter p. [1]

The prior distribution of p can be assumed to be uniform over the interval [0, 1]. $f_{prior}(p) = 1$, $0 <= p <= 1$ [0.5]

Likelihood function: L (p) = P(X = m) = $(1-p)^{m-1}.p$ [0.5]

Alternate:
Likelihood function: L (p) = P(X = m) = $(1-p)^{m}.p$
Posterior distribution of p is proportional to $f_{prior}(p).L(p)$ [0.5]
Therefore posterior distribution of p is proportional to $(1-p)^{m-1}.p$ [0.5]

Alternate:
Therefore posterior distribution of p is proportional to $(1-p)^{m}.p$
Posterior distribution of p is Beta (2, m) [1]

Alternate:
Posterior distribution of p is Beta (2,m+1) [1]

[4]

ii) Again, X|p is a Type 1 Geometric distribution with parameter p
With an added observation the likelihood function changes to:
L (p) = $(1-p)^{m+n-2}.p^2$
Alternate: L (p) = $(1-p)^{m+n}.p^2$

[1.5]

The prior distribution of p can be assumed to be uniform over the interval [0, 1]. $f_{prior}(p) = 1$, $0 <= p <= 1$ [0.5]

Posterior distribution of p is proportional to $f_{prior}(p).L(p)$ [0.5]

Therefore posterior distribution of p is proportional to (1-p) $^{m+n-2}$.p^2

Alternate: Therefore posterior distribution of p is proportional to (1-p) $^{m+n}$.p^2

[0.5]

Posterior distribution of p is Beta (3, m+n-1)

Alternate: Posterior distribution of p is Beta (3, m+n+1) [1]

[4]

[8 Marks]

Solution 6:

i) The likelihood function is given by

$$L(p) = C[(1-p)^4]^{70}.[p(1-p)^3]^{120}.[p^2(1-p)^2]^{201}.[p^3(1-p)^1]^{80}.[(p)^4]^{29}$$

[1]

$L(p) = C[p^{878}(1-p)^{1122}]$ [1]

Taking logs and differentiating w.r.t p:

$$\frac{\mathrm{d}}{\mathrm{dp}}\ln\big(L(p)\big) = \frac{\mathrm{d}}{\mathrm{dp}}(\ln(C) + 878\ln(p) + 1122\ln(1-p)) = \frac{878}{p} - \frac{1122}{1-p}$$

[1]

Equating to zero

$\frac{878}{p} = \frac{1122}{1-p}$ $or\ 878 = p(1122+878) or\ p = 0.439$ [1]

Checking for maximum:

$\frac{\mathrm{d}^2}{\mathrm{dp}^2}\ln\big(L(p)\big) = -\frac{878}{p^2} - \frac{1122}{(1-p)^2} < 0 \Rightarrow Maximum$ [1]

[5]

ii)

a) $\widehat{\theta_1} = \frac{70+12}{70+120+201+80+} = 0.3800$ [1]

And $\widehat{\theta_2} = \frac{40+100}{40+100+160+170+7} = 0.2593$ [1]

b) For samples from independent binomial distributions, we know that

$$\frac{(\widehat{\theta_1} - \widehat{\theta_2}) - (\theta_1 - \theta_2)}{\sqrt{\frac{\widehat{\theta_1}(1-\widehat{\theta_1})}{n_1} + \frac{\widehat{\theta_2}(1-\widehat{\theta_2})}{n_2}}} \sim N(0,1)$$

[1]

Substituting values of $\widehat{p_1}\ and\ \widehat{p_2}$ in the above equation,

$\frac{(0.1207)-(\theta_1-\theta_2)}{0.02876} \sim N(0,1)$ [1]

Critical value for given confidence level is 1.96. [0.5]

Hence the confidence interval is calculated as:

$(0.1207 - 1.96 \times 0.02876\ ,\ 0.1207 + 1.96 \times 0.02876\ )$ [0.5]

i.e. (0.0643, 0.1771) [1]

c) As the interval does not contain zero, there is significant difference in the proportion of cars plying with less than 2 passengers in both the cities. [1]

[7]

[12 Marks]

Solution 7:

i) Type I error - Event of Rejecting the hypothesis when it is true [1]

Let X be the random variable denoting the total number of claims on the portfolio. X thus follows Poi(nμ) i.e. Poi(3000) where μ is the Poisson parameter.

Null hypothesis H_0 is thus $X{\sim}Poi(3000)$ [0.5]

$P(reject\ H_0\ when\ H_0\ is\ true) = P(X > 3100\ when\ X{\sim}Poi(3000))$ [1]

Using normal approximation (as nλ is large enough) [0.5]

$$X{\sim}N(3000{,}3000)$$

$P(X > 3100) = P\left(Z > \frac{3100-300}{\sqrt{3000}}\right) = 1 - P(Z < 1.825) = 3.39\%$ [1]

[4]

ii) Type II error - Event of Accepting the hypothesis when it is false [1]

iii) Power of a test - Probability of Rejecting the hypothesis when it is false [1]

In terms of μ it is given by:

$P(reject\ H_0\ when\ H_0\ is\ false) = P(X > 3100\ when\ X{\sim}Poi(n\mu){\sim}N(n\mu, n\mu))$

$P(X > 3100) = 1 - \left(Z < \frac{3100-n\mu}{\sqrt{n\mu}}\right)$ [0.5]

The value of power of test will depend on the value of parameter under alternate hypothesis. [0.5]

[2]

iv) If $\hat{\mu}$ is the estimator of μ (the poisson parameter), $\hat{\mu}$ follows $N(\mu, \hat{\mu}/n)$ [0.5]

Hence $\frac{\hat{\mu}-\mu}{\sqrt{\hat{\mu}/n}}$ follows N(0,1) [0.5]

Or $P\left(-2.5758 < \frac{2.9-\mu}{\sqrt{2.9/1000}} < 2.5758\right) = 0.99$ [0.5]

Hence the confidence interval for μ is (2.7613, 3.0387) [0.5]

[2]

[9 Marks]

Solution 8:

i) X and Y can take whole values in the sample space (1,6) and (0,2) respectively. The joint mass function of X and Y is as follows:

	X=1	X=2	X=3	X=4	X=5	X=6
Y=0	(1/6).(5/6)	(1/6).(5/6)	(1/6).0	(1/6).(5/6)	(1/6).(5/6)	(1/6).(5/6)
Y=1	(1/6).(1/6)	(1/6).(1/6)	(1/6).(5/6)	(1/6).(1/6)	(1/6).(1/6)	(1/6).(1/6)
Y=2	(1/6).0	(1/6).0	(1/6).(1/6)	(1/6).0	(1/6).0	(1/6).0

OR

	X=1	X=2	X=3	X=4	X=5	X=6
Y=0	5/36	5/36	0	5/36	5/36	5/36
Y=1	1/36	1/36	5/36	1/36	1/36	1/36
Y=2	0	0	1/36	0	0	0

(0.25 marks for each non zero values,1 marks for all zero values)

[4]

ii) $Var(X+Y) = var(X) + var(Y) + 2cov(X+Y)$

$Var(X+Y) = E(X^2) - E(X)^2 + E(Y^2) - E(Y)^2 + 2\big(E(XY) - E(X)E(Y)\big)$ [1]

$E(XY)$ is calculated by taking sum of all the values in following table. Each entry is $x.y.P(X=x).P(Y=y)$:

	X=1	X=2	X=3	X=4	X=5	X=6
Y=0	0	0	0	0	0	0
Y=1	1/36	2/36	15/36	4/36	5/36	6/36
Y=2	0	0	6/36	0	0	0

Hence $E(XY) = \frac{39}{36}$ [4]

$Var(X+Y) = \frac{91}{6} - \left(\frac{21}{6}\right)^2 + \frac{14}{36} - \left(\frac{1}{3}\right)^2 + 2\left(\frac{39}{36} - \frac{21}{6}.\frac{1}{3}\right) = 3.027$ [1]

[6]

[10 Marks]

Solution 9:

i) The least square estimates of the regression coefficients are the values of α & β for which:

$q = \sum_{i=1}^{n} e_i^2$ [1]

$= \sum_{i=1}^{n} [y_i - (\alpha + \beta.x_i)]^2$ [1]

is a minimum

[2]

ii) The least squares estimate of the regression coefficient γ is the value of γ for which:

$q = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} [y_i - \gamma.e^{x_i}]^2$ [0.5]

is a minimum.

$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(y_i^2 - 2.\gamma e^{x_i}.y_i + \gamma^2 e^{2x_i})$ [1]

In order to find the minimum value, differentiate with respect to γ and set it equal to zero:

$\frac{dq}{d\gamma} = 2\gamma \sum_{i=1}^{n} e^{2x_i} - 2.\sum_{i=1}^{n} y_i.e^{x_i} = 0$ [1]

Therefore, $\hat{\gamma} = \frac{\sum_{i=1}^{n} y_i.e^{x_i}}{\sum_{i=1}^{n} e^{2x_i}}$ [1]

$\left(\frac{d^2q}{d\gamma^2} = 2\sum_{i=1}^{n} e^{2x_i} > 0, therefore\ minimum\right)$ [0.5]

[4]

iii) The least squares estimate of the regression coefficient γ is the value of γ for which:

$q = \sum_{i=1}^{n} w_i * e_i^2 = \sum_{i=1}^{n} \frac{[y_i - \gamma.e^{x_i}]^2}{x_i}$ [1]

is a minimum.

$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \frac{(y_i^2 - 2.\gamma e^{x_i}.y_i + \gamma^2 e^{2x_i})}{x_i}$ [1]

In order to find the minimum value, differentiate with respect to γ and set it equal to zero:

$\frac{dq}{d\gamma} = 2\gamma \sum_{i=1}^{n} \frac{e^{2x_i}}{x_i} - 2.\sum_{i=1}^{n} \frac{y_i.e^{x_i}}{x_i} = 0$ [1]

Therefore, $\hat{\gamma} = \frac{\sum_{i=1}^{n}\left(y_i.e^{x_i}/x_i\right)}{\sum_{i=1}^{n}\left(e^{2x_i}/x_i\right)}$ [0.5]

$\left(\frac{d^2q}{d\gamma^2} = 2\sum_{i=1}^{n} e^{2x_i}/x_i > 0, therefore\ minimum\right)$ [0.5]

[4]

[10 Marks]

Solution 10:

i) $S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 4789.42$ [0.5]

$S_{xy} = 2176.84$ [0.5]

$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = \frac{2176.84}{4789.42} = 0.4545$ [1]

$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = 17.895 - 0.4545 * 15.83 = 10.79$ [0.5]

The fitted regression equation is $\hat{y} = 10.79 + 0.4545 * x$ [0.5]

[3]

ii) $S_{xx} = 4789.42$ from result of part i

$S_{yy} = \sum y^2 - \frac{(\sum y)^2}{n} = 1189.21$ [0.5]

$S_{xy} = 2176.84$ from result of part i

$\hat{\sigma}^2 = \frac{1}{n-2}\left(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right) = \frac{1}{9} * \left(1189.21 - \frac{2176.84^2}{4789.42}\right) = 22.20$ [2]

$s.e.(\hat{\beta}) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} = \sqrt{\frac{22.20}{4789.42}} = 0.0681$ [1]

To test $H_o: \beta = 0 \ v \ H_1: \beta \neq 0, the\ test\ statistic\ is$

$\frac{\hat{\beta}-0}{s.e.(\hat{\beta})} = \frac{0.4545}{0.0681} = 6.674$ [1]

Under the assumption that the errors of the regression are i.i.d $N(0, \sigma^2)$ random variables, beta has a t distribution with n-2 degrees of freedom. [0.5]

Critical value for t-distribution with 9 degrees of freedom: $t_{9,0.025}$ = 2.68. [0.5]

Since the critical value at 95% level of significance is less than the test statistic, there is sufficient evidence to reject the null hypothesis. Hence, it cannot be concluded that there is no statistically significant relationship between x and y. [0.5]

[6]

iii) Pearson's correlation coefficient is computed as: $\frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$ [0.5]

S_{yy} = 1189.21

$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{2176.84}{\sqrt{4789.42 * 1189.21}} = 0.91$ [1.5]

[2]

iv) The estimated value of y corresponding to $x^3 = 25\ is\ 10.79 + 0.4545 * 25 = 22.15$ [1]

$\hat{\sigma}^2 = 22.20$ … from earlier workings

The variance of the estimator of the mean response is given by
$\left[\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}\right]\hat{\sigma}^2 = \left[\frac{1}{11} + \frac{84.0889}{4789.42}\right] * 22.20 = 2.41$ [2]

The variance of the estimator of the individual response is given by

$\left[1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}\right]\hat{\sigma}^2 = [1 + 0.1085] * 22.20 = 24.61$ [2]

Using t_9 distribution, the 95% confidence intervals for mean and individual responses are:

22.20 $\pm$ 2.262 $*$sqrt(2.41) and 22.20 $\pm$ 2.262 $*$sqrt(24.61) = (18.7,25.7) & (11.0,33.4) [2]

[7]

v) The residual plot shows a definite pattern. Although the correlation coefficient is high, the model does not seem to be appropriate. [1]
Using this model leads to underestimation of premium rates at low and high mortality ratings. [1]

[2]
[20 Marks]

****