Institute of Actuaries of India

Subject CT6 – Statistical Methods

November 2010 Examinations

INDICATIVE SOLUTIONS

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.

Question 1

(i) The deductible D satisfies the equation

$$1-\left(\frac{\lambda}{\lambda+D^{\gamma}}\right)^{\alpha}=0.25,$$

i.e., $1-\left(\frac{1000}{1000+D^{0.75}}\right)^{2}=0.25$.

Thus,

$$\left(\frac{1000}{1000+D^{0.75}}\right)^{2}=0.75.$$

$$\frac{1000}{1000+D^{0.75}}=\sqrt{0.75}=0.8660.$$

$$1+\frac{D^{0.75}}{1000}=\frac{1}{0.8660}.$$

$$\frac{D^{0.75}}{1000}=\frac{1}{0.8660}-1=0.1547.$$

$$D=154.7^{1/0.75}=830.5.$$

(ii) The density of the losses resulting in claims is the following truncated version of the Burr density.

$$\frac{f(x)}{1-F(D)}=\frac{\dfrac{\alpha\gamma\lambda^{\alpha}x^{\gamma-1}}{\left(\lambda+x^{\gamma}\right)^{\alpha+1}}}{\left(\dfrac{\lambda}{\lambda+D^{\gamma}}\right)^{\alpha}}=\frac{\alpha\gamma\left(\lambda+D^{\gamma}\right)^{\alpha}x^{\gamma-1}}{\left(\lambda+x^{\gamma}\right)^{\alpha+1}},\ x>D.$$

The probability density function of the claims (net of deductible) is the density of $Y=X-D$, where X has the above density. Thus,

$$f_Y(y)=\frac{\alpha\gamma\left(\lambda+D^{\gamma}\right)^{\alpha}(y+D)^{\gamma-1}}{\left\{\lambda+(y+D)^{\gamma}\right\}^{\alpha+1}},\ y>0.$$

If the claims data (net of deductible) are $Y_1, Y_2, \ldots, Y_{1250}$, then the likelihood function to be maximized with respect to the parameters α, γ and λ is

$$L(\alpha,\gamma,\lambda)=\prod_{i=1}^{1250}f_Y(Y_i)=\prod_{i=1}^{1250}\frac{\alpha\gamma\left(\lambda+D^{\gamma}\right)^{\alpha}(Y_i+D)^{\gamma-1}}{\left\{\lambda+(Y_i+D)^{\gamma}\right\}^{\alpha+1}}.$$

(iii) The MLE of the true fraction of the losses that result in no claim to the insurer is

$$F(D)\big|_{\alpha=\hat{\alpha},\gamma=\hat{\gamma},\lambda=\hat{\lambda}}=1-\left(\frac{\hat{\lambda}}{\hat{\lambda}+D^{\hat{\gamma}}}\right)^{\hat{\alpha}},$$

where $\hat{\alpha}$, $\hat{\gamma}$ and $\hat{\lambda}$ are the respective MLE's of the parameters α, γ and λ, obtained by maximizing the likelihood described in part (ii), and $D=830.5$, as determined from part (i).

[9]

Question 2

(i) Let $p=\dfrac{\theta}{10}$. Then the prior mean is $E(\theta)=10E(p)$, where the density of p is $\dfrac{(n_1+n_2+1)!\,p^{n_1}(1-p)^{n_2}}{n_1!n_2!}$.

$$E(p)=\int_0^1 p\times\frac{(n_1+n_2+1)!\,p^{n_1}(1-p)^{n_2}}{n_1!n_2!}\,dp$$

$$=\int_0^1\frac{n_1+1}{n_1+n_2+2}\times\frac{(n_1+n_2+2)!\,p^{n_1+1}(1-p)^{n_2}}{(n_1+1)!\,n_2!}\,dp=\frac{n_1+1}{n_1+n_2+2}.$$

The integral is evaluated by making use of the fact that $\dfrac{(n_1+n_2+2)!\,p^{n_1+1}(1-p)^{n_2}}{(n_1+1)!\,n_2!}$ is a probability density function (similar to the above density of p, but n_1 is replaced by n_1+1).

Thus, the prior mean is $10E(p)=\dfrac{10(n_1+1)}{n_1+n_2+2}$.

(ii) The likelihood is $\displaystyle\prod_{i=1}^{3}\binom{10}{X_i}\left(\frac{\theta}{10}\right)^{X_i}\left(1-\frac{\theta}{10}\right)^{10-X_i}$.

The log-likelihood is a constant plus

$$(X_1+X_2+X_3)\ln\left(\frac{\theta}{10}\right)+(30-X_1-X_2-X_3)\ln\left(1-\frac{\theta}{10}\right).$$

The MLE of θ, obtained by maximizing the log-likelihood, is $\dfrac{X_1+X_2+X_3}{3}=\overline{X}$.

(iii) The Bayes estimator of θ under the squared error loss function is the mean of the posterior distribution.

The prior density of θ is $\dfrac{1}{10}\times\dfrac{(n_1+n_2+1)!}{n_1!n_2!}\times\left(\dfrac{\theta}{10}\right)^{n_1}\left(1-\dfrac{\theta}{10}\right)^{n_2}$.

The likelihood is as given in part (ii).

Therefore, the posterior density of p is proportional to

$$\frac{1}{10}\times\frac{(n_1+n_2+1)!}{n_1!n_2!}\times\left(\frac{\theta}{10}\right)^{n_1}\left(1-\frac{\theta}{10}\right)^{n_2}\times\prod_{i=1}^{3}\binom{10}{X_i}\left(\frac{\theta}{10}\right)^{X_i}\left(1-\frac{\theta}{10}\right)^{10-X_i},$$

i.e., proportional to $\left(\dfrac{\theta}{10}\right)^{n_1+X_1+X_2+X_3}\left(1-\dfrac{\theta}{10}\right)^{n_2+30-X_1-X_2-X_3}$.

Comparison with the prior density reveals that the posterior density is

$$\frac{1}{10}\times\frac{(n_1+n_2+31)!\left(\frac{\theta}{10}\right)^{n_1+X_1+X_2+X_3}\left(1-\frac{\theta}{10}\right)^{n_2+30-X_1-X_2-X_3}}{(n_1+X_1+X_2+X_3)!\,(n_2+30-X_1-X_2-X_3)!},$$

which is similar to the prior density, but has different parameters. By comparing this density with the prior and the prior mean computed in part (i), we conclude that the posterior mean is $\dfrac{10(n_1 + X_1 + X_2 + X_3 + 1)}{n_1 + n_2 + 32}$.

(iv) The posterior mean can be written as

$$\frac{30}{n_1 + n_2 + 32} \times \frac{X_1 + X_2 + X_3}{3} + \frac{n_1 + n_2 + 2}{n_1 + n_2 + 32} \times \frac{10(n_1 + 1)}{n_1 + n_2 + 2},$$

which is of the form $Z \times \overline{X} + (1 - Z) \times E(\theta)$, with credibility factor $Z = \dfrac{30}{n_1 + n_2 + 32}$.

(v) $m(\theta) = E(X|\theta) = \theta = 10p$. Note that

$$E(p) = \frac{n_1 + 1}{n_1 + n_2 + 2},$$

$$E(p^2) = \int_0^1 p^2 \times \frac{(n_1 + n_2 + 1)!}{n_1! n_2!} p^{n_1} (1-p)^{n_2} dp$$

$$= \int_0^1 \frac{(n_1 + 2)(n_1 + 1)}{(n_1 + n_2 + 3)(n_1 + n_2 + 2)} \times \frac{(n_1 + n_2 + 3)!}{(n_1 + 2)! n_2!} p^{n_1 + 2} (1-p)^{n_2} dp = \frac{(n_1 + 2)(n_1 + 1)}{(n_1 + n_2 + 3)(n_1 + n_2 + 2)}.$$

Hence, $V(p) = E(p^2) - [E(p)]^2 = \dfrac{(n_1 + 2)(n_1 + 1)}{(n_1 + n_2 + 3)(n_1 + n_2 + 2)} - \dfrac{(n_1 + 1)^2}{(n_1 + n_2 + 2)^2}$

$$= \frac{(n_1 + 1)}{(n_1 + n_2 + 2)} \times \frac{(n_1 + 2)(n_1 + n_2 + 2) - (n_1 + 1)(n_1 + n_2 + 3)}{(n_1 + n_2 + 2)(n_1 + n_2 + 3)} = \frac{(n_1 + 1)(n_2 + 1)}{(n_1 + n_2 + 2)^2 (n_1 + n_2 + 3)}.$$

It follows that

$$E(m(\theta)) = \frac{10(n_1 + 1)}{n_1 + n_2 + 2}, \; V(m(\theta)) = \frac{100(n_1 + 1)(n_2 + 1)}{(n_1 + n_2 + 2)^2 (n_1 + n_2 + 3)}.$$

Further, $s^2(\theta) = V(X \mid \theta) = \theta\left(1 - \dfrac{\theta}{10}\right)$, and

$$E(s^2(\theta)) = 10 \times \frac{n_1 + 1}{n_1 + n_2 + 2} - 10 \times \frac{(n_1 + 2)(n_1 + 1)}{(n_1 + n_2 + 3)(n_1 + n_2 + 2)} = 10 \times \frac{(n_1 + 1)(n_2 + 1)}{(n_1 + n_2 + 2)(n_1 + n_2 + 3)}.$$

The EBCT Model 1 credibility factor is

$$Z = \frac{n}{n + \dfrac{E(s^2(\theta))}{V(m(\theta))}} = \frac{3}{3 + \dfrac{\dfrac{10(n_1 + 1)(n_2 + 1)}{(n_1 + n_2 + 2)(n_1 + n_2 + 3)}}{\dfrac{100(n_1 + 1)(n_2 + 1)}{(n_1 + n_2 + 2)^2 (n_1 + n_2 + 3)}}} = \frac{3}{3 + \dfrac{(n_1 + n_2 + 2)}{10}} = \frac{30}{n_1 + n_2 + 32},$$

and the corresponding credibility estimate of θ is

$$\frac{30}{n_1 + n_2 + 32} \times \frac{X_1 + X_2 + X_3}{3} + \frac{n_1 + n_2 + 2}{n_1 + n_2 + 32} \times \frac{10(n_1 + 1)}{n_1 + n_2 + 2},$$ same as in part (iii).

[16]

Question 3

(i) $E(S)=50\times(0.25\times 2500+0.5\times 5000+0.25\times 7500)=250{,}000$.
$V(S)=50\times(0.25\times 2500^2+0.5\times 5000^2+0.25\times 7500^2)=37{,}500^2$.

(ii) We have to find U such that

$$P(U+(1+0.1)E(S)<S)=0.05,$$

i.e., $P\left(\dfrac{S-E(S)}{\sqrt{V(S)}}>\dfrac{U+(1+0.1)E(S)-E(S)}{\sqrt{V(S)}}\right)=0.05$,

i.e., $\dfrac{U+(1+0.1)E(S)-E(S)}{\sqrt{V(S)}}=1.645$,

i.e., $\dfrac{U+25{,}000}{37{,}500}=1.645$, i.e., $U=36{,}682$.

(iii) We have to find U_R such that

$$P(U_R+(1+0.1)E(S)-(1+0.15)\times 0.25E(S)<(1-0.25)S)=0.05,$$

i.e., $P\left(\dfrac{0.75S-0.75E(S)}{0.75\sqrt{V(S)}}>\dfrac{U_R+(1+0.1)E(S)-(1+0.15)\times 0.25E(S)-0.75E(S)}{0.75\sqrt{V(S)}}\right)=0.05$,

i.e., $\dfrac{U_R+(1+0.1)E(S)-(1+0.15)\times 0.25E(S)-0.75E(S)}{0.75\sqrt{V(S)}}=1.645$,

i.e., $\dfrac{U_R+15{,}625}{28{,}125}=1.645$, i.e., $U=30{,}637$.

[9]

Question 4

(i) The Development factors are
$f_{01}=\dfrac{2749+3278+3716}{2463+3013+3321}=1.1075$, $f_{12}=\dfrac{3529+3608}{2749+3278}=1.1842$, $f_{23}=\dfrac{3980}{3529}=1.1278$.

Calculation of reserve for the year 2009:
The cumulative development factor applicable to the year 2009 is
$f=f_{01}\times f_{12}\times f_{23}=1.4791$.
$1-1/f=0.3239$.
The earned premium is Rs. 6,472,000 (given).
The assumed ultimate loss ratio is 0.88.
The emerging liability is $0.3239\times 6{,}472{,}000\times 0.88=1{,}844{,}000$.
The reported liability is Rs. 3,953,000 (from table).
The Ultimate liability is Rs. 1,844,000 + Rs. 3,953,000 = Rs. 5798,000.
Paid claims are Rs. 1,731,000 (given).
Reserve needed = Rs. 5,798,000 – Rs. 1,731,000 = Rs. 4,066,000.

(ii) The assumptions are:

- Payments from each accident year develop in the same way, i.e., for each accident year, the amount of claims paid in each development year is a constant proportion of the total claims paid from that accident year.
- Weighted average of the past inflation would be repeated in the future (this assumption holds trivially if the rate of inflation is constant).
- The estimated loss ratio is appropriate.

[11]

Question 5

(i) The characteristic polynomial is $(1-\alpha B^2)$, and the magnitude of its roots is $1/\sqrt{|\alpha|}$. For stationarity of the process, these roots should have magnitude greater than 1, which can happen if and only if $|\alpha|<1$.

(ii) $(1-\alpha B^2)Y_t = Z_t$.

Hence, $Y_t = (1-\alpha B^2)^{-1} Z_t = \{1+\alpha B^2 + (\alpha B^2)^2 + (\alpha B^2)^3 + \cdots\} Z_t = \sum_{j=0}^{\infty} \alpha^j Z_{t-2j}$,

i.e., $a_j = \begin{cases} \alpha^{j/2}, & \textit{for } j \textit{ even,} \\ 0, & \textit{for } j \textit{ odd.} \end{cases}$

(iii) $V(Y_t) = V\left(\sum_{j=0}^{\infty} \alpha^j Z_{t-2j}\right) = \sum_{j=0}^{\infty} V(\alpha^j Z_{t-2j}) = \sum_{j=0}^{\infty} \alpha^{2j} V(Z_{t-2j}) = \sum_{j=0}^{\infty} \alpha^{2j}\sigma^2 = \frac{\sigma^2}{1-\alpha^2}$.

(iv) The above expression is identical with that of the variance of an AR(1) process with parameter α.
This is more than a coincidence. The odd and even samples of the given time series separate into two completely independent sequences, each being an AR(1) process with parameter α.

[9]

Question 6

(i) The portmanteau test statistic computed from m sample ACF values, $r_1,\ldots,r_m$, is $n(n+2)\sum_{k=1}^{m}\frac{r_k^2}{n-k}$, n being the sample size. If the time series indeed comes from white noise, then this statistic has the chi-square distribution with m degrees of freedom. This holds for any fixed m.

In this case, $n = 500$, and we can choose $m = 10$.
The portmanteau test statistic turns out to be 922.4.
This is a very large value in relation to the null distribution of χ^2_{10}, indicating rejection of the null hypothesis of white noise.
[*A properly justified use of the portmanteau test for smaller values of m should also fetch full credit. The correct value of the statistic for m = 9,…,1 are 908.8, 884.4, 846.5, 801.0, 749.9, 690.9, 615.5, 498.0 and 305.5, respectively. All the tests are highly significant.*]

(ii) The ACF sequence becomes zero after lag q in the case of an MA(q) time series. Likewise, the PACF sequence becomes zero after lag p in the case of an AR(p) process. One can check whether *sample* ACF or PACF samples are significantly different from 0, by comparing their absolute value with the threshold $2/\sqrt{n}$.
In the present case, the threshold is 0.0894.

The sample ACF sequence has many values which are larger than this value in magnitude. Therefore, an MA(q) model does not appear to be appropriate.
On the other hand, the sample PACF sequence has smaller absolute values after lag 1. Therefore, an AR(1) model is indicated.

[9]

Question 7
Typical perils are:

- Fire
- Explosion
- Lightning
- Theft
- Storm
- Flood
- Earthquake **[3]**

[*Any six from the above list will suffice. Credit should be given for other reasonable answers.*]

Question 8

(i) The completed table is as follows (the index i indicates the age group and has values from 1 to 10; the index j indicates the occupation category and has values from 1 to 6).

Model	Linear predictor	No of parameters	Scaled deviance
SA	$\alpha + \beta x$	2	238.4
SA + AG	$\alpha + \beta x + \gamma_i$	11	206.7
SA + AG + SA * AG	$\alpha_i + \beta_i x$	20	178.3
SA * AG + OC	$\alpha_i + \beta_i x + \theta_j$	25	166.2
SA * AG * OC	$\alpha_{ij} + \beta_{ij} x$	120	58.9

(ii) The differences in deviance and degree of freedom in models of successively higher complexity are as follows.

Model	Degree of freedom	Scaled deviance	Change in degree of freedom	Change in scaled deviance
SA	2	238.4		
SA + AG	11	206.7	9	31.7
SA + AG + SA * AG	20	178.3	9	28.4
SA * AG + OC	25	166.2	5	12.1
SA * AG * OC	120	58.9	95	107.3

Going by the thumb-rule reduction threshold of two per degree of freedom, each of the successive levels of model sophistication, except for the last one, is justified. Therefore, the chosen model is SA * AG + OC.

(iii) One needs to analyse the residuals in order to check for violation of model assumptions before making recommendation about the appropriate choice of the model. A goodness of fit test may also be conducted.

[9]

Question 9

(i) The output X is equal to zero if and only if the first generated value of Z happens to be greater than or equal to 1. The probability of this even is

$P(X=0)=P(Z\geq 1)=P(-\ln Y\geq \lambda)=P(Y\leq e^{-\lambda})=e^{-\lambda}$,
which is the correct value of probability of a sample from the Poisson distribution (with mean λ) being equal to 0.

(ii) The successively assigned values of different variables are as follows.

Y	$\frac{-\ln Y}{\lambda}$	Z	X	Output
0.564	0.2864	0.2864	0	-
0.505	0.3416	0.6279	1	-
0.756	0.1399	0.7678	2	-
0.610	0.2471	1.0150	3	$X=2$
0.046	1.5395	1.5395	0	$X=0$

Thus, the generated values are 3 and 0.

(iii) The following table gives some values of the probability function and the cumulative distribution function for the Poisson distribution with mean $\lambda=2$.

x	$P(X=x)$	$P(X\leq x)$
0	0.1353	0.1353
1	0.2707	0.4060
2	0.2707	0.6767
3	0.1804	0.8571
4	0.0902	0.9473
5	0.0361	0.9834
6	0.0120	0.9955

Accordingly, *X* would be assigned values according to the following rule.

Range of generated sample from $U(0,1)$	Assigned value of *X*
(0, 0.1353)	0
(0.1353,0.4060)	1
(0.4060,0.6767)	2
(0.6767,0.8571)	3
(0.8571,0.9473)	4
(0.9473,0.9834)	5
(0.9834,0.9955)	6

Obviously the above tables are incomplete, but these are adequate for the purpose, since the given random numbers are within the intervals listed here.
As per the above table, the five random numbers correspond to the following values of *X*: 2, 2, 3, 2, 0.

[12]

Question 10

(i) The individual risk model is a model for the risk arising from a portfolio of a fixed number of individual risks. According to this model, the aggregate claims from the portfolio is
$S=Y_1+Y_2+\cdots+Y_n$,
where *n* is number of risks, and for $i=1,2,\ldots,n$, Y_i is the claim amount under the i^{th} risk.

The assumptions underlying this model are:

- the risks are independent,
- the number of risks does not change over the period of insurance cover,
- the number of claims from each risk is either 0 or 1.

The claim numbers and claim sizes for different risks are *not* assumed to have identical distribution.

(ii) This model differs from the collective risk model in three major ways.

- In the individual risk model, the *number of risks is fixed*, and stays the same over the period of cover, whereas in the collective risk model, this number may be random, and may also vary.
- Unlike the case of the individual risk model, there is *no restriction on the number of claims* arising from individual risks under the collective risk model.
- In the individual risk model, the *individual risks* are assumed to be independent, while in the collective risk model, the *individual claim amounts* are assumed to be independent.

[8]

Question 11

(i) Nature's choices in the present case are the number of faults, θ. The losses associated with the decisions to sign the AMC (d_1) or not to sign it (d_2) are as given below.

	θ					
	0	1	• 2	3	4	5
D_1	1000	1000	1000	1000	1000	1000
D_2	0	300	600	900	1200	1500

(ii) The maximum losses under d_1 and d_2 are Rs. 1000 and Rs. 1500, respectively. The minimax decision is to sign the AMC (d_1).

(iii) The average loss under d_1 is Rs. 1000.
The average loss under d_2 is
$0.1\times 0 + 0.1\times 300 + 0.2\times 600 + 0.3\times 900 + 0.2\times 1200 + 0.1\times 1500 =$ Rs. 810.
Therefore, the Bayes' decision is not to sign the AMC (d_2).

[5]

[Total Marks 100]

****