The Institute of Actuaries of India

Subject CT6 – Statistical Methods

18^{th} 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

Q.1)

i) For a risk to be insurable:

- The policyholder must have an interest in the risk being insured, to distinguish between insurance and a wager,
- A risk must be of a financial and reasonably quantifiable nature.

ii) Ideally risk events need to meet the following criteria if they are to be insurable:

- Individual risk events should be independent of each other.
- The probability of the event should be relatively small.
- Large numbers of potentially similar risks should be pooled in order to reduce the variance and hence achieve more certainty.
- There should be an ultimate limit on the liability undertaken by the insurer.
- Moral hazards should be eliminated as far as possible because these are difficult to quantify, result in selection against the insurer and lead to unfairness in treatment between one policyholder and another.

(Full credit will be given to those who list any three of the five criteria.)

[5]

Q.2)

i) The prior density is $2p$.

The likelihood (from the geometric distribution) is $(1-p)^4 p$.

The posterior density is proportional to $p^2 (1-p)^4$.

This distribution can be easily recognized as Beta (3,5).

The mean of the distribution is 3/(3+5) or 3/8.

ii) a) $E(X_{n+1}) = E\left[E\left(X_{n+1} \mid \theta\right)\right] = E(\theta) = \mu$

b) $\text{Var}(X_{n+1}) = E\left[\text{Var}\left(X_{n+1} \mid \theta\right)\right] + \text{Var}\left[E\left(X_{n+1} \mid \theta\right)\right]$

$$= E(\theta^2) + \text{Var}(\theta) = 2\mu^2 + \mu^2 = 3\mu^2$$

c) Following a standard derivation, the linear estimator of θ which minimizes the mean squared error is $(1-Z)E(\theta) + Z\overline{X}$, where $\overline{X} = (X_1 + \ldots\ldots + X_n)/n$

and

$$Z = \frac{n}{n + E\left[\text{Var}\left(X_{n+1} \mid \theta\right)\right] / \text{Var}\left[E\left(X_{n+1} \mid \theta\right)\right]}$$

$$= \frac{n}{n + 2\mu^2/\mu^2}$$

$$= \frac{n}{n+2}$$

Thus, the estimator is $\left(2\mu + n\overline{X}\right)/(n+2)$

[12]

Q.3)
i)

a) Liability cover provides indemnity where the insured, owing to some form of negligence, is legally liable to pay compensation to a third party.

Examples of Liability cover are:

Employers' Liability, Motor third party liability, Public liability, Product liability, Professional Indemnity (any two examples required).

b) Property cover indemnifies the policyholder against loss of or damage to his/her own material property.

Examples of Property cover are:

Residential building, Moveable property, commercial building, land vehicles, marine craft, aircraft (any 2 example required)

ii) Development factors are

3,991/3,819 = 1.04504 and 7,833/5,329 = 1.46988.

1-1/f = 1-1/ (1.04504*1.46988) = 0.3490.

Year 2006: Emerging liability = 5,012 *0.85 *0.3490 = 1,487.

Reported liability = 3,217.

Ultimate liability = Reported liability + Emerging liability

= 3217+ 1487 = 4,704.

Reserve = Ultimate liability – paid claims =4,704 - 1,472 = 3,232.

[10]

Q.4) In statistical inference, decisions about populations, such as what are the mean or variance of some characteristic, are based on sample data. Statistical inference can be regarded as a game between Nature, which controls the relevant features of a population, and the statistician, who is trying to make a decision about the population.
One way in which statistical games differ from game theory is that in game theory each player chooses a strategy without knowing what the opponent will do. In a statistical game the statistician has some sample data, which will give some information about Nature's choice.

Q.5) (i) Denote:

0	just had a claim
0*	1 claim free year after accident or new customer
1	25%
2	50%

	Premiums if no claim	*Premiums if claim*	*Difference*
0	750, 562.50, 375	750, 750, 562.50	375
0*	562.50, 375, 375	750, 750, 562.50	750
1	375, 375, 375	750, 750, 562.50	937.50
2	375, 375, 375	750, 750, 562.50	937.50

So minimum claim in state 0 is 375, in state 0* is 750 and in states 1 and 2 is 937.50.

ii)

$$P(\text{Claim}) = P(\text{Claim} \mid \text{Accident}) \,.\, P(\text{Accident})$$
$$= P(X > x) \times P(\text{Accident})$$

Where X is the loss and x is the minimum loss for which a claim will be made.

$$E(x) = \exp(\mu + \tfrac{1}{2}\sigma^2) = 1{,}451$$

$$\text{Var}(x) = \exp(2(\mu + \tfrac{1}{2}\sigma^2)) \exp((\sigma^2) - 1) = 604.4^2$$

Therefore, $\exp(\sigma^2) - 1 = 604.4^2 / 1{,}451^2$

$$\exp(\sigma^2) = 1.1735$$

$$\sigma^2 = 0.16$$

$$\sigma = 0.4$$

$$\mu = 7.2$$

$$P(X > 375) = 1 - \Phi\left(\frac{\ln 375 - 7.2}{0.4}\right) = 0.99927$$

$$P(X > 750) = 1 - \Phi\left(\frac{\ln 750 - 7.2}{0.4}\right) = 0.9264$$

$$(X > 937.50) = 1 - \Phi\left(\frac{\ln 937.50 - 7.2}{0.4}\right) = 0.8138$$

So the transition matrix is

$$\begin{bmatrix} 0.2498 & 0.7502 & 0 & 0 \\ 0.2316 & 0 & 0.7684 & 0 \\ 0.1628 & 0 & 0 & 0.8372 \\ 0.0814 & 0 & 0 & 0.9186 \end{bmatrix}$$

[6]

iii) (iii)

$$0.2498\pi_0 + 0.2316\pi_0{}^* + 0.1628\pi_1 + 0.0814\pi_2 = \pi_0$$
$$0.7502\pi_0 = \pi_0{}^*$$
$$0.7684\pi_0{}^* = \pi_1$$
$$0.8372\pi_1 + 0.9186\pi_2 = \pi_2$$
$$\pi_0 + \pi_0{}^* + \pi_1 + \pi_2 = 1$$
$$\pi_1 = 0.7502 \times 0.7684 \times \pi_0 = 0.5766\pi_0$$
$$0.8372\pi_1 = \pi_2(1 - 0.9186)$$
$$\pi_2 = 10.2850\pi_1$$
$$\pi_0 + 0.7502\pi_0 + 0.5766\pi_0 + 10.2850 \times (0.5766\pi_0) = 1$$
$$8.2556\pi_0 = 1$$
$$\pi_0 = 0.1211$$
$$\pi_0{}^* = 0.0909$$
$$\pi_1 = 0.7684 \times 0.0909 = 0.0698$$
$$\pi_2 = 1 - \pi_0 - \pi_0{}^* - \pi_1 = 0.7182$$

(iv) Average premium across portfolio

$$750 \times (0.1211 + 0.0909 + 0.0698 \times 0.75 + 0.7182 \times 0.5) = \text{Rs.}467.59$$

(v) Intention is to automatically premium rate with NCD system. Small number of categories and the relatively low discount result in high proportion of policyholders in maximum discount category. Many more categories and higher discount levels would be required to correctly rate such a heterogeneous population.

Q.6)

i) Let N be the number of claims.

$$P(N = n) = \int_0^\infty P(N = n|\lambda)\, 2e^{-2\lambda} d\lambda$$

$$= \int_0^\infty \frac{e^{-\lambda}\lambda^n}{n!}\, 2e^{-2\lambda} d\lambda$$

$$= \frac{2}{3^{n+1}} \int_0^\infty \frac{3^{n+1}\, \lambda^n\, e^{-3\lambda}}{n!}\, d\lambda$$

$$= \underline{\frac{2}{3^{n+1}}} \qquad n = 0, 1, 2, \ldots$$

ii) The claim amounts, X follow Gamma Distribution with mean 2 and variance 2,

For a gamma(α,λ), $\left.\begin{array}{l}\text{mean} = \alpha/\lambda = 2\\ \text{variance} = \alpha/\lambda^2 = 2\end{array}\right\} \Rightarrow \alpha = 2 \quad \lambda = 1$

So $\quad M_X(t) = \left(\dfrac{1}{1-t}\right)^2$

The moment generating function of S is

$$\begin{aligned} M_S(t) &= E(e^{St}) = E[E(e^{St}\,|\,N)] \\ &= E[\{E(e^{Xt})\}^N] \\ &= G_N(M_X(t)) \end{aligned}$$

N has pgf $\displaystyle\sum_0^\infty s^n \frac{2}{3^{n+1}} = \frac{2}{3}\sum_0^\infty \left(\frac{s}{3}\right)^n = \frac{2}{3-s}$

So, the moment generating distribution function of the aggregate claims distribution, S will be,

$$\text{So} \quad M_S(t) = \frac{2}{3 - \dfrac{1}{(1-t)^2}} = \frac{2(1-t)^2}{3(1-t)^2 - 1} = \underline{\frac{2(1-t)^2}{3t^2 - 6t + 2}}$$

[8]

Q.7)

i) If X represents the size of an individual claim, we have

$$E(X) = \frac{\lambda k}{\alpha - 1} = 160 \qquad \text{and} \qquad \text{var}(X) = \frac{\lambda^2 k(k+\alpha-1)}{(\alpha-1)^2(\alpha-2)} = 14,400$$

Then the amount paid by the insurer is $Y = 0.8X$, and:

$$E(Y) = 0.8 \times 160 = 128 \qquad \text{and} \qquad \text{var}(Y) = 0.8^2 \times 14,400 = 9,216$$

Similarly, the amount paid by the reinsurer is $Z = 0.2X$, and:

$$E(Z) = 0.2 \times 160 = 32 \qquad \text{and} \qquad \text{var}(Z) = 0.2^2 \times 14,400 = 576$$

ii) Let α be the proportion of loss retained by the direct insurer under the proportional reinsurance arrangement.

$$\begin{aligned}
\alpha\mu &= \int_0^{\mu} \frac{x}{\mu} e^{-x/\mu} dx \\
&= \mu \left[\int_0^1 y e^{-y} dy + \int_1^{\infty} e^{-y} dy\right] \\
&= \mu \left[-y e^{-y} \Big|_0^1 + \int_0^{\infty} e^{-y} dy\right] \\
&= \mu \left[-e^{-1} + 1\right].
\end{aligned}$$

Hence, $\alpha = 1 - 1/e$.

Under proportional reinsurance, the variance of the direct insurer's share of loss is $\alpha^2 \mu^2$ or $(1 - 1/e)^2 \mu^2$.

Under excess of loss reinsurance, the second moment of the direct insurer's share of loss is

$$\begin{aligned}
&\int_0^{\mu} \frac{x^2}{\mu} e^{-x/\mu} dx + \mu^2 \int_{\mu}^{\infty} \frac{1}{\mu} e^{-x/\mu} dx \\
&= \mu^2 \left[\int_0^1 y^2 e^{-y} dy + \int_1^{\infty} e^{-y} dy\right] \\
&= \mu^2 \left[-y^2 e^{-y} \Big|_0^1 + 2\int_0^1 y e^{-y} dy + e^{-1}\right] = 2\mu^2 \int_0^1 y e^{-y} dy \\
&= 2\mu^2 \left[-y e^{-y} \Big|_0^1 + \int_0^1 e^{-y} dy\right] \\
&= 2\mu^2 \left[-e^{-1} + 1 - e^{-1}\right] = 2(1 - 2/e)\mu^2.
\end{aligned}$$

Hence, the variance is

$$\begin{aligned}
&2(1 - 2/e)\mu^2 - (1 - 1/e)^2 \mu^2 \\
&= \left[2 - \frac{4}{e} - 1 + \frac{2}{e} - \frac{1}{e^2}\right] \mu^2 \\
&= \left(1 - \frac{2}{e} - \frac{1}{e^2}\right) \mu^2 \\
&< \left(1 - \frac{2}{e} + \frac{1}{e^2}\right) \mu^2 = (1 - 1/e)^2 \mu^2.
\end{aligned}$$

Hence, the variance of the direct insurer’s share of loss will be less under the excess of loss reinsurance.

[11]

Q.8)

i) The probability density function must integrate to 1

ii) The distribution function for this distribution is:

$$F(x) = \int_{-1}^{x} \frac{1}{1+e^{-x}}\,dx = \log\left(\frac{1+e^{x}}{1+e^{-1}}\right) \quad (-1 < x < 1)$$

We can find the inverse of this function by solving the equation $F(x) = y$ for x:

$$\log\left(\frac{1+e^{x}}{1+e^{-1}}\right) = y \quad \Rightarrow x = F^{-1}(y) = \log[(1+e^{-1})e^{y} - 1]$$

The pseudo-random values can now be generated by applying the function $F^{-1}()$ to the uniform pseudo-random values given. This gives:

–0.938, 0.650, 0.786, 0.282 and –0.121

iii) Based on the statistics given, the sample mean is:

$$\bar{X} = \frac{165.681}{1000} = 0.166$$

and the sample variance is:

$$S^2 = \frac{1}{999}\left(339.275 - \frac{165.681^2}{1000}\right) = 0.312 \quad \Rightarrow S = 0.559$$

So a 95% confidence interval for the mean is:

$$0.166 \pm 1.96 \times \frac{0.559}{\sqrt{1000}} = 0.166 \pm 0.035 = (0.131, 0.201)$$

[11]

Q.9)

i) The mean is

The mean is:

$$E(X_t) = E(\varepsilon_t + 0.25\varepsilon_{t-1} + 0.5\varepsilon_{t-2} + 0.25\varepsilon_{t-3}) = 0 \qquad [1]$$

The variance is:

$$\begin{aligned}\text{var}(X_t) &= \text{var}(\varepsilon_t + 0.25\varepsilon_{t-1} + 0.5\varepsilon_{t-2} + 0.25\varepsilon_{t-3})\\ &= \text{var}(\varepsilon_t) + 0.25^2\,\text{var}(\varepsilon_{t-1}) + 0.5^2\,\text{var}(\varepsilon_{t-2}) + 0.25^2\,\text{var}(\varepsilon_{t-3})\\ &= 1.375\sigma^2 \qquad [1]\end{aligned}$$

ii) Auto correlation function

We need to find the first few autocovariances γ_0, γ_1, γ_2 and γ_3.

γ_0 is just the variance of the series, which we've already worked out *ie* $\gamma_0 = 1.375\sigma^2$.

$$\begin{aligned}\gamma_1 &= \text{cov}(X_t, X_{t-1})\\ &= \text{cov}(\varepsilon_t + 0.25\varepsilon_{t-1} + 0.5\varepsilon_{t-2} + 0.25\varepsilon_{t-3}, \varepsilon_{t-1} + 0.25\varepsilon_{t-2} + 0.5\varepsilon_{t-3} + 0.25\varepsilon_{t-4})\\ &= 0.25\sigma^2 + (0.5)(0.25)\sigma^2 + (0.25)(0.5)\sigma^2\\ &= 0.5\sigma^2\end{aligned}$$

Similarly $\gamma_2 = 0.5625\sigma^2$, $\gamma_3 = 0.25\sigma^2$ and $\gamma_k = 0$ for all $k \geq 4$.

Dividing these by γ_0 (and noting that $\rho_{-k} = \rho_k$) gives us the autocorrelations:

$$\begin{aligned}&\rho_0 = 1\\ &\rho_1 = \rho_{-1} = \tfrac{4}{11} \quad (= 0.364)\\ &\rho_2 = \rho_{-2} = \tfrac{9}{22} \quad (= 0.409)\\ &\rho_3 = \rho_{-3} = \tfrac{2}{11} \quad (= 0.182)\\ &\rho_k = \rho_{-k} = 0 \text{ for } |k| \geq 4.\end{aligned}$$

[4]

iii) Note that

$$\begin{aligned}(1 - 0.5B - 0.5B^2)Y_t &= (1 + 0.5\,B)(1 - B)\,Y_t\\ &= X_t\\ &= (1 + 0.25B + 0.5B^2 + 0.25B^3)\,\varepsilon_t\end{aligned}$$

It is found by inspection that neither (1 + 0.5 B) nor (1 – B) is a factor of

($1 + 0.25B + 0.5B^2 + 0.25B^3$). Further, the only root of (1 + 0.5 B) is larger than 1.

Clearly, $p = 1$, $d = 1$, $q = 3$, i.e., the process is ARIMA (1,1

[10]

Q.10)

i) $e^{-\lambda} = e^{-1}$

ii) $\lambda e^{-\lambda} = e^{-1}$

iii) P(ruin within first year)

P(ruin in first year) = 1 – P(no ruin in first year)

= 1 – P(no ruin | 0 claim in first year) x P(0 claim)

– P(no ruin | 1 claim in first year) x P(1 claim)

$= 1 - 1.e^{-1} - \text{P(Claim after 8/12 year) } e^{-1} = 1 - e^{-1}(1 + e^{-2/3})$.

[5]

Q.11)

(i) We need to express the density function in the form $\exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right\}$.

We can write the density function as:

$$f(y) = \exp\left\{-\frac{y}{\mu} - \log\mu\right\}$$

So if we take:

$\theta = -1/\mu$ $\qquad b(\theta) = -\log(-\theta)$

$\phi = 1$ $\qquad a(\phi) = 1$ $\qquad c(y,\phi) = 0$

Then the density function will have an appropriate form.

(The choices $\theta = 1/\mu$, $b(\theta) = -\log(\theta)$ and $a(\phi) = -1$ should also fetch full credit.)

ii) The variance function is $b''(\theta)$. Differentiating $b(\theta)$ twice, we find that

$$b''(\theta) = 1/\theta^2 = \mu^2.$$

So the variance function is μ^2.

The variance is $a(\phi)\, b''(\theta) = \mu^2$.

iii) The density of Y is

$$e^{-(y-\alpha)}\, I(y > \alpha).$$

This expression cannot be written in the exponential family form, since the parameter- and data-dependent parts cannot be separated by factoring.

[10]

****