Institute of Actuaries of India

Subject CT6 – Statistical Methods

May 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) Inflation adjusted chain ladder method

Assumptions:

- Undiscounted Claim reserves are required (i.e. express answers in terms of amount payable)
- A constant proportion of the total claim amount, in real terms, arising from each origin year is paid is each development area
- Claim inflation in the past are as shown. There is a period of 12 months up to the middle of the given year.
- Payments are assumed to be made in the middle of each year.

Claim payments:

Year of Accident	Claim payment in the year of development			
	1	2	3	4
2006	141	94	54	17
2007	137	97	58	
2008	139	101		
2009	143			

Converting claim payments to year 2009 values:

Year of Accident	Claim payment			
	1	2	3	4
2006	**172.4**	**108.2**	**58.1**	17
2007	**157.7**	**104.4**	58	
2008	**149.6**	101		
2009	143			

where 141 * 1.062 * 1.07 * 1.076 = 172.4,

94 * 1.07 * 1.076 = 108.2 , etc

Cumulative claims:

Year of Accident	Cumulative Claims			
	1	2	3	4
2006	172.4	280.6	338.7	355.7
2007	157.7	262.1	320.1	
2008	149.6	250.6		
2009	143.0			

Calculating the development factors

Development factor for DY4: = 355.7 / 338.7 = 1.050188

Development factor for DY3: (338.7+320.1)/(280.6+262.1) = 1.213927

Development factor for DY2: (280.6+262.1+250.6)/(172.4+157.7+149.6) = 1.653739

Completing the run-off triangle

Year of Accident	Cumulative Claims			
	1	2	3	4
2006	172.4	280.6	338.7	355.7
2007	157.7	262.1	320.1	336.2
2008	149.6	250.6	304.2	319.4
2009	143.0	236.5	287.1	301.5

Dis-accumulating

Year of Accident	Claim amount			
	1	2	3	4
2006				
2007				16.1
2008			53.6	15.3
2009		93.5	50.6	14.4

Using future inflation rate of 7% pa

Year of Accident	Claim amount			
	1	2	3	4
2006				
2007				17.2
2008			57.4	17.5
2009		100.0	57.9	17.6

Outstanding claim reserve = 17.2 + 57.4 + 17.5 + 100.0 + 57.9 + 17.6 = 267.6 (thousand).

(ii) Inflation adjusted average cost per claim method

Assumptions:

- Undiscounted claim reserves are required (i.e. express answers in terms of amounts payable).
- The average amount of claim payments (in real terms) is a constant for each development year.
- A constant proportion of the total number of claims from each origin year are settled in each development year.

Inflation adjusted claim amounts:(2009 values)

Year of Accident	Claim payment (in Rs.'000)			
	1	2	3	4
2006	**172.4**	**108.2**	**58.1**	17
2007	**157.7**	**104.4**	58	
2008	**149.6**	101		
2009	143			

Number of claims:

Year of Accident	No. of claims settled in the year of development			
	1	2	3	4
2006	240	139	67	11
2007	248	145	71	
2008	251	141		
2009	246			

Average claim amounts: (claim amounts in 2009 values divided by no. of claims)

Year of Accident	Inflation adjusted average claim amount (in Rs.'000)			
	1	2	3	4
2006	0.7183	0.7786	0.8672	1.5455
2007	0.6360	0.7198	0.8169	
2008	0.5959	0.7163		
2009	0.5813			
Average	0.6329	0.7382	0.8421	0.15455

Applying Basic chain ladder method to number of claims

Required development factors are

Development factor for DY4: = 11/67 = 0.164179.

Development factor for DY3: (67+71) / (139+145) = 0.485915.

Development factor for DY2: (139+145+141) / (240+248+251) = 0.575101.

Projections of Number of claims

Year of Accident	Number of claims and projected number of claims in the year of development			
	1	2	3	4
2006	240	139	67	11
2007	248	145	71	**11.7**
2008	251	141	**68.5**	**11.2**
2009	246	**141.5**	**68.7**	**11.3**

Projected claim size (projected claim number times average claim size)

Year of Accident	Projected claim amount (in Rs.'000)			
	1	2	3	4
2006				
2007				18.0
2008			57.7	17.4
2009		104.4	57.9	17.4

Projected claim size after incorporating future inflation rate of 7% pa

Year of Accident	Inflation adjusted claim amount (in Rs.'000)			
	1	2	3	4
2006				
2007				19.3
2008			61.7	19.9
2009		111.8	66.3	21.4

Outstanding claim reserve (in Rs.'000)
= 19.3 + 61.7 + 19.9 + 111.8 + 66.3 + 21.4 = 300.3.

[17]

Question 2

(i) Truly random numbers are generated from physical processes. This necessitates a hardware attachment to the computer, which is inconvenient.

Another drawback is that these numbers cannot be reproduced through the same process.

For ease of reproduction, these numbers can be stored in a table for quick access. However, the table may be too short for practical use, particularly for large scale simulations.

(ii) (a) We need to find a density *f* and a constant *C* such that $g(x) \le Cf(x)$ for all $x \ge 0$.

Since $\frac{1}{(1+x)^2} \le 1$ for all $x \ge 0$, we have $g(x) = \frac{ke^{-3x}}{(1+x)^2} \le ke^{-3x}$ for all $x \ge 0$. Therefore, we can use $f(x) = 3e^{-3x}$, which is the density of the exponential distribution with parameter 3, and the constant $C = k/3$.

The distribution function *F* corresponding to the density *f* is given by $F(x) = 1 - e^{-3x}$. This function has an explicit inverse, i.e., $u = F(x)$ if and only if $x = -\frac{\ln(1-u)}{3}$. Therefore, a pseudo-random sample *X* from this distribution can be obtained by generating a sample *U* from the uniform distribution over the interval [0,1], and setting $X = -\frac{\ln(1-U)}{3}$.

The threshold for acceptance is $\frac{g(x)}{Cf(x)} = \frac{1}{(1+x)^2}$.

The procedure for generating a sample from *g* is:

1. Generate a pseudo-random sample X from the distribution $F(x)=1-e^{-3x}$, as indicated above.

2. Generate another pseudo-random sample U from the uniform distribution over the interval [0,1].

3. If $U<\dfrac{g(X)}{Cf(X)}$, i.e., if $U<\dfrac{1}{(1+X)^2}$, then accept the value X; otherwise reject it and return to step 1.

(b) Step 3 is similar to the tossing of a coin until Head turns up. The number of tosses required for a Head is a random variable with geometric distribution. The expected number is $1/p$, where p is probability of Head in each toss. In general,

$$p=P\left(U<\frac{g(X)}{Cf(X)}\right)=E\left[P\left(U<\frac{g(X)}{Cf(X)}\middle|X\right)\right]=E\left[\frac{g(X)}{Cf(X)}\right]=\int_0^\infty\frac{g(x)}{Cf(x)}f(x)dx=\frac{1}{C}.$$ The

refore, the expected number of iterations needed to generate a single sample is

$$\frac{1}{p}=C=\frac{k}{3}.$$

[Full credit should be given if a candidate gives the correct number of iterations without deriving the result.]

Each such iteration requires two uniform pseudo-random variables. So the answer is $\dfrac{2k}{3}$.

[9]

Question 3

(i)

(a) A Covariate is a variable in a linear model about which we have some information. We try to predict the values of the response variable by looking at the values taken by the covariates. Covariates may take numerical values or categorical values.

(b) A Linear Predictor is a function of the covariates that is used in the model. It is the linear function of the parameters, whose value we try to estimate. It may not necessarily be a linear function of the covariates themselves.

(c) A Link Function is the function that relates the response variable to a linear function involving the covariates, in which case the link function will be the identity function. In more usual examples we try to express some function of the response variable in linear form. The function applied to the response variable is called the link function. In many cases there is a natural form for the link function to take. This is called the canonical link function.

(ii)

(a) $f(x,\mu)=\dfrac{e^{-\mu}\mu^x}{x!}$.

A distribution for a random variable X belongs to an exponential family if its density function can be express in the following form:

$$f(x,\theta,\phi)=\exp\left[\frac{x\theta-b(\theta)}{a(\phi)}+c(x,\phi)\right]$$

$f(x, \mu)$ can be written as

$f(x,\mu) = \exp[x \ln \mu - \mu - \ln(x!)]$.

By comparing the expressions of $f(x,\mu)$ and $f(x,\theta,\varphi)$, we find that the latter is a special case of exponential family, with

$\theta = \ln \mu,$

$b(\theta) = \mu,$ i.e., $b(\theta) = e^{\theta},$

$\phi = 1,$

$a(\phi) = 1,$

$c(x,\phi) = -\ln(x!).$

(b) The canonical link function is defined by $g(\mu) = \theta$. Since $\theta = \ln \mu$, we have $g(\mu) = \ln \mu$.

(c) The likelihood is

$$f(\alpha,\beta) = \prod_{i=1}^{3} \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} \text{ with } g(\mu_i) = \eta_i = \alpha + \beta x_i.$$

The log-likelihood is

$$\ln f(\alpha,\beta) = \left[-\sum_{i=1}^{3} \mu_i + \sum_{i=1}^{3} y_i \ln \mu_i - \sum_{i=1}^{3} \ln(y_i!) \right]\Bigg|_{\ln(\mu_i)=\alpha+\beta x_i}$$

$$= -\sum_{i=1}^{3} \exp(\alpha + \beta x_i) + \sum_{i=1}^{3} y_i(\alpha + \beta x_i) - \sum_{i=1}^{3} \ln(y_i!).$$

Therefore, the likelihood equations are

$$\frac{\partial \ln f(\alpha + \beta x_i)}{\partial \alpha} = -\sum_{i=1}^{3} \exp(\alpha + \beta x_i) + \sum_{i=1}^{3} y_i = 0 \text{ and}$$

$$\frac{\partial \ln f(\alpha + \beta x_i)}{\partial \beta} = -\sum_{i=1}^{3} x_i \exp(\alpha + \beta x_i) + \sum_{i=1}^{3} x_i y_i = 0,$$

i.e., $e^{\alpha+2\beta} + e^{\alpha+4\beta} + e^{\alpha+6\beta} = 15$ and $2e^{\alpha+2\beta} + 4e^{\alpha+4\beta} + 6e^{\alpha+6\beta} = 72.$

Let $a = e^{\alpha+2\beta}$ and $b = e^{2\beta}$. Then the likelihood equations are

$a(1 + b + b^2) = 15$ and $2a(1 + 2b + 3b^2) = 72.$

After dividing the second equation by the first, one gets $\dfrac{2(1 + 2b + 3b^2)}{1 + b + b^2} = \dfrac{72}{15},$

i.e., $5 + 10b + 15b^2 = 12 + 12b + 12b^2$, i.e., $3b^2 - 2b - 7 = 0$. The only positive solution to this quadratic equation is $b = \dfrac{2 + \sqrt{4 + 4 \times 3 \times 7}}{6}. = \dfrac{\sqrt{22} + 1}{3} = 1.8968$. It follows that $a = \dfrac{15}{1 + b + b^2} = 2.3096.$

From the values of a and b, one can obtain the maximum likelihood estimates of α and β as $\hat{\alpha} = \ln(a/b) = 0.1969$; $\hat{\beta} = \ln(b)/2 = 0.3201$.

[12]

Question 4

	1/3	1/3	1/3			
	θ1	θ2	θ3	min	max	Expected profit
d1	25	19	7	7	25	**17**
d2	10	30	8	**8**	30	16
d3	0	2	34	0	**34**	12

(i) Minimax means minimizing the maximum loss, which is the same as maximizing the minimum profit. Hence, the minimax solution is d_2.

(ii) The Sales Director would choose d_3.

(iii) The Bayes criterion solution is d_1.

(iv) Let $P(\theta_1) = p$ and $P(\theta_3) = q$. Here, the assumption is that $p + q < 1/3$.

Expected profit of d_2 – Expected profit of d_1

$= 10p + 30(1 - p - q) + 8q - 25p - 19(1 - p - q) - 7q$

$= 11 - 26p - 10q = 10(1/3 - p - q) + 16(1/3 - p) + 7/3 > 0.$

Likewise, Expected profit of d_2 – Expected profit of d_3

$= 10p + 30(1 - p - q) + 8q - 0p - 2(1 - p - q) - 34q$

$= 28 - 18p - 54q = 6(1 - 3p - 3q) + 12(1 - 3q) + 10 > 0.$

Therefore, the Bayes criterion solution is d_2.

[9]

Question 5

(i) The model can be written in the form:
$(1 + \alpha B - \alpha^2 B^2)Y_t = Z_t$

So the characteristic equation is:

$1 + \alpha x - \alpha^2 x^2 = 0.$

Applying the quadratic formula, we see that this has roots:

$x = (1 \pm \sqrt{5}) / 2\alpha.$

Roots must lie outside the unit circle, so we require that $(\sqrt{5}-1)/2.|\alpha| > 1$ and $(\sqrt{5}+1)/2.|\alpha| > 1$, which means, $|\alpha| < (\sqrt{5} - 1) / 2$.

(ii) $Y_t = -\alpha Y_{t-1} + \alpha^2 Y_{t-2} + Z_t$

The Yule-Walker equations are:

$\text{Cov}[Y_t, Y_t] = \gamma_0 = -\alpha.\gamma_1 + \alpha^2 \gamma_2 + \sigma^2$. ----- Equation (1)

$\text{Cov}[Y_t, Y_{t-1}] = \gamma_1 = -\alpha.\gamma_0 + \alpha^2 \gamma_1$. ----- Equation (2)

$\text{Cov}[Y_t, Y_{t-2}] = \gamma_2 = -\alpha \gamma_1 + \alpha^2 \gamma_0$. ----- Equation (3)

From equation (2), $\gamma_1 = -\alpha.\gamma_0 / (1-\alpha^2)$. ----- Equation (4)

Substituting the value of γ_1 in equation (3), we have

$\gamma_2 = -\alpha. [-\alpha.\gamma_0 / (1-\alpha^2)] + \alpha^2. \gamma_0 = (2\alpha^2 - \alpha^4). \gamma_0 / (1-\alpha^2)$. ----- Equation (5)

Substituting the values of γ_1 and γ_2 in equation (1), we have

$\gamma_0 = -\alpha(-\alpha.\gamma_0 / (1-\alpha^2)) + \alpha^2 .(2\alpha^2 - \alpha^4). \gamma_0 / (1-\alpha^2) + \sigma^2$.

Thus,

$\gamma_0 = \sigma^2. (1-\alpha^2) / (1 - 2\alpha^2 - 2\alpha^4 + \alpha^6)$.

Substituting the value of γ_0 in equations (4) and (5), we have

$\gamma_1 = -\sigma^2. \alpha / (1 - 2\alpha^2 - 2\alpha^4 + \alpha^6)$.

$\gamma_2 = \sigma^2. (2\alpha^2 - \alpha^4) / (1 - 2\alpha^2 - 2\alpha^4 + \alpha^6)$.

[9]

Question 6

(i) The mean aggregate claim is 1,000,000. The mean amount of an individual claim (*X*) is

$$E(X) = e^{\mu+\sigma^2/2} = e^6 = 403.43.$$

So the mean of the claim frequency, *E*(*N*), satisfies the equation

$$1{,}000{,}000 = E(N) \times E(X).$$

This gives *E*(*N*) = 2478.75.

(ii) The mean amount paid by the direct insurer over an individual claim (*X*) is

$$E[\min\{X, 1000\}] = \int_0^{1000} xf(x)dx + \int_{1000}^{\infty} 1000 f(x)dx$$

$$= \int_{-\infty}^{[\ln(1000)-4]/2} e^{4+2z}\phi(z)dz + 1000P(X > 1000)$$

$$= \int_{-\infty}^{[\ln(1000)-4]/2} \frac{1}{\sqrt{2\pi}} e^{-z^2/2+2z+4}dz + 1000P(e^{4+2Z} > 1000)$$

$$= \int_{-\infty}^{[\ln(1000)-4]/2} \frac{1}{\sqrt{2\pi}} e^{-(z-2)^2/2+6}dz + 1000P\left(Z > \frac{\ln(1000)-4}{2}\right)$$

$$= e^6 \int_{-\infty}^{[\ln(1000)-8]/2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}dy + 1000 - 1000\Phi\left(\frac{\ln(1000)-4}{2}\right)$$

$$= e^6\Phi\left(\frac{\ln(1000)-8}{2}\right) + 1000 - 1000\Phi\left(\frac{\ln(1000)-4}{2}\right)$$

$$= e^6\Phi(-0.54612) + 1000 - 1000\Phi(1.453878)$$

$$= e^6 \times 0.292491 + 1000 - 1000 \times 0.92701 = 190.99.$$

The mean amount paid by the reinsurer over an individual claim (X) is

$$E(X) - E[\min\{X, 1000\}] = 403.43 - 190.99 = 212.44.$$

The reinsurance premium is $E(N)\times 212.44\times 1.15 = 605573.50$.

(iii) If the random variable X represents a typical claim size in the current year, the reinsurance premium in the current year is $1.15\times E(N)\times[E(X)-E\min\{X,1000\}]$, while the reinsurance premium in the next year is $1.15\times E(N)\times[E(1.1X)-E\min\{1.1X,1000\}]$, Therefore, the question is, whether $1.15\times E(N)\times[E(1.1X)-E\min\{1.1X,1000\}]$ is more than 1.1 times $1.15\times E(N)\times[E(X)-E\min\{X,1000\}]$, i.e., whether

$$E(1.1X)-E\min\{1.1X,1000\} > 1.1\times[E(X)-E\min\{X,1000\}].$$

Note that, for every positive x,

$$\min\{1.1x,1000\}\le \min\{1.1x,1100\}.$$

Since X has a continuous distribution, we have

$$E\min\{1.1X,1000\} < 1.1\times E\min\{X,1000\},$$

and therefore,

$$E(1.1X)-E\min\{1.1X,1000\} > 1.1\times[E(X)-E\min\{X,1000\}].$$

It follows that the reinsurance premium will increase by *more than* 10%.

[9]

Question 7

Let us denote the observed values for the two sets of claim sizes for the two groups of drivers as X_1, X_2, X_3, X_4 and Y_1, Y_2, Y_3, Y_4, respectively. Then the likelihood for these data is

$$\left(\prod_{i=1}^{4}\lambda e^{-\lambda X_i}\right)\times\left(\prod_{i=1}^{4}\frac{\lambda}{2}e^{-\lambda Y_i/2}\right)=\frac{\lambda^8}{16}e^{-\lambda[(X_1+X_2+X_3+X_4)+(Y_1+Y_2+Y_3+Y_4)/2]}.$$

The prior density of λ is $100e^{-100\lambda}$. Therefore, the posterior density is proportional to

$$\frac{\lambda^8}{16}e^{-\lambda[(X_1+X_2+X_3+X_4)+(Y_1+Y_2+Y_3+Y_4)/2]}\times 100e^{-100\lambda} = \frac{25\lambda^8}{4}e^{-\lambda[100+(X_1+X_2+X_3+X_4)+(Y_1+Y_2+Y_3+Y_4)/2]}.$$

Therefore the posterior density is a gamma density of the form $\frac{1}{\Gamma(\alpha)}\lambda^{\alpha-1}\beta^{\alpha}e^{-\lambda\beta}$, with parameters

$$\alpha=9,\ \beta=100+(X_1+X_2+X_3+X_4)+(Y_1+Y_2+Y_3+Y_4)/2=100+850+730=1680.$$

The posterior mean is $\frac{\alpha}{\beta}=\frac{9}{1680}=0.00536$.

[5]

Question 8

(i) The posterior density is proportional to the product of the likelihood and the prior density:

$$\binom{100}{N}p^N(1-p)^{100-N}\times\frac{(k+l-1)!}{(k-1)!(l-1)!}p^{k-1}(1-p)^{l-1}$$

$$\propto p^{N+k-1}(1-p)^{100-N+l-1}.$$

By comparison with the prior, this is seen to have the form of the beta density, with the parameters k and l replaced by $N + k$ and $100 - N + l$, respectively.

The complete expression for the density is

$$\frac{(100+k+l-1)!}{(N+k-1)!(100-N+l-1)!}\,p^{N+k-1}(1-p)^{100-N+l-1}.$$

(ii) The mean of the prior density is

$$\int_0^1 p\times\frac{(k+l-1)!}{(k-1)!(l-1)!}\,p^{k-1}(1-p)^{l-1}\,dp=\int_0^1\frac{(k+l-1)!}{(k-1)!(l-1)!}\,p^{k}(1-p)^{l-1}\,dp$$

$$=\frac{k}{k+l}\int_0^1\frac{(k+l)!}{k!(l-1)!}\,p^{k}(1-p)^{l-1}\,dp=\frac{k}{k+l}.$$

(iii) By comparing the expressions for the prior and posterior densities, it can be seen that the posterior mean is

$$\frac{N+k}{100+k+l}.$$

This can be written in the credibility form:

$$Z\times\frac{N}{100}+(1-Z)\times\frac{k}{k+l},$$

where $N/100$ is the estimate of p based only on the data, $k/(k+l)$ is the prior mean and the credibility factor is $Z=\dfrac{100}{100+k+l}$.

[7]

Question 9

(i) (a) Correct.

(b) Incorrect. One only need to know the expected value of the function $m(\theta)=E[X_1|\theta]$, which is estimated from collateral data.

(c) Incorrect. Only the expectations of certain functions of θ are needed, and all these are estimated from collateral data.

(d) Incorrect. The parameter θ is assumed to take three different values for the data sets $\{Y_{11}, Y_{21}, \ldots, Y_{81}\}$, $\{Y_{12}, Y_{22}, \ldots, Y_{82}\}$, and $\{Y_{13}, Y_{23}, \ldots, Y_{83}\}$. Had it not been

for this fact, it would have been impossible to calculate the variance of $m(\theta)$ from the collateral data.

(ii) Let θ be the underlying random parameter. The credibility premium under the EBCT Model 1 is

$$Z\overline{X}+(1-Z)E[m(\theta)],$$

where

$$\overline{X}=\frac{X_1+X_2+X_3+X_4+X_5}{5},$$

$$Z=\frac{5}{5+\dfrac{E[s^2(\theta)]}{\operatorname{var}[m(\theta)]}},$$

$m(\theta)$ is the conditional mean of the observations given θ, and

$s^2(\theta)$ is the conditional variance of the observations given θ.

The following estimates, computable from the collateral data, may be used.

$$E[m(\theta)] = \overline{Y} = \frac{1}{3}\sum_{j=1}^{3}\overline{Y}_j, \qquad \text{where } \overline{Y}_j = \frac{1}{8}\sum_{i=1}^{8}Y_{ij};$$

$$E[s^2(\theta)] = \frac{1}{3}\sum_{j=1}^{3}\left[\frac{1}{7}\sum_{i=1}^{8}\left(Y_{ij} - \overline{Y}_j\right)^2\right];$$

$$\text{var}[m(\theta)] = \frac{1}{2}\sum_{j=1}^{3}\left(\overline{Y}_j - \overline{Y}\right)^2 - \frac{1}{24}\sum_{j=1}^{3}\frac{1}{7}\sum_{i=1}^{8}\left(Y_{ij} - \overline{Y}_j\right)^2.$$

[8]

Question 10

The aggregate claim at time t is $S(t) = \sum_{i=1}^{N(t)} X_i$, where X_i is the size of the ith claim, and $N(t)$ is the number of claims till time t. The claims are independent and have the discrete distribution described in the question.

$$E(X_1) = 0.25\times100 + 0.5\times200 + 0.25\times250 = 187.5.$$

$$E(X_1^2) = 0.25\times100^2 + 0.5\times200^2 + 0.25\times250^2 = 38125.$$

$E[S(t)] = E[N(t)]\times E(X_1) = \lambda t\times187.5 = 3750t$. In particular, $E[S(3)] = 11250$.

$\text{var}[S(t)] = E[N(t)]\times E(X_1^2) = \lambda t\times38125 = 762{,}500t$. In particular, $\text{var}[S(3)] = 1512.45^2$.

The probability of ruin at the end of time t is

$$P[U + (1+\theta)E\{S(3)\} - S(t) < 0],$$

where the initial surplus, U, is 1000. It is required that

$$P[U + (1+\theta)E\{S(3)\} - S(3) < 0] \le 0.05,$$

i.e., $$P\left[\frac{S(3) - E\{S(3)\}}{\sqrt{\text{var}\{S(3)\}}} > \frac{U + \theta E\{S(3)\}}{\sqrt{\text{var}\{S(3)\}}}\right] \le 0.05.$$

Under the normal approximation, this inequality reduces to

$$\Phi\left(\frac{U + \theta E\{S(3)\}}{\sqrt{\text{var}\{S(3)\}}}\right) \ge 0.95.$$

This happens if and only if

$$\theta \ge \frac{1.645\times\sqrt{\text{var}[S(3)]} - U}{E[S(3)]} = \frac{1.645\times1512.45 - 1000}{11250} = 0.1322.$$

Thus, the loading has to be at least 13.22%.

[7]

Question 11

(i) The density of the claim size distribution experienced by the reinsurer is

$$\frac{f(x+M)}{1-F(M)} = \frac{\frac{1}{\theta}e^{-\frac{x+M}{\theta}}}{e^{-\frac{M}{\theta}}} = \frac{1}{\theta}e^{-\frac{x}{\theta}} = f(x),$$ which is the same as the density of the original claims.

(ii) The number of claims experienced by the reinsurer is

$I_1 + I_2 + \cdots + I_N$, where N has the Poisson distribution with mean λ, and I_1, I_2, …are independent and have the Binomial distribution with parameters 1 and $e^{-1000/\theta}$. Therefore, the MGF of this random variable is

$$E\left[E\left(e^{(I_1+I_2+\cdots+I_N)t}\middle|N\right)\right] = E\left[\left\{E\left(e^{I_1 t}\right)\right\}^N\right] = E\left[\left\{1 - e^{-1000/\theta} + e^{-1000/\theta} \times e^t\right\}^N\right].$$

Since $E\left(e^{Ns}\right) = e^{\lambda\left(e^s - 1\right)}$, the above expression simplifies to $e^{\lambda\left(1-e^{-1000/\theta}+e^{-1000/\theta}\times e^t - 1\right)}$, or $e^{\left[\lambda e^{-1000/\theta}\left(e^t-1\right)\right]}$. This can be recognized as the Poisson distribution with mean $\alpha = \lambda e^{-1000/\theta}$.

(iii) It follows from part (i) that the maximum likelihood estimate (MLE) of the parameter θ is the sample mean, 700.

There is only one observation from the distribution derived in part (ii), which is 4. Therefore, the MLE of the parameter α is 4.

Hence, the MLE of $\lambda = \alpha e^{1000/\theta}$ is $4 \times e^{1000/700} = 16.69$.

(iv) The MLE of the mean aggregate claims ($\lambda\theta$) is $16.69 \times 700 = 11683$.

[6]

Question 12

Any two of the following should suffice.

Bilinear model: $X_n + \alpha(X_{n-1} - \mu) = \mu + e_n + \beta e_{n-1} + b(X_{n-1} - \mu)e_{n-1}$.

Threshold autoregressive model: $X_n = \begin{cases} \mu + \alpha_1(X_{n-1} - \mu) + e_n & if\ X_{n-1} \le d, \\ \mu + \alpha_2(X_{n-1} - \mu) + e_n & if\ X_{n-1} > d. \end{cases}$

Autoregressive model with conditional heteroscedasticity: $X_t = \mu + e_t\sqrt{\alpha_0 + \sum_{k=1}^{p}\alpha_k\left(X_{t-K} - \mu\right)^2}$.

[2]

Total Mark { 100 }

****