Master in Actuarial Science
Rate Making and Experience Rating

Exam 1, 11/01/2021, 15:00-17:30, Room AF4
Time allowed: 2:30

Instructions:

1. This paper contains 4 groups of questions and comprises three pages including the title page;
2. Enter all requested details on the cover sheet;
3. You have 10 minutes of reading time. You must not start writing your answers until instructed to do so;
4. Number the pages of the paper where you are going to write your answers;
5. Attempt all questions;
6. Begin your answer to each of the 4 question groups on a new page;
7. Marks are shown in brackets. Total marks: 200;
8. Show calculations where appropriate;
9. An approved calculator may be used. No mobile phones or other communication devices are permitted;
10. The distributed *Formulary* and the *Formulae and Tables for Actuarial Examinations* (the 2002 edition) may be used. Note that the parametrization used for the different distributions is that of the distributed *Formulary*, unless stated clearly.

1. Consider an insurance portfolio where the chief actuary admits that their policies are classified into three broad categories, not easily identified from observed each policy annual claims. These categories are identified by the parameter θ, and labeled as A, B and C. He also admits that members of C are half of B, and B is half of A.

   Per year each risk in the portfolio can produce a number of claims summarized as 0, 1, or 2 (or more). Probabilities are shown in Table 1. **[85]**

Cat.	0	1	2
A	0.40	0.35	0.25
B	0.35	0.40	0.25
C	0.25	0.35	0.40

Table 1: Claim count probabilities for each category

   Suppose that a risk category is chosen at random, and two risks are chosen with replacement, randomly and independently, from that class. Suppose that a total of three claims was observed from those two risks drawn. Two more risks are then drawn with replacement from the same category, and it is of interest to predict the total on these next two.

   Let X be the total of claims from the two risks drawn and $f_X(.)$ its probability function.

   (a) Determine the conditional probability function $f_{X|\theta}(x|\theta)$ for the totals of claims for the two risks drawn when $\theta = A, B, C$. (15)

   (b) Calculate the probability of the total claims of the first two risks drawn equals 3: $\Pr(X_1 = 3)$. (5)

   (c) Calculate the posterior distribution $\pi_{\Theta|X_1}(\theta|3)$. (10)

   (d) Determine the predictive p.f., $f_{X_2|X_1}(x_2|3)$, of the total X_2 of the next double draw given that $X_1 = 3$ was observed in the previous double draw. Also, calculate the Bayesian premium. (25)

   (e) Compute the structural parameters $\mu = E(\mu(\Theta))$, $\upsilon = E(\upsilon(\Theta))$ and $a = V(\mu(\Theta))$. (10)

   (f) Compute Bühlmann's credibility premium. Conclude about *exact credibility.* (15)

   (g) From above items (d) and (f), choose your preferred estimate choice for the risk premium. Explain Briefly. (5)

2. Consider a risk, denoted by X, that brings associated a given parameter θ that cannot be observed but is considered to be an outcome of a random variable denoted by Θ. We assume that, given $\Theta = \theta$, $X|\theta$ is exponentially distributed with mean $1/\theta$, and that $\Theta \frown Gamma(4, 0.001)$.Consider that risk X was observed, independently, and reported claims of amounts $X_1 = 150$, $X_2 = 400$ and $X_3 = 950$. The unconditional joint density of (X_1, X_2, X_3) is given by **[35]**

$$f(150, 400, 950) = \frac{1,000^4}{6} \frac{6!}{2,500^7} .$$

   Let $X_4 = x_4$ be the fourth observation for the risk.

   (a) Show that:

      i. The unconditional joint density is given by

$$f(150, 400, 950, x_4) = \frac{1,000^4}{6} \frac{7!}{(2,500 + x_4)^8} ;$$

(10)

      ii. The predictive distribution is given by

$$f(x_4|150, 400, 950) = \frac{7\,(2500)^7}{(2,500 + x_4)^8} ;$$

(7.5)

      iii. And the posterior distribution is given by

$$\pi(\theta|150, 400, 950) = \theta^6 e^{-2500\theta} 2500^7/6! .$$

(7.5)

(b) Calculate the following: Risk premium, colective premium, Bayesian premium, credibility premium and empirical risk mean. Compare the different means. (10)

3. A working party is set to study a *bonus* system to be based on the claims frequency to rate the risk of some given motor insurance portfolio. The decision is to set a Markovian system, and is establishing the following rules: [**50**]

   (i) A 25% discount in case of no claims for two consecutive years;

   (ii) A 20% penalty if the policy has one claim in the last year, except for policies in the highest penalty class;

   (iii) A 50% penalty in the case of two or more claims in the last year. If it is already there, it stays in the case of one or more claims.

   (iv) Entry class corresponds to a class with no *bonus* and no penalty. Base index premium is set as 100.

   Premia, indices, are set on an annual basis. Answer the following questions:

   (a) The system is not directly Markovian. Explain briefly, and discuss change procedures. (12.5)

   (b) Rearrange and define the set of classes under a Markov framework, as well as the corresponding premia vector. (12.5)

   (c) Build a table with the rule behaviour of the (new) system. (12.5)

   (d) Write down the transition rule matrix. (12.5)

4. A group of students is working on a project that, among other things it involves modelling a tariff for a given existing motor insurance portfolio. A wide variety of risk factors affecting both the claim count and size are being tested to have influence on different premium levels. They have a data base available with reliable and enough data, from a stable year, allowing to study and estimate the influence of each risk factor provided. [**30**]

   (a) Discuss briefly the importance of getting data "from a stable year". (7.5)

   (b) For the estimation of the pure premium, the study group is studying a multiplicative model. A member of the group suddendly commented: *That's because Pure premium = (Claim frequency mean) × (Average claim size)...* Comment briefly. (7.5)

   (c) Another student said: *Since it is a motor insurance portfolio we should focus on the estimation of the severity mean as we already use a Markovian "bonus-malus" system based on claim counts.* Comment. (7.5)

   (d) Suppose that the group is using a "GLM Gamma model" for the claim size, with log link, and three rating factors with two, three and four levels, labeled respectively $F11$, $F12$, $F21$, $F22$, $F23$, $F31$, $F32$, $F33$ and $F34$. For the different factors, consider base levels as 1, 2 and 3, respectively. Write the estimated mean for a policy included in ordered factors in levels 1, 2 and 3. Interpret the intercept estimate. (7.5)

Solutions

1.

(a) Conditional p.f. $f_{X|\theta}(x|\theta)$ for $\theta = A, B, C$.

θ	A	B	C
$\Pr(X=0\|\theta)$	0.1600	0.1225	0.0625
$\Pr(X=1\|\theta)$	0.2800	0.2800	0.1750
$\Pr(X=2\|\theta)$	0.3225	0.3350	0.3225
$\Pr(X=3\|\theta)$	0.1750	0.2000	0.2800
$\Pr(X=4\|\theta)$	0.0625	0.0625	0.1600
Total	1.0	1.0	1.0

(b) Calculate the probability of the total claims of the first two risks drawn equals 3:

$$\Pr(X_1 = 3) = \frac{1}{7}(0.175(4) + 0.2(2) + 0.28) \simeq 0.19714286$$

(c) Calculate the posterior distribution $\pi_{\Theta|X_1}(\theta|3)$.

θ	A	B	C	
$\pi(\theta)$	4/7	2/7 1/7		
$\Pr(X_1=3\|\theta)$	0.175	0.2	0.28	
$\Pr(X_1=3\|\theta)\ \pi(\theta)$	0.1	0.057142857	0.04	
$\Pr(\Theta=\theta\|X_1=3)$	0.507246377	0.289855072	0.202898551	1

(d) Determine the predictive p.f., $f_{X_2|X_1}(x_2|3)$, of the total X_2 of the next double draw given that $X_1 = 3$ was observed in the previous double draw. Also, calculate the Bayesian premium.

$\pi(\theta)$	0.571428571	0.285714286	0.142857143	
θ	A	B	C	
$\pi_{\Theta\|X_1}(\theta\|3)$	0.507246377	0.289855072	0.202898551	
x_2				$\Pr(X_2=x_2\|X_1=3) = \sum \Pr(X_2=x_2\|X_1=3,\theta)\,\pi_{\Theta\|X_1}(\theta\|3)$
0	0.08115942	0.035507246	0.012681159	0.129347826
1	0.142028986	0.08115942	0.035507246	0.258695652
2	0.163586957	0.097101449	0.065434783	0.326123188
3	0.088768116	0.057971014	0.056811594	0.203550725
4	0.031702899	0.018115942	0.032463768	0.082282609
	0.507246377	0.289855072	0.202898551	1

$$\begin{aligned} E[X_2|X_1=3] &= 0(0.129347826) + 1(0.258695652) + 2(0.326123188) + 3(0.203550725) + 4(0.082282609) \\ &= 0.258695652 + 0.652246377 + 0.610652174 + 0.329130435 = \mathbf{1.850724638}. \end{aligned}$$

(e) Compute the structural parameters $\mu = E(\mu(\Theta))$, $\upsilon = E(\upsilon(\Theta))$ and $a = V(\mu(\Theta))$.

θ	A	B	C
$\mu(\theta)$	1.7	1.8	2.3
$\upsilon(\theta)$	1.255	1.18	1.255

μ	υ	a
1.814285714	1.233571429	0.04122449

(f) Compute Bühlmann's credibility premium.

$$\begin{aligned} k &= 1.233571429/0.04122449 = 29.92326733 \\ z &= \frac{1}{1 + 29.92326733} = 0.032338109 \\ P_c &= 0.032338109\,(3) + (1 - 0.032338109)\,1.814285714 = \mathbf{1.852629473} \end{aligned}$$

Although close, $\mathbf{1.852629473} \neq \mathbf{1.850724638}$ so, no exact credibility...

(g) The Credibility premium uses the same minimizing loss function as the Bayesian one but it imposes a constraint: A linear function on the observations. The Bayesian estimate is a better choice,

2. (a)

i.

$$\begin{aligned} f(150,400,950) &= \int_0^\infty f(150|\theta)f(400|\theta)f(950|\theta)\pi(\theta)d\theta \\ &= \frac{1,000^4}{6}\frac{6!}{2,500^7}\underbrace{\int_0^\infty \frac{2,500^7}{6!}\theta^6 e^{-2500\theta}d\theta}_{1} \\ &= \frac{1,000^4}{6}\frac{6!}{2,500^7}\,. \end{aligned}$$

Similarly, we get

$$\begin{aligned} f(150,400,950,x_4) &= \int_0^\infty f(150|\theta)f(400|\theta)f(950|\theta)f(x_4|\theta)\pi(\theta)d\theta \\ &= \frac{1,000^4}{6}\frac{7!}{(2500+x_4)^8}\,. \end{aligned}$$

ii.

$$f(x_4|150,400,950) = \frac{f(x_4,150,400,950)}{f(150,400,950)} = \frac{7(2500)^7}{(2500+x_4)^8}\,.$$

$\to Pareto(7,2500):\ 2500/6.$

iii.

$$\begin{aligned} \pi(\theta|150,400,950) &= \frac{f(150,400,950|\theta)\pi(\theta)}{f(150,400,950)} \\ &= \theta^6 e^{-2500\theta}2500^7/6!\,. \end{aligned}$$

(b) Posterior is $Gamma(7, 2500^{-1})$, with mean 7/2500. We have conjugate distributions and exact credibility. Then Credibility Premium equals Bayesian Premium and equal mean of $Pareto(7,1500):\ 2500/6$.

$$\begin{aligned} P_c &= E[X_4|(150,400,950)] = 2500/6 = 416.6(6) \\ \mu(\theta) &= 1/\theta \\ \mu &= E[1/\theta] = \frac{1,000^4}{3!}\frac{2!}{1000^3}\underbrace{\int_0^\infty \frac{1000^3}{2!}\theta^2 e^{-1000\theta}d\theta}_{1} = 1000/3 \\ \bar{X} &= 500\,. \end{aligned}$$

Means ordered: $1000/3 < 416.(6) < 500$...

3.

(a) The system is Markovian of order two, or semi-Markovian. We have to split classes, premia are set in an annual basis and to get a bonus we need to wait for two years, we have to consider situations where there are no claims in the first year... Also beware in the case of being placed with the highest premium class, you can olny expect to go down if you spend one year without any claim.

(b)

$$\mathbf{b} = (75, 100, 100, 120, 120, 150, 150)$$

We build a system with seven classes:

C_1 25% *bonus class*

C_2 Class with neither bonus nor penalty, with no claims in the last year

C_3 Entry class

C_4 20% *penalty* and no claims last yr

C_5 20% *penalty*

C_6 Policies with 50% *penalty* and no claims last yr

C_7 Policies with 50% *penalty* and claims last yr

(c)

		New step after claims		
Step	Pr.	0	1	2^+
1	75	1	5	7
2	100	1	5	7
3	100	2	5	7
4	120	1	5	7
5	120	4	5	7
6	150	1	5	7
7	150	6	7	7

(d) Write down the transition rules matrix

$$\begin{array}{c} \\ 1\\2\\3\\4\\5\\6\\7 \end{array}\begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \begin{pmatrix} \{0\} & & & & \{1\} & & \{2^+\} \\ \{0\} & & & & \{1\} & & \{2^+\} \\ & \{0\} & & & \{1\} & & \{2^+\} \\ \{0\} & & & & \{1\} & & \{2^+\} \\ & & & \{0\} & \{1\} & & \{2^+\} \\ \{0\} & & & & \{1\} & & \{2^+\} \\ & & & & & \{0\} & \{1^+\} \end{pmatrix} \end{array}$$

4.

(a) Classical BM systems are built under the assumption of a long term behaviour, where risk conditions are well established, known, risk factors are well identified, information is neat, clean, and especially reliable. Particular risk behaviour features can be adjusted “a posteriori”... according to each risk record.

(b) Wrong. The model being proposed is a multiplicative model where each factor connects by a product. We can linearize calculating the logarithm...

(c) It is an ignorant statement: GLM is used to build a tariff for the pure premium which has two components: claim frequency and severity mean. One can model risk factors for the two components separately, or all together considering an aggregate claim mean model. In both cases we're talking about prior ratemaking. BMS is posterior ratemaking, or experience rating, it means adjusting premia previously set.

(d) In the intercept we have included the estimated premium for a policy with factor levels $F11$, $F22$ and $F33$. They are the base factor levels. We have

$$\exp\{intercept\}\,.$$