LISBON SCHOOL OF ECONOMICS & MANAGEMENT
UNIVERSIDADE DE LISBOA

Master in Actuarial Science
Rate Making and Experience Rating

Exam 2, 05/02/2021, 9:00-11:30, Room 006 F1
Time allowed: 2:30

Instructions:

1. This paper contains 4 groups of questions and comprises 3 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.

1. Consider a certain insurance portfolio where each risk can be individualized by a specific characteristic. Denote the risk by the random variable X and the individual characteristic by parameter θ which quantifies the individual characteristic. For a given θ, let $X|\theta \frown \exp(\theta)$, with mean $1/\theta$. Let θ be the outcome of a random variable $\Theta \frown Gamma(4, 0.002)$, mean 0.008. That is

$$\begin{aligned} f_x(x|\theta) &= \theta e^{-\theta x}, \;\; x,\theta > 0\,, \\ \pi(\theta) &= \theta^3 e^{-500\theta} 500^4/6, \; \theta > 0\,. \end{aligned}$$

Assume that the usual hypothesis of the Bayesian credibility theory are fulfilled.

Consider that in the past a given risk had reported claim amounts of amounts 100, 300 and 100. Consider the prediction or estimation of a next claim X_4. **[62.5]**

(a) Calculate: (25)

i. $f_X(100, 300, 100)$; and

ii. $f(250, 800, 450, x_4)$.

(b) Calculate the predictive density, for x_4, and compute the Bayesian premium. (10)

(c) Determine: (17.5)

i. Bühlmann's structural parameters, $\mu = E(\mu(\Theta))$, $\upsilon = E(\upsilon(\Theta))$ and $a = V(\mu(\Theta))$.

ii. Bühlmann's credibility premium, denoted as P_c. Compare this with the Bayesian premium. Comment briefly.

(d) Show that $E[\mu(\Theta|(100, 300, 100))]$, the expectation with respect to the posterior distribution, equals the predictive mean. (10)

2. The following table shows data generated by a motor insurance portfolio in some year corresponding to a "regular" year of exposure. **[57.5]**

No. of Claims	0	1	2	3	4	5	6^+	Total
No. of policies	102 994	14 575	1 876	259	44	8	0	119 756

Let N be the number of claims per year, and suppose that $N \frown$Poisson(θ). The parameter θ is unknown. Consider that the usual hypothesis in credibility theory may be applicable to the risk portfolio under study, θ is the associated risk parameter and you are a promising actuary, expert on ratemaking and experience rating. Observations from the past n years for the risk are available.

(a) For the calculation of an estimate of Bühlmann's credibility (pure) premium we need to estimate the structural parameters. Based on the data available above, calculate these estimates. (7.5)

(b) Based on the calculated estimates can you infer that the portfolio is not fully homogeneous? Explain your answer briefly. (5)

(c) Calculate an estimate for Bühlmann's credibility premium (denoted as P_c) for a policy that produced an average of 1.5 in an observed 5-year period. What is called this estimate? (7.5)

(d) Is it appropriate to say that *on average, Bühlmann's credibility premium premium coincides with the collective premium*? Also, Show that $Var[P_c] = z\,a$, where z is the credibility factor and $a = V(\mu(\Theta))$. Write all apropriate calculation. (15)

(e) Consider a bonus-malus system based on Bühlmann's credibility model, with the estimates calculated above in item 2a. Complete Table 1 on next page, where n is year and k is claim counts (figures are rounded to the unit). Write clearly computational formula. (22.5)

3. A certain insurer is studying a *bonus-malus* system (briefly BMS) based on the individual's annual claim counts record to rate each individual risk in a given motor insurance portfolio. **[50]**

(a) Suppose that the insurer wanted to consider a BMS where claim amounts could be introduced for premium adjustment. Could you give a suggestion for a model that could be used (from those you studied)? Comment briefly. (12.5)

(b) Bonus systems are usually based on Markov chain analysis where premiums are computed and adjusted annually. Suppose you have a 3-state system with an entry class, a sole bonus class (40% bonus) and a remainder penalty class (50% penalty). From any class, you could reach a next year bonus by having no claims in last two consecutive years only. The system is not directly Markovian. How would you solve the problem under a Markov chain framework, yearly based, as you learned in class? Write down classes and corresponding premia. (12.5)

$n\backslash k$	0	1	2	3
0	100			
1			248	
2	77	148		290
3		133		
4			178	
5	57			216

Table 1: B-M Table, values in %.

(c) Consider a *bonus* system that evolves according to what shown in Table 2. Considering a Poisson($\lambda = 1/8$) distribution for the claim counts, build the associated transition probability matrix. (12.5)

(d) Suppose that for some 3-state *bonus* system and some given θ, the steady state premium distribution is given by vector $(\theta^2, \theta(1-\theta), 1-\theta)$, where θ is the probability of not getting any claim in one year. Number of claims is Poisson distributed and parameter can take values $\lambda = 0.125, 0.15, 0.2$ with probabilities 0.6, 0.3, 0.1, respectively. (12.5)

4. In a study for modelling a tariff for a worker's compensation line of business, a study group tried a GLM model with several factors supposed to have impact in both the claim frequency and the claim size. [**30**]

(a) For the response variables the analysts are studying *key ratios*. The following table show several pairs of response variables (labeled X) and exposure (w). The last column contains numbers to label the corresponding *key ratio* variable. (7.5)

w	X	Key Ratio
Duration	Number of Claims	1
Duration	Claims Cost	2
Number of Claims	Claims Cost	3
Earned Premium	Claims Cost	4
Number of Claims	Number of large claims	5

Identify and assign the key ratio variable to the corresponding number.

(b) For the estimation of the pure premium, the study group is analising separately the means. Comment briefly. (7.5)

(c) In analysing the claim severity what sort of GLM model would you likely consider as, say, a first choice for the analysis. Comment briefly. (7.5)

(d) Suppose we have a Poisson model for the claim frequency with three rating factors with two, three and three levels, labeled F11, F12, F21, F22, F23, F31, F32, F33. Write the estimated mean for the three rating actors with levels 1, 2 and 2, respectively. (7.5)

Step	%	New step after claims			
		0	1	2	3+
1	30	1	5	9	12
2	35	1	6	10	12
3	30	2	6	10	12
4	40	3	7	11	12
5	50	4	8	12	12
6	60	5	9	12	12
7	70	6	10	12	12
8	80	7	11	12	12
9	90	8	12	12	12
10	100	9	12	12	12
11	120	10	12	12	12
12	150	11	12	12	12

Table 2: Rules and premium percentages

Solutions

1. (a) i.

$$f(100,300,100) = \int_0^\infty f(100|\theta)f(300|\theta)f(100|\theta)\pi(\theta)d\theta = \frac{500^4}{1000^7}\frac{6!}{3!}.$$

ii. Similarly, we get

$$f(250,800,450,x_4) = \frac{7!}{3!}\frac{10^{12}}{(2500+x_4)^8}$$

.

(b) The predictive density comes

$$f(x_4|100,300,100) = \frac{f(x_4,100,300,100)}{f(100,300,100)} = \frac{7(1000)^7}{(1000+x_4)^8},$$

and

$$E[\mu(\Theta|(100,300,100))] = \frac{1000}{6} = \frac{500}{3}.$$

(c)

i. Structural parameters:

- $\mu(\theta) = 1/\theta \rightarrow$

$$\mu = E[1/\Theta] = \int_0^\infty \theta^2 e^{-500\theta}\frac{500^4}{6}\,d\theta = \frac{500}{3};$$

- $\upsilon(\theta) = 1/\theta^2 \rightarrow$

$$\upsilon = E[1/\Theta^2] = \int_0^\infty \theta e^{-500\theta}\frac{500^4}{6}\,d\theta = \frac{500^2}{6};$$

- $E[\mu(\Theta)^2] = E[1/\Theta^2] = 500^2/6 \rightarrow$

$$a = V[\mu(\Theta)] = \frac{500^2}{6} - \frac{500^2}{9} = \frac{500^2}{18}.$$

ii. For P_c, we have exact credibility, since Bayesian premium is linear on the observations. So, $P_c = \frac{500}{3}$. We can compute (to check): $n = 3$, $k = \upsilon/a = 3$, $\bar{x} = 500/3$, $z = 1/2$, then we get

$$P_c = \frac{1}{2}\frac{500}{3} + \frac{1}{2}\frac{500}{3} = 166.(6).$$

(d) Posterior is $Gamma(7, 1000^{-1})$, mean 7/1000:

$$\pi(\theta|(100,300,100)) = \frac{f(100,300,100|\theta)\,\pi(\theta)}{f(100,300,100)} = \theta^6\, e^{-1000\,\theta}\frac{1000^7}{6!},$$

then

$$E[\mu(\Theta)|(100,300,100)] = E[1/\Theta|(100,300,100)] = \frac{500}{3}.$$

2. (a) Structural parameters are: $\mu = E(\mu(\Theta))$, $\upsilon = E(\upsilon(\Theta))$ and $a = V(\mu(\Theta))$. Since we have a Poisson with $\mu(\theta) = \upsilon(\theta) = \theta$, we set $\hat{\mu} = \hat{\upsilon} = \bar{N} \simeq 0.161328034$; $\hat{Var}[N] = S^2 = 0.185353071$ then $\hat{a} = 0.185353071 - 0.161328034 = 0.024025037$.

(b) If the portfolio were fully homogeneous we could fit a Poisson distribution, however we can see that the empirical variance S^2 is 15% above the mean $\bar{N}$ suggesting a negative binomial or other mixed Poisson to be a fitting candidate. This means heterogeneity.

(c)

$$P_c = \frac{5}{5+6.714996126}1.5 + \frac{6.714996126}{5+6.714996126}0.161328034 = 0.73267776$$

It is called "Empirical (Bühlmann's) credibility premium" (it is an empirical Bayes estimate).

(d) Yes, since

$$\begin{aligned} E[P_c] &= zE[E[\bar{N}|\Theta]] + (1-z)\mu \\ &= zE[\mu(\Theta)] + (1-z)\mu = \mu\,. \end{aligned}$$

Also,

$$\begin{aligned} Var[P_c] &= z^2 Var[\bar{N}] = z^2\left(Var[\mu(\Theta)] + E[\upsilon(\theta)]/n\right) \\ &= z^2\left(a + \upsilon/n\right) = z\,a\,. \end{aligned}$$

(e)

$$\tilde{\theta} = \frac{n}{n + 6.714996126}\,\bar{N} + \frac{6.714996126}{n + 6.714996126}\,0.161328034 = \frac{\sum_{i=1}^{n} N_i + 6.714996126(0.161328034)}{n + 6.714996126}$$

Then

$$100\,\frac{\tilde{\theta}}{0.161328034} = \frac{\sum_{i=1}^{n} + 1.083317121}{0.161328034\,n + 1.083317121}$$

$n\backslash k$	0	1	2	3
0	100			
1	**87**	**167**	248	**328**
2	77	148	**219**	290
3	**69**	133	**197**	**261**
4	**63**	**121**	178	**236**
5	57	**110**	**163**	216

3. (a) They could try credibility, since credibility modelling can use claim frequency or amounts, more precisely aggregate claim amounts.

(b) The system is semi-Markovian, or Markov of order 2. Solution is to split classes. We could build a system with 5 classes, a bonus class, an entry class , a *wating bonus from entry* class,... and two penalty classes. However the system keeps with three different premia, still.

(c)

$$\begin{array}{cc} & \begin{array}{cccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \end{array} \\ \begin{array}{c} 1\\2\\3\\4\\5\\6\\7\\8\\9\\10\\11\\12 \end{array} & \begin{pmatrix} 0.882496903 & & & & 0.110312113 & & & & 0.006894507 & & & 0.000296478 \\ 0.882496903 & & & & & 0.110312113 & & & & 0.006894507 & & 0.000296478 \\ & 0.882496903 & & & & 0.110312113 & & & & 0.006894507 & & 0.000296478 \\ & & 0.882496903 & & & & 0.110312113 & & & & 0.006894507 & 0.000296478 \\ & & & 0.882496903 & & & & 0.110312113 & & & & 0.007190985 \\ & & & & 0.882496903 & & & & 0.110312113 & & & 0.007190985 \\ & & & & & 0.882496903 & & & & 0.110312113 & & 0.007190985 \\ & & & & & & 0.882496903 & & & & 0.110312113 & 0.007190985 \\ & & & & & & & 0.882496903 & & & & 0.117503097 \\ & & & & & & & & 0.882496903 & & & 0.117503097 \\ & & & & & & & & & 0.882496903 & & 0.117503097 \\ & & & & & & & & & & 0.882496903 & 0.117503097 \end{pmatrix} \end{array}.$$

(d)

$$\begin{aligned} E[e^{-2\lambda}] &= 0.6\,e^{-2(0.125)} + 0.3\,e^{-2(0.15)} + 0.1\,e^{-2(0.2)} = 0.756557941 \\ E[e^{-\lambda}] &= 0.6\,e^{-0.125} + 0.3\,e^{-0.15} + 0.1\,e^{-0.2} = 0.86958361 \\ \left(E[\theta^2],\,E[\theta(1-\theta)],\,E[1-\theta]\right) &= (0.756557941;\,0.113025669;\,0.13041639) \\ & \ 0.756557941 + 0.113025669 + 0.13041639 = 1 \end{aligned}$$

4.

(a)

w	X	Key Ratio
Duration	Number of Claims	1. Claim frequency
Duration	Claims Cost	2. Pure premium
Number of Claims	Claims Cost	3. Claim Severity
Earned Premium	Claims Cost	4. Loss ratio
Number of Claims	Number of large claims	5. Proportion of large claims

(b) The pure premium has two components: The Claim Frequency and the Claim Severity, that can be studied separately. Indeed, Claim counts and severities are usually assumed to be independent, then

$$Pure\ Premium = Claim\ Frequency \times Severity$$

... However, one could use a Tweedie GLM model that works a compound Poisson distribution and estimate the Pure Premium on aggregate. Here, there's no need to separate the two premium components.

(c) A Gamma model may be suitable, with a canonical link or logarithm function... It is a classical choice. For this example may be suitable an Inverse Gaussian (with canonical Inverse-square link).

(d) Suppose we have a Poisson model for the claim frequency with three rating factors with three, two and three levels, labeled F11, F12, F13, F21, F22, F31, F32, F33. Write the estimated mean for level for the three rating factors with levels 1, 2 and 2, respectively. E.g.

$$\exp\{intercept + \hat{F22} + \hat{F32}\}\,,$$

and explain where is the base level...