

准备金发展年相关的贝叶斯估计

陈明镜

(西华师范大学数学与信息学院,四川 南充 637002)

摘要:文章基于广义线性混合模型对准备金运用贝叶斯原理进行预测,该预测模型有三个特点:一是使得各个发展年的赔付之间具有相关性;二是各个事故年的期望总赔付是随机变量;三是在已知其他参数的条件下,各个事故年的后验期望总赔付是先验期望和实际数据的加权平均,且随发展年增长实际数据被赋予更多权重。最后用 MCMC 方法给出了参数值。

关键词:广义线性混合模型;贝叶斯;准备金;MCMC

中图分类号:O212

文献标识码:A

由于保险赔付滞后的原因,准备金预测是精算学的重要研究领域之一,即根据表 1“准备金流量三角”左上方已知的赔付数据(本文使用字母 D_{ij} 表示)预测右下方未来发生的数据“NA”,其中字母 D_{ij} ($i = 1, \dots, n, j = 1, \dots, n - i + 1$) 表示在事故年 i 发生的赔案在发展

年 j 的赔付,比如 $D_{23} = 934$ 。众多文献都假设同一事故年下各个发展年的赔付相互独立^[1-2],而对于发展年之间的相关性最近才有为数不多的涉及^[3-4]。文章通过 Markov 链建立了不同发展年赔付之间的相关性,使用贝叶斯方法得到了准备金的预测值。

表 1 准备金流量三角(单位:千)

	1	2	3	4	5	6	7	8	9	10
1	358	767	611	483	527	574	146	140	227	68
2	352	884	934	1183	446	321	528	266	425	NA
3	291	1002	926	1017	751	147	496	280	NA	NA
4	311	1108	776	1562	272	352	206	NA	NA	NA
5	443	693	992	769	505	471	NA	NA	NA	NA
6	396	937	847	805	706	NA	NA	NA	NA	NA
7	441	848	1131	1063	NA	NA	NA	NA	NA	NA
8	359	1062	1443	NA	NA	NA	NA	NA	NA	NA
9	377	987	NA	NA	NA	NA	NA	NA	NA	NA
10	344	NA	NA	NA	NA	NA	NA	NA	NA	NA

本文基于广义线性混合模型,使用贝叶斯原理^[5]建立了同一事故年下不同发展年的赔付之间的相关性,并且各个事故年的后验期望总赔付是先验期望和实际数据的加权平均。

1 模型建立

文献[1]假设 D_{ij} 相互独立,并且 $D_{ij} \sim poisson(x_i y_j)$, 其中 $\sum_{j=1}^n y_j = 1$,那么参数 x_i 可以理解为第 i 个事故年终极赔付的期望 y_j ($j = 1, 2, \dots, n$) 可以理解为第 j 个发展年占终极赔付的比例。

本文使用与 Verrall(2000)^[1]相近的假设,但是参数 x_i 不再是确定的值,而是服从分布 $Gamma(\alpha, \beta)$,这样 x_i 的期望是 α/β ,方差是 α/β^2 。这样的假设有两方面的作用:一是使 x_i 成为随机变量,那么第 i 个事故年终极赔付的期望 x_i 好像是从分布 $Gamma(\alpha, \beta)$ 的一次实现,这符合实际情况,不同事故年的终极赔付不可能一成不变,而是随机的;二是不同发展年的赔付之间具有相关性,详细分析如下。

记 $y = (y_1, \dots, y_n)$, $x_i \sim Gamma(\alpha, \beta)$, 则 $D_{ij} | y, \alpha, \beta$ 服从负二项分布:

$$f(D_{ij} | y, \alpha, \beta) =$$

$$\int \frac{(x_i y_j)^{D_{ij}} \exp(-x_i y_j)}{D_{ij}!} \frac{\beta^\alpha x_i \exp(-\beta x_i)}{\Gamma(\alpha)} dx_i = \frac{\Gamma(D_{ij} + \alpha)}{D_{ij}! \Gamma(\alpha)} \left(\frac{y_j}{y_j + \beta}\right)^{D_{ij}} \left(\frac{\beta}{y_j + \beta}\right)^\alpha$$

从而

$$E(D_{ij} | y, \alpha, \beta) = \frac{\alpha}{\beta} y_j$$

$$Var(D_{ij} | y, \alpha, \beta) = \frac{\alpha}{\beta} y_j + \frac{\alpha}{\beta^2} y_j$$

$$Cov(D_{ij}, D_{ik} | y, \alpha, \beta) = \frac{\alpha}{\beta^2} y_j y_k$$

从而, 在已知参数 y, α, β 的条件下, D_{ij} 服从负二项分布, 不同发展年的 D_{ik} 和 $D_{ij} (j \neq k)$ 具有相关性, 并且这种相关性和比例, y_j 与 y_k 有关。

2 贝叶斯估计

记 $D = (D_{ij}, i + j \leq n + 1), x = (x_1, \dots, x_n)$, 使用贝叶斯原理估计参数 x, y, α, β :

$$f(D, x, y, \alpha, \beta) = \pi(y) \pi(\alpha) \pi(\beta)$$

$$\prod_{i=1}^n \left\{ \prod_{j=1}^{n-i+1} \frac{(x_i y_j)^{D_{ij}} \exp(-x_i y_j)}{D_{ij}!} \right\}$$

$$\frac{\beta^\alpha x_i \exp(-\beta x_i)}{\Gamma(\alpha)}$$

其中 $\pi(y), \pi(\alpha), \pi(\beta)$ 分别是 y, α, β 的先验分布。这样可以得到联合后验密度 $f(x, y, \alpha, \beta | D) \propto f(D, x, y, \alpha, \beta)$, 从而得到边际后验密度 $f(y, \alpha, \beta | D)$ 。

一般来讲, (边际) 后验密度没有简单的表示, 本文使用 MCMC 技术通过 Gibbs 抽样得到, 这时就需要各个参数的全条件分布。为了尽量使参数的全条件分布有解析式, 取 $\pi(y) = Dirichlet(c_1, \dots, c_n)$, 对 α 的先验取 $\pi(\alpha) = Gamma(a, b)$, 对 β 的先验取 $\pi(\beta) = Gamma(d, \rho)$ 。

这里主要分析 x_i 的全条件分布 $f(x_i | D, y, \alpha, \beta)$, 其它参数的全条件分布类似。

注意到事故年之间相互独立, 从而 x_i 只与第 i 个事故年的数据有关, 则

$$f(x_i | D, y, \alpha, \beta) \propto x_i$$

$$\exp\left(-\left(\sum_{j=1}^{n-i+1} y_j + \beta\right) x_i\right)$$

已知 D, y, α, β , 的条件下, $x_i | D, y, \alpha, \beta$ 服从 $Gamma\left(\sum_{j=1}^{n-i+1} D_{ij} + \alpha, \sum_{j=1}^{n-i+1} y_j + \beta\right)$, 这样得到 $x_i | D, y, \alpha, \beta$ 的期望是

$$E(x_i | D, y, \alpha, \beta) = \frac{\beta}{\sum_{j=1}^{n-i+1} y_j + \beta} \frac{\alpha}{\beta} +$$

$$\frac{\sum_{j=1}^{n-i+1} y_j}{\sum_{j=1}^{n-i+1} y_j + \beta} \frac{\sum_{j=1}^{n-i+1} D_{ij}}{\sum_{j=1}^{n-i+1} y_j}$$

注意这是 $\frac{\alpha}{\beta}$ 和 $\frac{\sum_{j=1}^{n-i+1} D_{ij}}{\sum_{j=1}^{n-i+1} y_j}$ 的加权平均, 随着数据的发展, $\sum_{j=1}^{n-i+1} y_j \uparrow 1$, 权重 $\frac{\beta}{\sum_{j=1}^{n-i+1} y_j + \beta}$ 越来越小, 所以赋予实现数

据, $\frac{\sum_{j=1}^{n-i+1} D_{ij}}{\sum_{j=1}^{n-i+1} y_j}$ 更多的权重。

全条件分布 $f(x_i | D, y, \alpha, \beta)$ 说明, 如果得到了 y, α, β 的值, 那么 x_i 的估计可以基于分布 $f(x_i | D, y, \alpha, \beta)$ 的特征, 比如分布的均值; 而 y, α, β 的值是通过贝叶斯方法 $f(y, \alpha, \beta | D)$ 得到的, 所以, x, y, α, β 的值是相互影响的。因此

$$Cov(D_{ij}, D_{ik} | D, y, \alpha, \beta) = \frac{\sum_{j=1}^{n-i+1} D_{ij} + \alpha}{\left(\sum_{j=1}^{n-i+1} y_j + \beta\right)^2} y_j y_k$$

3 预测

对准备金建模的一个中心任务就是预测表 1 “准备金流量三角”右下方 $i + j > n + 1$ 未来发生的数据, 对 $k \geq n - i + 2$ 预测分布可以分部积分为

$$f(D_{ik} | D) = \iiint f(D_{ik} | x_i, y)$$

$$f(x_i, y, \alpha, \beta | D) dx_i dy d\alpha d\beta = \iiint [f(D_{ik} | x_i, y)$$

$$f(x_i | D, y, \alpha, \beta) dx_i] f(y, \alpha, \beta | D) dy d\alpha d\beta =$$

$$\int f(D_{ik} | x_i, y) f(x_i | D, y, \alpha, \beta) dx_i$$

$$f(D_{ik} | D, y, \alpha, \beta) =$$

$$\frac{\Gamma(D_{ik} + \sum_{j=1}^{n-i+1} D_{ij} + \alpha)}{D_{ik}! \Gamma\left(\sum_{j=1}^{n-i+1} D_{ij} + \alpha\right)} \left(\frac{y_k}{y_k + \sum_{j=1}^{n-i+1} y_j + \beta}\right)^{D_{ik}}$$

$$\left(\frac{\sum_{j=1}^{n-i+1} y_j + \beta}{y_k + \sum_{j=1}^{n-i+1} y_j + \beta}\right)^{\sum_{j=1}^{n-i+1} D_{ij} + \alpha}$$

从而

$$E(D_{ik} | D, y, \alpha, \beta) = \left(\frac{\beta}{\sum_{j=1}^{n-i+1} y_j + \beta}\right) \frac{\alpha}{\beta} +$$

$$\frac{\sum_{j=1}^{n-i+1} y_j}{\sum_{j=1}^{n-i+1} y_j + \beta} \frac{\sum_{j=1}^{n-i+1} D_{ij}}{\sum_{j=1}^{n-i+1} y_j} y_k$$

注意, 这也是加权平均。由于边际后验密度 $f(y, \alpha, \beta | D)$, 已经得到, 这样得到预测均值:

$$E(D_{ik} | D) = \iiint E(D_{ik} | D, y, \alpha, \beta)$$

$$f(y, \alpha, \beta | D) dy d\alpha d\beta$$

由于没有简单的解析式, 所以只有通过 MCMC 方法得到。

4 应用

使用本文提出的方法对表1“准备金流量三角”进行预测,该表来自 Verrall 和 England(2005)^[2]。取 $c_1 = \dots = c_n = 1$,对 α 和 β 使用低信息的先验分布, $a = 10$, $b = 0.0001$ $d = 0.01$ $e = 0.0001$,从而, α β 的先验方

差分别是 10^9 和 10^7 。

在表2中 $fit[i]$ 表示第 i 个事故年未来赔付总额, total 是所有事故年未来赔付的总额;表3是第9个事故年的从第3个发展年开始每个发展年赔付额的相关系数阵。

表2 参数估计值和准备金

参数	均值	MC error	2.50%	97.50%	参数	均值	MC error	2.50%	97.50%
alpha	183.1	2.986	93.86	283.8	x[5]	4872	0.5517	4715	5029
beta	0.03456	5.67E-04	0.01764	0.05363	x[6]	5114	0.5694	4947	5281
fit[2]	93340	91.34	6.60E+04	1.24E+05	x[7]	5633	0.8207	5440	5831
fit[3]	466100	212.9	4.10E+05	5.25E+05	x[8]	6657	2.135	6396	6923
fit[4]	705700	248	6.37E+05	7.77E+05	x[9]	5587	1.379	5292	5890
fit[5]	984200	281.6	9.03E+05	1.07E+06	x[10]	5075	1.769	4621	5548
fit[6]	1.42E+06	369.5	1.32E+06	1.52E+06	y[1]	0.06931	9.17E-06	0.06699	0.0717
fit[7]	2.16E+06	530.5	2.03E+06	2.29E+06	y[2]	0.173	1.65E-05	0.1691	0.1769
fit[8]	3.84E+06	1442	3.63E+06	4.05E+06	y[3]	0.181	1.47E-05	0.177	0.1851
fit[9]	4.23E+06	1197	3.97E+06	4.51E+06	y[4]	0.193	1.32E-05	0.1886	0.1974
fit[10]	4.72E+06	1679	4.28E+06	5.19E+06	y[5]	0.1068	1.09E-05	0.1032	0.1104
total	1.861E+7	5431.0	1.79E+7	1.934E+7	y[6]	0.07486	1.04E-05	0.07157	0.0783
x[1]	3948	0.8261	3826	4072	y[7]	0.06868	1.25E-05	0.06519	0.0723
x[2]	5427	0.4451	5282	5571	y[8]	0.04657	1.15E-05	0.04321	0.05
x[3]	5373	0.4834	5221	5526	y[9]	0.06954	1.76E-05	0.06441	0.0748
x[4]	5294	0.5291	5140	5452	y[10]	0.01719	1.43E-05	0.01337	0.0214

表3 第9个事故年未来每个发展年赔付额的相关系数阵

	D[9_3]	D[9_4]	D[9_5]	D[9_6]	D[9_7]	D[9_8]	D[9_9]	D[9_10]
D[9_3]	1	0.40315	0.32817	0.28676	0.27486	0.22042	0.25188	0.10718
D[9_4]		1	0.33414	0.28542	0.27857	0.23221	0.26665	0.11415
D[9_5]			1	0.24967	0.23498	0.17988	0.20204	0.09242
D[9_6]				1	0.2043	0.16572	0.17941	0.08227
D[9_7]					1	0.151	0.18772	0.0778
D[9_8]						1	0.14277	0.06392
D[9_9]							1	0.07303
D[9_10]								1

5 结束语

本文运用贝叶斯原理对准备金预测建模,分析了模型三个特点,从这些特点可以看出,该模型具有良好的性质。并且最后使用 MCMC 的方法给出了参数估计。

参考文献:

[1] Verrall R J. An investigation into stochastic claims reserving models and the chain-ladder technique [J]. Insurance: Mathematics and Economics 2000(26): 91-99.

[2] Verrall ,England. Incorporating expert opinion into a stochastic model for the chain-ladder technique [J]. Insurance: Mathematics and Economics ,2005 (37) : 355-370.
 [3] de Jong P. Forecasting runoff triangles [J]. North American Actuarial Journal 2006 ,10(2) : 28-38.
 [4] de Alba ,Nieto-Barajas. Claims reserving: A correlated Bayesian model [J]. Insurance: Mathematics and Economics 2008(43) : 368-376.
 [5] 陈明镜. 马氏转移对数正态模型参数置信区间的 Bootstrap 估计 [J]. 四川理工学院学报: 自然科学版 2010 23(3) : 267-269.

Bayesian Correlated Reserving for Development Years

CHEN Ming-jing

(College of Mathematics and Information , West China Normal University , Nanchong 637002 , China)

Abstract: Based on Generalized Linear Mixed Model we present a model for reserving using Bayesian technique. This model has three profiles: firstly , the payments in different development year are correlated; secondly , the mean of total payment in different accident years is random; thirdly , the posterior estimation of this mean is weight average of prior mean and realized data.

Key words: Generalized Linear Mixed Model; Bayesian; reserving; MCMC