精算建模：信度理论参考答案及批改评述（Chap 5）

庄源

日期：2025 年 9 月 21 日

目录

1 Question 5: Classical credibility on compound Poisson

1.1 原题

Annual aggregate claims are modeled by a compound Poisson distribution where the claim distribution is well approximated by a gamma distribution with shape parameter $2(X \sim \Gamma(2,\beta))$. Full credibility in estimating the pure premium $\theta = E(S)$ is defined by having a 95% chance of the estimator differing from the true value by at most 0.04θ. What partial credibility (using the square root rule) would be assigned to experience from 1200 claims?

1.2 参考答案

本题中的总索赔分布为复合 Poisson 分布，将 $k = 0.04$，$\alpha = 0.05$ 代入课本 162 面式 5.3 后可知，要达到经验数据的完全可信，所需的索赔数据量应满足：

$$r\lambda \geqslant \frac{Z^2_{1-\frac{\alpha}{2}}}{k^2}\frac{E(X^2)}{E^2(X)} = \left(\frac{1.96}{0.04}\right)^2 \times \frac{\left(\frac{2}{\beta}\right)^2 + \frac{2}{\beta^2}}{\left(\frac{2}{\beta}\right)^2} = 3601.5$$

已有索赔数据量为 1200 的情况下，部分信度因子为

$$Z = \sqrt{\frac{r\lambda}{n_{F(k,\alpha)}}} = \sqrt{\frac{1200}{3601.3677}} = 0.5772$$

1.3 给分标准与批改评价

表 1: Question 5 给分标准（共 15 分）

采分点	分值
计算最低所需数据量	7.5
使用平方根法则计算信度因子	7.5

本题完成情况很好，部分同学把方差当成了二阶矩代入公式造成错误。

2 Question 6: Mixture distribution

2.1 原题

The probability that an insured individual will give rise to no claims next years is $e^{-\theta}$, where θ varies by individuals according to the density function $f_\Theta(\theta) = 25\theta e^{-5\theta}$. What is a probability that a randomly selected individual will give rise to no claims next year?

2.2 参考答案

设个体索赔次数为随机变量 X，则有：

$$P(X=0|\Theta=\theta)=e^{-\theta} \quad (e^{-\theta}\in[0.1],\theta\geqslant 0)$$

$$f_{\Theta}(\theta)=25\theta e^{-5\theta}$$

所以，次年无索赔概率为

$$\begin{aligned}P(X=0)&=\int_0^{\infty}P(X=0|\Theta=\theta)f_{\Theta}(\theta)\mathrm{d}\theta\\&=\int_0^{\infty}e^{-\theta}\cdot 25\theta e^{-5\theta}\mathrm{d}\theta\\&=\int_0^{\infty}25\theta e^{-6\theta}\mathrm{d}\theta\\&=\frac{25}{36}\end{aligned}$$

2.3 给分标准与批改评价

表 2: Question 6 给分标准（共 5 分）

采分点	分值
正确使用条件概率公式	5

这道题运用了概率论中最简单的条件概率原理，完成情况较好。但是，这道题的含义很深刻：当没有后验信息（也就是观测样本），仅仅有条件分布和先验分布时，参数的分布可以看作是一个连续情况下的混合分布。当有了后验信息以后，我们才可以算后验分布。

有些同学把这道题想得太复杂，这道题不需要去计算 θ 的均值，直接使用条件概率公式就可以了。

3 Question 9: Bayesian credibility

3.1 原题

Claims in an automobile portfolio are modeled by a Burr distribution with density function given by

$$f(x\mid\theta)=\begin{cases}\dfrac{2\theta x}{(1+x^2)^{\theta+1}} & \text{for } x>0\\ 0 & \text{otherwise,}\end{cases}$$

where θ is an unknown parameter. Prior information suggests that a gamma distribution $\Gamma(42,20)$ is appropriate for θ. If a sample of $n=200$ claims is taken, determine the form of the Bayesian estimator for θ, assuming a quadratic loss function is used. Give an expression for a 90% Bayesian belief interval for θ.

3.2 参考答案

下面简单复习一下怎样算出后验分布和贝叶斯估计。

在我们以前学过的统计学中，我们认为，分布的参数本来就在那里，因此要把未知参数看作固定值。但贝叶斯统计学家们不这么看，他们认为未知参数本身也服从一个分布。随着我们了解到更多信息，我们将更清楚地认识到这个未知参数服从的分布。

一开始，我们会假设待研究的对象 x 满足一个分布 $f(x\mid\theta)$。只要 θ 给定，x 满足的分布就完全知道了。这里的 x 可以是单个案件的索赔额，也可以是发生案件的个数。

就像贝叶斯学派说的那样，θ 也服从一个分布。这个分布的确定具有主观性，也就是说，一开始 θ 服从的分布需要自己来确定。一开始确定的 θ 的分布被称为先验分布 $f_\Theta(\theta)$（Prior）。

现在，新信息，也就是观测到的样本来了，这群样本被写为 $\mathbf{X}$。知道了 $\mathbf{X}$ 以后，我们对于 θ 的看法又会有怎样的更新呢？也就是说，我们要知道在了解 $\mathbf{X}$ 的情况下，θ 的条件分布，即 $f_{\Theta|\mathbf{X}}(\theta\mid\mathbf{x})$。这时，贝叶斯定理就可以派上用场了。

$$f_{\Theta|\mathbf{X}}(\theta\mid\mathbf{x})=\frac{f_{\Theta,\mathbf{X}}(\theta,\mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})}=\frac{f_{\mathbf{X}|\Theta}(\mathbf{x}\mid\theta)f_\Theta(\theta)}{\int f_{\mathbf{X}|\Theta}(\mathbf{x}\mid\theta)f_\Theta(\theta)d\theta}\tag{*}$$

在上面的公式中，$f_{\Theta|\mathbf{X}}(\theta\mid\mathbf{x})$ 是 $f(x\mid\theta)$ 的乘积，$f_\Theta(\theta)$ 是先验分布，一般给定。计算(*)式并化简，我们便可以知道这个更新后的分布是什么，这个分布被称为后验分布（Posterior）。

然而，(*)式中的那一坨积分一般来说非常难算，还要涉及上下分子分母的约分，一道题就得算一天。不过没关系，我们可以通过观察，“猜”出后验分布到底是哪种分布：

$$f_{\Theta|\mathbf{X}}(\theta\mid\mathbf{x})\propto f_{\mathbf{X}|\Theta}(\mathbf{x}\mid\theta)f_\Theta(\theta)$$

上面的公式非常显然，但是却有更深的含义：(*)式的分母是 $\int f_{\mathbf{X}|\Theta}(\mathbf{x}\mid\theta)f_\Theta(\theta)d\theta$，积分完成以后，分母与 θ 没有关系。换言之，去观察(*)式分子中与 θ 有关的那部分，我们就可以知道后验分布中与 θ 有关的部分长什么样。因此，我们可以不用管(*)式分母那部分积分，先观察分子部分和 θ 有关的那部分内容。$\propto$ 指“正比于”。

由题，$\theta\sim\Gamma(42,20)$，所以有

$$\begin{aligned}
f_{\Theta|\mathbf{X}}(\theta|\mathbf{X}) &\propto \left[\prod_{i=1}^n f(x_i|\theta)\right]f_\Theta(\theta)\\
&=\frac{2^n\theta^n\prod_{i=1}^n x_i}{\left[\prod_{i=1}^n(1+x_i^2)\right]^{\theta+1}}\times\frac{20^{42}}{\Gamma(42)}e^{-20\theta}\theta^{41}\\
&\propto\frac{\theta^{n+41}e^{-20\theta}}{\left[\prod_{i=1}^n(1+x_i^2)\right]^{\theta+1}}\cdots\cdots\text{与 }\theta\text{ 无关的项都可以略掉，方便我们观察}\\
&=\frac{\theta^{241}e^{-20\theta}}{e^{(\theta+1)\sum_{i=1}^n\log(1+x_i^2)}}\cdots\cdots\text{本题 }n=200\text{，并做一个简单的指数等价变换}\\
&\propto\theta^{241}e^{-\left[\sum_{i=1}^n\log(1+x_i^2)+20\right]\theta}
\end{aligned}$$

可知后验分布为 $f_{\Theta|\mathbf{X}}(\theta|\mathbf{X})\sim Gamma\left(242,20+\sum_{i=1}^{240}\log(1+x_i^2)\right)$

在平方损失函数下，θ 的估计值为 $\hat\theta=E(\theta\mid\mathbf{X})=\dfrac{242}{20+\sum_{i=1}^{200}\log(1+x_i^2)}$，

关于 θ 的 90% 的置信区间（双侧 5% 区间）为 $\hat\theta\pm z_{0.95}\sqrt{Var(\theta|\mathbf{X})}$

即 $\left[\hat\theta-1.645\times\dfrac{\sqrt{242}}{20+\sum_{i=1}^{200}\log(1+x_i^2)},\hat\theta+1.645\times\dfrac{\sqrt{242}}{20+\sum_{i=1}^{200}\log(1+x_i^2)}\right]$。

3.3 给分标准与批改评价

表 3: Question 9 给分标准（共 20 分）

采分点	分值
给出后验分布的类型	10
计算平方损失下 θ 的估计值	5
得出 θ 的 90% 置信区间	5

这道题有点难度，但是大家做得非常出色。很多同学因为置信区间扣分，求 90% 置信区间，两侧应该各留 5%，也就是看 $z_{0.95}$ 而不是 $z_{0.90}$。大家不要人云亦云呀（笑）。

4 Question 16: Bühlmann credibility

4.1 原题

Annual claim numbers X_i are modeled by a Poisson distribution with unknown parameter λ. Prior knowledge about λ can be summarized by a uniform distribution on the interval $[60, 120]$.

(a) Determine the form if the Bühlmann credibility estimator of the pure premium $E(X_{n+1})$ based on a random sample $\mathbf{x}$ of claim numbers over n years.

(b) What would the estimate be if $X = (106, 105, 110, 98, 101, 113)$?

(c) Generate for yourself six observations from a Poisson distribution with mean 80, and determine the Bühlmann credibility estimate.

4.2 参考答案

4.2.1 a：信度因子的计算

要想计算信度因子，就必须算出 $E[s^2(\Theta)]$ 和 $Var[m(\Theta)]$。首先，

$$m(\Theta) = \lambda$$

$$s^2(\Theta) = \lambda$$

进而有

$$E[m(\Theta)] = \frac{1}{2} \times (60 + 120) = 90$$

$$Var[m(\Theta)] = \frac{1}{12} \times (120 - 60)^2 = 300 \cdots\cdots \text{Bernoulli Shortcut}$$

$$E[s^2(\Theta)] = \frac{1}{2} \times (60 + 120) = 90$$

由上式可得：$K = \dfrac{E[s^2(\Theta)]}{Var[m(\Theta)]} = \dfrac{90}{300} = 0.3$，所以 $Z = \dfrac{n}{n+0.3}$。因此

$$\begin{aligned} E(X_{n+1} \mid \mathbf{X} = \mathbf{x}) &= \frac{n}{n+0.3} \times \bar{x} + \frac{0.3}{n+0.3} \times E[m(\Theta)] \\ &= \frac{n}{n+0.3} \times \bar{x} + \frac{0.3}{n+0.3} \times 90 \end{aligned}$$

4.2.2 b&c：算例

由 a 可得，$n=6$，所以 $\bar{x}=\frac{1}{6}\times(106+105+110+98+101+103)=105.5$ 时，有

$$E(X_{n+1}\mid \mathbf{X}=\mathbf{x})=\frac{6}{6+0.3}\times 105.5+\frac{0.3}{6+0.3}\times 90=104.7619$$

使用 R 或 EXCEL，生成 6 个 $\lambda=80$ 的泊松分布随机数 $X=(83,85,96,90,76,68)$，有 $\bar{x}=\frac{1}{6}\times(83+85+96+90+76+68)=83$（同学们生成的随机数可与答案不同）。

$$E(X_{n+1}\mid \mathbf{X}=\mathbf{x})=\frac{6}{6+0.3}\times 83+\frac{0.3}{6+0.3}\times 90=83.33$$

4.3 给分标准与批改评价

表 4: Question 16 给分标准（共 20 分）

采分点	分值
分别给出 $E[s^2(\Theta)]$，$E[m(\Theta)]$ 和 $Var[m(\Theta)]$	12，每个 4 分
计算 $E(X_{n+1}\mid \mathbf{X}=\mathbf{x})$	5
生成随机数后再次计算 $E(X_{n+1}\mid \mathbf{X}=\mathbf{x})$	3，列式正确即给分

这道题的回答情况不佳。这道题的 a 问叫大家去寻找 Bühlmann credibility，大家只用把 $E[s^2(\Theta)]$，$E[m(\Theta)]$ 和 $Var[m(\Theta)]$ 算出来，然后根据信度因子加权就可以了。有很多同学在计算 $E[s^2(\Theta)]$ 的时候出错。

部分同学用贝叶斯方法去计算后验分布，但是这没有必要，这道题在题干中已明确说是 Bühlmann credibility 问题，不需要知道后验分布的类型也可以做得出来。还有些同学不知道怎么生成 Poisson 分布的随机数。在 R 中，可以使用 `rpois` 函数生成随机数；在 EXCEL 中，数据 → 数据分析 → 随机数发生器可以生成 Poisson 随机数。

5 Question 23: Empirical Bayes approach to credibility theory

注. 本题制作了计算所用的 EXCEL 表格，可下载学习。[下载]

5.1 原题

The data in Table 5 give the aggregate claims for household damage insurance in six successive years by four separate (regional) groups of policyholders. Assume that the claim amounts have been adjusted to remove any effect of inflation and that the unit of money is millions of dollars. Using empiritical Bayes credibility Model 1 , calculate estimates of the pure premiums for the coming year for each of the regions.

表 5: Household damage claims

	Year →	1	2	3	4	5	6
Region	A	206	146	271	178	136	162
	B	144	284	310	218	266	301
	C	64	57	43	97	132	110
	D	204	186	248	222	188	204

5.2 参考答案

设 x_{ij} 为风险类 i 在第 j 年的总索赔（其中，$i = A,\ B,\ C,\ D;\ j = 1, 2, \cdots, 6$）

设 θ_i 为 X_i 的先验分布参数，由表中数据可知，

$$\bar{x}_A = \frac{1}{6}\sum_{j=1}^{6} x_{A_j} = 183.1667 \qquad s_A^2 = \frac{1}{6-1}\sum_{j=1}^{6}(x_{A_j} - \bar{x}_A)^2 = 2463.3667$$

$$\bar{x}_B = \frac{1}{6}\sum_{j=1}^{6} x_{B_j} = 253.8333 \qquad s_B^2 = \frac{1}{6-1}\sum_{j=1}^{6}(x_{B_j} - \bar{x}_B)^2 = 3956.9667$$

$$\bar{x}_C = \frac{1}{6}\sum_{j=1}^{6} x_{C_j} = 83.8333 \qquad s_C^2 = \frac{1}{6-1}\sum_{j=1}^{6}(x_{C_j} - \bar{x}_C)^2 = 1191.7667$$

$$\bar{x}_D = \frac{1}{6}\sum_{j=1}^{6} x_{D_j} = 208.6667 \qquad s_D^2 = \frac{1}{6-1}\sum_{j=1}^{6}(x_{D_j} - \bar{x}_D)^2 = 541.8667$$

所以有

$$\widehat{E[m(\Theta)]} = \frac{1}{4}(\bar{x}_A + \bar{x}_B + \bar{x}_C + \bar{x}_D) = 182.375$$

进而有，

$$\widehat{E[s^2(\Theta)]} = \frac{1}{4}(s_A^2 + s_B^2 + s_C^2 + s_D^2) = 2038.4917$$

$$\widehat{Var[m(\Theta)]} = \frac{\sum(\bar{x}_i - \bar{x})^2}{N-1} - \frac{\widehat{E[s^2(\Theta)]}}{n} = 5169.544 - 339.7486 = 4829.7954$$

由上式得，$K = \dfrac{\widehat{E[s^2(\Theta)]}}{\widehat{Var[m(\Theta)]}} = 0.4221$，$\mathbf{Z} = \dfrac{6}{6+0.4221} = 0.9343$。

各区域的未来预期纯保费分别为：

$$A : \mathbf{Z} \cdot \bar{x}_A + (1-\mathbf{Z}) \times \widehat{E[m(\Theta)]} = 183.115$$

$$A : \mathbf{Z} \cdot \bar{x}_B + (1-\mathbf{Z}) \times \widehat{E[m(\Theta)]} = 249.137$$

$$A : \mathbf{Z} \cdot \bar{x}_C + (1-\mathbf{Z}) \times \widehat{E[m(\Theta)]} = 90.310$$

$$A : \mathbf{Z} \cdot \bar{x}_D + (1-\mathbf{Z}) \times \widehat{E[m(\Theta)]} = 206.939$$

5.3 给分标准与批改评价

表 6: Question 23 给分标准（共 20 分）

采分点	分值
分别给出 $\widehat{E[s^2(\Theta)]}$，$\widehat{E[m(\Theta)]}$ 和 $\widehat{Var[m(\Theta)]}$	12，每个 4 分
计算 K 和 $\mathbf{Z}$	4，每个 2 分
计算未来预期纯保费	4，每个 1 分

这道题大部分同学都做得很不错。部分同学在计算 $\widehat{Var[m(\Theta)]}$ 的时候出了错，忘了使用 $\frac{\widehat{E[s^2(\Theta)]}}{n}$ 做调整。

6 Question 24: Empirical Bayes approach to credibility theory

6.1 原题

Summary statistics on aggregate claims for each of four car rental company risks over five years are given in Table 7 below. Here X_{ij} denotes aggregate claims in year j for risk i. Using Model 1 of empirical Bayes credibility, calculate the credibility factor and credibility premium for risk 1. In your opinion, is Model 1 suitable to calculate credibility premiums for these risks?

表 7: Car rental company risks

Risk	$\bar{X}_i$	$\sum_{j=1}^{5}\left(X_{ij}-\bar{X}_i\right)^2$
1	$6,132$	$5,321,643$
2	$7,465$	$5,974,212$
3	$4,927$	$4,321,615$
4	$23,416$	$41,271,314$

6.2 参考答案

设各类风险的先验分布参数为 $\theta_i (i=1,2,3,4)$, 有:

$$\widehat{E[m(\Theta)]} = \frac{1}{4}\times(6132+7465+4927+23416) = 10485$$

$$\widehat{E[s^2(\Theta)]} = \frac{1}{4}\times(\frac{5321643}{5-1}+\frac{5974212}{5-1}+\frac{4321615}{5-1}+\frac{41271314}{5-1}) = 3555549$$

$$\begin{aligned}\widehat{Var[m(\Theta)]} = &\frac{1}{4-1}\left[(6132-10485)^2+(7465-10485)^2+(4927-10485)^2+(23416-10485)^2\right]\\ &-\frac{1}{5}\widehat{E[s^2(\Theta)]} = 74679268.2\end{aligned}$$

所以 $\mathbf{Z} = \dfrac{5}{5+3555549/74679268.2} = 0.9906$,

因此，风险 1 的预期保费为 $\mathbf{Z}\cdot 6132+(1-\mathbf{Z})\cdot\widehat{E[m(\Theta)]} = 6173.06$。

但是，使用 Model 1 来计算信度保费并不合适。从表格中可以看到，多年来风险 4 总索赔的平均值为 23416，相比于其它三种风险高了非常多。算出的 $\widehat{Var[m(\Theta)]}$ 相对于 $\widehat{E[s^2(\Theta)]}$ 很大，

进而导致信度因子 **Z** 的值接近 1。在 Model 1 中，不同风险共享同一个信度因子 **Z**，风险组 4 高 $\bar{X}_i$ 的存在让其它三个风险计算得到的信度保费几乎完全依赖于 $\bar{X}_i$，这是明显不合理的。可以使用 Model 2 的风险量加权方式，令不同风险适用不同信度因子。

6.3 给分标准与批改评价

表 8: Question 24 给分标准（共 20 分）

采分点	分值
分别给出 $E\widehat{[s^2(\Theta)]}$，$E\widehat{[m(\Theta)]}$ 和 $Var\widehat{[m(\Theta)]}$	12，每个 4 分
计算 **Z**	2
计算未来预期纯保费	2
评论 Model 1 是否适用本例	4

这道题中，大部分同学都能计算出未来纯保费，但是出现了下面这些错误：

- “Credibility premiums” 的中文翻译是“信度保费”，不是什么“风险溢价”！Premium 在英文中有“溢价”和“保费”两种翻译，在保险精算中一般指保费，在收益率曲线中一般指溢价。
- 部分同学没有评论 Model 1 是否适合计算信度保费，无法得到这一部分的分数。
- 有些同学认为，$\bar{X}_i$ 与最终保费很相近，所以 Model 1 非常适合，这不对。事实上，$\bar{X}_i$ 和最终保费相不相近，取决于信度因子 **Z** 有多大，如果太接近了，就说明信度因子很大。一个较大的信度因子可能表示我们有非常充足的数据，但现在这道题中只有 5 年的数据，并不算多。因此大的信度因子有可能反映了数据深层结构有关的问题。
- 批改时，只要同学们能够指出“K 值太小”、“信度因子太大”或“风险 4 代表的风险可能与其它风险异质”，都可得分。

7 批改评述总结

本次作业共五题，各题分值总结如下：

表 9: 各题分值分布及班级卷面平均分

题号	分值	全班均分（不含迟交、漏交和严重抄袭）
5	15	13.84
6	5	4.16
9	20	16.48
16	20	14.41
23	20	17.05
24	20	15.00
总分	100	81.42

同学们对于 Bühlmann 信度及其经验贝叶斯估计的了解需要加强。