经验贝叶斯参数估计

问题的提出：当我们对结构函数$\pi(\theta)$和条件分布$f_{X_i|\Theta}(x_i|\theta)$不是很清楚时，如何去计算 a, v, 和$\mu$。

分三种情况讨论

（1） 非参数估计：结构函数$\pi(\theta)$和条件分布$f_{X_i|\Theta}(x_i|\theta)$形式未知

（2） 半参数估计：结构函数$\pi(\theta)$未知，但条件分布$f_{X_i|\Theta}(x_i|\theta)$形式已知

（3） 参数估计：结构函数$\pi(\theta)$、条件分布$f_{X_i|\Theta}(x_i|\theta)$形式已知，但参数未知

下面我们将讨论（1）和（2）两种情况。

记号和假设

（1）损失

假设有 r 个投保人，

X_{ij} 表示第 i 个投保人，第 j 年的每风险单位的损失

$X_i=(X_{i1},X_{i2},...,X_{in_i})'$，表示第 i 个投保人的经验损失，$n_i$ 表示第 i 个投保人的观测值的个数

$\mathbf{X}=(X_1,X_2,...,X_r)$，表示全部投保人的损失，r 表示投保人的个数。

（2）风险参数

Θ_i 表示第 i 个投保人的风险参数，$\Theta_1,...,\Theta_r$ 独立同分布，

$\pi(\theta_i)$，$i=1,...,r$ 表示风险参数 Θ_i 的分布

给定 Θ_i，$X_{ij}, j=1,...,n_i$ 是相互独立的，分布为 $f_{X_{ij}|\Theta}(x_i\,|\,\theta_i)$

（3）风险单位数

m_{ij} 表示第 i 个投保人(团体)第 j 年的风险单位数，一般约定为 1

$m_i = \sum_{j=1}^{n_i} m_{ij}$，第 i 个投保人(团体)的风险单位数总和

$m = \sum_{i=1}^{r} m_i$，所有投保人(团体)的风险单位数的总和

（4）平均损失

$\bar{X}_i = \frac{1}{m_i} \sum_{j=1}^{n_i} m_{ij} X_{ij}$ 表示第 i 个投保人(团体)的每风险单位的平均损失

$\bar{X} = \frac{1}{m} \sum_{i=1}^{r} m_i \bar{X}_i = \frac{1}{m} \sum_{i=1}^{r} \sum_{j=1}^{n_i} m_{ij} X_{ij}$ 每风险单位的平均损失

例 10.19

年份	投保人									
	1	2	3	4	5	6	7	8	9	10
1									1	
2	1	1							1	
3	1		1						1	
4			1						1	
5					1	1			1	
6		1								
7	1	1		1	1					
8	1			1						
9	1									
10	1									
$\overline{S}_i$	0.6	0.3	0.2	0.2	0.2	0.1	0	0	0.7	0
S_i	0.23									

17. You are given the following commercial automobile policy experience:

	Company	Year 1	Year 2	Year 3
Losses Number of Automobiles	I	50,000 100	50,000 200	? ?
Losses Number of Automobiles	II	? ?	150,000 500	150,000 300
Losses Number of Automobiles	III	150,000 50	? ?	150,000 150

Determine the nonparametric empirical Bayes credibility factor, Z, for Company III.

Buhlmann－Straub 模型的记号

假定：在给定$\Theta_i=\theta_i$条件下，$X_{i1},\cdots,X_{in_i}$是相互独立的，具有相同的条件均值

$$\mu(\theta_i)=E(X_{ij}\mid\Theta_i=\theta_i) \tag{11}$$

但条件方差不相同

$$\operatorname{var}(X_{ij}\mid\Theta_i=\theta_i)=v(\theta_i)/m_{ij} \tag{12}$$

结构参数

$$\mu=E(\mu(\Theta_i)),$$
$$v=E(v(\Theta_i)),$$
$$a=\operatorname{var}(\mu(\Theta_i))$$

如何估计 a，v 和μ，计算$z_i=m_i/(m_i+k)$，k＝v/a

非参数估计

1、Bulhmann 模型

假设对任意 i，$n_i = n > 1$，$m_{ij} = 1(\forall i, j)$。假设第 i 个投保人的损失为

$$X_i = (X_{i1}, X_{i2}, ..., X_{in})' ,\quad i = 1, ..., r$$

给定$\Theta_i = \theta_i$，$X_{ij}, j = 1, ..., n_i$是相互独立的，且

$$\mu(\theta_i) = E(X_{ij} | \Theta = \theta_i)$$

$$v(\theta_i) = Var(X_{ij} | \Theta = \theta_i)$$

另外还假设两个不同的投保人之间损失是相互独立的，即如果$i \neq l$，则X_{ij}和X_{lt}是相互独立，求 Buhlmann 信度估计中 $a, v,$ 和 μ 的无偏估计。

（1） μ 的无偏估计

考虑统计量

$$\bar{X}=r^{-1}\sum_{i=1}^{r}\bar{X}_i=(rn)^{-1}\sum_{i=1}^{r}\sum_{j=1}^{n}X_{ij}$$

下面证明 $\bar{X}$ 是无偏估计

$$E(\bar{X})=(rn)^{-1}\sum_{i=1}^{r}\sum_{j=1}^{n}E(X_{ij})=(rn)^{-1}\sum_{i=1}^{r}\sum_{j=1}^{n}E(E(X_{ij}\mid\Theta_i))$$

$$=(rn)^{-1}\sum_{i=1}^{r}\sum_{j=1}^{n}E(\mu(\Theta_i))=(rn)^{-1}\sum_{i=1}^{r}\sum_{j=1}^{n}\mu=\mu$$

（2）v 的无偏估计

考虑统计量

$$\hat{v}_i = \frac{1}{n-1}\sum_{j=1}^{n}(X_{ij} - \bar{X}_i)^2$$

由于给定$\Theta_i = \theta_i$， $X_{ij}, j = 1,...,n$是相互独立的， $v(\theta_i) = Var(X_{ij} \mid \Theta = \theta_i)$，

因此给定$\Theta_i = \theta_i$， $\hat{v}_i$是$v(\theta_i)$的无偏估计，且

$$E(\hat{v}_i) = E(E(\hat{v}_i \mid \Theta_i)) = E(v(\Theta_i)) = v$$

因此$\hat{v}_i$是v的无偏估计。于是统计量

$$\hat{v} = \frac{1}{r}\sum_{i=1}^{r}\hat{v}_i = \frac{1}{r(n-1)}\sum_{i=1}^{r}\sum_{j=1}^{n}(X_{ij} - \bar{X}_i)^2$$

是v的无偏估计。

（3）a 的无偏估计

由于

$$E(\bar{X}_i \mid \Theta_i = \theta_i) = n^{-1}\sum_{j=1}^{n} E(X_{ij} \mid \Theta_i = \theta_i) = n^{-1}\sum_{j=1}^{n} \mu(\theta_i) = \mu(\theta_i)$$

因此

$$E(\bar{X}_i) = E(E(\bar{X}_i \mid \Theta_i)) = E(\mu(\Theta_i)) = \mu$$

$$\begin{aligned} Var(\bar{X}_i) &= Var(E(\bar{X}_i \mid \Theta_i)) + E(Var(\bar{X}_i \mid \Theta_i)) \\ &= Var(\mu(\Theta_i)) + E(v(\Theta_i)/n)) = a + v/n \end{aligned}$$

因此，$\bar{X}_1,...,\bar{X}_r$ 相互独立，具有相同的均值 μ 和方差 $a+v/n$。从而统计量

$$(r-1)^{-1}\sum_{i=1}^{r}(\bar{X}_i - \bar{X})^2$$ ，其中 $\bar{X} = r^{-1}\sum_{i=1}^{r}\bar{X}_i$ 是样本均值

是 $a+v/n$ 的无偏估计。于是 a 的无偏估计等于

$$\hat{a}=\frac{1}{r-1}\sum_{i=1}^{r}(\bar{X}_i-\bar{X})^2-\frac{\hat{v}}{n}$$

$$=\frac{1}{r-1}\sum_{i=1}^{r}(\bar{X}_i-\bar{X})^2-\frac{1}{rn(n-1)}\sum_{i=1}^{r}\sum_{j=1}^{n}(X_{ij}-\bar{X}_i)^2$$

直观解释

$\hat{v}$ ：组内均方误差

$\hat{a}$ 中的第一项：组间均方误差

当组间均方误差 $\hat{v}$ 相对组内均方误差较小时，即 $\hat{a}$ 小于 $\hat{v}$，k 较大，$\hat{Z}$ 将接近于 0，$\bar{X}_i$ 信度较小。

计算出来的 $\hat{a}$ 可能小于 0，这时令 $\hat{a}=\hat{Z}=0$。这种情况等价于方差分析中 F 检验统计量小于 1，导致无法拒绝等均值假设。

例 10.20：假设有两个投保人，他们在 3 年内的理赔额分别为 $\vec{x}_1=(3,5,7)^T$， $\vec{x}_2=(6,12,9)^T$，求他们各自的 Buhlmann 信度估计。

解：

$$\begin{aligned}\bar{X}_1&=(3+5+7)/3=5\\ \bar{X}_2&=(6+12+9)/3=9\end{aligned} \Longrightarrow \bar{X}=(5+9)/2=7$$

$$\hat{v}_1=[(3-5)^2+(5-5)^2+(7-5)^2]/2=4$$

$$\hat{v}_2=?$$

$$\hat{v}=(4+9)/2=13/2$$

$$\begin{aligned}\hat{a}&=\frac{1}{r-1}\sum_{i=1}^{r}(\bar{X}_i-\bar{X})^2-\frac{\hat{v}}{n}\\ &=[(5-7)^2+(9-7)^2]-\frac{1}{3}\times\frac{13}{2}=\frac{35}{6}\end{aligned}$$

$$\hat{k}=\frac{\hat{v}}{\hat{a}}=\frac{39}{35},\hat{z}=\frac{3}{3+k}=\frac{35}{48}$$

投保人 1 的信度估计 $\hat{z}\overline{X}_1+(1-z)\hat{\mu}=\left(\frac{35}{48}\right)\times 5+\frac{13}{48}\times 7=\frac{133}{24}$

投保人 2 的信度估计 $\hat{z}\overline{X}_2+(1-z)\hat{\mu}=\left(\frac{35}{48}\right)\times 9+\frac{13}{48}\times 7=\frac{203}{24}$

课堂讨论:

例 10.21: 假设某保险公司开发一新险种，保单组合由 10 位投保人构成。开始，由于没有任何理赔经验数据，只能先验地假定他们具有相似的风险水平。然后假定每一投保人每年至多引发一次理赔，且理赔额为 1。最初，根据同行业的损失水平，估计这一保单组合的保费为 0.2，我们称这种保费为先验保费，或集体保费。

这样的估计是否符合实际情形，需要经验数据来验证。为了搜集足够的理赔数据，保险公司连续追踪十年，采集的全部数据显示于下表。

请分析下表的数据，说明保费收取是否合理，该如何改进。

年份	投保人									
	1	2	3	4	5	6	7	8	9	10
1									1	
2	1	1							1	
3	1		1						1	
4			1						1	
5					1	1			1	
6		1								
7	1	1		1	1					
8	1			1						
9	1									
10	1									
$\bar{S}_i$	0.6	0.3	0.2	0.2	0.2	0.1	0	0	0.7	0
S_i	0.23									

2、Buhlmann－Straub 模型

假设

$$E(X_{ij}\mid\Theta_i=\theta_i)=\mu(\theta_i)$$

$$\operatorname{var}(X_{ij}\mid\Theta_i=\theta_i)=v(\theta_i)/m_{ij}\text{，}\quad j=1,\ldots,n_i$$

下面求 a，v 和 μ 的无偏估计。

（1）μ 的无偏估计

考虑统计量

$$\bar{X}=\frac{1}{m}\sum_{i=1}^{r}m_i\bar{X}_i=\frac{1}{m}\sum_{i=1}^{r}\sum_{j=1}^{n_i}m_{ij}X_{ij}$$

由于 $E(X_{ij})=E(E(X_{ij}\mid\Theta_i))=E(\mu(\Theta_i))=\mu$，因此

$$E(\bar{X}_i\mid\Theta_i)=\sum_{j=1}^{n_i}\frac{m_{ij}}{m_i}E(X_{ij}\mid\Theta_i)=\sum_{j=1}^{n_i}\frac{m_{ij}}{m_i}\mu(\Theta_i)=\mu(\Theta_i)$$

从而

$$E(\bar{X}_i) = E(E(\bar{X}_i \mid \Theta_i)) = E(\mu(\Theta_i)) = \mu$$

$$E(\bar{X}) = \frac{1}{m}\sum_{i=1}^{r} m_i E(\bar{X}_i) = \frac{1}{m}\sum_{i=1}^{r} m_i \mu = \mu$$

所以 $\bar{X}$ 是 μ 的无偏估计。

（2）v 的无偏估计

考虑统计量

$$\hat{v}_i = \frac{\sum_{j=1}^{n_i} m_{ij}(X_{ij} - \bar{X}_i)^2}{n_i - 1}, \quad i = 1, \ldots, r$$

由于

$$\sum_{j=1}^{n_i} m_{ij}(X_{ij}-\bar{X}_i)^2$$

$$=\sum_{j=1}^{n_i} m_{ij}(X_{ij}-\mu(\Theta_i)-(\bar{X}_i-\mu(\Theta_i)))^2$$

$$=\sum_{j=1}^{n_i} m_{ij}(X_{ij}-\mu(\Theta_i))^2-2\sum_{j=1}^{n_i} m_{ij}(X_{ij}-\mu(\Theta_i))(\bar{X}_i-\mu(\Theta_i))+\sum_{j=1}^{n_i} m_{ij}(\bar{X}_i-\mu(\Theta_i))^2$$

$$=\sum_{j=1}^{n_i} m_{ij}(X_{ij}-\mu(\Theta_i))^2-m_i(\bar{X}_i-\mu(\Theta_i))^2$$

因此

$$
\begin{aligned}
&E\left[\sum_{j=1}^{n_i} m_{ij}(X_{ij}-\bar{X}_i)^2 \mid \Theta_i\right] \\
&=\sum_{j=1}^{n_i} m_{ij}E\left[(X_{ij}-\mu(\Theta_i))^2 \mid \Theta_i\right]-m_iE\left[(\bar{X}_i-\mu(\Theta_i))^2 \mid \Theta_i\right] \\
&=\sum_{j=1}^{n_i} m_{ij}Var(X_{ij} \mid \Theta_i)-m_iVar(\bar{X}_i \mid \Theta_i) \\
&=\sum_{j=1}^{n_i} m_{ij}v(\Theta_i)/m_{ij}-v(\Theta_i)=v(\Theta_i)(n_i-1)
\end{aligned} \tag{*}
$$

因此

$$E(\hat{v}_i \mid \Theta_i)=v(\Theta_i)$$

$$E(\hat{v}_i)=E(E(\hat{v}_i \mid \Theta_i))=E(v(\Theta_i))=v$$

所以$\hat{v}_i$是 v 的无偏估计。

令 $\hat{v}=\Sigma_{i=1}^{r}w_i\hat{v}_i$，其中 w_i 表示权重，$0<w_i<1, \Sigma_{i=1}^{n}w_i=1$，则 $\hat{v}$ 是 v 的无偏估计。这里我们选择 $w_i=(n_i-1)/\Sigma_{i=1}^{n}(n_i-1)$，则

$$\hat{v}=\frac{\sum_{i=1}^{r}\sum_{j=1}^{n_i}m_{ij}(X_{ij}-\bar{X}_i)^2}{\sum_{i=1}^{r}(n_i-1)}$$

是 v 的无偏估计

（3）a 的无偏估计

由于固定 i，$X_{i1},X_{i2},\ldots,X_{in_i}$ 是关于 Θ_i 是条件独立的，因此

$$Var(\bar{X}_i \mid \Theta_i) = Var(\frac{1}{m_i}\sum_{j=1}^{n_i} m_{ij} X_{ij} \mid \Theta_i)$$

$$= \sum_{j=1}^{n_i} \left(\frac{m_{ij}}{m_i} \right)^2 Var(X_{ij} \mid \Theta_i)$$

$$= \sum_{j=1}^{n_i} \left(\frac{m_{ij}}{m_i} \right)^2 \frac{v(\Theta_i)}{m_{ij}}$$

$$= \frac{v(\Theta_i)}{m_i^2} \sum_{j=1}^{n_i} m_{ij} = \frac{v(\Theta_i)}{m_i}$$

$$Var(\bar{X}_i) = Var(E(\bar{X}_i \mid \Theta_i)) + E(Var(\bar{X}_i \mid \Theta_i))$$

$$= Var(\mu(\Theta_i)) + E(v(\Theta_i / m_i))$$

$$= a + v / m_i$$

则 $\bar{X}_1, ..., \bar{X}_r$ 相互独立，具有相同的均值 μ 和方差 $a + v / m_i$ 。

令

$$\bar{X}=\frac{1}{m}\sum_{i=1}^{r}m_i\bar{X}_i$$

利用(*)类似的方法，可以证明

$$E(\sum_{i=1}^{r}m_i(\bar{X}_i-\bar{X})^2)=a\left(m-m^{-1}\sum_{i=1}^{r}m_i^2\right)+v(r-1)$$

a 的无偏统计量为

$$\hat{a}\quad=\left(m-m^{-1}\sum_{i=1}^{r}m_i^2\right)^{-1}\left(\sum_{i=1}^{r}m_i(\bar{X}_i-\bar{X})^2-\hat{v}(r-1)\right)$$

17. You are given the following commercial automobile policy experience:

	Company	Year 1	Year 2	Year 3
Losses	I	50,000	50,000	?
Number of Automobiles		100	200	?
Losses	II	?	150,000	150,000
Number of Automobiles		?	500	300
Losses	III	150,000	?	150,000
Number of Automobiles		50	?	150

Determine the nonparametric empirical Bayes credibility factor, Z, for Company III.

解答

The subscripts denote the three companies.

$$x_{I1}=\frac{50,000}{100}=500,\quad x_{I2}=\frac{50,000}{200}=250,\quad x_{II1}=\frac{150,000}{500}=300$$

$$x_{II2}=\frac{150,000}{300}=500,\quad x_{III1}=\frac{150,000}{50}=3,000,\quad x_{III2}=\frac{150,000}{150}=1,000$$

$$\overline{x}_{I}=\frac{100,000}{300}=333.33,\quad \overline{x}_{II}=\frac{300,000}{800}=375,\quad \overline{x}_{III}=\frac{300,000}{200}=1,500,\quad \overline{x}=\frac{700,000}{1,300}=538.46$$

$$\hat{v} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{n_i} m_{ij}(X_{ij} - \bar{X}_i)^2}{\sum_{i=1}^{r}(n_i - 1)}$$

$$\hat{v} = \frac{\begin{array}{c}100(500-333.33)^2 + 200(250-333.33)^2 + 500(300-375)^2 + 300(500-375)^2 \\ + 50(3,000-1,500)^2 + 150(1,000-1,500)^2\end{array}}{(2-1)+(2-1)+(2-1)}$$

$$= 53,888,888.89$$

$$\hat{a} = \left(m - m^{-1}\sum_{i=1}^{r} m_i^2\right)^{-1}\left(\sum_{i=1}^{r} m_i(\bar{X}_i - \bar{X})^2 - \hat{v}(r-1)\right),$$ $m_1=300, m_2=800, m_3=200$

$$\hat{a} = \frac{300(333.33-538.46)^2 + 800(375-538.46)^2 + 200(1,500-538.46)^2 - 53,888,888.89(3-1)}{1,300 - \dfrac{300^2 + 800^2 + 200^2}{1,300}}$$

$$= 157,035.60$$

$$k=\frac{53,888,888.89}{157,035.60}=343.1635,\ Z=\frac{200}{200+343.1635}=0.3682$$

几种特殊情况的讨论：

$\hat{a}<0$

$n_i=1$

$r=1$，

μ 已知

a，v 和 μ 的无偏估计以及更深入的讨论，请阅读 loss model

平衡调整

值得注意的是，虽然所有参数的估计值都是无偏的，但这并不一定就说明经验信度估计值也是无偏的。事实上我们可以求出所有被保险人的总损失 $TL=\sum_{i=1}^{r} m_i \bar{X}_i$。依照上述模型，如果在过去也按照上面计算出的信度保费收取保费，总保费应该是

$$
\begin{aligned}
TP &= \sum_{i=1}^{r} m_i [\hat{Z}_i \bar{X}_i + (1-\hat{Z}_i)\hat{\mu}] \\
&= \sum_{i=1}^{r} m_i \bar{X}_i - \sum_{i=1}^{r} m_i [(1-\hat{Z}_i)\bar{X}_i - (1-\hat{Z}_i)\hat{\mu}] \\
&= TL - \sum_{i=1}^{r} \frac{m_i \hat{k}}{m_i + \hat{k}} (\bar{X}_i - \hat{\mu})
\end{aligned}
$$

理想情况是TL等于TP。因为任何增加保费的行为要想得到监管机构的同意，必须要以过去总的索赔水平为依据。信度保费需要有着理论和现实意义，同时如果还能够保持总保费收入与总损失相匹配就更好了。这应该有

$$\sum_{i=1}^{r}\frac{m_i\hat{k}}{m_i+\hat{k}}(\bar{X}_i-\hat{\mu})=0$$

即$\sum_{i=1}^{r}\hat{Z}_i(\bar{X}_i-\hat{\mu})=0$。因此

$$\hat{\mu}=\frac{\sum_{i=1}^{r}\hat{Z}_i\bar{X}_i}{\sum_{i=1}^{r}\hat{Z}_i}$$

μ 的估计是个体样本均值的信度加权平均，而非是直接的按风险量加权平均。它的好处在于估计先验均值时考虑到了每个个体样本的可信度，使得总保费等于总损失。

例 10.22：假设两组被保险人在过去三年的经验数据和第四年的被保险人数如下所示：

被保险人		保单年度			
		1	2	3	4
1	总损失	-	11,000	18,000	-
	团体人数	-	50	80	70
2	总损失	20,000	25,000	24,000	-
	团体人数	100	120	125	100

求（1）对两组被保险人第四年应收取的信度保费。

（2）使用信度加权平均估计μ，两组被保险人各自应缴纳的第四年的信度保费。

（3）如果第一组保险人的第二年总损失数据不是18,000而是15,000，

重新计算上面问题。

解：1）根据数据，可以看出 $r=2$，$n_1=2$，$n_2=3$。对于第一组被保险人团体，下标表示哪一年对运算没有影响，方便起见，我们下标采取的年份和表中对应，

$$m_{12}=50\text{，}\quad X_{12}=\frac{11000}{50}=220\text{，}\quad m_{13}=80\text{，}\quad X_{13}=\frac{18000}{80}=225\text{，}$$ 以及

$$m_1=m_{12}+m_{13}=130\text{，}\quad \bar{X}_1=\frac{11000+18000}{130}=223.08$$

对于第二组被保险人团体 $m_{21}=100$，$X_{21}=\dfrac{20000}{100}=200$，$m_{22}=120$，

$$X_{22} = \frac{25000}{120} = \frac{625}{3}, \quad m_{23} = 125, \quad X_{23} = \frac{24000}{125} = 192,$$

$$m_2 = m_{21} + m_{22} + m_{23} = 345, \quad m = m_1 + m_2 = 475$$

以及 $\bar{X}_2 = \dfrac{20000 + 25000 + 24000}{345} = 200$

于是 $\hat{\mu} = \bar{X} = \dfrac{223.08 \times 130 + 200 \times 345}{130 + 345} = 206.31$

$$\hat{\upsilon} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{n_i} m_{ij}(X_{ij} - \bar{X}_i)^2}{\sum_{i=1}^{r}(n_i - 1)}$$

$$= \frac{\begin{array}{c} 50\times(220-223.08)^2 + 80\times(225-223.08)^2 + 100\times(200-200)^2 \\ +120\times(625/3-200)^2 + 125\times(192-200)^2 \end{array}}{1+2}$$

$$= 5700.855$$

$$\hat{a}=[\sum_{i=1}^{r}m_i(\bar{X}_i-\bar{X})^2-\hat{\upsilon}(r-1)]/(m-\frac{\sum_{i=1}^{r}m_i^{\ 2}}{m})$$

$$=\frac{130\times(223.08-206.31)^2+345\times(200-206.31)^2-5700.855\times1}{475-(130^2+345^2)/475}$$

$$=236.15$$

因此$\hat{k}=\frac{\hat{\upsilon}}{\hat{a}}=24.15$， $\hat{Z}_1=\frac{m_1}{m_1+\hat{k}}=0.84$， $\hat{Z}_2=\frac{m_2}{m_2+\hat{k}}=0.93$

被保险人团体 1 和被保险人团体 2 的个体平均信度保费估计值分别为

$$\hat{Z}_1\bar{X}_1+(1-\hat{Z}_1)\hat{\mu}=0.84\times223.08+(1-0.84)\times206.31=220.45$$

$$\hat{Z}_2\bar{X}_2+(1-\hat{Z}_2)\hat{\mu}=0.93\times 200+(1-0.93)\times 206.31=200.41$$

于是，被保险人团体 1 和被保险人团体 2 下一年要交的总的信度保费估计分别为 $220.45\times 70=15431.5$ 和 $200.41\times 100=20041$

（2）从上题中我们知道 $\hat{Z}_1=0.84$， $\hat{Z}_2=0.93$,，得到

$$\hat{\mu}=\frac{\sum_{i=1}^{r}\hat{Z}_i\bar{X}_i}{\sum_{i=1}^{r}\hat{Z}_i}=\frac{0.84\times 223.08+0.93\times 200}{0.84+0.93}=210.95$$

于是被保险人团体 1 和被保险人团体 2 的个体平均信度保费估计值分别为

$$\hat{Z}_1\bar{X}_1+(1-\hat{Z}_1)\hat{\mu}=0.84\times 223.08+(1-0.84)\times 210.95=221.18$$

$$\hat{Z}_2\bar{X}_2+(1-\hat{Z}_2)\hat{\mu}=0.93\times 200+(1-0.93)\times 206.31=200.72$$

于是，被保险人团体 1 和被保险人团体 2 下一年要交的总的信度保费估计分别为 $221.18\times 70=15482.6$ 和 $200.72\times 100=20072$

（3）此时，$X_{13}=\dfrac{15000}{80}=187.5$，$\bar{X}_1=\dfrac{11000+15000}{130}=200$，其他不变，经计算可得 $\hat{\mu}=200$，$\hat{\upsilon}=16888.57$，

$$\hat{a} = [\sum_{i=1}^{r} m_i (\bar{X}_i - \bar{X})^2 - \hat{\upsilon}(r-1)] / (m - \frac{\sum_{i=1}^{r} m_i^2}{m})$$

$$= \frac{130 \times (200-200)^2 + 345 \times (200-200)^2 - 16888.57 \times 1}{475 - (130^2 + 345^2)/475}$$

$$= -89.43$$

此时 $\hat{a} < 0$，因此取 $\hat{Z}_1 = \hat{Z}_2 = 0$，两个团体中的个体平均保费估计均等于 $\hat{\mu} = 200$。

两个被保险人团体总的信度保费估计值分别为 14000 和 200000。

半参数估计

假设 $f_{X_{ij}|\Theta}(x_{ij}|\theta_i)$ 的分布已知，但 $\pi(\theta_i)$ 的参数未知。根据 $f_{X_{ij}|\Theta}(x_{ij}|\theta_i)$ 的情况具体分析。

考虑理赔次数模型，假设给定 $\Theta_i=\theta_i$，第 i 个投保人在第 j 年的理赔次数 $m_{ij}X_{ij}$ 服从 Poisson 分布，参数为 $m_{ij}\theta_i$，则

$$E(m_{ij}X_{ij}|\Theta_i)=Var(m_{ij}X_{ij}|\Theta_i)=m_{ij}\Theta_i$$

$$\mu(\Theta_i)=v(\Theta_i)=\Theta_i$$

$$\mu=v$$

因此，我们可以用 $\hat{\mu}=\bar{X}$ 去估计 v。

例 10.23：某汽车险的一个被保险人的理赔次数服从 Poisson 分布，参数为θ， θ随被保险人变化而变化。得到去年的理赔记录如下：

理赔次数	被保险人数
0	1563
1	271
2	32
3	7
4	2
合计	1875

对每个被保险人，求理赔次数的信度估计。

解：r＝1875，n_i＝1，m_{ij}＝1，

对投保人 i，（$i=1,\ldots,1875$），$X_{i1}\,|\,\Theta_i=\theta_i$服从 Poisson 分布，均值为$\theta_i$，

$\mu=v$

$$\bar{X}=\frac{1}{1875}\left(\sum_{i=1}^{1875}X_{i1}\right)$$

$$=\frac{0(1563)+1\times 271+2\times 32+3\times 7+4\times 2}{1875}=0.194$$

$$Var(X_{i1})=Var(E(X_{i1}\,|\,\Theta_i))+E(Var(X_{i1}\,|\,\Theta_i))$$

$$=Var(\mu(\Theta_i))+E(v(\Theta_i))=a+v=a+\mu$$

$a+v$的无偏估计是样本方差

$$\frac{\sum_{i=1}^{1875}(X_{i1}-\bar{X})^2}{1874}=\frac{1563(0-0.194)^2+271(1-0.194)^2+...+2(4-0.194)^2}{1874}$$

$$=0.226$$

因此

$a=0.226-0.194=0.032$， $k=0.194/0.032=6.06$

$$z=1/(1+6.06)=0.14$$

所以一个被保险人的理赔次数的信度估计为

$$0.14X_{i1}+0.86\times0.194$$

例 10.24：考虑某团体保险的个体理赔发生概率。设θ_i表示保单持有人（团体）i 的理赔发生概率。X_{ij}表示理赔发生，即

$$X_{ij}=\begin{cases}1 & \text{理赔发生}\\ 0 & \text{理赔不发生}\end{cases}$$

m_{ij}表示保单持有人（团体）i 的第 j 年的个体数，则$m_{ij}X_{ij}$表示保单持有人（团体）i 在第 j 年发生的理赔总数，建立信度模型。

解：显然$m_{ij}X_{ij}$服从参数为m_{ij},θ_i的二项分布，因此

$$E(m_{ij}X_{ij}\mid\Theta_i)=m_{ij}\Theta_i$$

$$Var(m_{ij}X_{ij}\mid\Theta_i)=m_{ij}\Theta_i(1-\Theta_i)$$

因此

$$\mu(\Theta_i)=\Theta_i,v(\Theta_i)=\Theta_i(1-\Theta_i)$$

于是

$$\mu=E(\Theta_i),$$

$$v=\mu-E(\Theta_i)^2$$

$$a=Var(\Theta_i)=E(\Theta_i)^2-\mu^2=\mu-v-\mu^2$$

37. For a portfolio of motorcycle insurance policyholders, you are given:

(i) The number of claims for each policyholder has a conditional Poisson distribution.

(ii) For Year 1, the following data are observed:

Number of Claims	Number of Policyholders
0	2000
1	600
2	300
3	80
4	20
Total	3000

Determine the credibility factor, Z, for Year 2.

这是半参数模型
解答

The sample mean is $\frac{0(2000)+1(600)+2(300)+3(80)+4(20)}{3000}=0.5066667=\hat{\mu}=\hat{v}$ and the sample variance is

$$\frac{2000(0-\hat{\mu})^2+600(1-\hat{\mu})^2+300(2-\hat{\mu})^2+80(3-\hat{\mu})^2+20(4-\hat{\mu})^2}{2999}=0.6901856.$$ Then,

$\hat{a}=0.6901856-0.5066667=0.1835189$, $k=\frac{0.5066667}{0.1835189}=2.760842$ and

$$Z=\frac{1}{1+2.760842}=0.2659.$$

7. The following information comes from a study of robberies of convenience stores over the course of a year:

(i) X_i is the number of robberies of the i^{th} store, with $i = 1, 2, ..., 500$.

(ii) $\sum X_i = 50$

(iii) $\sum X_i^2 = 220$

(iv) The number of robberies of a given store during the year is assumed to be Poisson distributed with an unknown mean that varies by store.

Determine the semiparametric empirical Bayes estimate of the expected number of robberies next year of a store that reported no robberies during the studied year.

解

$\hat{v} = \bar{x} = \dfrac{50}{500} = 0.1$ and

$$\hat{a} = s^2 - 0.1 = \frac{\left[220 - 500(0.1)^2\right]}{499} - 0.1 = 0.33086.$$

Then $\hat{Z} = \dfrac{1}{1 + \dfrac{0.1}{0.33086}} = 0.76791.$

The credibility estimate is then $0.76791(0) + 0.23209(0.1) = 0.023209.$

15.-16. *(Repeated for convenience) Use the following information for questions 15 and 16.*

Survival times are available for four insureds, two from Class A and two from Class B. The two from Class A died at times $t=1$ and $t=9$. The two from Class B died at times $t=2$ and $t=4$.

16. *For question 16 only, you are also given*:

Nonparametric Empirical Bayes estimation is used to estimate the mean survival time for each class. Unbiased estimators of the expected value of the process variance and the variance of the hypothetical means are used.

Estimate Z, the Bühlmann credibility factor.

第 15 题的解

$$W(1,b)=2+2e^b,\ W(2,b)=1+2e^b,\ \hat{H}_0(3)=\frac{1}{2e^b+2}+\frac{1}{2e^b+1}.$$

第 16 题的解

$$\hat{v}=\frac{1}{2(2-1)}\left[(1-5)^2+(9-5)^2+(2-3)^2+(4-3)^2\right]=17.$$

$$\hat{a}=\frac{1}{2-1}\left[(5-4)^2+(3-4)^2\right]-\frac{17}{2}=-6.5<0.$$

作业：

27. You are given the following information on towing losses for two classes of insureds, adults and youths:

Exposures

Year	Adult	Youth	Total
1996	2000	450	2450
1997	1000	250	1250
1998	1000	175	1175
1999	1000	125	1125
Total	5000	1000	6000

Pure Premium

Year	Adult	Youth	Total
1996	0	15	2.755
1997	5	2	4.400
1998	6	15	7.340
1999	4	1	3.667
Weighted average	3	10	4.167

You are also given that the estimated variance of the hypothetical means is 17.125.

Determine the nonparametric empirical Bayes credibility premium for the youth class, using the method that preserves total losses.