第二节 贝叶斯方法在信度模型中的应用

一、贝叶斯模型

设风险子集的一个投保人的风险水平可以通过一个参数θ来描述，但是θ的取值随投保人的不同而不同。这样，通过不同取值θ，我们可以区分不同投保人的风险水平的差异。θ具有如下特点：

θ是存在的，但θ是不可观察的，而且我们永远不知道θ的真实值。

θ随投保人变化，可以看作随机变量Θ的观察值，在风险子集内存在一个关于Θ的概率分布$\pi(\theta)$

假定$\pi(\theta)$是已知的，但每个投保人的具体风险参数θ的值都是未知的。

记号：

设投保人的风险参数为Θ，则

$F_{X|\Theta}(x|\theta)=F(x;\theta)$ 为投保人的损失分布函数

$\mu(\Theta)=E(X|\Theta)$ 为投保人的期望损失，称为风险保费。

$\sigma^2(\Theta)=\mathrm{var}(X|\Theta)=E[(X-\mu(\Theta))^2|\Theta]$ 为投保人的方差。

设$\pi(\Theta)$为Θ的分布密度。则任意选取一个投保人的损失 X 的分布为

$$F_X(x)=\int F_{X\Theta}(x|\theta)d\pi(\theta)$$

投保人的平均损失为

$$\mu=E[X]=E[\mu(\Theta)]$$

投保人的方差为

$$\mathrm{var}(X)=E(\mathrm{var}(X\mid\Theta))+\mathrm{var}(E(X\mid\Theta))=v+a$$

其中

- $v=E[\sigma^2(\Theta)]$ 为组内方差的均值（average variance within subgroups）
- $a=\mathrm{var}[\mu(\Theta)]$ 为各组均值的方差(variance between subgroup averages)
- μ称为集体保费（手册保费），可视为不知投保人任何信息时，对其征收的理想保费。

二、如何续收保费

假设我们已经知道了某投保人过去经验值 $\vec{\mathbf{X}}=\vec{\mathbf{x}}=(x_1,\cdots,x_n)$，如何预测明年的平均损失额?

（1）如果 θ 已知，那么明年的损失 X_{n+1} 的期望为

$$u_{n+1}(\theta)=E(X_{n+1}\,|\,\Theta=\theta)=\int X_{n+1}f_{X_{n+1}|\Theta}(x_{n+1}\,|\,\theta)dx_{n+1}$$

因此，对于一个已知风险水平为 θ 的投保人，保险人收取它的风险保费（risk premium）$\mu_{n+1}(\theta)$。$\mu_{n+1}(\theta)$ 最能反映投保人的损失水平。

注：$\mu_{n+1}(\theta)$ 有时也称为 individual premium。

（2）如果保险人对投保人一无所知，那么保险人就收取平均保费（collective premium）

$$\mu_{n+1} = E(E(X_{n+1} | \Theta)) = E(\mu_{n+1}(\Theta))$$

（3）如果保险人知道投保人的过去的损失经验 $\bar{\mathbf{X}} = (x_1, \cdots, x_n)$，但风险水平未知。那么保险人将会使用

$$E(X_{n+1} | \bar{\mathbf{X}} = \bar{\mathbf{x}})$$

做为对 $\mu_{n+1}(\theta)$ 的预测，这种估计称为贝叶斯估计(Bayesian premium)。

三、贝叶斯保费的计算

假设$X_1, X_2 \cdots, X_n$独立，投保人的风险参数为θ，θ未知。

设给定$\Theta=\theta$时，X_j的分布为$f_{X_j|\Theta}(x_j|\theta)$，则$\vec{X}=(X_1, X_2 \cdots, X_n)$的条件分布密度为$\prod_{j=1}^{n} f_{X_j|\Theta}(x_j|\theta)$。

设Θ的分布密度为$\pi(\Theta)$，则$(X_1, X_2 \cdots, X_n, \Theta)$的联合密度为

$$f_{\vec{X}|\Theta}(x_1, x_2 \cdots x_n|\theta)\pi(\theta)=[\prod_{j=1}^{n} f_{X_j|\Theta}(x_j|\theta)\pi(\theta)]$$

$\vec{X}=(X_1, X_2 \cdots, X_n)$的边缘密度为

$$f_{\vec{X}}(x_1, x_2 \cdots, x_n) = \int f_{\vec{X},\Theta}(\vec{x}, \theta) d\theta$$

$$= \int \prod_{j=1}^{n} f_{X_j|\Theta}(x_j \mid \theta)\pi(\theta) d\theta$$

于是由条件分布公式, 可得到已知 $\vec{X} = \vec{x}$ 条件下, Θ 的后验分布(posterior distribution)

$$\pi_{\Theta|\bar{X}}(\theta \mid \bar{x}) = \frac{f_{\vec{X},\Theta}(\vec{x}, \theta)}{f_{\vec{X}}(\vec{x})} = \frac{1}{f_{\bar{X}}(\bar{x})}\left(\prod_{j=1}^{n} f_{X_j|\Theta}(x_j \mid \theta)\pi(\theta)\right)$$

这时，给定 $(X_1, X_2, \cdots, X_n)$， X_{n+1} 的分布(the predictive distribution)可以写为

$$
\begin{aligned}
f_{X_{n+1}|\vec{X}}(x_{n+1}\,|\,\vec{x}) &= \frac{1}{f_{\vec{X}}(\overline{x})} f_{X_{n+1},\vec{X}}(x_{n+1},\vec{x}) \\
&= \frac{1}{f_{\vec{X}}(\vec{x})} \int \prod_{j=1}^{n+1} f_{X_j|\Theta}(x_j\,|\,\theta)\pi(\theta)d\theta \\
&= \int f_{X_{n+1}|\Theta}(x_{n+1}\,|\,\theta)\pi_{\Theta|\vec{X}}(\theta\,|\,\vec{x})d\theta
\end{aligned}
$$

这是一个混合分布。因此贝叶斯估计$E(X_{n+1}\,|\,\vec{X}=\vec{x})$为

$$
\begin{aligned}
E(X_{n+1}\,|\,\vec{X}=\vec{x}) &= \int x_{n+1} f_{X_{n+1}|\vec{X}}(x_{n+1}\,|\,\vec{x})dx_{n+1} \\
&= \iint x_{n+1} f_{X_{n+1}|\Theta}(x_{n+1}\,|\,\theta)\pi_{\Theta|\vec{X}}(\theta\,|\,\vec{x})d\theta dx_{n+1} \\
&= \int \mu_{n+1}(\theta)\pi_{\Theta|\vec{X}}(\theta\,|\,\vec{x})d\theta
\end{aligned}
$$

例 10.12： 某投保人的理赔额 X 的分布密度是

$$f(x|b)=\frac{2x}{b^2},0<x<b$$

其中 b 的先验分布是 $g(b)=\frac{1}{b^2},1<b<\infty$。已知该投保人上一次的理赔额为 2。求他下次理赔额的期望是多少。

解： 根据全概率公式，X_1 的密度为

$$f(x)=\int_1^{\infty} f(x|b)g(b)db=\int_2^{\infty}\frac{2x}{b^2}\frac{1}{b^2}db=\frac{x}{12}$$

注意积分区域是从 2 到∞，因为如果 $b<2$，X_1 不可能等于 2。给定 $X_1=2$，b 的后验分布为

$$\pi_{b|X_1}(b|2)=\frac{f(2|b)\pi(b)}{f(2)}=\frac{4\cdot 6}{b^4}=\frac{24}{b^4},b>2$$

$$\pi_{b|X_1}(b|2)=0,b\le 2$$

X 的条件期望为

$$E(X\,|\,b)=\int_0^b xf(x|b)dx=\int_0^b x\cdot\frac{2x}{b^2}db=\frac{2b}{3}$$

因此，他下次理赔额的期望为

$$E(X_2\,|\,X_1=2)=\int_2^\infty \frac{2b}{3}\cdot\frac{24}{b^4}\,db=2$$

练习：例 10.13

45. You are given:

(i) The amount of a claim, X, is uniformly distributed on the interval $[0,\theta]$.

(ii) The prior density of θ is $\pi(\theta)=\dfrac{500}{\theta^2}$, $\theta>500$.

Two claims, $x_1=400$ and $x_2 600$, are observed. You calculate the posterior distribution as:

$$f(\theta\,|\,x_1,x_2)=3\left(\frac{600^3}{\theta^4}\right),\quad \theta>600$$

Calculate the Bayesian premium, $E(X_3\,|\,x_1,x_2)$.

解：

$$E(X \mid \theta) = \theta / 2$$

$$E(X_3 \mid 400, 600) = \int_{600}^{\infty} E(X \mid \theta) f(\theta \mid 400, 600) d\theta = \int_{600}^{\infty} \frac{\theta}{2} 3 \frac{600^3}{\theta^4} d\theta$$

$$= \frac{3(600^3)}{2} \frac{\theta^{-2}}{-2} \Bigg|_{600}^{\infty} = \frac{3(600^3)(600^{-2})}{4} = 450.$$

例 10.14：假设有某类投保人团体，其中每个人的每次理赔额都为常数，但理赔次数都服从参数为θ的泊松分布。假设该团体在第j年的人数为m_j，因此该团体的总理赔次数 N_j 服从均值为$m_j\theta$的泊松分布，即

$$P(N_j = n \mid \Theta = \theta) = \frac{(m_j\theta)^n e^{-m_j\theta}}{n!} \quad n = 0,1,2\cdots$$

风险参数Θ服从 gamma 分布

$$\pi(\theta) = \frac{\theta^{\alpha-1} e^{-\theta/\beta}}{\Gamma(\alpha)\beta^{\alpha}} \quad \theta > 0$$

如果每年的理赔额的通货膨胀率为$100r\%$，即 0 时的损失c，在t时的值为$(1+r)^t c$，假设人均保费等于人均理赔额期望，求在$n+1$年有m_{n+1}个投保人的团体的总贝叶斯保费。

解：不失一般性，设 X_j 表示第 j 年该投保团体的人均理赔总额，

$$X_j = c(1+r)^j N_j / m_j \qquad j=1,2,\cdots,n$$

因此， X_j 的条件分布为

$$f_{X_j|\Theta}(x_j \mid \theta) = P(N_j = c^{-1}(1+r)^{-j} m_j x_j \mid \Theta = \theta)$$

$$= \frac{[m_j\theta]^{c^{-1}(1+r)^{-j} m_j x_j} e^{-m_j\theta}}{[c^{-1}(1+r)^{-j} m_j x_j]!}$$

则 $(X_1, X_2 \cdots, X_n, \Theta)$ 的联合分布密度为

$$f_{\bar{X},\Theta}(\vec{x},\theta) = (\prod_{j=1}^{n} f_{X_j|\Theta}(x_j \mid \theta)\pi(\theta)$$

Θ 后验分布为

$$\pi_{\Theta|\vec{X}}(\theta\,|\,\vec{x}) = \frac{f_{\vec{X},\Theta}(\vec{x},\theta)}{f_{\vec{x}}(\vec{x})}$$

$$= \frac{\left(\prod_{j=1}^{n}[m_j\theta]^{c^{-1}(1+r)^{-j}m_jx_j}e^{-m_j\theta}\right)\theta^{\alpha-1}e^{-\theta/\beta}}{\prod_{j=1}^{n}[c^{-1}(1+r)^{-j}m_jx_j]!\,\Gamma(\alpha)\beta^{\alpha}\cdot f_{\vec{X}}(\vec{x})}$$

$$= A\theta^{\alpha+c^{-1}\sum_{j=1}^{n}(1+r)^{-j}m_jx_j-1}e^{-\theta\left(\frac{1}{\beta}+\sum_{j=1}^{n}m_j\right)}$$

$$\propto \theta^{\alpha+c^{-1}\sum_{j=1}^{n}(1+r)^{-j}m_jx_j-1}e^{-\theta\left(\frac{1}{\beta}+\sum_{j=1}^{n}m_j\right)}$$

其中 A 是与 θ 无关的常数。请问 $\pi_{\Theta|\vec{X}}(\theta\,|\,\vec{x})$ 是什么分布？

从后验分布的形式可以看出，$\pi_{\Theta|\bar{X}}(\theta\,|\,\vec{x})$ 是一个 gamma 分布，

$$\pi_{\Theta|\bar{X}}(\theta\,|\,\vec{x})=\frac{\left(\beta_*\right)^{-\alpha_*}}{\Gamma\left(\alpha_*\right)}\theta^{\alpha_*-1}e^{-\theta/\beta_*}$$

其中

$$\alpha_*=\alpha+c^{-1}\sum_{j=1}^{n}(1+r)^{-j}m_jx_j \qquad \beta_*=(\frac{1}{\beta}+\sum_{j=1}^{n}m_j)^{-1}$$

同学们可验证

$$A=\frac{1}{\left[\int_0^{\infty}\theta^{\alpha+c^{-1}\sum_{j=1}^{n}(1+r)^{-j}m_jx_j-1}e^{-\theta(\frac{1}{\beta}+\sum_{j=1}^{n}m_j)}d\theta\right]}=\frac{\left(\beta^*\right)^{-\alpha^*}}{\Gamma\left(\alpha^*\right)}\text{。}$$

另外，X_j 的条件期望为

$$\mu_j(\theta) = E(X_j \mid \Theta = \theta) = E(\frac{c(1+r)^j N_j}{m_j} \mid \Theta = \theta)$$

$$= \frac{c(1+r)^j}{m_j} E(N_j \mid \Theta = \theta)$$

$$= c(1+r)^j \theta$$

n＋1 年的人均风险保费为

$$\mu_{n+1}(\theta) = E(X_{n+1} \mid \Theta = \theta) = c(1+r)^{n+1}\theta$$

$$\mu_{n+1} = E(X_{n+1}) = E(c(1+r)^{n+1}\Theta) = c(1+r)^{n+1}\alpha\beta$$

人均贝叶斯保费为：

$$
\begin{aligned}
E(X_{n+1} \mid \vec{X} = \vec{x}) &= \int_0^{\infty} \mu_{n+1}(\theta)\pi_{\Theta|\vec{x}}(\theta \mid \vec{x})d\theta \\
&= E(\mu_{n+1}(\theta) \mid \vec{X} = \vec{x}) \\
&= E(c(1+r)^{n+1}\Theta \mid \vec{X} = \vec{x}) \\
&= c(1+r)^{n+1}\alpha_* \beta_* \\
&= c(1+r)^{n+1}(\alpha + c^{-1}\sum_{j=1}^{n}(1+r)^{-j} m_j x_j)(\frac{1}{\beta} + \sum_{j=1}^{n} m_j)
\end{aligned}
$$

例题讨论

上式可化简为

$$E(X_{n+1} \mid \vec{X} = \vec{x}) == z\bar{x} + (1-z)\mu_{n+1}$$

其中， $z = \dfrac{m}{m+\beta^{-1}}$ $\bar{x} = m^{-1}\sum_{j=1}^{n}(1+r)^{n+1-j} m_j x_j$ 为该团体在 n 年内的经

验人均保费，$\mu_{n+1}=E(\mu_{n+1}(\Theta))=c(1+r)^{n+1}\alpha\beta$ 为任意投保人的纯保费，因此贝叶斯保费可以看作是该团体经验保费与纯保费的加权平均。

特别的，若 c＝1，r＝1，m_j＝1，则 $X_j\equiv N_j, j=1,\cdots,n$。此时

$$E(X_{n+1}\mid\vec{X}=\vec{x})=z\bar{x}+(1-z)\mu$$

其中 $z=n/(n+\beta^{-1})$，$\bar{x}=n^{-1}\sum_{j=1}^{n}x_j$，$\mu=\alpha\beta$。

m_{n+1} 个体的贝叶斯保费即为：

$$m_{n+1}E(X_{n+1}\mid\vec{X}=\vec{x})$$

结论

贝叶斯信度估计的做法是用经验数据的信息代替风险参数，求 $E(X_{n+1}|\overrightarrow{X}=\vec{x})$。从直观上，我们可以想象贝叶斯信度估计值优于条件期望的期望，因为它更加充分地利用了已知的信息。事实上，贝叶斯信度估计值是所有利用经验数据估计 X_{n+1} 中误差最小的估计量，见下面的定理。

定理在所有利用 $\overrightarrow{X}$ 来估计 X_{n+1} 的估计量中，贝叶斯信度估计值的均方误差最小，即 $E\{[X_{n+1}-E(X_{n+1}|\overrightarrow{X})]^2\}$ 小于任何其他估计量的均方误差。