贝叶斯估计

一、条件分布

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。

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

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

$\sigma^2(\Theta)=\mathrm{var}(X\mid\Theta)=E[(X-\mu(\Theta))^2\mid\Theta]$ 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

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

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。

例1：假设有两个骰子，

D_1 的2面有标记，另外4面没有标记，

D_2 的3面有标记，另外的3面没有标记。

有标记面出现，表示损失事故发生，反之，则说明损失没有发生。

还假设有两个转盘，

S_1　有三扇标记着5，另外两扇标记着10；

S_2 有一扇标记着5，另外四扇标记着10。

当损失发生，则转动转盘，指数指向的扇数所标记的值为损失事件的损失额（索赔额）。

设 X 表示投保人的损失，求 X 的条件分布。

解：每个投保人的风险特征都可以用一个骰子和一个转盘来表示。因

此，投保人的风险可以分为 4 个等级

$$\theta_{11}=(D_1,S_1),\theta_{12}=(D_1,S_2),\theta_{21}=(D_2,S_1),\theta_{22}=(D_2,S_2)$$

假设每个转盘和骰子的选取都是随机的，那么，$P(\Theta=\theta_{ij})=\frac{1}{2}\frac{1}{2}=\frac{1}{4}$。

设 X 表示投保人的损失，则我们可以计算得到对于上述四类风险特征的人的损失分布如下：

$$P(X=0|\Theta=\theta_{11})=\frac{4}{6}=\frac{10}{15}$$

$$P(X=5|\Theta=\theta_{11})=\frac{2}{6}\frac{3}{5}=\frac{3}{15}$$

$$P(X=10|\Theta=\theta_{11})=\frac{2}{6}\frac{2}{5}=\frac{2}{15}$$

二、参数的后验分布

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

$$\begin{aligned}f_{\vec{X}}(x_1,x_2\cdots,x_n)&=\int f_{\vec{X},\Theta}(\vec{x},\theta)d\theta\\&=\int\prod_{j=1}^{n} f_{X_j|\Theta}(x_j|\theta)\pi(\theta)d\theta\end{aligned}$$

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)

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

例 2

You are given:

(i) The probability that an insured will have exactly one claim is θ.

(ii) The prior distribution of θ has probability density function:

$$\pi(\theta)=\frac{3}{2}\sqrt{\theta},\quad 0<\theta<1$$

A randomly chosen insured is observed to have exactly one claim.

Calculate the posterior probability that θ is greater than 0.60.

解：

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

$$=\theta(1.5\theta^{0.5})\propto\theta^{1.5}$$

因为$\int_0^1\theta^{1.5}d\theta=0.4$，所以

$$\pi_{\Theta|\bar{X}}(\theta\,|\,1)=2.5\theta^{1.5}$$

$$\Pr(\theta>0.6\,|\,1)=\int_{0.6}^{1}2.5\theta^{1.5}d\theta=\theta^{2.5}\Big|_{0.6}^{1}=1-0.6^{2.5}=0.721$$

三、预测分布和贝叶斯估计

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

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

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

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

例3 例1（续）：假设有两个骰子:

D_1 的2面有标记，另外4面没有标记，

D_2 的3面有标记，另外的3面没有标记。

有标记面出现，表示损失事故发生，反之，则说明损失没有发生。

还假设有两个转盘:

S_1　有三扇标记着5，另外两扇标记着10;

S_2 有一扇标记着5，另外四扇标记着10。

当损失发生，则转动转盘，指数指向的扇数所标记的值为损失事件的损失额（索赔额）。

假设我们已知某投保人的上一次的损失 $X_1＝5$，求他下次的损失额的期望，即$E(X_2|X_1)$.

解：根据例 1，经计算得到 X 的条件分布为

对于$\Theta=\theta_{11}$，

$$P(X=0|\Theta=\theta_{11})=\frac{10}{15},\quad P(X=5|\Theta=\theta_{11})=\frac{3}{15},\quad P(X=10|\Theta=\theta_{11})=\frac{2}{15}$$

对于$\Theta=\theta_{12}$

$$P(X=0|\Theta=\theta_{12})=\frac{10}{15},\quad P(X=5|\Theta=\theta_{12})=\frac{1}{15},\quad P(X=10|\Theta=\theta_{12})=\frac{4}{15}$$

对于$\Theta=\theta_{21}$

$$P(X=0|\Theta=\theta_{21})=\frac{5}{10},\quad P(X=5|\Theta=\theta_{21})=\frac{3}{10},\quad P(X=10|\Theta=\theta_{21})=\frac{2}{10}$$

对于$\Theta=\theta_{21}$

$P(X=0|\Theta=\theta_{22})=\frac{5}{10}$, $P(X=5|\Theta=\theta_{21})=\frac{1}{10}$, $P(X=10|\Theta=\theta_{21})=\frac{4}{10}$

$$\pi(\theta_{ij})=\frac{1}{4} \qquad i=1,2,\ j=1,2$$

$$f_{X_1}(x_1)=\sum_{i=1}^{2}\sum_{j=1}^{2} f_{X_1|\Theta}(x_1|\theta_{ij})\pi(\theta_{ij})=\frac{1}{4}\sum_{i=1}^{2}\sum_{j=1}^{2}P(X=x_1|\Theta=\theta_{ij})$$

将 X 的条件分布代入可得

$$f_{X_1}(0)=\frac{1}{4}(\frac{10}{15}+\frac{10}{15}+\frac{5}{10}+\frac{5}{10})=\frac{7}{12}$$

$$f_{X_2}(5)=\frac{1}{4}(\frac{3}{15}+\frac{1}{15}+\frac{3}{10}+\frac{1}{10})=\frac{2}{12}$$

$$f_{X_3}(13)=\frac{1}{4}(\frac{2}{15}+\frac{4}{15}+\frac{2}{10}+\frac{4}{10})=\frac{3}{12}$$

因此，由贝叶斯公式有

$$\pi_{\Theta|X_1}(\theta \mid x_1) = \frac{f_{X_1|\Theta}(x_1 \mid \theta)\pi(\theta)}{f_{X_1}(x_1)}$$

当 $X_1 = 5$ 时，

$$\pi_{\Theta|X_1}(\theta_{11} \mid 5) = \frac{5}{13}\frac{1}{4} / \frac{2}{12} = \frac{6}{20}$$

$$\pi_{\Theta|X_1}(\theta_{12} \mid 5) = \frac{1}{15}\frac{1}{4} / \frac{2}{12} = \frac{2}{20}$$

$$\pi_{\Theta|X_1}(\theta_{21} \mid 5) = \frac{3}{10}\frac{1}{4} / \frac{2}{12} = \frac{9}{20}$$

$$\pi_{\Theta|X_1}(\theta_{22} \mid 5) = \frac{1}{10}\frac{1}{4} / \frac{2}{12} = \frac{3}{20}$$

$$u_{n+1}(\theta_{11}) = 0\times\frac{10}{15} + 5\times\frac{3}{15} + 10\times\frac{2}{15} = \frac{7}{3}$$

$$u_{n+1}(\theta_{12}) = 0\times\frac{10}{15} + 5\times\frac{1}{15} + 10\times\frac{4}{15} = 3$$

类似可以计算$u_{n+1}(\theta_{21}) = \dfrac{7}{2}, u_{n+1}(\theta_{22}) = \dfrac{9}{2}$，则

纯保费 $u_2 = E(X_2) = \sum_{i=1}^{2}\sum_{j=1}^{2} u_{n+1}(\theta_{ij})\pi(\theta_{ij}) = \dfrac{1}{4}(\dfrac{7}{3} + 3 + \dfrac{7}{2} + \dfrac{9}{2}) = \dfrac{10}{3}$

贝叶斯保费为

$$E(X_2 \mid X_1 = 5) = \left(\frac{7}{3}\right)\left(\frac{6}{20}\right) + 3\left(\frac{2}{20}\right) + \left(\frac{7}{2}\right)\left(\frac{9}{20}\right) + \left(\frac{9}{2}\right)\left(\frac{3}{20}\right) = \frac{13}{4}$$

例 4：设两个坛中装有标记为 0 或 1 的球，球的比例如下

	0 的比例	1 的比例
1	0.6	0.4
2	0.8	0.2

先随机选取一个坛子，然后在该坛子中抽取 3 个球（有放回），标记的数总和为 2，求如果再一次从该坛子有放回的抽取 2 个球，则标记的总和的期望等于多少？

解：$\pi(\theta_1)=\pi(\theta_2)=\dfrac{1}{2}$

X_1 表示第一次抽取 3 个球的标记数的总和

X_2 表示第二次抽取的 2 个球的标记数的和

Y_i 表示第 i 个球的标记， $i=1,2\cdots 5$ ，显然 Y_i 是独立同分布

下面求 $E(X_2 \mid X_1)$

显然有 $X_1 = Y_1 + Y_2 + Y_3, X_2 = Y_4 + Y_5$ ， 给定 Θ ， X_1 ， X_2 服从二项分布。

$$f_{X_1|\Theta}(2 \mid \theta_1) = P(X_1 = 2 \mid \Theta = \theta_1) = \binom{3}{2} \times 0.4^2 \times 0.6 = 0.288$$

$$f_{X_2|\Theta}(2 \mid \theta_2) = P(X_1 = 2 \mid \Theta = \theta_2) = \binom{3}{2} \times 0.2^2 \times 0.8 = 0.096$$

$$f_{X_1}(2) = P(X_1 = 2 \mid \Theta = \theta_2)\pi(\theta_1) + P(X_1 = 2 \mid \Theta = \theta_2)\pi(\theta_2)$$
$$= 0.288 \times \frac{1}{2} + 0.096 \times \frac{1}{2} = 0.192$$

给定 X_1＝2，Θ的后验分布为

$$\pi_{\Theta|X_1}(\theta_1 \mid 2) = \frac{f_{X_1|\Theta}(2 \mid \theta_1)\pi(\theta_1)}{f_{X_2}(2)} = \frac{0.288 \times \frac{1}{2}}{0.192} = 0.75$$

$$\pi_{\Theta|X_2}(\theta_2 \mid 2) = \frac{f_{X_1|\Theta}(2 \mid \theta_2)\pi(\theta_2)}{f_{X_2}(2)} = \frac{0.096 \times \frac{1}{2}}{0.192} = 0.25$$

下面计算 X_2 的贝叶斯估计。由于

$$E(Y_4 \mid \Theta = \theta_1) = 0 \times 0.6 + 1 \times 0.4 = 0.4$$

$$E(Y_5 \mid \Theta = \theta_2) = 0 \times 0.8 + 1 \times 0.2 = 0.2$$

因此

$$u_2(\theta_i) = E(X_2 \mid \Theta = \theta_i) = E(Y_4 \mid \Theta = \theta_i) + E(Y_5 \mid \Theta = \theta_i)$$

$$u_2(\theta_1) = E(X_2 \mid \Theta = \theta_1) = 2E(Y_4 \mid \Theta = \theta_1) = 0.8$$

$$u_2(\theta_2) = E(X_2 \mid \Theta = \theta_2) = 2E(Y_4 \mid \Theta = \theta_2) = 0.4$$

$$\begin{aligned} E(X_2 \mid X_1 = 2) &= u_2(\theta_1)\pi_{\Theta|X_1}(\theta_1 \mid 2) + u_2(\theta_2)\pi_{\Theta|X_1}(\theta_2 \mid 2) \\ &= 0.8 \times 0.75 + 0.4 \times 0.25 \\ &= 0.7 \end{aligned}$$

课堂练习

You are given:
(i) The probability that an insured will have at least one loss during any year is p.

(ii) The prior distribution for p is uniform on [0, 0.5].

(iii) An insured is observed for 8 years and has at least one loss every year.

Calculate the posterior probability that the insured will have at least one loss during Year 9.

Question #15
Key: A

$$\pi(p \mid 1,1,1,1,1,1,1,1) \propto \Pr(1,1,1,1,1,1,1,1 \mid p)\pi(p) \propto p^8$$

$$\pi(p \mid 1,1,1,1,1,1,1,1) = \frac{p^8}{\int_0^{0.5} p^8 dp} = \frac{p^8}{0.5^9 / 9} = 9(0.5^{-9})p^8$$

$$\Pr(X_9 = 1 \mid 1,1,1,1,1,1,1,1) = \int_0^{0.5} \Pr(X_9 = 1 \mid p)\pi(p \mid 1,1,1,1,1,1,1,1)dp$$

$$= \int_0^{0.5} p9(0.5^{-9})p^8 dp = 9(0.5^{-9})(0.5^{10}) / 10 = 0.45.$$

例 5：(SOA 1104-5)You are given:

(i) Two classes of policyholders have the following severity distributions:

Claim Amount	Probability of Claim Amount for Class 1	Probability of Claim Amount for Class 2
250	0.5	0.7
2,500	0.3	0.2
60,000	0.2	0.1

(ii) Class 1 has twice as many claims as Class 2.

A claim of 250 is observed.

Determine the Bayesian estimate of the expected value of a second claim from the same policyholder.

解：设θ_1和θ_2分别表示 class1 和 class2， $\pi(\theta_1)=\dfrac{2}{3},\pi(\theta_2)=\dfrac{1}{3}$

给定$X_1=250$， Θ的后验分布为

$$\pi_{\Theta|X_1}(\theta_1\,|\,250)=\frac{f_{X_1|\Theta}(250\,|\,\theta_1)\pi(\theta_1)}{f_{X_1}(250)}$$

$$=\frac{0.5(2/3)}{0.5(2/3)+0.7(1/3)}=\frac{10}{17}$$

类似可计算出

$$\pi_{\Theta|X_1}(\theta_2\,|\,250)=?$$

$$\mu(\theta_1)=0.5\times 250+0.3\times 2500+0.2\times 60000=12875$$

$$\mu(\theta_2)=0.7\times 250+0.2\times 2500+0.1\times 60000=6675$$

从而，X 的贝叶斯估计为

$$E(X_2 \mid X_1 = 250) = u_2(\theta_1)\pi_{\Theta|X_1}(\theta_1 \mid 250) + u_2(\theta_2)\pi_{\Theta|X_1}(\theta_2 \mid 250)$$

$$= \frac{10}{17} \times 12875 + \frac{7}{17} \times 6675 = 10322$$

例 6：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

$$E(X \mid b)=\int_0^b x f(x \mid b) dx=\int_0^b x \cdot \frac{2x}{b^2} db=\frac{2b}{3}$$

因此，他下次理赔额的期望为

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