贝叶斯估计

贝叶斯方法是最古老的构造估计量的方法，由托马斯·贝叶斯在 18 世纪末提出。

这一方法建立在一种**表达不确定性的哲学**基础上。

在这种哲学的形式中，统计模型不包含唯一的参数值来对应“真实”状态，而是每个参数值都有一个概率，该概率在必要时可以通过主观的、个人的方式确定。这一方法的主观成分引发了许多批评。

然而，20 世纪 90 年代以来，贝叶斯方法的实际应用越来越受到欢迎，因为最初的计算问题现在可以通过计算机模拟解决。

贝叶斯方法的起点是除了指定统计模型（或似然函数）外，还需要在参数空间 Θ 上指定一个所谓的**先验概率分布**。

先验分布要么通过特定的假设选取，要么作为对不同参数值的概率的先验、可能带有主观性的估计。例如，对于具有参数 $\theta \in [0,1]$ 的二项分布变量 X，可以选择均匀分布作为 θ 的先验分布。

然后，通过应用概率论中的贝叶斯法则将此先验分布调整为当前数据，即后验概率分布。我们将首先介绍构造估计量方法的贝叶斯方法，并在之后详细描述先后验概率分布的调整。

为简单起见，我们取先验分布为具有密度 π 的连续分布，即在 Θ 上的任意概率密度。估计量 T 对于实值参数 $g(\theta)$ 的贝叶斯风险定义为带有权重 π 的均方误差 $\mathrm{MSE}(\theta;T)$ 的加权平均：

$$R(\pi;T) = \int \mathrm{E}_\theta(T - g(\theta))^2 \pi(\theta) d\theta$$

这是一种对估计量 T 的"质量"的衡量方法，并且对先验中更可能的值 θ 赋予更高的权重，换句话说这些 θ 上的误差会更加重要一些。

贝叶斯估计量定义为在该准则下最优的估计量。目标是找到一个使 $\mathrm{MSE}(\theta;T)$ (和平常不一样的是这是一个带有先验分布权重的 MSE) 对所有 θ 都较小的估计量；通过对不同的 θ 值赋予权重使准则更具体化。

贝叶斯估计量

相对于先验密度 π 的贝叶斯估计量是所有估计量 T 中使 $R(\pi;T)$ 最小的 T。

在下述定理中，贝叶斯估计量用两个积分的商表示。设 $x \mapsto p_\theta(x)$ 为随机向量 X 的概率密度。

贝叶斯估计定理

相对于先验密度 π 的 $g(\theta)$ 的贝叶斯估计为：

$$T(x) = \frac{\int g(\theta) p_\theta(x) \pi(\theta) d\theta}{\int p_\vartheta(x) \pi(\vartheta) d\vartheta}$$

因此，贝叶斯估计依赖于似然函数 $\theta \mapsto p_\theta(x)$ 和先验密度 π。最大似然估计量定义为似然函数达到最大值的点，而贝叶斯估计则是该函数的一种加权平均。

例 指数分布

令 $X=(X_1,\ldots,X_n)$ 为来自具有未知参数 θ 的指数分布的样本。我们也将指数分布作为 θ 的先验分布，但其参数 λ 已知。基于 $x=(x_1,\ldots,x_n)$ 且相对于给定先验分布的 θ 的贝叶斯估计 $T_\lambda(x)$ 为

$$\frac{\int_0^\infty \theta\left(\prod_{i=1}^n \theta e^{-\theta x_i}\right)\lambda e^{-\lambda\theta}d\theta}{\int_0^\infty\left(\prod_{i=1}^n \vartheta e^{-\vartheta x_i}\right)\lambda e^{-\lambda\vartheta}d\vartheta}=\frac{\int_0^\infty \theta^{n+1}\lambda e^{-\theta\left(\lambda+\sum_{i=1}^n x_i\right)}d\theta}{\int_0^\infty \vartheta^n\lambda e^{-\vartheta\left(\lambda+\sum_{i=1}^n x_i\right)}d\vartheta}$$

直接计算该分子和分母中的积分并不是求得 $T_\lambda(x)$ 的最佳方式。我们将看到，如果我们首先确定后验密度，则计算会更简便；参见例指数分布续例。在该例中，我们推导出
$T_\lambda(x)=(n+1)/\left(\lambda+\sum_{i=1}^n x_i\right)$ 是贝叶斯估计。

当 n 值较大时，贝叶斯估计 $T_\lambda(X)$ 与最大似然估计 $\hat\theta=1/\bar X$ 大致相等。

贝叶斯估计定理的证明涉及条件分布的操作。以下“贝叶斯”符号和概念在这里很有用，并且本身也很重要。它们描述了一个更广泛的贝叶斯方法框架，其中所谓的后验分布构成了分析的终点。

通常，我们将参数 θ 视为确定的，且存在一个“真实”参数值决定了观测值 X 的密度 $x\mapsto p_\theta(x)$。在本节中，我们偏离此观点，将 p_θ 视为变量 X 在（假设的）随机变量 $\bar\Theta$ 取值为 θ 时的条件密度 $p_{X|\bar\Theta=\theta}$。我们赋予该量 $\bar\Theta$ 以（边际）概率密度 π。则 $(X,\bar\Theta)$ 的联合密度为

$$p_{X,\bar\Theta}(x,\theta)=p_{X|\bar\Theta=\theta}(x)p_{\bar\Theta}(\theta)=p_\theta(x)\pi(\theta)$$

在这种贝叶斯设定下，通过对 θ 积分，我们获得 X 的边际密度，从而得到

$$p_X(x)=\int p_{X,\bar\Theta}(x,\theta)d\theta=\int p_\theta(x)\pi(\theta)d\theta$$

因此，给定 $X=x$ 时 $\bar\Theta$ 的条件密度为

$$p_{\bar\Theta|X=x}(\theta)=\frac{p_{X,\bar\Theta}(x,\theta)}{p_X(x)}=\frac{p_\theta(x)\pi(\theta)}{\int p_\vartheta(x)\pi(\vartheta)d\vartheta}$$

定义 3.35 后验密度

$\bar\Theta$ 的后验密度为

$$p_{\bar\Theta|X=x}(\theta)=\frac{p_\theta(x)\pi(\theta)}{\int p_\vartheta(x)\pi(\vartheta)d\vartheta}$$

后验密度中的分母项只是一个归一化常数，使得

$$\int p_{\bar\Theta|X=x}(\theta)d\theta=1$$

在观测值已知之前，我们为 $\bar\Theta$ 赋予先验密度 π。一旦知道观测值，后验密度便提供了调整后的概率分布。这样，观测值引导我们调整有关参数的假设。

这些计算表明，定理 3.33 中的表达式 $T(x)$ 正是后验概率分布下 $g(\bar{\Theta})$ 的期望值，即给定 $X = x$ 时 $g(\bar{\Theta})$ 的条件期望。因此，我们可以将定理重新表述如下。