# 贝叶斯统计

汪燕敏

2022/3/30

## 一、频率学派 VS 贝叶斯学派

基本观点之争：未知参数可以被看作随机变量吗？

基本方法之争：参数估计和推断是否利用先验信息？

## 从一个例子讲起

设某事件 A 在一次实验中发生的概率为 $\theta$。为估计 $\theta$，对实验进行了 $n$ 次独立观测，其中事件 A 发生了 $x$ 次，显然 $x|\theta \sim b(n,\theta)$, 我们可以分别用频率学派和贝叶斯学派的方法来求参数。

## 二、贝叶斯学派的起源：贝叶斯定理

贝叶斯定理是关于随机事件 A 和 B 的条件概率的一则定理。

$$P(B|A) = \frac{P(A)P(A|B)}{P(B)}$$

其中 $A$ 和 $B$ 为随机事件，且 $P(B)$ 不为零。$P(A \mid B)$ 是指在事件 $B$ 发生的情况下事件 $A$ 发生的概率。

在贝叶斯定理中，每个名词都有约定俗成的名称：

$P(A \mid B)$ 是已知 $B$ 发生后，$A$ 的条件概率。也称作 $A$ 的事后概率。
$P(A)$ 是 $A$ 的先验概率（或边缘概率）。其不考虑任何 $B$ 方面的因素。

$P(B \mid A)$ 是已知 $A$ 发生后 $B$ 的条件概率。也可称为 $B$ 的后验概率。统计学里表示似然函数，因为 $P(A \mid B) = L(B \mid A)$。这里 A 是未知参数 $\theta$ ，$B$ 是样本数据。$P(B)$ 是 $B$ 的边际概率函数，与未知参数无关，是个常数。按这些术语，贝叶斯定理可表述为：

后验概率 = (似然函数 * 先验概率)/标准化常量

$$P(\theta \mid data) \propto L(\theta \mid data) * \pi(\theta)$$

也就是说，后验概率与先验概率和似然函数的乘积成正比。

## 三、什么时候用贝叶斯统计？

### 小概率事件

某工厂在一次产品质量检验中出现次品的概率为 $\theta$。为估计 $\theta$，对检验了 100 个产品，其中次品发现了 0 个。求次品率，取先验为 $U(0,1)$。

### 小样本或无样本场合

某人第一次进赌场，要对未来 $n$ 次赌博中他赢钱的次数 $z$ 做出预测。若一次赌博中成功的概率为 $\theta$。则可得 $z$ 的先验预测分布，取先验为 $U(0,1)$。

## 四、如何确定先验分布

### 共轭先验

在贝叶斯统计中，如果后验分布与先验分布属于同类，则先验分布与后验分布被称为共轭分布，而先验分布被称为似然函数的共轭先验。
共轭先验的好处主要在于代数上的方便性，可以直接给出后验分布的封闭形式，否则的话只能数值计算。共轭先验也有助于获得关于似然函数如何更新先验分布的直观印象。在二项模型中，经常使用贝塔分布作为似然函数的共轭先验。在上面的例子中，假设 $\theta$ 的先验分布服从贝塔分布，表示为 $\theta \sim Beta(\alpha, \beta)$, 那么可以得到 $\theta \sim Beta(X+\alpha, n-X+\beta)$

### 无信息先验

在许多情形下，我们几乎不知道分布应该具有的形式。这时，我们可能需找一种被称为无信息先验 (noninformative prior) 的先验分布。这种先验分布的目的是尽可能的后验分布产生小的影响 (Jeffreys, 1946; Box and Tiao, 1973; Bernardo and Smith, 1979)。这有时也被称为 “让数据自己说话”。无信息先验有很多，这里介绍 Jeffreys 先验。其求解思路是找到总体分布 (样本分布) 的费歇尔信息量，其行列式的平方根就是 Jeffreys 先验。

### 弱信息先验

一般将先验分布的方差设的很大，甚至于没有方差，比如柯西分布。

## 五、贝叶斯计算

除了共轭先验和一些简单的情形，对于实际问题，通常都是很难得到参数的贝叶斯估计或后验分布的显式表示。这时一般采用 MCMC(Markov Chain Monte Carlo) 方法进行统计分析。