贝叶斯估计

统计推断领域分为两个大的分支，分别是频率学派和贝叶斯学派。频率学派不在这里说明，这里只重点阐述贝叶斯学派的估计方法。贝叶斯估计是综合利用三种信息（总体信息，样本信息，先验信息）的估计。它与传统频率学派的主要差别在于它利用了先验信息，注重样本的信息。

贝叶斯估计的一个基本思想是：任何一个未知数θ都可以看成是一个随机变量，可以用一个概率分布去描述θ。这个概率分布是采样前就有的关于θ的先验信息的描述，称为先验概率分布。任何一个未知量都有不确定性，用概率分布描述不确定性再合适不过了。

整个贝叶斯估计的核心理论其实只有一个公式，其余所有的公式都是基于此衍生出来的，这就是贝叶斯公式。贝叶斯公式的原始形式:$\mathrm{P}(A_i|\mathrm{B})=\frac{P(A_iB)}{P(B)}, \mathrm{i}=1,2,3\ldots\mathrm{n}$。令 n=1，取其特殊形式，即：$\mathrm{P}(\mathrm{A}|\mathrm{B})=\frac{P(AB)}{P(B)}=\frac{P(A)P(B|A)}{P(B)}$---- （1）。

基于公式(1)，贝叶斯学派有一些具体的想法：

1，依赖于参数θ的概率密度函数在频率学派中记为：$\mathrm{f}(\mathrm{x}:\theta)$,它表示在参数空间{θ}中不同的θ 对应的分布。在贝叶斯统计中记为：$\mathrm{f}(\mathrm{x}|\theta)$。

2, 根据参数θ的先验信息确定先验分布$\pi(\theta)$。这也是贝叶斯估计重要的工作。

3，从贝叶斯观点来看，样本$\mathbf{x}=\{x_1,x_2,x_3\ldots\}$的产生分为两步，首先从先验分布$\pi(\theta)$产生一个$\theta'$，在从总体分布$\mathrm{f}(\mathrm{x}|\theta)$产生样本$\mathbf{x}=\{x_1,x_2,x_3\ldots x_n\}$。次样本发生的概率与以下联合概率密度成正比，$\mathrm{f}(\mathbf{x}|\theta')=\prod_{i=1}^{n}f(x_i|\theta')$。这个密度函数又称为似然函数。

4，θ'只是θ的一个特殊值，要想把先验信息整合到对参数θ的估计，光考虑一个θ'是不够的，而应对θ的所有值进行考虑，所以应该把θ的先验分布$\pi(\theta)$整合

进去。这样，样本$\mathbf{x}$和θ的联合分布为：$f(\mathbf{x},\theta)=f(\mathbf{x}|\theta)\pi(\theta)$。这样三种可用的信息都利用进去了。

5，我们要做的是估计未知数θ。在没有样本信息时，只能依靠先验信息进行估计，在有样本$\mathbf{x}=\{x_1,x_2,x_3\dots x_n\}$之后，我们应该依据联合分布$f(\mathbf{x},\theta)$估计θ。为此，我们需要对$f(\mathbf{x},\theta)$进行分解：$f(\mathbf{x},\theta)=f(\theta|\mathbf{x})f(\mathbf{x})$。其中 f($\mathbf{x}$)是 $\mathbf{x}$ 的边缘密度函数。$f(x)=\int_{-\infty}^{+\infty}f(\boldsymbol{x},\theta)d\theta$。注意，从$f(\mathbf{x})$的表达式可以看出，它与θ无关（因为θ都被积分了）。也就是说$f(x)$不包含θ的任何信息，因此能用作θ的估计的只有$f(\theta|\mathbf{x})$。它的表达式是：$f(\theta|\mathbf{x})=\frac{f(x,\theta)}{f(x)}=\frac{f(x|\theta)\pi(\theta)}{\int_{-\infty}^{+\infty}f(x,\theta)d\theta}$----(2)。（2）式就是贝叶斯公式的概率密度形式。这个在样本给定下θ的分布称为后验概率分布。它集中了总体，样本，先验三种信息。

可以这样理解，先验信息是在采样前对未知数的了解情况，后验概率分布则是在有了样本信息之后重新认识未知数的结果。这也再次说明了贝叶斯估计是基于先验信息并且注重样本的估计。