

浅谈频率学派与贝叶斯学派的区别

参考了@[段子手实习生](https://www.cnblogs.com/shixisheng/p/7136890.html)的博客<https://www.cnblogs.com/shixisheng/p/7136890.html>上整理的相关内容。

根据对概率本质的理解不同，概率论中存在两大学派：频率学派和贝叶斯学派。理解这两大学派的思想对进一步理解先验、后验以及贝叶斯学习的概念有很大帮助。

1. 对于随机性的认知

频率学派与贝叶斯学派探讨「不确定性」这件事时的出发点与立足点不同。

频率学派从「自然」角度出发，试图直接为「事件」本身建模，即事件 A 在独立重复试验中发生的频率趋于极限 p ，那么这个极限就是该事件的概率。举例而言，想要计算抛掷一枚硬币时正面朝上的概率，我们需要不断地抛掷硬币，当抛掷次数趋向无穷时正面朝上的频率即为正面朝上的概率。**贝叶斯学派并不从试图刻画「事件」本身，而从「观察者」角度出发**。贝叶斯学派并不试图说「事件本身是随机的」，或者「世界的本体带有某种随机性」，而只是从「观察者知识不完备」这一出发点开始，构造一套在贝叶斯概率论的框架下可以对不确定知识做出推断的方法。

2. 对于参数空间的认知

对于参数空间的不同认知是频率学派和贝叶斯学派的最大差异。所谓参数空间，就是我们所关心的某个参数的取值范围，或者说参数的分布。

频率学派并不关心参数空间的所有细节，他们相信数据都是在这个空间的“某个”固定参数下产生的（虽然不知道具体是哪个参数），所以他们的方法论一开始就是从“哪个参数值最有可能是真实值”这个角度出发的。于是就有了最大似然（maximum likelihood）和置信区间（confidence interval）这样的概念，从名字就可以看出，他们所关心的是有多大把握找出那个唯一的真实参数。

贝叶斯学派的观点恰恰相反，他们关心参数空间里的每一个值，因为他们认为无法确定哪个参数是真实值。所以参数空间的每个值都有可能是真实模型使用的值，区别只是概率不同而已。于是才会引入先验分布（prior distribution）和后验分布（posterior distribution）这样的概念来设法找出参数空间上的每个值的概率。

最好诠释这种差别的例子是想象一个具有双峰的后验分布，频率学派的方法会去选这两个峰中较高的那一个对应的值作为他们的最好猜测，而贝叶斯学派则会同时报告这两个值，并给出对应的概率。

简单总结一下：

- 频率学派把需要推断的参数 θ 视作**固定且未知的常数**，而样本 X 是随机的，其着眼点在**样本空间**，有关的概率计算都是针对 X 的分布；
- 贝叶斯学派把参数 θ 视作**随机变量**，而样本 X 是固定的，其着眼点在**参数空间**，重视参数 θ 的分布，固定的操作模式是先通过参数的先验分布结合样本信息得到参数的后验分布。

3. 两者的优缺点

频率学派的优点在于其没有假设一个先验分布，因此更加客观，也更加无偏，在一些保守的领域（比如制药业、法律）比贝叶斯方法更受到信任。但是，频率学派过于看重事实，以至于容易被现实欺骗，比如掷硬币，掷了无数次都是正面，从频率学派的角度就会认为正面出现的概率为1。

贝叶斯学派因为将所有的参数都视为随机变量，都有分布，因此可以使用一些基于采样的方法（如MCMC）使得我们更容易构建复杂模型。但是贝叶斯学派过于幻想，以至于想象中的很多东西很难实现，例如很难准确判断参数的先验分布，如果先验选的不好，那么后验分布的结果就会有很大偏差。