贝叶斯分析：抛硬币的概率真的是1/2吗

原创 张磊 机器学习算法那些事 2018-10-21

前言

前面两文介绍了贝叶斯学派的思想和先验分布、后验分布的相关知识，古典频率学派认为抛硬币的概率是常数，本文从贝叶斯学派的角度看待抛硬币的概率问题。本文详细介绍了 β分布，重述贝叶斯思想，对于抛硬币的概率问题作各种情况的分析，最后总结本文。

目录

1、为什么选择β分布作为先验分布

本节详细介绍β分布的定义及解释选择β分布作为先验分布的原因。

1、β分布

β函数的定义：

$$B(\alpha,\beta)=\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx$$

其中α，β > 0，对等式两边各除以B(α,β)，字母p代替x，得：

$$\int_0^1 \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}\,dp=1$$

选积分项作为β分布函数，由积分项可知β分布已完成标准化（总积分等于1）。
因此，β分布：

$$B(p|\alpha,\beta)=\frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}$$

β分布的期望和方差：

$$E(p) = \frac{\alpha}{\alpha+\beta}$$

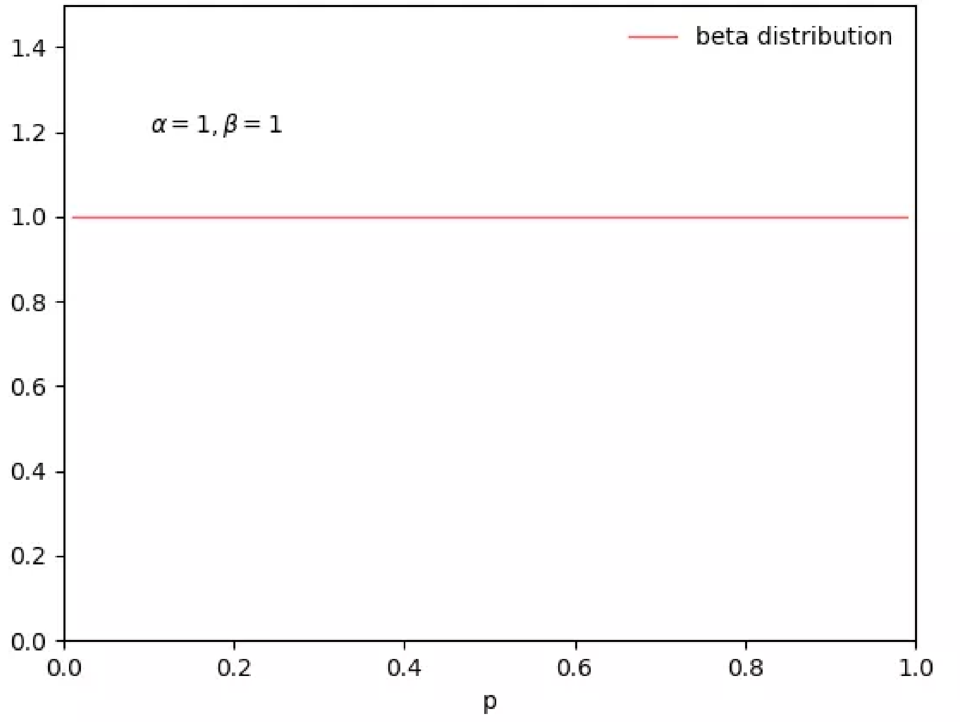
$$D(p) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

2、β分布作为先验分布的原因

由β分布定义可知，β分布是概率分布的分布，β分布常作为先验分布的原因：

（1）、贝叶斯对参数的估计与先验分布的选择有很重要的关系，先验分布不同，贝叶斯对参数的估计也不同。先验分布往往是人们根据以往经验去设计，β分布是概率分布的分布，涵盖了所有参数空间出现的概率大小，并通过设置参数α和β，可以使先验分布与你的先验经验基本符合。

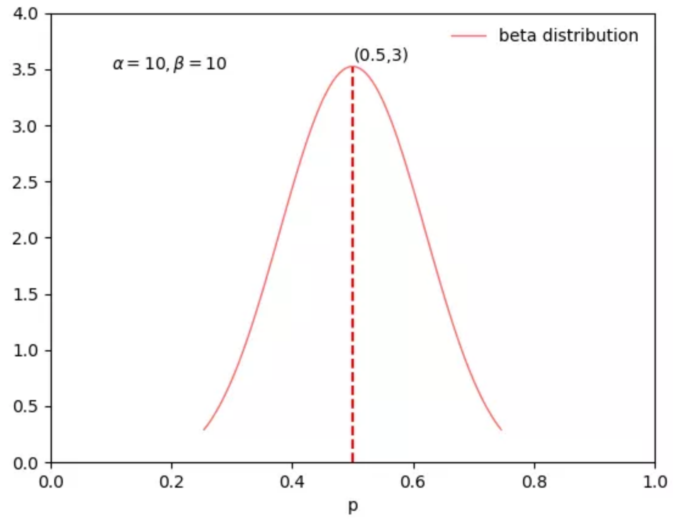
i) α=1，β=1



由上图可知，α=1，β=1，β分布符合均匀分布，即参数空间所有取值的概率相等。

因此，**当你对参数没有任何的先验知识时，建议你假设先验参数符合均匀分布，参数的后验分布由你的实际观测数据决定。**

ii) α=10，β=10



由上图可知，α=10，β=10时，β分布符合高斯分布，且在概率为0.5取得最大值，由β分布期望和方差的公式可知期望和方差分别等于0.5和0.01。

假设参数的先验分布是高斯分布，设置参数α和β相等（α>1）使β分布成为高斯分布，α越大方差越小。

因此，设置α和β使参数的先验分布符合你对参数的先验认知。

（2）、上节已提到，参数的先验分布是β分布时，则先验分布和后验分布形式一样，且可以形成先验链，方便分析问题。

2、重述贝叶斯思想

因人而异，因阅历而异

关于频率学派和贝叶斯学派对频率的理解可以参考我前面的文章《浅谈频率学派和贝叶斯学派》。

贝叶斯思想是量化事件发生的不确定性，是主观评价。不同人评价同一事件发生的概率不同，因为不同人的生活经历不同，对某一事件的先验知识很可能不同，比如一个博士生和一个小学生对某一事件的看法可能不同；同一个人对同一事件发生的概率也随着自身阅历的增加而不同，例如某个人做了九件好事，你评估他是好人的概率为0.9，当他做了一件大逆不道的事情后，你评估他是好人的概率降到了0.1。**贝叶斯评价事件发生的概率带有主观性，因人而异，因阅历而异。**

凡事要讲数据

我们根据自己的阅历对某一事件作一个先验假设，先验假设是否正确需要经过时间的检验，即是否有足够多的观测数据符合先验假设。先验假设和观测数据是影响后验假设的两个因素，若观测数据不符合先验假设，则后验假设在先验假设的基础上开始向观测的数据偏斜，若观测的数据为无穷大时，则

先验假设可以忽略不计，直接通过观测数据来估计后验假设。因此，**贝叶斯思想评价事件发生概率的准则是凡事要讲数据。**

PS： 有点绕口，希望大家看完笔者介绍抛硬币的例子，再来悟一悟这几句话，若还有疑问请微信我😊

3、抛硬币问题的多情况分析

抛硬币问题的公式说明

由于《浅谈先验分布和后验分布》已经通过例子推导了抛硬币正面向上的后验概率，因此，本文不做推论，具体可参考上篇文章，**若有疑问请微信我**。本文只引用一些结论性的公式。
假设硬币正面向上的概率为u，正面向上记为1，反面向上记为0。
则硬币正面向上的先验分布如下：

$$\mathrm{Beta}(\mu|a,b)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu^{a-1}(1-\mu)^{b-1}$$

硬币正面向上的期望：

$$\mathbb{E}[\mu]=\frac{a}{a+b}$$

其中a，b表示虚拟的硬币正面向上的次数和反面向上的次数，根据自己的先验知识来设置a，b值。

若后续的观测结果为**m**次正面向上，**l**次反面向上，共N次。
则硬币正面向上的后验分布如下：

$$p(\mu|m,l,a,b)=\frac{\Gamma(m+a+l+b)}{\Gamma(m+a)\Gamma(l+b)}\mu^{m+a-1}(1-\mu)^{l+b-1}$$

硬币为正面向上的概率：

$$p(x=1|\mathcal{D})=\int_0^1 p(x=1|\mu)p(\mu|\mathcal{D})\,\mathrm{d}\mu=\int_0^1 \mu p(\mu|\mathcal{D})\,\mathrm{d}\mu=\mathbb{E}[\mu|\mathcal{D}]$$

$$p(x=1|\mathcal{D})=\frac{m+a}{m+a+l+b}$$

多情况的抛硬币问题

（1）第1次抛硬币为正面向上的概率；
（2）9次硬币正面向上，1次反面向上，第十一次硬币正面向上的概率；
（3）90次硬币正面向上，10次硬币反面向上，求101次正面向上的概率；
（4）900次硬币正面向上，100次硬币反面向上，求第1001次硬币正面向上的概率。

解：
贝叶斯的后验分布受先验分布的影响，不同的先验分布会有不同的后验分布。**请参考第一节**，假设硬币正面向上的分布符合高斯分布（a=10，b=10），高斯分布符合大部分人的思想，认为硬币为正面向上的概率在0.5达到最大，方差表示先验分布的确定程度，若你坚信硬币向上的概率肯定是0.5，那么可以调大a和b值。

笔者就先验分布为高斯分布来解答抛硬币的四个问题。其他先验分布可通过调节a，b的值来实现，后面的计算过程一致。

正面向上的后验概率：

$$p(x=1|\mathcal{D})=\frac{m+a}{m+a+l+b}$$

a，b，m，l分别表示先验分布的正面向上次数，反面向上次数，已观测数据的正面向上次数，反面向上次数。

先验分布为高斯分布：

（1）由于没有任何观测数据，因此第一次正面向上的分布为先验分布，先验分布在在参数为0.5时，概率最大，即正面向上的概率为0.5。
（2）正面向上的概率为：

$$P(X=1\mid D)=\frac{\mathrm{m}+a}{m+a+l+b}=\frac{9+10}{9+10+1+10}=0.63$$

（3）计算过程与（2）一样，正面向上的概率：0.83
（4）正面向上的概率：0.89

讨论：

频率学派认为硬币向上的概率是0.5，与观测数据无关。贝叶斯学派是通过数据来主观评价硬币向上的概率，由例子可知，即使先验分布符合高斯分布且正面向上的概率在0.5达到最大，但是如果观测数据倾向于正面向上，则最终的判断结果会倾向于正面向上，贝叶斯思想有点像是风往哪边吹树就往

哪边倒的意思😂。**当观测结果的正面向上次数远远大于正面向下次数，也远远大于先验分布的正面向下次数，则判断下次为正面向上的概率无限接近1**（若不理解请参考公式）。

4、总结

本文首先详细介绍了β分布，通过调节参数a和b使β分布符合假设的先验分布，β分布使后验分布和先验分布为共轭分布，形成先验链，便于分析问题。后面讲的内容是贝叶斯思想，贝叶斯是主观评价事件发生的概率，根据先验知识来假设先验分布，若观测的数据符合先验分布，则后验分布与先验分布类似；若观测的数据不符合先验分布，则后验分布开始向观测数据倾斜，若观测数据为无穷大时，那么前验分布可以忽略不计，最大似然函数估计参数与后验分布估计参数相同，直接可以用最大似然函数来估计参数。

参考：

Christopher M.Bishop <<Pattern Reconition and Machine Learning>>

推荐阅读文章

浅谈频率学派和贝叶斯学派

浅谈先验分布和后验分布

模型优化的风向标：偏差与方差

机器学习模型性能评估（三）：代价曲线

机器学习模型性能评估（二）：P-R曲线与ROC曲线

机器学习模型性能评估（一）：错误率与精度

-END-