維基百科

后验概率

维基百科，自由的百科全书

在贝叶斯统计中，一个随机事件或者一个不确定事件的后验概率（Posterior probability）是在考虑和给出相关证据或数据后所得到的条件概率。同样，后验概率分布是一个未知量（视为随机变量）基于试验和调查后得到的概率分布。“后验”在本文中代表考虑了被测试事件的相关证据。

目录

定义

后验概率是在给定证据X后，参数θ的概率：$p(\theta|X)$。

与似然函数相对，其为在给定了参数θ后，证据X的概率：$p(X|\theta)$。

两者有以下联系：

首先定义先验概率服从以下概率分布函数，$p(\theta)$，则样本x的可能性为$p(x|\theta)$，那么后验概率可以定义为

$$p(\theta|x)=\frac{p(x|\theta)p(\theta)}{p(x)}$$ [1]

后验概率可以写成更易记忆的形式，为

后验概率 $\propto$ 可能性 $\times$ 先验概率。

实例

假设一个学校裡有60%男生和40%女生。女生穿裤子的人数和穿裙子的人数相等，所有男生穿裤子。一个人在远处随机看到了一个穿裤子的学生。那么这个学生是女生的概率是多少？

使用贝叶斯定理，事件A是看到女生，事件B是看到一个穿裤子的学生。我们所要计算的是P(A|B)。

P(A)是忽略其它因素，看到女生的概率，在这里是40%

P(A')是忽略其它因素，看到不是女生（即看到男生）的概率，在这里是60%

P(B|A)是女生穿裤子的概率，在这里是50%

P(B|A')是男生穿裤子的概率，在这里是100%

P(B) 是忽略其它因素，学生穿裤子的概率，P(*B*) = P(*B*|*A*)P(*A*) + P(*B*|*A*')P(*A*')，在这里是0.5×0.4 + 1×0.6 = 0.8.

根据贝叶斯定理，我们计算出后验概率P(A|B)

$$P(A|B)=\frac{P(B|A)P(A)}{P(B)}=\frac{0.5\times 0.4}{0.8}=0.25。$$

可见，后验概率实际上就是条件概率。

计算

根据贝叶斯定理，一个随机变量在给定另一随机变量值之后的后验概率分布可以通过先验概率分布与似然函数相乘并除以归一化常数求得

$$f_{X|Y=y}(x)=\frac{f_X(x)L_{X|Y=y}(x)}{\int_{-\infty}^{\infty}f_X(u)L_{X|Y=y}(u)\,du}$$

上式为给出了随机变量X在给定数据$Y=y$后的后验概率分布函数，式中

- $f_X(x)$为X的先验密度函数，
- $L_{X|Y=y}(x)=f_{Y|X=x}(y)$为$x$的似然函数，
- $\int_{-\infty}^{\infty}f_X(u)L_{X|Y=y}(u)\,du$为归一化常数，
- $f_{X|Y=y}(x)$为考虑了数据$Y=y$后X的后验密度函数。

置信区间

后验概率是考虑了一系列随机观测数据的条件概率。对于一个随机变量来说，量化其不确定性非常重要。其中一个实现方法便是提供其后验概率的置信区间。

参见

- 经验贝叶斯方法
- 边缘分布
- Lindley's 悖论

引用

1. Christopher M. Bishop. Pattern Recognition and Machine Learning. Springer. 2006: 21–24. ISBN 978-0-387-31073-2.