

1 相关理论

1.1 全概率公式

全概率公式为概率论中的重要公式，它将对一复杂事件A的概率求解问题转化为了在不同情况下发生的简单事件的概率的求和问题。

如果事件 B_1 、 B_2 、 $B_3 \cdots B_n$ 构成一个完备事件组，即它们两两互不相容，其和为全集；并且 $P(B_i) > 0$ ，则对任一事件A有

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \dots + P(A|B_n)P(B_n).$$

特别地，对于任意两随机事件A和B，有如下成立：

1.2 先验概率

先验概率（prior probability）是指根据以往经验和分析得到的概率。

意思是说我们人有一个常识，比如骰子，我们都知道概率是1/6，而且无数次重复实验也表明是这个数，这是一种我们人的常识，也是我们在不知道任何情况下必然会说出一个值。而所谓的先验概率是我们人在未知条件下对事件发生可能性猜测的数学表示！

1.3 后验概率

事情已经发生，要求这件事情发生的原因是由某个因素引起的可能性的大小

1.4 举个例子

举一个简单的例子：一口袋里有3只红球、2只白球，采用不放回方式摸取，求：

(1) 第一次摸到红球（记作A）的概率；

(2) 第二次摸到红球（记作B）的概率；

(3) 已知第二次摸到了红球，求第一次摸到的是红球的概率。

解：

(1) $P(A)=3/5$ ，这就是先验概率；

(2) $P(B)=P(A)P(B|A)+P(A^c)P(B|A^c)=3/5$

(3) $P(A|B)=P(A)P(B|A)/P(B)=1/2$ ，这就是后验概率。

2 最大似然估计

最大似然估计的核心思想是：找到参数 θ 的一个估计值，使得当前样本出现的可能性最大，俗话说是“谁大像谁”。

假设有一组独立同分布(i. i. d)的随机变量 X ，给定一个概率分布 D ，假设其概率密度函数为 f ，以及一个分布的参数 θ ，从这组样本中抽出 x_1, x_2, \dots, x_n ，那么通过参数 θ 的模型 f 产生上面样本的概率为：

$$f(x_1, x_2, \dots, x_n | \theta) = f(x_1 | \theta) \times f(x_2 | \theta) \times \dots \times f(x_n | \theta)$$

最大似然估计会寻找关于 θ 的最可能的值，即在所有可能的 θ 取值中，寻找一个值使这个采样的“可能性”最大化！

因为是”模型已定，参数未知”，此时我们是根据样本采样 x_1, x_2, \dots, x_n 取估计参数 θ ，定义似然函数为：

$$L(\theta | x_1, x_2, \dots, x_n) = f(x_1, x_2, \dots, x_n | \theta) = \prod f(x_i | \theta)$$

实际使用中，因为 $f(x_i | \theta)$ 一般比较小，而且 n 往往会比较大，连乘容易造成浮点运算下溢。所以一般我们用对数似然函数：

$$\ln L(\theta|x_1, x_2, \dots, x_n) = \sum_{i=1}^n f(x_i|\theta)$$

$$\hat{l} = \frac{1}{n} \ln L$$

那最终 θ 的估计值为：

$$\hat{\theta}_{MLE} = \arg \max_{\theta} \hat{l}(\theta|x_1, x_2, \dots, x_n)$$

根据前面的描述，总结一下求最大释然估计值的步骤：

1. 写似然函数
2. 一般对似然函数取对数，并将对数似然函数整理
3. 对数似然函数求导，令导数为0，求得似然方程
4. 根据似然方程求解，得到的参数即为所求估计值

3 贝叶斯估计

统计学里有两个大的流派，一个是频率派，一个是贝叶斯派。时至今日，这两派还未就各自的观点达成统一。我们前面提到的最大似然估计就是频率派的典型思路，接下来再看看贝叶斯派的思路，到底跟频率派估计有何不同。

先来看几个相关的小公式：

两个随机变量 x, y 的联合概率 $p(x, y)$ 的乘法公式：

$$p(x, y) = p(x|y)p(y) = p(y|x)p(x)$$

如果 x, y 是独立随机变量，上面的式子可以表示为：

$$p(x, y) = p(x)p(y) = p(y)p(x)$$

那么条件概率就可以表示为：

$$p(x|y) = \frac{p(x, y)}{p(y)}, p(y|x) = \frac{p(x, y)}{p(x)}$$

对于一个完备事件组 y_1, y_2, \dots, y_n ，可以使用全概率公式：

$$p(x) = \sum_{i=1}^n p(y_i)p(x|y_i), \text{ 其中 } \sum_{i=1}^n p(y_i) = 1$$

由以上这些，可以得出贝叶斯公式：

$$p(y_i|x) = \frac{p(x, y_i)}{p(x)} = \frac{p(y_i)p(x|y_i)}{p(x)}$$

其中， $p(y_i|x)$ 是后验概率。 $p(x|y_i)$ 是条件概率，或者说似然概率，这个概率一般都可以通过历史数据统计得出。而 $p(y_i)$ 是先验概率，一般也是根据历史数据统计得出或者认为给定的，贝叶斯里的先验概率，就是指 $p(y_i)$ 。对于 $p(x)$ ，我们前面提到可以用全概率公式计算得出，但是在贝叶斯公式里面我们一般不care这个概率，因为我们往往只需要求出最大后验概率而不需要求出最大后验的具体值。

2, 3部分内容来自：

[最大似然估计MLE与贝叶斯估计](#)

4 区别

- 理解1:

最大似然估计和贝叶斯估计最大区别便在于估计的参数不同，最大似然估计要估计的参数 θ 被当作是固定形式的一个未知变量，然后我们结合真实数据通过最大化似然函数来求解这个固定形式的未知变量！

贝叶斯估计则是将参数视为是有某种已知先验分布的随机变量，意思便是这个参数他不是一个固定的未知数，而是符合一定先验分布如：随机变量 θ 符合正态分布等！那么在贝叶斯估计中除了类条件概率密度 $p(x|w)$ 符合一定的先验分布，参数 θ 也符合一定的先验分布。我们通过贝叶斯规则将参数的先验分布转化成后验分布进行求解！

- 理解2:

简而言之，最大似然估计认为参数的所有可能取值都是一样可能的。而贝叶斯方法认为还存在一个先验估计，有些取值更有可能，有些取值更加没有可能。

- 理解3:

最大似然是对点估计，贝叶斯推断是对分布估计。

即，假设求解参数 θ ，最大似然是求出最有可能的 θ 值，而贝叶斯推断则是求解 θ 的分布。

在公式上，贝叶斯推断还引入了先验，通过先验和似然来求解后验分布，而最大似然直接使用似然函数，通过最大化其来求解。

参考资料

- [全概率公式](#)
- [贝叶斯公式的直观理解\(先验概率/后验概率\)](#)
- [最大似然估计MLE与贝叶斯估计](#)
- [贝叶斯估计和极大似然估计到底有何区别](#)