最大似然估计(MLE)与最大后验概率(MAP)

最大似然估计(Maximum likelihood estimation，简称MLE)和最大后验概率(Maximum a posteriori estimation，简称MAP)是很常用的两种参数估计方法，如果不理解这两种方法的思路，就很容易弄混它们。下面我们详细说明MLE和MAP的思路与区别。

**概率与统计**

概率(probability)和统计(statistics)看似两个相近的概念，其实研究的问题刚好相反。

概率研究的问题是，已知一个模型和参数，怎么去预测这个模型产生的结果的特性（例如均值，方差，协方差等）。

统计研究的问题相反。统计是，有一堆数据，要利用这堆数据去预测模型和参数。

一句话总结就是，概率是已知模型和参数，推数据。统计是已知数据，推模型和参数。

显然本文要介绍的MLE和MAP都是统计领域的问题，它们都是用来推测参数的方法。

**先验分布和后验分布**

先验(**一般认为用p()表示**)是指“来自先前的东西”，拓展为“在经验之前”，一般认为先验指无需经验或先于经验获得的知识。他通常与后验知识相比较，后验(**一般用p()来表示**)意指“在经验之后”，需要经验。一般也说，从原因到结果的论证称为先验的，而从结果到原因的论证称为后验的。

先验概率的计算比较简单，没有使用贝叶斯公式；而后验概率的计算，要使用贝叶斯公式。

先验概率是在缺乏某个事实的情况下描述一个变量；而后验概率是在考虑了一个事实之后的条件概率。

**似然函数**

这里我们在简单介绍下似然(likelihood)函数，详细的可以看似然函数那个文档。

对于函数P(),它的输入有两个：x表示某一个具体的数据；表示模型的参数。

如果是已知确定的，x是变量，那么这个函数叫做概率函数，它描述对于不同的样本点x，其出现的概率是多少。

如果x是已知确定的，是变量，这个函数叫做似然函数，它描述对于不同的模型参数，出现x这个样本点的概率是多少。

最大似然函数可以理解为是让已知的样本x出现的概率最大。

**用一个例子理解先验分布、后验分布、似然估计**

一个小哥想去公园，他可以选择步行、骑自行车或者开车，然后通过其中一种方式花费一段时间到达公园。这个时间中采取哪种交通方式是因，花费了多少时间是果。

**后验概率（知果求因）**：

假设我们已经知道小哥花了一个小时到了公园，那么我们猜测他是怎么去的（走路or开车or自行车），事实上我们不能百分百确定他的交通方式，我们正常人的思路是他很大可能是骑自行车过去的，当然也不排除他是开车过去的却由于堵车花费了过长的时间，当然也有可能他是一个运动员一路跑过去的。

假设我们知道小哥是花了三个小时才到公园，这个时候我们猜的时候会觉得他很大可能是静静地走路过去的，但是如果已知小哥是只花了20分钟才到公园，那么正常人都会觉得它最大可能是开车去的。

这种事先已经知道结果（路上花的时间），然后根据结果估计原因（交通方式）的概率分布叫做后验概率。

例子公式化：

改成一般的公式：

公式正规化：

这里的表示因，x表示果。

**先验概率（由历史求因）**

换个场景，我们不再考虑小哥去公园的结果了。假设小哥还没去，大早上刚起床，打算吃完早饭再去。假设我们比较了解小哥的个人习惯，小哥是个健身爱好者，就喜欢跑步运动，这个时候我们可以猜测他更可能倾向于走路过去。如果小哥是个死肥宅，很懒，那么我们猜测他更可能倾向于坐车过去，骑自行车的可能性都不大。

这个场景中隔壁小哥的交通方式选择与花费时间不再相关。因为我们是在结果发生前就开始猜的，根据历史规律确定原因（交通方式）的概率分布即先验概率。

例子公式化：

一般化：

正规化：

**似然估计（由因求果）**

换个场景，我们先重新考虑小哥去公园的交通方式。

假设小哥步行走路去，15公里的路到公园，一般情况下小哥大概要花两个小时，当然很小的可能性是小哥是飞毛腿，跑步过去用了1个小时左右，极为小的可能是小哥是隐藏的高手，10分钟就轻功跑到了。也有可能小哥决定开车，到公园半个小时是非常可能的，非常小的概率小哥是因为途经的路上有车祸堵了三个小时。

这种先定下来原因（交通方式），根据原因来估计结果的概率分布即似然估计。根据原因来统计各种可能结果的概率即似然函数。

似然函数问题公式化：

一般化：

正规化：

**贝叶斯公式**

我们熟知的贝叶斯公式是这样的：

但在这里我们可以采用如下的形式：

即：

注：P(x)即evidence，小哥去了公园很多次，忽略交通方式是什么，只统计每次到达公园的时间x，于是得到了一组时间的概率分布。这种不考虑原因。只看结果的概率分布即evidence，它也称为样本发生的概率分布的证据。

**最大似然估计**

**可以理解为最大化该事件发生的可能性。最大似然估计就是求参数，使似然函数**最大。

**最大后验估计**

**最大后验估计，从数学上来说，它的表达式如下：**

这里的目标也是求一个使得最大。但是因为给定样本x后，P(x)后在空间上为一个定值，所以可以理解为最大后验估计就是最大化**，他求得的不仅让似然函数大，自己出现的先验概率也得大。**

**最大似然估计和最大后验估计的区别**

**根据上文，我们可以得出MLE和MAP的区别是：MAP就是多了一个作为因子的先验概率。或者也可以反过来，认为MLE是把先验概率认为等于1。**

**通俗的理解，最大似然估计就是已知参数，求在当前参数的情况下，哪种x出现的概率最大；最大后验估计就是已知当前的x分布，求参数。**