如何通俗地理解“最大似然估计法”?

我们假设硬币有两面，一面是“花”，一面是“字”。

一般来说，我们都觉得硬币是公平的，也就是“花”和“字”出现的概率是差不多的。

如果我扔了100次硬币，100次出现的都是“花”。

在这样的事实下，我觉得**似乎**硬币的参数不是公平的。你硬要说是公平的，那就是侮辱我的智商。

这种通过事实，反过来猜测硬币的情况，就是**似然**。

而且，我觉得最有可能的硬币的情况是，两面都是“花”：

通过事实，推断出最有可能的硬币情况，就是**最大似然估计**。

1 概率vs似然

让我们先来比较下概率和似然。

为了避免和我们想讨论的概率混淆，我们把硬币的“花”出现的概率称为硬币的参数。

1.1 概率

已知硬币的参数，就可以去推测抛硬币的各种情况的可能性，这称为**概率**。

比如已知硬币是公平的，也就是硬币的参数为0.5。

那么我们就可以推测，扔10次硬币，出现5次“花”朝上的**概率**为（抛硬币遵循二项分布，这个就不多解释了）：

C_{10}^{5}

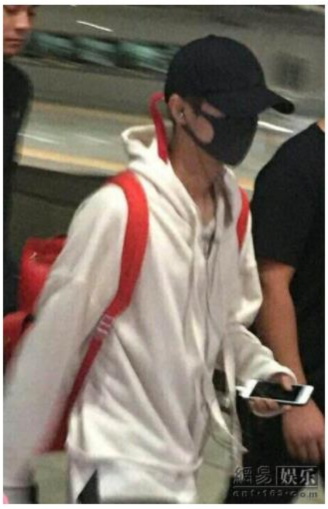
$$\binom{10}{5}0.5^5(1-0.5)^5 \approx 0.25$$

1.2 似然

正如开头所说，我们对硬币的参数并不清楚，要通过抛硬币的情况去推测硬币的参数，这称为**似然**。

可以再举不那么恰当（主要模型不好建立）的例子，蹭下热点。

比如我们发现，鹿晗和关晓彤戴同款手链，穿同款卫衣：



我们应该可以推测这两人关系的“参数”是“亲密”。

进一步发现，两人在同一个地方跨年：

似乎，关系的“参数”是“不简单”。

最后，关晓彤号称要把初吻留给男友，但是最近在荧幕中献出初吻，对象就是鹿晗：

我觉得最大的可能性，关系的“参数”是“在一起”。

通过证据，对两人的关系的“参数”进行推断，叫做**似然**，得到最可能的参数，叫做**最大似然估计**。

2 最大似然估计

来看看怎么进行最大似然估计。

2.1 具体的例子

我们实验的结果是，10次抛硬币，有6次是"花"。

所谓最大似然估计，就是假设硬币的参数，然后计算实验结果的概率是多少，概率越大的，那么这个假设的参数就越可能是真的。

我们先看看硬币是否是公平的，就用0.5作为硬币的参数，实验结果的概率为：

$$\binom{10}{6}0.5^6(1-0.5)^4 \approx 0.21$$

单独的一次计算没有什么意义，让我们继续往后面看。

再试试用0.6作为硬币的参数，实验结果的概率为：

$$\binom{10}{6}0.6^6(1-0.6)^4 \approx 0.25$$

之前说了，单次计算没有什么意义，但是两次计算进行比较就有意义了。

可以看到：

$$\frac{0.25}{0.21} \approx 1.2$$

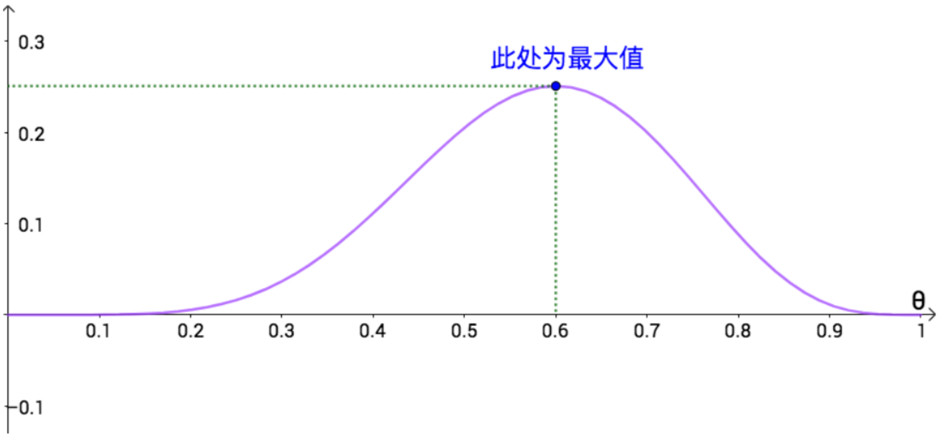
我们可以认为，0.6作为参数的可能性是0.5作为参数的可能性的1.2倍。

2.2 作图

我们设硬币的参数为θ，可以得到似然函数为：

$$L(\theta) = \binom{10}{6}\theta^6(1-\theta)^4$$

这样我们就可以作图了：



我们可以从图中看出两点：

- 参数为0.6时，概率最大
- 参数为0.5、0.7也是有可能的，虽然可能性小一点

所以更准确的说，似然（现在可以说似然函数了）是推测参数的分布。

而求最大似然估计的问题，就变成了求似然函数的极值。在这里，极值出现在0.6。

2.3 更多的实验结果

如果实验结果是，投掷100次，出现了60次"花"呢？

似然函数为：

$$L(\theta)=\binom{100}{60}\theta^{60}(1-\theta)^{40}$$

用0.5作为硬币的参数，实验结果的概率为：

$$\binom{100}{60}0.5^{60}(1-0.5)^{40}\approx 0.01$$

再试试用0.6作为硬币的参数，实验结果的概率为：

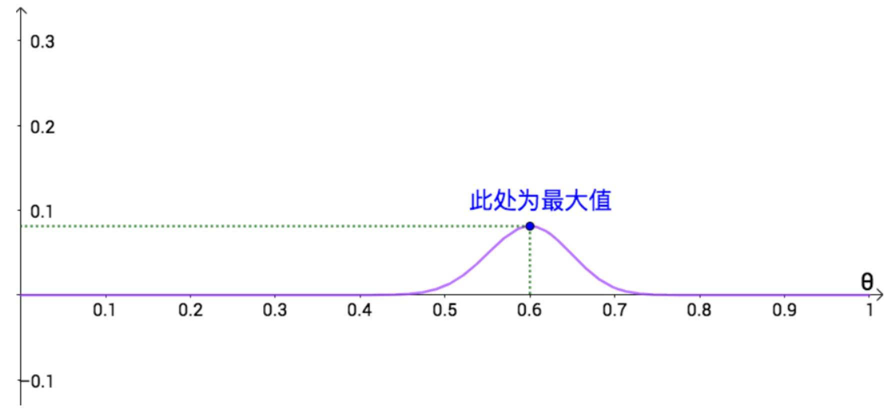
$$\binom{100}{60}0.6^{60}(1-0.6)^{40}\approx 0.08$$

此时：

$$\frac{L(0.6)}{L(0.5)}=\frac{0.08}{0.01}=8$$

此时，0.6作为参数的可能性是0.5作为参数的可能性的8倍，新的实验结果更加支持0.6这个参数。

图像为：



很明显图像缩窄了，可以这么解读，可选的参数的分布更集中了。越多的实验结果，让参数越来越明确。

2.4 更复杂一些的最大似然估计

2.4.1 数学名词

下面提升一点难度，开始采用更多的数学名词了。

先说一下数学名词：

- 一次实验：抛硬币10次，出现6次"花"，就是一次实验。
- 二项分布：抛硬币10次，出现6次"花"的概率为0.25，出现5次"花"的概率为0.21，所有的可能的结果（比如抛硬币10次，出现11次"花"，这就是不可能）的概率，放在一起就是二项分布

2.4.2 多次实验

之前的例子只做了一次实验。只做一次实验，没有必要算这么复杂，比如投掷100次，出现了60次"花"，我直接：

$$\frac{60}{100} = 0.6$$

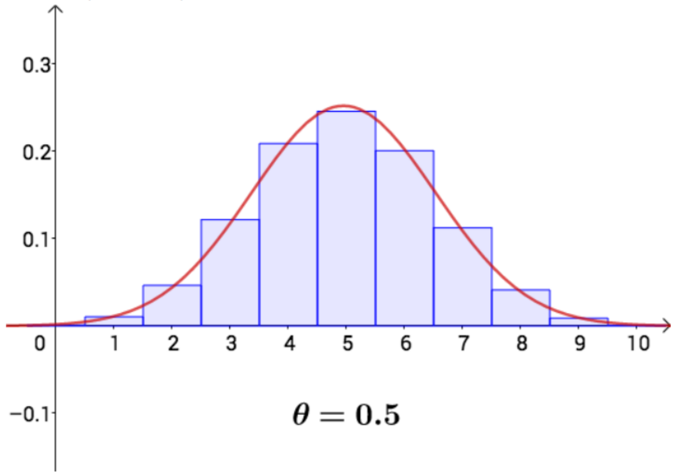
不就好了？

最大似然估计真正的用途是针对多次实验。

2.4.3 上帝视角

为了说清楚这个问题，我引入一个上帝视角。

比如，我有如下的二项分布，θ为出现"花"的概率（硬币抛10次）：

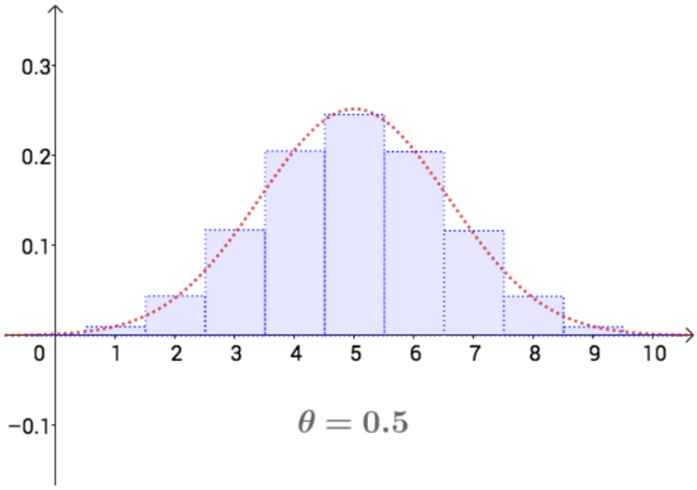


在实际生活中，θ往往是不知道的，这里你可以看得到，就好像你是上帝一样。

要提醒大家注意的一点，上面的图像只有上帝才能看到的，包括：

- 二次分布的柱状图
- 二次分布的曲线图
- θ值为多少

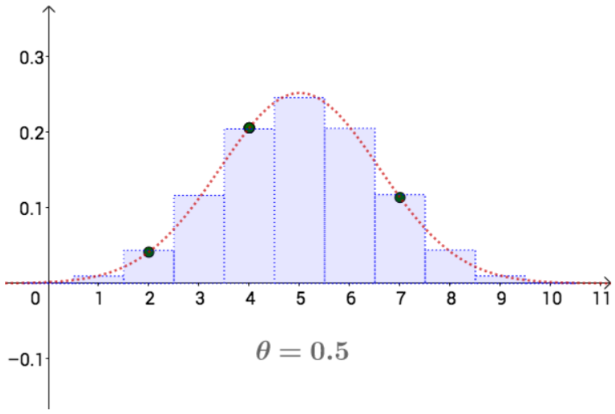
我把只有上帝能看到的用虚线表示，θ用淡一点的颜色表示：



2.4.4 通过多次实验进行最大似然估计

上面的二项分布用通俗点的话来说，就是描述了抛10次硬币的结果的概率，其中，“花”出现的概率为 θ 。

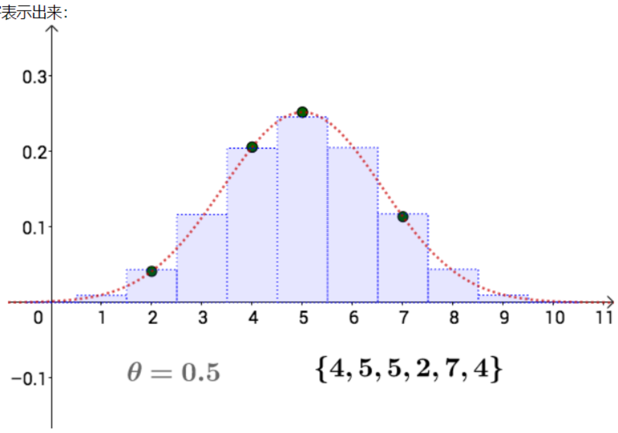
根据上面的二项分布，我进行了6次实验（也就是总共6次，每次抛10次硬币），把实验结果用点的形式标记在图像上（从技术上讲，这6个点是根据二项分布随机得到的）：



这个实验结果，也就是图上的点，是我们“愚蠢的人类”可以看见的了。

可以看到，虽然进行了6次实验，但是却没有6个点，这是因为有的实验结果是一样的，就重合了。

为了方便观察，我把6个点的值用文字表示出来：



上图中的 $\{4, 5, 5, 2, 7, 4\}$ 就是6次实验的结果，分别表示：

- 第一次实验，4次出现“花”
- 第二次实验，5次出现“花”
- 第三次实验，5次出现“花”
- 以此类推

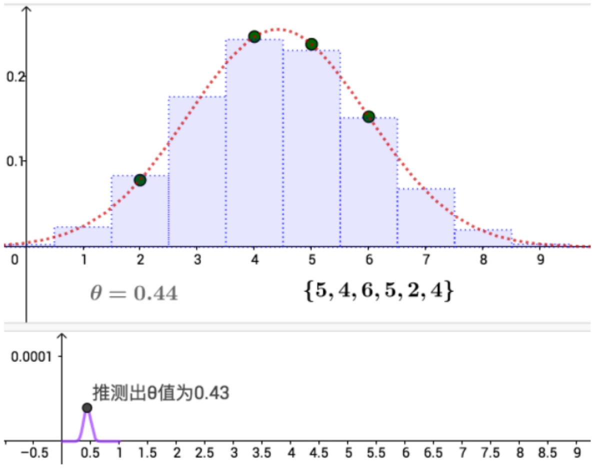
我们用 $x_1, x_2, \cdots, x_n$ 表示每次实验结果，因为每次实验都是独立的，所以似然函数可以写作（得到这个似然函数很简单，独立事件的联合概率，直接相乘就可以得到）：

$$L(\theta) = f(x_1 \mid \theta) f(x_2 \mid \theta) \cdots f(x_n \mid \theta)$$

$f(x_n \mid \theta)$ 表示在同一个参数下的实验结果，也可以认为是条件概率。

$\binom{10}{4}\theta^4(1-\theta)^6 \times \cdots \times$

下面这幅图，分为两部分，上面除了实验结果外，都是上帝看到的，而下面是通过实验结果，利用似然函数对θ值的推断：



可以看出，推断出来的θ值和上帝看到的差不多。之所以有差别是因为实验本身具有二项随机性，相信试验次数越多，推测会越准确。