

Cesco: a minimal blog

似然与极大似然估计

⌚ 2016-08-18 | ⌂ [thinkstats](#)

Likelihood & Maximum likelihood

似然与概率

在统计学中，似然函数（likelihood function，通常简写为likelihood，似然）是一个非常重要的内容，在非正式场合似然和概率（Probability）几乎是一对同义词，但是在统计学中似然和概率却是两个不同的概念。概率是在特定环境下某件事情发生的可能性，也就是结果没有产生之前依据环境所对应的参数来预测某件事情发生的可能性，比如抛硬币，抛之前我们不知道最后是哪一面朝上，但是根据硬币的性质我们可以推测任何一面朝上的可能性均为50%，这个概率只有在抛硬币之前才是有意义的，抛完硬币后的结果便是确定的；而似然刚好相反，是在确定的结果下去推测产生这个结果的可能环境（参数），还是抛硬币的例子，假设我们随机抛掷一枚硬币1,000次，结果500次人头朝上，500次数字朝上（实际情况一般不会这么理想，这里只是举个例子），我们很容易判断这是一枚标准的硬币，两面朝上的概率均为50%，这个过程就是我们运用出现的结果来判断这个事情本身的性质（参数），也就是似然。

结果和参数相互对应的时候，似然和概率在数值上是相等的，如果用 θ 表示环境对应的参数， x 表示结果，那么概率可以表示为：

$$P(x|\theta)$$

$p(x|\theta)$ 是条件概率的表示方法， θ 是前置条件，理解为在 θ 的前提下，事件 x 发生的概率，相对应的似然可以表示为：

$$\mathcal{L}(\theta|x)$$

可以理解为已知结果为 x ，参数为 θ （似然函数里 θ 是变量，这里说的参数和变量是相对与概率而言的）对应的概率，即：

$$\mathcal{L}(\theta|x) = P(x|\theta)$$

需要说明的是两者在数值上相等，但是意义并不相同， \mathcal{L} 是关于 θ 的函数，而 P 则是关于 x 的函数，两者从不同的角度描述一件事情。

举个例子

以伯努利分布（Bernoulli distribution，又叫做两点分布或0-1分布）为例：

$$f(x; p) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases}$$

也可以写成以下形式：

$$f(x; p) = p^x (1 - p)^{1-x} \quad \text{for } x \in \{0, 1\}$$

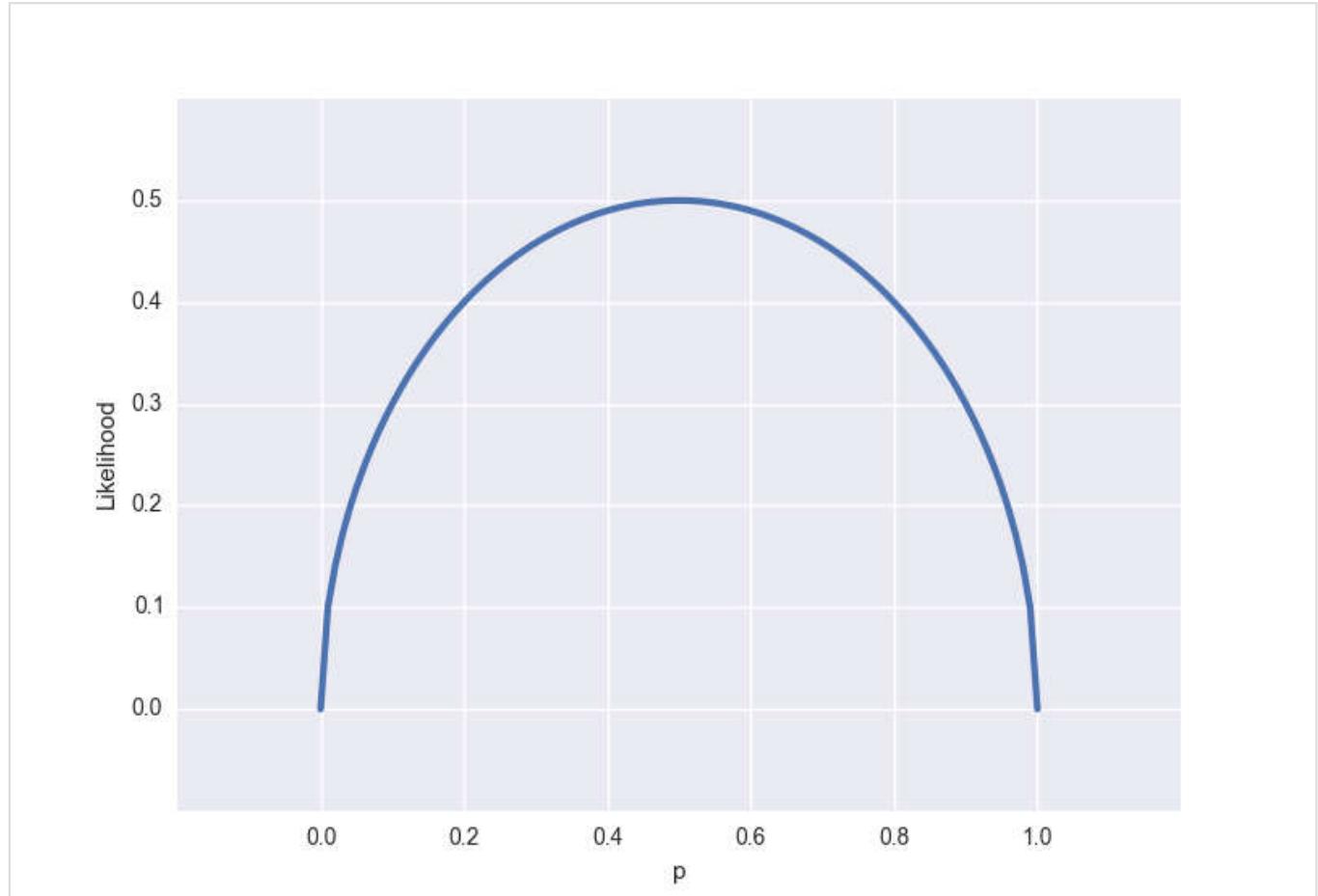
这里注意区分 $f(x; p)$ 与前面的条件概率的区别，引号后的 p 仅表示 f 依赖于 p 的值， p 并不是 f 的前置条件，而只是这个概率分布的一个参数而已，也可以省略引号后的内容：

$$f(x) = p^x (1 - p)^{1-x} \quad \text{for } x \in \{0, 1\}$$

对于任意的参数 p 我们都可以画出伯努利分布的概率图，当 $p = 0.5$ 时：

$$f(x) = 0.5$$

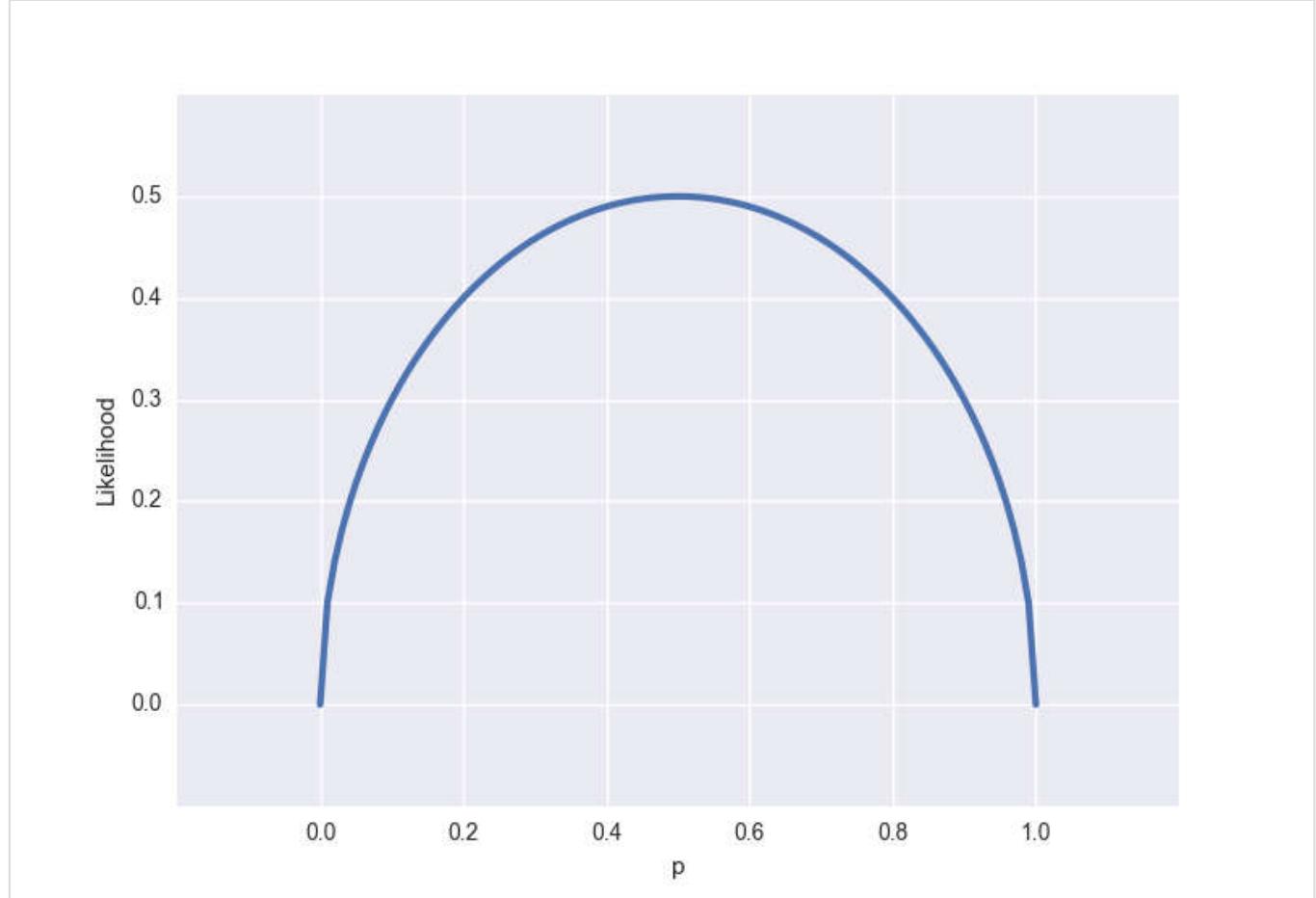
我们可以得到下面的概率密度图：



从似然的角度出发，假设我们观测到的结果是 $x = 0.5$ （即某一面朝上的概率是50%，这个结果可能是通过几千次几万次的试验得到的，总之我们现在知道这个结论），可以得到以下的似然函数：

$$\mathcal{L}(p|x=0.5) = p^{0.5}(1-p)^{0.5}$$

对应的图是这样的：



与概率分布图不同的是，似然函数是一个(0, 1)内连续的函数，所以得到的图也是连续的，我们很容易看出似然函数的极值（也是最大值）在 $p = 0.5$ 处得到，通常不需要做图来观察极值，令似然函数的偏导数为零即可求得极值条件。

ps. 似然函数里的 p 描述的是硬币的性质而非事件发生的概率（比如 $p = 0.5$ 描述的是一枚两面均匀的硬币）。为了避免混淆，可以用其他字母来表示这个性质，如果我们用 π 来表示，那么似然函数就可以写成：

$$\mathcal{L}(\pi|x=0.5) = \pi^{0.5}(1-\pi)^{0.5}$$

似然函数的最大值

似然函数的最大值意味着什么？让我们回到概率和似然的定义，概率描述的是在一定条件下某个事件发生的可能性，概率越大说明这件事情越可能会发生；而似然描述的是结果已知的情况下，该事件在不同条件下发生的可能性，似然函数的值越大说明该事件在对应的条件下发生的可能性越大。

现在再来看看之前提到的抛硬币的例子：

$$\mathcal{L}(p|x=0.5) = p^{0.5}(1-p)^{0.5}$$

上面的 p (硬币的性质) 就是我们说的事件发生的条件, \mathcal{L} 描述的是性质不同的硬币, 任意一面向上概率为 50% 的可能性有多大, 是不是有点绕? 让我们来定义 A:

A=事件的结果=任意一面向上概率为50%

那么 \mathcal{L} 描述的是性质不同的硬币, A 事件的可能性有多大, 这么一说是清楚多了?

在很多实际问题中, 比如机器学习领域, 我们更关注的是似然函数的最大值, 我们需要根据已知事件来找出产生这种结果最有可能的条件, 目的当然是根据这个最有可能的条件去推测未知事件的概率。在这个抛硬币的事件中, p 可以取 $[0, 1]$ 内的所有值, 这是由硬币的性质所决定的, 显而易见的是 $p = 0.5$ 这种硬币最有可能产生我们观测到的结果。

对数化的似然函数

实际问题往往要比抛一次硬币复杂得多, 会涉及到多个独立事件, 在似然函数的表达式中通常都会出现连乘:

$$\mathcal{L} = \prod_{i=1}^N p_i$$

对多项乘积的求导往往非常复杂, 但是对于多项求和的求导却要简单的多, 对数函数不改变原函数的单调性和极值位置, 而且根据对数函数的性质可以将乘积转换为加减式, 这可以大大简化求导的过程:

$$\log(\mathcal{L}) = \log\left(\prod_{i=1}^N p_i\right) = \sum_{i=1}^N \log(p_i)$$

在机器学习的公式推导中, 经常能看到类似的转化。

Tec # Stats # MachineLearning

◀ BeautifulSoup 实战 (二) :获取文章详细信息

Think Bayes ➤

2条评论

Cescfangs

 推荐 2 推文 分享

评分最高



加入讨论...

通过以下方式登录

或注册一个 DISQUS 帐号 

姓名



月光 Music · 6个月前

我這邊數學式子只持續5秒左右 就壞了...

  · 回复 · 分享,

Cesc 管理员 → 月光 Music · 6个月前

tex公式需要加载 mathjax, 我这边都是可以的, 挂一下代理?

  · 回复 · 分享,

在 CESCFANGS 上还有

Numpy 进阶应用——最大最小值

2条评论 · 2年前



Cesc — 比我当时写的好太多了

我所理解的SVM(支持向量机)2——核函数的应用

2条评论 · 2年前



Cesc — 其实是对的, 图中的 gamma 是 sigma^2 的倒数, sigma^2 越大即 gamma 越小

PyQt5学习笔记&简明教程 (一) by Cescfangs

2条评论 · 4年前



Cesc — 我这条真是有先见之明啊

 订阅  在您的网站上使用 Disqus 添加 Disqus 添加  Disqus 隐私政策 隐私政策 隐私© 2015 - 2018  Ces FangsPowered by [Hexo](#) | Theme - [NexT.Muse](#)