

# 极大似然估计

最大似然估计 (Maximum Likelihood Estimation, MLE) 是一种确定模型参数值的方法。找到参数值以使它们最大化由模型描述的过程产生实际观察到的数据的可能性。

## 1、引例一

先看一个简单的例子：某位同学与一位猎人一起外出打猎，一只野兔从前方窜过。只听一声枪响，野兔应声到地，如果要你推测，这一发命中的子弹是谁 ( $\theta$ ) 打的？

- 直觉和简短的分析告诉我们，应该是猎人( $\theta$ =猎人) 打的这一发子弹
- 原因在于猎人的命中率要远高于一名学生的命中率，即  
 $P(x = \text{命中}; \theta = \text{猎人}) > P(x = \text{命中}; \theta = \text{学生})$ ，因此我们更倾向于认为猎人打了这发子弹

$$P(x = \text{命中}; \theta = \text{猎人}) > P(x = \text{命中}; \theta = \text{学生})$$

## 2、引例二

再考虑一种情况：前方有许多野兔，这时有人打了 $n$ 发子弹，并且命中与否的记录为( $x_1 = \text{命中}, x_2 = \text{未命中}, \dots, x_n = \text{命中}$ )，那么这次的命中记录是谁打的呢？

- 根据第一个例子，我们知道应该将 $m$ 次射击命中记录一起来考虑，即需要验证这次的命中纪录是由谁产生的
- 由于每次射击的命中率是独立时间，因此可以考虑验证下面的不等式：

$$P(x_1, x_2, \dots, x_n; \theta = \text{猎人}) > P(x_1, x_2, \dots, x_n; \theta = \text{学生})$$

其中

$$\begin{aligned} P(x_1, x_2, \dots, x_n; \theta = \text{猎人}) &= P(x_1; \theta = \text{猎人}) \cdot P(x_2; \theta = \text{猎人}) \cdots P(x_n; \theta = \text{猎人}) \\ &= \prod_{i=1}^n P(x_i; \theta = \text{猎人}) \end{aligned}$$

$$\begin{aligned} P(x_1, x_2, \dots, x_n; \theta = \text{学生}) &= P(x_1; \theta = \text{学生}) \cdot P(x_2; \theta = \text{学生}) \cdots P(x_n; \theta = \text{学生}) \\ &= \prod_{i=1}^n P(x_i; \theta = \text{学生}) \end{aligned}$$

## 3、引例三

再考虑一种情况：前方有许多野兔，并且有 $m$ 个候选射击人 ( $\alpha_1, \alpha_2, \dots, \alpha_m$ )，此时有人打了 $n$ 发子弹，并且命中与否的记录为( $x_1 = \text{命中}, x_2 = \text{未命中}, \dots, x_n = \text{命中}$ )，那么这次的命中记录如何确定是哪个候选射击人打的呢？

- 根据之前的分析，我们应该考察每个候选射击人产生本次命中记录的概率（似然），然后选取概率最大的那个（极大似然）
- 因此，直接考虑下面的公式进行推断

$$\arg \max_{\theta} P(x_1, x_2, \dots, x_n; \theta)$$

其中 ( $\theta \in \{\alpha_1, \alpha_2, \dots, \alpha_m\}$ )

## 4、极大似然函数

通常，我们把以上寻找最优参数 $\theta$ 的过程称为极大似然估计，并称 $\mathbf{L}(\theta|X)$ 为似然函数：

$$\begin{aligned}\hat{\theta}(X) &= \arg \max_{\theta} \mathbf{L}(\theta|x_1, x_2, \dots, x_n) \\ &= \arg \max_{\theta} P(x_1, x_2, \dots, x_n; \theta) \\ &= \arg \max_{\theta} \prod_{i=1}^n P(x_i; \theta)\end{aligned}$$

而通常计算联合概率时用到的连乘操作偶尔会导致变量越界，因此更通用的方法是对以上的似然函数求对数，获得对数似然函数，然后在进行求最优的操作。（想想为什么取对数可以保证最有条件是一样的？）

$$\begin{aligned}\hat{\theta}(X) &= \arg \max_{\theta} \ln \mathbf{L}(\theta|x_1, x_2, \dots, x_n) \\ &= \arg \max_{\theta} \ln P(x_1, x_2, \dots, x_n; \theta) \\ &= \arg \max_{\theta} \sum_{i=1}^n \ln P(x_i; \theta)\end{aligned}$$

## 5、求解极大似然值

- 至于如何寻找似然函数 $\mathbf{L}(\theta|X)$ 或者 $\log(\mathbf{L}(\theta|X))$ 的最优值，最简单的方法就是对参数求导 然后获得极值条件。
- 如果求导后不能通过方程组求解，那么可以借助梯度下降法来求解最优值。
- 对于某些更复杂的对数似然函数（例如含有隐变量的），可以借助其他优化方法，例如EM算法求解。

具体解法步骤：

1. 构造似然函数： $\mathbf{L}(\theta|X)$
2. 取对数： $\ln \mathbf{L}(\theta|X)$
3. 求导  $\ln \mathbf{L}(\theta|X)$ ，令  $\ln \mathbf{L}(\theta|X) = 0$
4. 解方程，得到最优估计  $\theta$

## 6、极大似然估计应用一（伯努利分布）

令 $X \in \text{Bernoulli}(\theta)$ ，即 $x_i = 1$ 的概率为 $\theta$ ， $x_i = 0$ 的概率为 $1 - \theta$ ，其中 $\theta \in [0, 1]$ 。则 $x_i$ 的概率分布函数（PDF）为：

$$f(x_i; \theta) = \theta^{x_i} (1 - \theta)^{1-x_i}$$

若 $x_1, x_2, \dots, x_n$ 是按照该分布生成的样本点，则它们的联合概率函数构造的似然函数为：

$$\begin{aligned}\mathbf{L}(\theta|X) &= \theta^{x_1} (1 - \theta)^{1-x_1} \dots \theta^{x_n} (1 - \theta)^{1-x_n} \\ &= \theta^{(x_1 + \dots + x_n)} (1 - \theta)^{n - (x_1 + \dots + x_n)}\end{aligned}$$

$$\begin{aligned}\ln \mathbf{L}(\theta|X) &= \ln(\theta \sum_i x_i) + \ln((1 - \theta)(n - \sum_i x_i)) \\ &= n(\bar{x} \ln \theta + (1 - \bar{x}) \ln(1 - \theta))\end{aligned}$$

$$\frac{\partial \ln \mathbf{L}(\theta|X)}{\partial \theta} = n \frac{\bar{x} - \theta}{\theta(1 - \theta)} = 0$$

即

$$\theta = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

## 7、极大似然估计应用二（正态分布）

**Example 4** (Normal data). Maximum likelihood estimation can be applied to a vector valued parameter. For a simple random sample of  $n$  normal random variables, we can use the properties of the exponential function to simplify the likelihood function.

$$\mathbf{L}(\mu, \sigma^2|\mathbf{x}) = \left( \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{-(x_1 - \mu)^2}{2\sigma^2} \right) \dots \left( \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{-(x_n - \mu)^2}{2\sigma^2} \right) = \frac{1}{\sqrt{(2\pi\sigma^2)^n}} \exp -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2.$$

The score function

$$\ln \mathbf{L}(\mu, \sigma^2|\mathbf{x}) = -\frac{n}{2}(\ln 2\pi + \ln \sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2.$$

$$\frac{\partial}{\partial \mu} \ln \mathbf{L}(\mu, \sigma^2|\mathbf{x}) = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = \frac{1}{\sigma^2} n(\bar{x} - \mu)$$

Because the second partial derivative with respect to  $\mu$  is negative,

$$\hat{\mu}(\mathbf{x}) = \bar{x}$$

is the maximum likelihood estimator. For the derivative of the score function with respect to the parameter  $\sigma^2$ ,

$$\frac{\partial}{\partial \sigma^2} \ln \mathbf{L}(\mu, \sigma^2|\mathbf{x}) = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n (x_i - \mu)^2 = -\frac{n}{2(\sigma^2)^2} \left( \sigma^2 - \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 \right).$$

Recalling that  $\hat{\mu}(\mathbf{x}) = \bar{x}$ , we obtain

$$\hat{\sigma}^2(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

## 8、极大似然估计在伯努利分布下的程序

In [1]:

```
from scipy.stats import bernoulli

# 生成样本, 假设样本服从p为1/2的bernoulli分布
p_1 = 1.0 / 2

# 产生伯努利随机变量
fp = bernoulli(p_1)

# 产生100个样本
xs = fp.rvs(100)

print(xs)
```

```
[0 0 1 0 0 0 1 0 0 1 0 0 1 1 0 1 0 0 1 0 0 0 1 1 1 1 1 0 1 0 1 1 1 0 0 1 0
 1 1 1 1 0 1 1 0 1 1 0 0 1 0 1 1 1 1 1 0 1 1 1 0 0 0 1 1 1 0 1 1 1 1 1 0 1
 1 1 0 1 0 0 1 1 0 1 0 1 0 0 1 0 0 1 0 0 1 1 0 1 0 1]
```

In [2]:

```
import sympy
import numpy as np

x, p, z = sympy.symbols('x p z', positive=True)

# 伯努利分布函数
pdf = p*x*(1-p)**(1-x)

# 似然函数
L = np.prod([pdf.subs(x, i) for i in xs]) # 似然函数

print(L)

# 对数似然函数
logL = sympy.expand_log(sympy.log(L))

# 求解极大似然函数估计
sol = sympy.solve(sympy.diff(logL, p), p)

print(sol)
```

```
p**56*(-p + 1)**44
[14/25]
```

## 本节作业：仿写程序

请仿照上述伯努利分布的构造并求解似然函数的过程，试图补全高斯分布下的极大似然估计过程，并求解 $\mu$ 的极大似然估计。其中高斯分布函数为：

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

▶

In [4]:

```
import sympy
import numpy as np

xs = np.random.normal(5, 1, 10)
print(xs)
sol = 0
x, delta, mu = sympy.symbols('x y z', positive=True)

# 此处写高斯分布函数

# 似然函数

# 对数似然函数

# 求解极大似然函数估计

print(sol)
```

```
[3.93497763 5.16724815 5.5487745  5.83080269 5.73972641 4.93836142
 5.00083258 5.22275963 3.49520998 7.43264482]
0
```