似然函数(Likelihood)、极大似然估计(MLE)及其他

Grissom, 2019/07/16

http://sunbo2019.github.io

目录：

- 样本
- 似然率（又叫“似然函数” Likelihood）
- 参数估计
  - ✓ 矩估计
  - ✓ 极大似然估计（MLE）

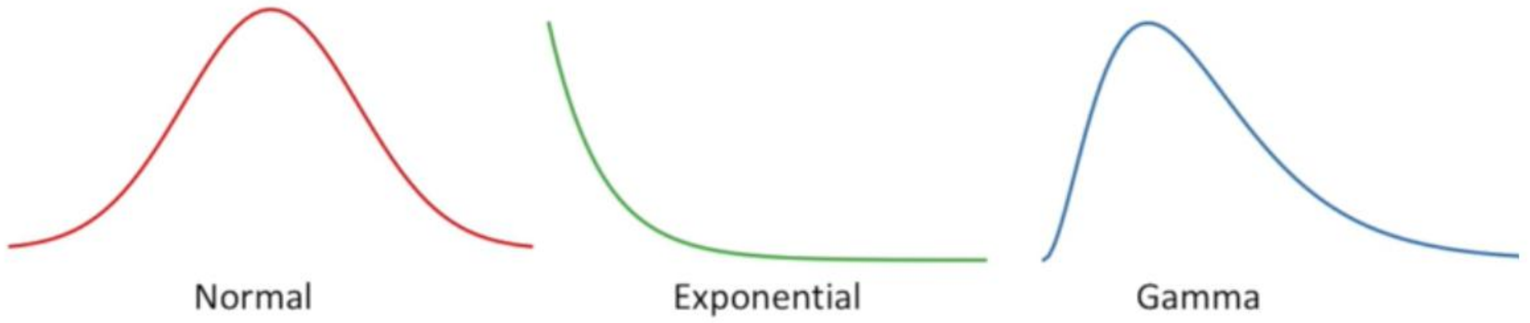
首先，明确几个概念：

（1）样本

我们这里所说的样本，都是简单随机样本，即：独立同分布（i.i.d）。

（2）似然率（又叫“似然函数” Likelihood）

假定数据是服从某一种分布的。在工程实践中，我们可以多尝试几种常用的分布。



（3）参数估计

A. 矩估计
B. 最大似然估计（又叫“极大似然估计”，MLE - Maximum Likelihood Estimate）

似然率（似然函数）

假设某事件的随机变量服从如下**分布率**：

$$x \sim \begin{pmatrix} 0 & 1 & 2 & 3 \\ \theta^2 & 2\theta(1-\theta) & \theta^2 & 1-2\theta \end{pmatrix}$$

*在工程实践上，我们一般是假设数据服从某一种分布，然后再去验证的。

* 本题来自新东方张宇的考研数学。

现在观察到一批数据如下：

$$3, 1, 3, 0, 3, 1, 2, 3$$

那么，观察到这批数据的概率为：

$$\begin{aligned} L(\theta) &= (1-2\theta)^4 \cdot [2\theta(1-\theta)]^2 \cdot \theta^2 \cdot \theta^2 \\ &= 4\theta^6(1-\theta)^2(1-2\theta)^4 \end{aligned}$$

因为这里的 θ 是变量，所以 $L(\theta)$ 是一个函数。我们称之为“**似然函数（Likelihood function）**”。

参数估计（parameter estimation）

参数估计的方法，有**点估计**（point estimation，得到具体的值）、**区间估计**（interval estimation，得到置信范围）。在这里，我们只讨论点估计。点估计又分为如下四种：

①矩估计法：

简言之，此法的精髓是：假定样本分布 = 总体分布。

②最大似然估计法（MLE）:

本质：当参数为何值时，似然函数可以取得最大值？

[能直接求解，则直接得到解集；若难以直接求解，则需要使用**梯度下降法**]

③最小二乘法：

一般用于“线性回归”。

④贝叶斯估计法

回到前面的那个例题：

$$L(\theta)=4\theta^6(1-\theta)^2(1-2\theta)^4$$

$$\Rightarrow \ln L(\theta)=\ln 4+6\ln\theta+2\ln(1-\theta)+4\ln(1-2\theta)$$

$$\Rightarrow \frac{d\ln(\theta)}{d\theta}=\frac{6}{\theta}-\frac{2}{1-\theta}-\frac{2\cdot 4}{1-2\theta}=0$$

$$\Rightarrow \hat{\theta}_{1,2}=\tfrac{1}{12}(7\pm\sqrt{3})$$

（1）用“**极大似然法**”来求解

1. 写出似然函数
2. 对似然函数取对数

   < 取对数的原因 >

   ✓ 简化计算（log函数将<u>乘法</u>转化为<u>加法</u>）

   ✓ 解决了当某一项概率为零时，整个结果的概率都变成零的问题（防止下溢）
3. 求导数

（2）用“**矩估计**”来求解，核心思想：**样本的均值 = 总体的期望**

样本均值：$\overline{x}=(3+1+3+0+3+1+2+3)\cdot\dfrac{1}{8}=2$

总体期望：$EX=0.\theta^2+1.2\theta(1-\theta)+2.\theta^2+3(1-2\theta)(3-4\theta)$

$\overline{x}=E(x)$ $\longrightarrow$ $2=3-4\theta\rightarrow\hat{\theta}=\frac{1}{4}$