

## 估计量

统计模型包含所有看似可能生成给定数据的概率分布。在构建合适的统计模型后，我们假设数据来自模型中的某一个分布。**下一步是确定模型中的哪个分布最符合数据。**

如果模型由一个参数描述，我们唯一不能确定的就是这个参数，有多少个不同的参数取值就代表这个模型中有多少不同的分布，当然根据设定只有一个正确的，在这样的情况下，确定一个最好的分布相当于确定最佳拟合参数值，求得的通常称为“真”参数。

在统计学中，这个过程称为“估计”。其他称呼比如在机器学习领域可能叫“拟合”和“学习”。

假设  $X$  的分布依赖于未知参数  $\theta$ ，因此统计模型为  $\{P_\theta: \theta \in \Theta\}$ ，其中  $P_\theta$  是当  $\theta$  是“真”参数时  $X$  的分布。

基于观察到的  $x$ ，我们希望估计  $\theta$  的真实值，或可以得到  $\theta$  的某个函数  $g(\theta)$  的值。在这里，“根据数据估计”意味着对  $\theta$  或  $g(\theta)$  做出如下形式的陈述：“我认为  $g(\theta)$  大约等于  $T(x)$ ”，其中  $T(x)$  是依赖于观察值  $x$  的某个值。

### 估计量

估计量或统计量是仅依赖于观察值  $X$  的随机向量  $T(X)$ 。对于具体的数据  $x$ ，相应的估计值是  $T(x)$ 。

这里的要点在于仅依赖于**观察值**，

根据这个定义，很多对象都可以是估计量。关键在于  $T(X)$  是一个仅依赖于  $X$  的函数，并且**不依赖于参数  $\theta$** ：我们必须能够从数据  $x$  中计算出  $T(x)$ 。

给定观察值  $x$ ，统计量  $T$  通过  $t = T(x)$  实现，并作为  $\theta$ （或  $g(\theta)$ ）的估计值使用。我们通常将  $T(X)$  简化为  $T$ 。数学上，“统计量”与“估计量”有着相同的定义，但在不同的上下文中使用。

估计量和  $\theta$  的估计值通常用  $\hat{\theta}$  表示。这个符号上的帽子表明  $\hat{\theta}$  是一个依赖于观察值的函数，但这个符号没有区分随机向量和其观察值： $\hat{\theta}$  可以同时表示  $\hat{\theta}(X)$  和  $\hat{\theta}(x)$ 。

有许多估计方法。将先讨论一些通用原则，例如最大似然法、矩估计法和贝叶斯方法。不过，我们首先要搭建一个框架，以便比较不同估计量的性能。这引入了均方误差。