

参数估计

总体、样本、统计量

矩估计

$$\alpha_m = \int_{-\infty}^{\infty} x^m f(x, \theta_1, \dots, \theta_k) dx$$
$$\left(\text{或} \sum_i x_i^m f(x_i, \theta_1, \dots, \theta_k) \right)$$

依赖于 $\theta_1, \dots, \theta_k$. 另一方面, 至少在样本大小 n 较大时, α_m 又应接近于样本原点矩 α_m . 于是

$$\alpha_m = \alpha_m(\theta_1, \dots, \theta_k) \approx a_m = \sum_{i=1}^n X_i^m / n$$

取 $m = 1, \dots, k$, 并让上面的近似式改成等式, 就得到一个方程组:

$$\alpha_m(\theta_1, \dots, \theta_k) = a_m, m = 1, \dots, k \quad (2.1)$$

解此方程组, 得其根 $\hat{\theta}_i = \hat{\theta}_i(X_1, \dots, X_n), i = 1, \dots, k$. 就以 $\hat{\theta}_i$ 作为 θ_i 的估计. $i = 1, \dots, k$. 如果要估计的是 $\theta_1, \dots, \theta_k$ 的某函数 $g(\theta_1, \dots, \theta_k)$, 则用 $\hat{g} = \hat{g}(X_1, \dots, X_n) = g(\hat{\theta}_1, \dots, \hat{\theta}_k)$ 去估计它. 这样定出的估计量就叫做矩估计.

极大似然估计

从上述分析就自然地导致如下的方法:应该用似然程度最大的那个点 $(\theta_1^*, \dots, \theta_k^*)$,即满足条件

$$\begin{aligned} L(X_1, \dots, X_n; \theta_1^*, \dots, \theta_k^*) \\ = \max_{\theta_1, \dots, \theta_k} L(X_1, \dots, X_n; \theta_1, \dots, \theta_k) \end{aligned} \quad (2.3)$$

的 $(\theta_1^*, \dots, \theta_k^*)$ 去作为 $(\theta_1, \dots, \theta_k)$ 的估计值,因为在已得样本 X_1, \dots, X_n 条件下,这个“看来最像”是真参数值.这个估计 $(\theta_1^*, \dots, \theta_k^*)$ 就叫做 $(\theta_1, \dots, \theta_k)$ 的“极大似然估计”.如果要估计的是 $g(\theta_1, \dots, \theta_k)$,则 $g(\theta_1^*, \dots, \theta_k^*)$ 是它的极大似然估计.

因为

$$\log L = \sum_{i=1}^n \log f(X_i; \theta_1, \dots, \theta_k) \quad (2.4)$$

且为使 L 达到最大,只须使 $\log L$ 达到最大,故在 f 对 $\theta_1, \dots, \theta_k$ 存在连续的偏导数时,可建立方程组(称为似然方程组):

$$\frac{\partial \log L}{\partial \theta_i} = 0, i = 1, \dots, k \quad (2.5)$$

如果这方程组有唯一的解,又能验证它是一个极大值点,则它必是使 L 达到最大之点,即极大似然估计.在几个常见的重要例子中这一点不难验证.可是,在较复杂的场合,方程组(2.5)可以有不止一组解,求出这些解很费计算,且不易判定那一个使 L 达到最大.

贝叶斯法

(知道的之前的概率以及现在的证据,推现在的概率)

设总体有概率密度 $f(X, \theta)$ (或概率函数, 若总体分布为离散的), 从这总体抽样本 X_1, \dots, X_n , 则这样本的密度为 $f(X_1, \theta) \cdots f(X_n, \theta)$. 它可视为在给定 θ 值时 (X_1, \dots, X_n) 的密度, 根据第二章 (3.5) 式及该式下的一段说明, $(\theta, X_1, \dots, X_n)$ 的联合密度为

$$h(\theta)f(X_1, \theta) \cdots f(X_n, \theta)$$

由此, 算出 (X_1, \dots, X_n) 的边缘密度为

$$p(X_1, \dots, X_n) = \int h(\theta)f(X_1, \theta) \cdots f(X_n, \theta) d\theta \quad (2.10)$$

积分的范围, 要看参数 θ 的范围而定. 如上例 θ 为废品率, 则 $0 \leq \theta \leq 1$. 若 θ 为指数分布中的参数 λ , 则 $0 < \theta < \infty$, 等等. 由 (2.10), 再根据第二章的公式 (3.4), 得到在给定 X_1, \dots, X_n 的条件下, θ 的条件密度为

$$h(\theta|X_1, \dots, X_n) = h(\theta)f(X_1, \theta) \cdots f(X_n, \theta) / p(X_1, \dots, X_n) \quad (2.11)$$

照贝叶斯学派的观点, 这个条件密度代表了我们现在 (即在取得样本 X_1, \dots, X_n 后) 对 θ 的知识, 它综合了 θ 的先验信息 (以 $h(\theta)$ 反映) 与由样本带来的信息. 通常把 (2.11) 称为 θ 的“后验 (或验后)”

	问 题	先验知识	当前知识	后验 (现在) 知识
贝叶斯公式	事件 B_1, \dots, B_n 中那一个发生了?	$P(B_1),$ $\dots, P(B_n)$	事件 A 发生了	$P(B_1 A), \dots,$ $P(B_n A)$
此处的问题	$\theta = ?$	$h(\theta)$	样本 X_1, \dots, X_n	后验密度 (2.11)