矩估计

矩估计法是最大似然估计法的替代方法。
这是由于矩估计法通常未能利用统计模型中的全部信息，矩估计量通常效率不如最大似然估计量。
然而，这个方法有时更容易执行。此外，矩估计法只需要知道矩的理论形式，而不需要观测值的完整概率分布。所以这些矩通常比完整的概率分布更容易建模，因此这一点可以成为一个显著优势。使用错误的模型来构建估计量带来的偏差的风险也因此降低。

矩和样本矩

设 X 是一个依赖于未知参数 θ 的随机变量，其 j 阶矩为 $\mathrm{E}_\theta\left(X^j\right)$，当然前提是该期望存在。独立同分布变量 $X_1,\ldots,X_n$ 的 j 阶样本矩为 $\overline{X^j}=n^{-1}\sum_{i=1}^n X_i^j$。

j 阶矩可以使用相同分布样本的 j 阶样本矩进行估计。根据大数定律，这对 $\mathrm{E}_\theta\left(X^j\right)$ 是一个良好的估计量。

矩估计量

设 $X_1,\ldots,X_n$ 是来自一个未知参数为 θ 的分布的样本。矩估计量 $\hat\theta$ 是使 j 阶矩与 j 阶样本矩相对应的值：

$$\mathrm{E}_{\hat\theta}\left(X^j\right)=\overline{X^j}$$

对于一个映射 $g:\Theta\to H$，$g(\theta)$ 的矩估计量为 $g(\hat\theta)$。

在实际中，我们通常选择尽可能小的 j 来得到矩估计量。对于一维参数 θ，只要边际分布的期望依赖于 θ，取 $j=1$ 即可。当第一阶矩不依赖于 θ 时，选择 $j=2$，依此类推。如果 θ 是维度大于 1 的参数，则需要多个方程来得到 $\hat\theta$ 的唯一解。在这种情况下，矩估计量 $\hat\theta$ 是通过 $j=1,\ldots,k$ 的方程组求解得到的，其中 k 是可以使得方程组有唯一解的最小整数。

例 指数分布

设 $X_1,\ldots,X_n$ 是来自参数未知的指数分布的样本。则 $\mathrm{E}_\lambda X_i=1/\lambda$。矩估计量 λ 可以通过解方程 $\bar X=1/\hat\lambda$ 得到 $\hat\lambda$。因此，$\hat\lambda=1/\bar X$ 是 λ 的矩估计量。该估计量也是 λ 的最大似然估计量。

例 均匀分布

设 $X_1,\ldots,X_n$ 是来自 $\mathrm{U}[0,\theta]$ 分布的样本，参数 θ 未知。则 $\mathrm{E}_\theta X_i=\theta/2$，因此 θ 的矩估计量为 $\hat\theta=2\bar X$。
不过我们还知道 θ 的最大似然估计量为 $X_{(n)}$。之前就知道，$X_{(n)}$ 的均方误差小于 $2\bar X$。因此，在此情况下我们更倾向于使用最大似然估计量。

例 正态分布

设 $X_1, \ldots, X_n$ 是来自 $\mathrm{N}\left(0, \sigma^2\right)$ 分布的样本，参数 $\sigma^2>0$ 未知。那么 $\mathrm{E}_{\sigma^2} X_i=0$，因此无法使用第一阶矩来确定 σ^2 的矩估计量。X_i 的第二阶矩为 $\mathrm{E}_{\sigma^2} X_i^2=\sigma^2$。因此，$\sigma^2$ 的矩估计量为 $\hat{\sigma}^2=\overline{X^2}$ 。如果 X_i 的期望值未知或不为零，那么我们会得到一个不同的 σ^2 的矩估计量。

例 伽马分布(多参数)

设 $X_1, \ldots, X_n$ 是具有未知形状参数 α 和尺度参数 λ 的伽马分布的随机变量。那么 $\mathrm{E}_{\alpha, \lambda} X_i=\alpha / \lambda$ 且 $\operatorname{var}_{\alpha, \lambda} X_i=\alpha / \lambda^2$，因此第二阶矩为 $\mathrm{E}_{\alpha, \lambda} X_i^2=\operatorname{var}_{\alpha, \lambda} X_i+\left(\mathrm{E}_{\alpha, \lambda} X_i\right)^2=\alpha(1+\alpha) / \lambda^2$。通过解以下方程可以得到 α 和 λ 的矩估计量:

$$
\begin{aligned}
\mathrm{E}_{\hat{\alpha}, \hat{\lambda}} X_i &= \hat{\alpha} / \hat{\lambda}=\bar{X} \\
\mathrm{E}_{\hat{\alpha}, \hat{\lambda}} X_i^2 &= \hat{\alpha}(1+\hat{\alpha}) / \hat{\lambda}^2=\overline{X^2}
\end{aligned}
$$

解得 $\hat{\alpha}$ 和 $\hat{\lambda}$ 为

$$
\hat{\alpha}=\frac{(\bar{X})^2}{\overline{X^2}-(\bar{X})^2} \quad \text{和} \quad \hat{\lambda}=\frac{\bar{X}}{\overline{X^2}-(\bar{X})^2}
$$

由于这个例子用最大似然估计量没有已知的显式表达式，均方误差也无法直接确定。为了基于估计量的性能（偏差和方差）进行选择，可以进行模拟。

续例 期望与方差 (正态普遍版)

设 $X_1, \ldots, X_n$ 是期望为 μ 且方差为 σ^2 的样本。通过解以下方程可得 $\hat{\mu}$ 和 $\hat{\sigma}^2$:

$$
\begin{aligned}
\mathrm{E}_{\hat{\mu}, \hat{\sigma}^2} X_i &= \hat{\mu}=\bar{X} \\
\mathrm{E}_{\hat{\mu}, \hat{\sigma}^2} X_i^2 &= \hat{\mu}^2+\hat{\sigma}^2=\overline{X^2}
\end{aligned}
$$

得到 $\hat{\mu}$ 和 $\hat{\sigma}^2$ 的矩估计量为:

$$
\hat{\mu}=\bar{X}, \quad \hat{\sigma}^2=\overline{X^2}-(\bar{X})^2=\frac{1}{n} \sum_{i=1}^n\left(X_i-\bar{X}\right)^2
$$

如果例子中此分布为 $N\left(\mu, \sigma^2\right)$ 分布，则这些 μ 和 σ^2 的矩估计量等于它们的最大似然估计量。