样本均值和方差的分布

Guangyao Zhao

2022-08-16

Contents

在之前学习统计学的时候，我可能没有真正的理解过『随机变量』这个概念。比如随机变量 X 现有 n 个样本，从中依次有放回的取出 n_1, n_2, n_3 个样本。即现有 3 个样本集 $x_{11}, x_{12}, ..., x_{1i}$，$x_{11}, x_{12}, ..., x_{1j}$，$x_{11}, x_{12}, ..., x_{1k}$。

也就是说每个样本集又有 μ_1, μ_2, μ_3 以及 $\sigma_1, \sigma_2, \sigma_3$ 三个均值和方差，那也就是说样本 n 的均值和方差也成了『随机变量。这就可以根据 μ_1, μ_2, μ_3 以及 $\sigma_1, \sigma_2, \sigma_3$ 对整体的 μ 和 σ 进行统计上的估计。

样本均值和样本方差的数字特征（并非正态分布）

假设整体 X 的 $E(X) = \mu$，$D(X) = \sigma^2$。$X_1, X_2, .., X_n$ 是来自总体 X 的样本，其中 n 是 n 个样本集。则：

- 均值的期望：$E(\bar{X}) = \mu$，n 均值的期望和整体的期望相等
- 均值的方差：$D(\bar{X}) = \frac{\sigma^2}{n}$
- 方差的期望：$E(S^2) = \sigma^2$，方差的期望和整体的方差相等

证明：

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(\bar{X}_i) = \frac{1}{n}n\mu = \mu$$

$$
\begin{aligned}
D(\bar{X}) &= D(\frac{1}{n}\sum_{i=1}^{n} X_i) \\
&= \frac{1}{n^2} D(X_1 + X_2 +, ..., X_n) \\
&= \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n},\, D(X+Y) = D(X) + D(Y)
\end{aligned}
$$

在证明第三个性质的时候有一个前提公式：$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\left(\sum_{i}^{n} X_i^2 - n\bar{X}^2\right)$

$$
\begin{aligned}
E(S^2) &= E\left(\frac{1}{n-1}\left(\sum_{i}^{n} X_i^2 - n\bar{X}\right)^2\right) \\
&= \frac{1}{n-1} E\left(\left(\sum_{i}^{n} X_i^2 - n\bar{X}\right)^2\right) \\
&= \frac{1}{n-1} E\left(E\left(\sum_{i}^{n} X_i^2\right) - E\left(n\bar{X^2}\right)\right) \\
&= \frac{1}{n-1} E\left(nE\left(X^2\right) - nE\left(\bar{X^2}\right)\right) \\
&= \frac{1}{n-1} E\left(n\left(D(X) + E^2(X)\right) - n\left(D(\bar{X}) + E^2(\bar{X})\right)\right) \\
&= \frac{1}{n-1}\left(n\left(\sigma^2 + \mu^2\right) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right) \\
&= \sigma^2
\end{aligned}
$$

注意，以上结论适用于任何分布。

正态样本均值和样本方差的分布（正态分布）

正态样本均值和样本方差的分布的性质符合上述的每一项。