核心结论：样本中心矩是对方差的最大似然估计

最大似然估计（MLE）的核心思想是 **通过观察的数据找到使数据出现概率最大的参数值**。具体步骤为：

1. **写出似然函数**：即样本的联合概率密度函数。
2. **对数似然函数**：简化最大化过程。
3. **求导数并解方程**：找到使似然函数最大化的参数值。

以 **正态分布** 为例，其概率密度函数为：

$$f(x_i|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

参数为均值 μ 和方差 σ^2。我们的目标是找到 μ 和 σ^2 的最大似然估计值 $\hat{\mu}$ 和 $\hat{\sigma}^2$。

写出似然函数

对于独立同分布的样本 $x_1, x_2, \ldots, x_n$，似然函数为：

$$\begin{aligned} L(\mu,\sigma^2) &= \prod_{i=1}^{n} f(x_i|\mu,\sigma^2) \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right) \end{aligned}$$

对数似然函数

取对数简化计算：

$$\ln L(\mu,\sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

求偏导数并解方程

对均值 μ 求偏导

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)$$

令导数为 0：

$$\sum_{i=1}^{n}(x_i-\mu) = 0 \quad \Rightarrow \quad \hat{\mu} = \bar{x}$$

因此，最大似然估计的均值为样本均值 $\hat{\mu}_{\text{MLE}} = \bar{x}$，这与样本中心矩一致。

对方差 σ^2 求偏导

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^{n} (x_i - \mu)^2$$

令导数为 0:

$$-\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^{n} (x_i - \mu)^2 = 0$$

整理得:

$$\hat{\sigma}^2_{\mathrm{MLE}} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$$

代入 $\mu = \hat{\mu}_{\mathrm{MLE}} = \bar{x}$，得到方差的 MLE:

$$\hat{\sigma}^2_{\mathrm{MLE}} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

这正是 **样本中心矩的二阶矩。**

关键区别：为何最大似然不修正为 $\frac{1}{n-1}$？

最大似然的目标与无偏性无关

- 最大似然的目标是最大化数据的似然函数（即找到最可能的参数值），而非消除估计的系统误差（即偏差）。因此:
  - 即使 MLE 的估计量有偏差（如这里的方差 MLE），只要它使似然最大，就被接受为“最佳”解。
  - 修正为无偏（如除以 $n-1$）不是 MLE 的关注点。

方差的无偏估计与 MLE 的对比

- **无偏方差估计**:

  $$\hat{\sigma}^2_{\text{无偏}} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

  这是通过 **调整分母为 $n-1$ 消除偏差**的结果（因为 $\bar{x}$ 是用样本估计的，导致丢掉一个自由度）。
- **方差的 MLE**:

  $$\hat{\sigma}^2_{\mathrm{MLE}} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

  这 **没有考虑自由度的损失**，因为 MLE 着眼于最大化似然，而非无偏性。

补充思考：无偏估计 vs MLE 的适用场景

- **选择无偏性时**:
  - 小样本场景需精确估计参数。
  - 对偏差敏感，更看重估计量的长期无系统误差。
- **选择 MLE 时**:
  - 更关注渐近性质（如大样本下的一致性、有效性）。

    - 接受短时有偏但更高效（方差小）的估计。
    - 计算简单且适用范围广。

在实际应用中，无偏估计和 MLE 是统计学中互补的方法，选择取决于具体问题和目标。