# Part III-B: Probability Theory and Mathematical Statistics

Lecture by 李漫漫
Note by THF

2024 年 11 月 19 日

## 目录



## Lecture 16

11.12

*Review:*

1. 抽样分布定理：定理 6.4.3
2. 表达式 $\frac{(n-1)S^2}{\sigma^2}$ 什么时候是统计量：$\sigma^2$ 已知

当 $\sigma^2$ 未知时：表达式符合 $\chi^2(n-1)$，称为**枢轴量**

**Corollary.** $X_1, X_2, \dots X_n$ 来自总体 $X \sim N(\mu, \sigma^2)$，样本均值和方差记为 $\bar{X}, S^2$，则：

*a.* 求 $ES^2$：$ES^2 = \sigma^2$
*b.* 求 $DS^2$：$DS^2 = \frac{2\sigma^4}{n-1}$
*c.* 构造 $t$ 分布：$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} / \sqrt{\frac{S^2}{\sigma^2}} = \frac{\bar{X}-\mu}{S}\sqrt{n} \sim t(n-1)$

（上述 $\sqrt{(n-1)S^2/\sigma^2} \sim \chi^2(n-1)$ 且 $(\bar{X}-\mu)/\frac{\sigma}{\sqrt{n}} \sim \Phi(x)$）

重点题目：例 6.4.4

**Corollary.** 两个正态总体的抽样分布定理：两个总体记为 $X, Y$，样本容量分别为 $m, n$，样本分布分别符合 $\mu = \mu_1, \sigma^2 = \sigma_1^2$ 和 $\mu = \mu_2, \sigma^2 = \sigma_2^2$ 的正态分布，...

# 1 参数估计

参数：*param*

- 点估计
  - 矩估计
  - 极大似然估计
- 区间估计

## 1.1 矩估计

使用样本矩 $\bar{X}$ 替代总体矩 $\hat{\mu}$

**Notation.** 总体矩不存在（无穷）时不能使用矩估计

**Example.** 总体 $X \sim U[a,b]$，$X_1, X_2, \ldots, X_n$ 为总体的样本，求 $a, b$ 的矩估计量

解：

$$\begin{cases} EX & = \frac{a+b}{2} \\ DX & = \frac{(b-a)^2}{12} \end{cases}.$$

求解得：$\begin{cases} a & = EX - \sqrt{3DX} \\ b & = EX + \sqrt{3DX} \end{cases}$，替代后为：$\begin{cases} \hat{a} & = \bar{X} - \sqrt{3M_2^\star} \\ \hat{b} & = \bar{X} + \sqrt{3M_2^\star} \end{cases}$，$\hat{a}$ 代表对 $a$ 的估计

## 1.2 极大似然估计

似然函数：样本的联合概率分布函数（P167）

**Example.** 同矩估计例题，求 $a, b$ 的极大似然估计量

解：写出联合密度函数：$L(\ldots) = \prod_{i=1}^{n} f_{X_i}(x_i)$，由于 $f_X(x) = \begin{cases} \frac{1}{b-a}, & x \in [a,b] \\ 0, & \text{Others} \end{cases}$，则联合密度函数为：

$$\begin{aligned} L &= \prod_{i=1}^{n} \frac{1}{b-a} I_{[a,b]}(x_i) \\ &= \frac{1}{(b-a)^n} \prod_{i=1}^{n} I_{[a,b]}(x_i). \end{aligned}$$

## Lecture 18

11.19

*Review:*

◦ 对参数（完全可观测数据）进行点估计：极大似然估计、矩估计

两种方法得到的结果可能一样或不一样

◦ 点估计的评价标准：

**（渐进）无偏** 估计的参数的期望 $E\hat{\theta}$ 求极限为 $\theta$
**有效性** 在无偏的前提下：$D\hat{\theta}$ 越小越有效
**相合性（一致性）** $\mathrm{MSE}\left(\hat{\theta},\theta\right)$：均方误差

**Example.** $\bar{X}$ 和 $\hat{W}=\sum_{i=1}^{n}a_iX_i$ 可以证明都是 $\mu$ 的无偏估计，称 $\bar{X}$ 为算术均值，$\hat{W}$ 为加权均值，且算术均值比加权均值更有效（均值不等式）

**Notation.** 均方误差：$\mathrm{MSE}\left(\hat{\theta},\theta\right)=E\left(\hat{\theta}-\theta\right)^2$

### 1.3 区间估计

**Definition.** 置信区间：$\alpha$ 为给定值，总体的分布函数为 $F\left(x,\theta\right)$ ，有两个从总体中抽取后构造的统计量 $T_1,T_2$ ，当：

$$P\left\{T_1<\theta<T_2\right\}=1-\alpha.$$

时：称 $P$ 为置信度，区间长度 $T_2-T_1$ 的数学期望 $E\left(T_2-T_1\right)$ 为精度

纽曼提出的准则：先确定 $\alpha$ 来确定置信度，再确定置信上下限

**Notation.** 已知标准正态分布 $X\sim N\left(\mu,\sigma^2\right)$ 中的 $\sigma^2$ ，置信度为 $1-\alpha$ 时参数 $\mu$ 的置信区间：

$$\left(T_1,T_2\right)=\left(\bar{X}-\frac{\sigma}{\sqrt{n}}u_{1-\frac{\alpha}{2}},\bar{X}+\frac{\sigma}{\sqrt{n}}u_{1-\frac{\alpha}{2}}\right).$$

如果 $\sigma^2$ 未知：通过 $t$ 分布可得：$T=\frac{\bar{X}-\mu}{S}\sqrt{n}\sim t\left(n-1\right)$ ，置信区间为：

$$\left(T_1,T_2\right)=\left(\bar{X}-\frac{S}{\sqrt{n}}t_{1-\frac{\alpha}{2}}\left(n-1\right),\bar{X}+\frac{S}{\sqrt{n}}t_{1-\frac{\alpha}{2}}\left(n-1\right)\right).$$

**Notation.** 卡方分布下：方差为 $\sigma^2$ ，置信度为 $1-\alpha$ 的置信区间：

$$\left(T_1,T_2\right)=\left(\frac{\left(n-1\right)S^2}{\chi^2_{1-\frac{\alpha}{2}}\left(n-1\right)},\frac{\left(n-1\right)S^2}{\chi^2_{\frac{\alpha}{2}}\left(n-1\right)}\right).$$

对应的标准差为 $\sigma$ ，置信度为 $1-\alpha$ 的置信区间：$T_1'=\sqrt{T_1},T_2'=\sqrt{T_2}$

如果 $\mu$ 已知，$\sigma^2$ 未知，令 $S_1^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\mu\right)^2$ ，因为 $\chi^2=\frac{nS_1^2}{\sigma^2}\sim\chi^2\left(n\right)$ ，可得置信度为 $1-\alpha$ 的方差的置信区间为：

$$\left(\frac{nS_1^2}{\chi^2_{1-\frac{\alpha}{2}}\left(n\right)},\frac{nS_1^2}{\chi^2_{\frac{\alpha}{2}}\left(n\right)}\right).$$

## 2 假设检验

**Notation.** 在医药研发大量应用

- 参数假设检验：假设效果
- 非参数假设检验

**Notation.** 下节课之前准备一个本专业的假设检验问题