# 数理统计 · 2025 Spring

## 期中复习

2025 年 4 月

# 考试重点

- 指数型分布族.
- 统计中的三大分布.
- 样本均值和样本方差的分布.
- 次序统计量的分布.
- 数据压缩: 充分统计量和完全统计量.
- 矩估计和最大似然估计.
- 一致最小方差无偏估计.
- 置信区间的计算与含义.
- 贝叶斯模型与贝叶斯估计.

总体分布 $\longrightarrow$ 统计模型

$\downarrow$

指数型分布族?

$\downarrow$

(极小) 充分统计量? 完全统计量?

$\downarrow$

矩估计、最大似然估计

$\downarrow$

一致最小方差无偏估计

$\downarrow$

点估计的相合性与渐近正态性

$\downarrow$

置信区间

$\downarrow$

贝叶斯点估计与可信区间

**Example 1 (2023 • 期末, 2024 • 期中, 可携带计算器)**

一个移动通讯公司随机抽取了共 900 个包月客户, 计算得知他们一个月平均使用时间是 220 分钟, 样本标准差是 90 分钟. 假设使用时间服从正态分布 $N(\mu, \sigma^2)$.

1. 求包月客户平均使用时间和标准差的 95% 置信区间.
2. 如果要求客户平均使用时间的 95% 置信区间的长度不超过 5 分钟, 应至少抽取多少个客户? 该公司的抽样规模是否满足要求?
3. 假设总体标准差为 90 分钟, 取 $\mu$ 的先验分布为无信息先验, 求 $\mu$ 的后验分布并据此给出 $\mu$ 的可信系数为 95% 的可信区间.
4. 解释 (1) 和 (3) 中关于平均使用时间所得区间的含义与区别.

附表: $u_{0.025} = 1.960$, $u_{0.05} = 1.645$, $\chi^2_{899}(0.025) = 984$, $\chi^2_{899}(0.975) = 817.8$.

**试题溯源**: 教材 4.2.2 节 (单个正态总体参数的置信区间); 例 4.5.1 (正态-正态共轭下的可信区间); 习题二 4、5、9、33.

**(6 分) 均值的置信区间:**

- 取合适的枢轴量并给出其分布: (3 分);
- 写出均值的置信区间: (2 分);
- 代入数值计算区间的实现: (1 分).

**(6 分) 标准差的置信区间:**

- 取合适的枢轴量并给出其分布: 3 分;
- 写出标准差的置信区间: 2 分;
- 代入数值计算区间的实现: 1 分.

**(6 分) 样本规模的计算:**

- 求出置信区间长度: 2 分;
- 按照长度要求反解得到样本规模的要求: 2 分;
- 代入数值并给出结论: 2 分.

**(5 分) 均值的后验分布:**

- 写出似然函数和均值的先验分布: 2 分;
- 计算均值的后验分布: 3 分.

**(3 分) 均值的可信区间:**

- 给出可信区间的形式: 2 分;
- 计算区间的实现: 1 分.

**(4 分) 均值的可信区间:**

- 置信区间的意义: 2 分;
- 可信区间的意义: 2 分.

**注:** 均值的置信区间应含有 $t$ 分布的分位数, 利用 $t$ 分布与正态分布的性质将分位数转化为正态分布分位数; 在计算样本规模时同理.

**Example 2 (2021 • 期中, 2024 • 期中, 可携带计算器)**

下表统计了某铁路局 122 个扳道员五年内由于操作失误引起的严重事故情况, 其中 $r$ 表示一扳道员某五年内引起严重事故的次数, $s$ 表示扳道员人数. 假设扳道员由于操作失误在五年内所引起的严重事故的次数服从 Poisson 分布. 求

| $r$ | 0 | 1 | 2 | 3 | 4 | 5 | $\geq 6$ |
|---|---|---|---|---|---|---|---|
| $s$ | 44 | 42 | 21 | 9 | 4 | 2 | 0 |

1. 一个扳道员在五年内未引起严重事故的概率 $p$ 的一致最小方差无偏估计 $\hat{p}_1$ 和最大似然估计 $\hat{p}_2$.
2. 说明 $p$ 不存在可以达到 C-R 不等式下界的无偏估计.
3. $p$ 的一个 (渐近) 95% 水平的置信上界.

**试题溯源**: 教材 3.3.2 节 (最大似然估计的求法); 教材 3.4.3 节 (充分完全统计量法); 教材 3.5.2 节 (单参数 C-R 不等式); 教材 4.3 节 (非正态总体参数的置信区间); 例 4.3.5 (泊松分布参数的置信区间).

**(9 分) 最大似然估计:**

- 似然函数及对数似然函数: 3 分;
- 利用对数似然方程得到 Poisson 分布参数的最大似然估计: 2 分;
- 利用最大似然估计的不变性, 得到 $p$ 的最大似然估计;
    - 计算 $p = \mathbb{P}(X_1 = 0) = \mathrm{e}^{-\lambda}$: 1 分;
    - 最大似然估计量 $\hat{p}_2$: 1 分.
- 代入数值计算得到估计值.
    - 样本均值 $\bar{x}$: 1 分;
    - 最大似然估计值 $\hat{p}_2$: 1 分.

**(6 分) 一致最小方差无偏估计:**

- 寻找充分完全统计量: 2 分;
- 将无偏估计 $I(X_1 = 0)$ 条件化得到估计量 $\hat{p}_1$: 3 分;
- 一致最小方差无偏估计值: 1 分.

**(5 分) 说明 $p$ 没有可达 C-R 下界的无偏估计:**

- 用 Cramér-Rao 不等式.
  - 计算 Fisher 信息量: 2 分;
  - 求 Cramér-Rao 下界和估计量方差并比较: 3 分.
- 直接用 Cramér-Rao 不等式等号成立条件.
  - 写出样本的密度函数并指出 $T$ 的形式: 3 分;
  - 指明等号成立的条件: 2 分.

**(5 分) 置信上界:**

- 构造枢轴量并给出其极限分布: 3 分;
- 据此给出置信上界及其值: 2 分.

**Example 3 (2023 • 期中)**

设从总体

| $X$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $\mathbb{P}$ | $3\theta/4$ | $\theta/4$ | $2\theta$ | $1-3\theta$ |

(其中 $0<\theta<1/3$) 中抽取的一个简单样本 $X_1,\ldots,X_n$, 试

1. 将样本分布表示为指数族自然形式, 指出自然参数及自然参数空间.
2. 求 $\theta$ 的最大似然估计量 $\hat{\theta}$, 其是否为一致最小方差无偏估计?
3. 证明 $\hat{\theta}$ 具有相合性和渐近正态性.

**试题溯源**: 教材 1.5.2 节 (指数族的自然形式及自然参数空间); 教材 3.1.3 节 (点估计的优良性准则); 教材 3.3.2 节 (最大似然估计的求法); 教材 3.4 节 (一致最小方差无偏估计); 例 1.5.7、例 3.3.8; 习题一 16、习题三 12.

**(5 分) 指数族自然形式表示:**

- 指数族的自然形式表示: 3 分;
- 自然参数: 1 分;
- 自然参数空间: 1 分.

**(3 分) 最大似然估计:**

- 似然函数或对数似然函数: 1 分;
- 求偏导得到最大似然估计并验证: 2 分.

**(4 分) 一致最小方差无偏估计:**

- 用充分完全统计量法.
  - 充分完全统计量: 2 分;
  - Lehmann-Scheffé 定理 (无偏性证明): 2 分.
- 用 Cramér-Rao 不等式.
  - 计算 Cramér-Rao 下界: 2 分;
  - 计算估计量的方差并比较: 2 分.

**(4 分) 相合性:**

- 若证强相合.
  - 使用强大数律: 4 分.
- 若证弱相合.
  - 使用切比雪夫不等式: 4 分.

**(4 分) 渐近正态性:**

- 利用课本定理 3.3.4 说明是 CAN 估计.
  - 计算 Fisher 信息量: 2 分;
  - 说明 CAN 估计并得出结果: 2 分.
- 利用中心极限定理证明.
  - 计算期望和方差: 2 分;
  - 应用中心极限定理: 2 分.

注: 证明过程过于简略会有相应程度的扣分.

**Example 4 (2023 • 期中)**

设 $X_1, \ldots, X_n$ i.i.d. $\sim N(\theta, 1)$, $Y_1, \ldots, Y_n$ i.i.d. $\sim N(2\theta, 1)$, $\theta$ 为未知参数. 又设合样本 $X_1, \ldots, X_n, Y_1, \ldots, Y_n$ 独立. 试

1. 求 $\theta$ 的 UMVUE.
2. 如果要求 $\theta$ 的置信系数为 $1-\alpha$ 的置信区间宽不超过指定的 $d$, 则样本量 $n$ 应该至少多大?

**试题溯源**: 教材 3.4 节 (一致最小方差无偏估计); 教材 4.1 节 (区间估计的相关概念); 习题四 6 (高度相似).

**(10 分) UMVUE:**

- 合样本的联合密度函数: 4 分;
- 指出充分完全统计量: 4 分;
- 根据 Lehmann-Scheffé 定理得到 UMVUE: 2 分.

**(10 分) 样本量的确定:**

- 枢轴量及其分布的给出: 4 分;
- 得到正确的置信区间: 2 分;
- 计算区间长度: 2 分;
- 按照长度要求得到样本量: 2 分.

注: 本题每小题逻辑过程正确均各有 2 分.

**Example 5 (2021 • 期中)**

设 $X_1, \ldots, X_n$ 是来自正态总体 $N(\mu, 1)$ 的一组简单随机样本.

1. 证明: 样本均值 $\overline{X}$ 与统计量 $X_1 - \overline{X}$ 相互独立.
2. 求概率 $\mathbb{P}(X_1 \leq 0)$ 的 UMVUE 及其置信水平为 $1-\alpha$ 的置信区间.

**试题溯源**: 教材 3.4.3 节 (充分完全统计量法); 教材 4.2 节与 4.3 节 (枢轴变量法); 定理 2.7.2 (Basu 定理); 习题二 53、例 3.4.7.

请不要完全针对该历年题进行复习，
仍需回到课本与作业题中！

预祝大家期中考取得好成绩！

Thanks for your time and attention!