# 数理统计第一次习题课

2023.4.1

# 目录

# 基本概念回顾

- **总体**：一个统计问题所研究的对象的全体
- **样本空间**：样本 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 可能取值的全体，记为 $\mathcal{X}$
- **简单随机样本 (简单样本或随机样本)**：独立同分布
- **样本分布**：样本作为随机变量服从的概率分布
- **参数**：样本分布中的未知的常数
- **参数空间**：参数的取值范围
- **样本分布族**：当参数取不同值时得到的不同分布构成的一个分布族
- **参数统计推断**：样本分布形式已知，确定未知参数的值
- **非参数统计推断**：样本分布形式未知，通过样本对总体的分布作推断

# 基本概念回顾

- **统计量**：由样本算出的量，是样本的函数

  注 统计量只与样本有关，不能与未知参数有关
- **样本均值**：$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
- **样本方差**：$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

  注 也可以定义为 $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$，但 $S^2$ 无偏
- **样本矩**
  - $k$ 阶原点矩：$a_{n,k} = \frac{1}{n}\sum_{i=1}^{n} X_i^k$
  - $k$ 阶中心矩：$m_{n,k} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k$
- **样本协方差**：$S_{XY} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$
- **样本变异系数**：$\hat{\nu} = \frac{S_n}{\bar{X}}$
- **样本偏度**：$\hat{\beta}_1 = \frac{m_{n,3}}{m_{n,2}^{3/2}} = \sqrt{n}\sum_{i=1}^{n}(X_i - \bar{X})^3 / \left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right)^{3/2}$
- **样本峰度**：$\hat{\beta}_2 = \frac{m_{n,4}}{m_{n,2}^2} - 3 = n\sum_{i=1}^{n}(X_i - \bar{X})^4 / \left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right)^2 - 3$

# 基本概念回顾

- **依概率收敛** ($X_n \xrightarrow{P} X$)：对 $\forall \epsilon$, $\lim_{n\to\infty} \mathrm{P}(|X_n - X| \geq \epsilon) = 0$
- **几乎处处收敛** ($X_n \xrightarrow{a.s.} X$)：$\mathrm{P}(\lim_{n\to\infty} X_n = X) = 1$
- **依分布收敛** ($X_n \xrightarrow{\mathcal{L}} X$)：$\lim_{n\to\infty} F_n(x) = F(x)$, $x$ 为 $F$ 的任一连续点
- $r$ **阶矩收敛** ($X_n \xrightarrow{L_r} X$)：$\lim_{n\to\infty} \mathrm{E}|X_n - X|^r = 0$
- **四种收敛关系**:
  - $X_n \xrightarrow{L_r \ or \ a.s.} X \Rightarrow X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{\mathcal{L}} X$
  - $X_n \xrightarrow{\mathcal{L}} c_0 \Rightarrow X_n \xrightarrow{P} c_0$
  - $X_n \xrightarrow{P} c_0$ 且 $g$ 在 $c_0$ 连续 $\Rightarrow g(X_n) \xrightarrow{P} g(c_0)$
- **连续映射定理**: $X_n \xrightarrow{\mathcal{L}} X$ 且 $g$ 连续 $\Rightarrow g(X_n) \xrightarrow{\mathcal{L}} g(X)$, 其中依分布收敛可换成依概率收敛和几乎处处收敛
- **Slutsky 定理**：$X_n \xrightarrow{\mathcal{L}} X$ 且 $Y_n \xrightarrow{P} c_0 \Rightarrow$
  - $X_n + Y_n \xrightarrow{\mathcal{L}} X + c_0$
  - $X_n Y_n \xrightarrow{\mathcal{L}} c_0 X$
  - $X_n / Y_n \xrightarrow{\mathcal{L}} X / c_0$, $c_0 \neq 0$

# 基本概念回顾

- **切比雪夫不等式**: $X$ 随机变量, $\forall b > 0$ 有
$$\mathrm{P}(|X - \mathrm{E}X| \geq b) \leq \frac{\mathrm{Var}(X)}{b^2}.$$
- **控制收敛定理**: $X_n \overset{a.s.}{\to} X$ 且对充分大的 $n$ 有 $|X_n| \leq Y$, 其中 $\mathrm{E}(Y) < \infty$, 则 $\lim_{n\to\infty} \mathrm{E}(X_n) = \mathrm{E}(X)$.
- **强大数律**: $X_1, \cdots, X_n$ 独立同分布且 $\mathrm{E}|X_1| < \infty$ $\Rightarrow \bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i \overset{a.s.}{\to} \mathrm{E}X$.
- **中心极限定理**: 设 $\{X_i\}$ 为一列独立同分布的随机变量序列, 有共同的均值 $\mu$ 和方差 $0 < \sigma^2 < \infty$, 则有 $\sqrt{n}(\bar{X}_n - \mu)/\sigma \overset{\mathcal{L}}{\to} N(0,1)$, 其中 $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.
- **Delta 方法**: $\sqrt{n}(X_n - \theta) \overset{\mathcal{L}}{\to} N(0, \sigma^2)$, 其中 $\sigma^2 > 0$ 和 $\theta$ 均为有限数, $g'(x)$ 在 $x = \theta$ 处连续非零, 则
$$\sqrt{n}\,[g(X_n) - g(\theta)] \overset{\mathcal{L}}{\to} N(0, [g'(\theta)]^2\sigma^2).$$

# 基本概念回顾

- **次序统计量**：设 $X_1, \cdots, X_n$ 为从总体 $F$ 中抽取的样本，按大小排列 $X_{(1)} \leq \cdots \leq X_{(n)}$，则 $(X_{(1)}, \cdots, X_{(n)})$ 为样本 $(X_1, \cdots, X_n)$ 的次序统计量
  注 $(X_{(1)}, \cdots, X_{(n)})$ 的任一部分也称为次序统计量
- **样本中位数**：

$$m_{1/2} = \begin{cases} X_{((n+1)/2)}, & n\text{为奇数} \\ \frac{1}{2}\left[X_{(n/2)} + X_{(n/2+1)}\right], & n\text{为偶数} \end{cases} \tag{0.1}$$

- **极值**：$X_{(1)}$ 和 $X_{(n)}$ 称为样本的极小值和极大值
- **样本极差**：$R = X_{(n)} - X_{(1)}$

# 基本概念回顾

**经验分布函数**：$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I\{X_i \leq x\}$

- $nF_n(x) \sim B(n, F(x))$ 且

$$\frac{\sqrt{n}(F_n(x) - F(x))}{\sqrt{F(x)(1-F(x))}} \xrightarrow{\mathcal{L}} N(0,1)$$

- Glivenko-Cantelli 定理：$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \to 0,\ a.s.$
- DKW 不等式：对任意 $\epsilon > 0$ 和 $n$, 有

$$P\left(\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| > \epsilon\right) \leq 2e^{-2n\epsilon^2}$$

# 基本概念回顾

- **单个次序统计量的分布**：$X_{(m)}$ 的分布, $1 \le m \le n$, cdf 为

$$F_m(x) = P\left(X_{(m)} \le x\right) = \sum_{i=m}^{n} \left( \begin{array}{c} n \\ i \end{array} \right) (F(x))^i (1 - F(x))^{n-i},$$

pdf 为

$$f_m(x) = F'_m(x) = m \left( \begin{array}{c} n \\ m \end{array} \right) (F(x))^{m-1} (1 - F(x))^{n-m} f(x).$$

- $n$ **个次序统计量的联合分布**：

$$g\left(y_1, y_2, \ldots, y_n\right) = \begin{cases} n! f\left(y_1\right) f\left(y_2\right) \cdots f\left(y_n\right), & y_1 < y_2 < \cdots < y_n \\ 0, & \text{其他}. \end{cases}$$

- **两个次序统计量** $\left(X_{(i)}, X_{(j)}\right)$ **的联合分布**：

$$f_{ij}(x, y) = \begin{cases} \frac{n!}{(i-1)!(j-i-1)!(n-j)!} (F(x))^{i-1} (F(y) - F(x))^{j-i-1} & \\ \times (1 - F(y))^{n-j} f(x) f(y), & x < y, \\ 0, & \text{其他}. \end{cases}$$

**极差的分布**

- 作下列变换:

$$\begin{cases} V = X_{(j)} - X_{(i)}, \\ Z = X_{(i)} \end{cases} \Longleftrightarrow \begin{cases} X_{(i)} = Z, \\ X_{(j)} = V + Z. \end{cases}$$

- Jacobi 行列式 $J = \left| \frac{\partial(X_{(i)}, X_{(j)})}{\partial V \partial Z} \right| = 1$
- $(V, Z)$ 的联合密度

$$g_{ij}(v, z) = \begin{cases} \frac{n!}{(i-1)!(j-i-1)!(n-j)!}(F(z))^{i-1}(F(v+z) - F(z))^{j-i-1} \\ \quad \times (1 - F(v+z))^{n-j} f(z) f(v+z), & v > 0, \\ 0, & \text{其他}. \end{cases}$$

- $V$ 的密度为 $g_V(v) = \int_{-\infty}^{\infty} g_{ij}(v, z) dz$.
- 取 $i = 1,\ j = n$ 即可得到极差 $R = X_{(n)} - X_{(1)}$ 的密度.

**样本分位数**

- 对于 $0 < p < 1$, 分布 $F$ 的 $p$ 分位数为
$$\xi(p) = F^{-1}(p) = \inf\{x : F(x) \geq p\}.$$
- **样本 $p$ 分位数**: $X_{\lceil np \rceil}$, 即 $\hat{\xi}_{np} = F_n^{-1}(p) = \inf\{x : F_n(x) \geq p\}$.
- $X_{(i)} = F_n^{-1}(\frac{i}{n})$, $i = 1, \cdots, n$
- $\hat{\xi}_{np} \xrightarrow{P} \xi_p$, 如果 $\xi_p$ 是 $F(x-) \leq p \leq F(x)$ 中 $x$ 的唯一解.
- Q-Q 图: $\left(x_{(i)}, F^{-1}(\frac{i}{n+1})\right)$, $i = 1, \cdots, n$.

# 基本概念回顾

## $\chi^2$ 分布

- 定义：$X_1, \cdots, X_n$ i.i.d. $\sim N(0,1)$, 则 $\xi = \sum_{i=1}^n X_i^2 \sim \chi_n^2$.
- $\chi^2$ 分布的 pdf：

$$g_n(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x > 0 \\ 0, & x \le 0. \end{cases}$$

- **Gamma 分布**：$\Gamma(\alpha, \lambda)$ 的 pdf 为

$$f(x; \alpha, \lambda) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x > 0, \\ 0, & x \le 0. \end{cases}$$

- $\Gamma(n/2, 1/2)$ 就是 $\chi_n^2$；若 $Y \sim \Gamma(\alpha, \lambda)$, 则 $2\lambda Y \sim \chi_{2\alpha}^2$.
- $\chi_n^2$ 的特征函数 $\varphi(t) = (1-2it)^{-n/2}$.
- $\mathrm{E}(\xi) = n$, $\mathrm{Var}(\xi) = 2n$.
- $Z_1 \sim \chi_{n_1}^2$, $Z_2 \sim \chi_{n_2}^2$, 且 $Z_1$ 和 $Z_2$ 独立, 则 $Z_1 + Z_2 \sim \chi_{n_1+n_2}^2$.

# 基本概念回顾

## $t$ 分布

- 定义：$X \sim N(0,1)$, 则 $Y \sim \chi_n^2$, 且 $X$ 和 $Y$ 独立, 则 $T = \frac{X}{\sqrt{Y/n}} \sim t_n$.
- $t$ 分布的 pdf：

$$t_n(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, -\infty < x < \infty.$$

- $\mathrm{E}(T^r)$ 只有 $r < n (n > 1)$ 时存在, 且

$$E(T^r) = \begin{cases} n^{\frac{r}{2}} \frac{\Gamma\left(\frac{r+1}{2}\right)\Gamma\left(\frac{n-r}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{1}{2}\right)}, & r \text{ 为偶数}, \\ 0, & r \text{ 为奇数}. \end{cases}$$

  特别, 当 $n \geq 2$ 时, $\mathrm{E}(T) = 0$, 当 $n \geq 3$ 时, $\mathrm{Var}(T) = n/(n-2)$.
- 当 $n = 1$ 时, $t$ 分布就是柯西分布 $t_1(x) = \frac{1}{\pi(1+x^2)}$, $-\infty < x < +\infty$.
- 当 $n \to \infty$ 时, $t$ 变量的极限分布为 $N(0,1)$.

# 基本概念回顾

## $F$ 分布

- 定义：$X \sim \chi^2_m$, 则 $Y \sim \chi^2_n$, 且 $X$ 和 $Y$ 独立, 则 $F = \frac{X/m}{Y/n} \sim F_{m,n}$.
- $F$ 分布的 pdf：

$$f_{m,n}(x) = \begin{cases} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)} m^{\frac{m}{2}} n^{\frac{n}{2}} x^{\frac{m}{2}-1}(n+mx)^{-\frac{m+n}{2}}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

- $Z \sim F_{m,n}$, 则 $1/Z \sim F_{n,m}$.
- 对 $r > 0$, 有

$$E(Z^r) = \left(\frac{n}{m}\right)^r \frac{\Gamma\left(\frac{m}{2}+r\right)\Gamma\left(\frac{n}{2}-r\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)}, \quad 2r < n.$$

- $T \sim t_n$ 则 $T^2 \sim F_{1,n}$.
- $F_{m,n}(1-\alpha) = 1/F_{n,m}(\alpha)$.

# 基本概念回顾

**正态总体性质**:

- $X_1, \cdots, X_n$ 独立且 $X_k \sim N(a_k, \sigma_k^2)$, $k = 1, \cdots, n$, $c_1, \cdots, c_n$ 是常数不全为零, 则

$$T = \sum_{k=1}^{n} c_k X_k \sim N\left(\sum_{k=1}^{n} c_k a_k, \sum_{k=1}^{n} c_k^2 \sigma_k^2\right).$$

- 若 $X_1, \cdots, X_n$ i.i.d. $\sim N(a, \sigma^2)$, 样本均值 $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$, 样本方差 $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$, 则
  - $\bar{X} \sim N(a, \sigma^2/n)$;
  - $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$;
  - $\bar{X}$ 和 $S^2$ 独立;
  - $T_n = \frac{\sqrt{n}(\bar{X}-a)}{S} \sim t_{n-1}$.

**非正态总体**可使用中心极限定理和 Delta 方法.

**1.4** 一物体的重量 $a$ 未知, 有两架天平可用, 其随机误差分别服从正态分布 $N(0,\sigma_1^2)$ 和 $N(0,\sigma_2^2)$, $\sigma_1^2$ 和 $\sigma_2^2$ 都未知. 先把物件在第一架天平上称两次得 $X_1, X_2$, 再在第二架天平上称两次得 $X_3, X_4$, 然后视 $|X_1 - X_2| \leq |X_3 - X_4|$ 或否而在第一架天平或第二架天平上再称 $n-4$ 次得 $X_5, \ldots, X_n$. 写出 $(X_1, \ldots, X_n)$ 的密度.

**1.5** 设总体 $X$ 服从两点分布 $b(1,p)$ (即 $\mathrm{P}(X=1)=p$, $\mathrm{P}(X=0)=1-p$), 其中 $p$ 是未知参数, $\boldsymbol{X}=(X_1,\dots,X_5)$ 为此次总体中抽取的简单样本,
(1) 写出样本空间 $\mathscr{X}$ 和 $\boldsymbol{X}$ 的概率分布;
(2) 指出 $X_1+X_2$, $\min\limits_{1\le i\le 5} X_i$, $X_5+2p$, $X_5-\mathrm{E}(X_1)$, $(X_5-X_1)^2/\mathrm{Var}(X_1)$ 哪些是统计量, 哪些不是, 为什么?
(3) 若样本观测值 $(X_1,\dots,X_n)$ 中有 $m$ 个 1, $n-m$ 个 0, 求此样本的经验分布函数.

**1.5** 设总体 $X$ 服从两点分布 $b(1, p)$ (即 $\mathrm{P}(X=1)=p$, $\mathrm{P}(X=0)=1-p$), 其中 $p$ 是未知参数, $\boldsymbol{X}=(X_1,\dots,X_5)$ 为此次总体中抽取的简单样本,

(1) 写出样本空间 $\mathcal{X}$ 和 $\boldsymbol{X}$ 的概率分布;

(2) 指出 $X_1+X_2$, $\min\limits_{1\le i\le 5} X_i$, $X_5+2p$, $X_5-\mathrm{E}(X_1)$, $(X_5-X_1)^2/\mathrm{Var}(X_1)$ 哪些是统计量, 哪些不是, 为什么?

(3) 若样本观测值 $(X_1,\dots,X_n)$ 中有 $m$ 个 1, $n-m$ 个 0, 求此样本的经验分布函数.

**注** 统计量只与样本有关, 不能与未知参数有关.

**经验分布函数**: $F_n(x)=\frac{1}{n}\sum_{i=1}^{n} I\{X_i\le x\}$

**1.11** 设 $X_1, \ldots, X_n$ 是取自总体 $X \sim N(\mu, \sigma^2)$ 的简单样本, 假如要以 99.7% 的概率保证偏差 $|\overline{X} - \mu| < 0.1$, 试问在 $\sigma^2 = 0.5$ 时, 样本容量 $n$ 应取多大?

**1.9** 在正态总体 $N(50, 6^2)$ 中抽取容量为 36 的简单样本, 求样本均值 $\overline{X}$ 落在 50.6 和 51.8 之间的概率.

**1.12** 利用切比雪夫不等式求一枚均匀硬币需抛掷多少次才能使样本均值 $\overline{X}$ 落在 0.4 和 0.6 之间的概率至少为 0.9 (此处 $X_i = 1$ 表示抛掷硬币出现正面, 否则 $X_i = 0$, $i = 1, \ldots, n$)? 若用中心极限定理计算这个问题, 需抛掷的次数又是多少?

**1.12** 利用切比雪夫不等式求一枚均匀硬币需抛掷多少次才能使样本均值 $\overline{X}$ 落在 0.4 和 0.6 之间的概率至少为 0.9 (此处 $X_i = 1$ 表示抛掷硬币出现正面, 否则 $X_i = 0$, $i = 1, \ldots, n$)? 若用中心极限定理计算这个问题, 需抛掷的次数又是多少?

**切比雪夫不等式**: $X$ 随机变量, $\forall b > 0$ 有 $\mathrm{P}(|X - \mathrm{E}X| \geq b) \leq \frac{\mathrm{Var}(X)}{b^2}$.
**中心极限定理**: 设 $\{X_i\}$ 为一列独立同分布的随机变量序列, 有共同的均值 $\mu$ 和方差 $0 < \sigma^2 < \infty$, 则有 $\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{\mathcal{L}} N(0,1)$, 其中 $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

**2.3** 设 $X_1, \ldots, X_n$ 为从下列总体中抽取的简单样本, 利用特征函数试求样本均值 $\overline{X}$ 的分布:

(1) 正态总体 $N(a, \sigma^2)$;

(2) 参数为 $\lambda$ 的 Poisson 总体 $P(\lambda)$;

(3) 参数为 $\lambda$ 的指数分布.

**2.3** 设 $X_1,\ldots,X_n$ 为从下列总体中抽取的简单样本, 利用特征函数试求样本均值 $\overline{X}$ 的分布:
(1) 正态总体 $N(a,\sigma^2)$;
(2) 参数为 $\lambda$ 的 Poisson 总体 $P(\lambda)$;
(3) 参数为 $\lambda$ 的指数分布.

- $\Gamma(\alpha,\lambda)$ 的 pdf 为

$$f(x;\alpha,\lambda)=\begin{cases}\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}, & x>0,\\ 0, & x\le 0.\end{cases}$$

- Gamma 分布的特征函数: $\varphi(t)=(1-\frac{it}{\lambda})^{-\alpha}$.
- 对于 Gamma 分布 $\Gamma(\alpha,\lambda)$, 当 $\alpha=1$ 时, 即为指数分布 $\mathrm{Exp}(\lambda)$.
- $X_1,\cdots,X_n$ i.i.d. $\sim\mathrm{Exp}(\lambda)$, 则 $\sum_{i=1}^n X_i\sim\Gamma(n,\lambda)$.
- Gamma 分布的可加性: $X\sim\Gamma(\alpha,\lambda)$, $Y\sim\Gamma(\beta,\lambda)$, 且 $X$ 和 $Y$ 独立, 则 $X+Y\sim\Gamma(\alpha+\beta,\lambda)$.

# 作业讲解

**2.4** 设 $X_1, \ldots, X_n$ 是从两点分布 $b(1, p)$ 中抽取的简单样本, $0 < p < 1$, $\overline{X}$ 和 $S_n^2$ 为样本均值和样本方差, 求 $S_n^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2/n$ 的分布. (提示: $S_n^2 = \overline{X}(1 - \overline{X})$)

# 作业讲解

**2.5** 设 $X_1, X_2$ 为取自正态总体 $X \sim N(0, \sigma^2)$ 的一个样本, 证明统计量 $X_1/X_2$ 和 $\sqrt{X_1^2 + X_2^2}$ 是相互独立的.

**2.6** 设 $X_1, \cdots, X_n$ i.i.d. $\sim F$, $(X_{(1)}, X_{(2)}, \cdots, X_{(n)})$ 是其次序统计量, 已知

$$\mathrm{P}(X_{(m)} < x) = \sum_{i=m}^{n} \binom{n}{i} (F(x))^i (1 - F(x))^{n-i},$$

证明恒等式:

$$\sum_{i=m}^{n} \binom{n}{i} (F(x))^i (1 - F(x))^{n-i} = m \binom{n}{m} \int_0^{F(x)} t^{m-1} (1 - t)^{n-m} \mathrm{d}t.$$

**2.14** 设 $X_1, \ldots, X_n$ 独立, $X_i \sim N(0, \sigma_i^2)$, $i = 1, \ldots, n$, 定义

$$\xi = \sum_{i=1}^{n} \frac{(X_i - Z)^2}{\sigma_i^2}, \quad \text{其中 } Z = \sum_{i=1}^{n} \frac{X_i}{\sigma_i^2} \Big/ \sum_{i=1}^{n} \frac{1}{\sigma_i^2},$$

求 $\xi$ 的分布.

# 作业讲解

**2.27** 设总体 $X$ 服从双指数分布, 其分布函数为

$$F(x) = \begin{cases} 1 - \exp\left\{-\dfrac{x-\mu}{\sigma}\right\}, & x > \mu, \\ 0, & x \leq \mu, \end{cases}$$

其中 $-\infty < \mu < +\infty$, $0 < \sigma < +\infty$, $X_{(1)} \leq \cdots \leq X_{(n)}$ 是样本 $X_1, \ldots, X_n$ 的次序统计量, 试证明 $\dfrac{2(n-i+1)}{\sigma}(X_{(i)} - X_{(i-1)})$ 服从自由度为 2 的 $\chi^2$ 分布 $(i = 2, \ldots, n)$.

**2.27** 设总体 $X$ 服从双指数分布, 其分布函数为

$$F(x) = \begin{cases} 1 - \exp\left\{-\dfrac{x-\mu}{\sigma}\right\}, & x > \mu, \\ 0, & x \le \mu, \end{cases}$$

其中 $-\infty < \mu < +\infty$, $0 < \sigma < +\infty$, $X_{(1)} \le \cdots \le X_{(n)}$ 是样本 $X_1, \dots, X_n$ 的次序统计量, 试证明 $\dfrac{2(n-i+1)}{\sigma}(X_{(i)} - X_{(i-1)})$ 服从自由度为 2 的 $\chi^2$ 分布 $(i = 2, \dots, n)$.

作下列变换:

$$\begin{cases} V = X_{(j)} - X_{(i)}, \\ Z = X_{(i)} \end{cases} \iff \begin{cases} X_{(i)} = Z, \\ X_{(j)} = V + Z. \end{cases}$$

从而 Jacobi 行列式 $J = \left|\frac{\partial(X_{(i)}, X_{(j)})}{\partial V \partial Z}\right| = 1$, 得到 $(V, Z)$ 的联合密度

$$g_{ij}(v,z) = \begin{cases} \frac{n!}{(i-1)!(j-i-1)!(n-j)!}(F(z))^{i-1}(F(v+z) - F(z))^{j-i-1} \\ \times(1 - F(v+z))^{n-j} f(z) f(v+z), & v > 0, \\ 0, & \text{其他}. \end{cases}$$

**2.34** 设 $\xi_n \sim \chi_n^2$, 证明: 当 $n \to \infty$ 时, $(\xi_n - n)/\sqrt{2n} \xrightarrow{\mathcal{L}} N(0,1)$. 利用这一事实给出 $\chi_n^2$ 的 $p$ 分位数点与 $N(0,1)$ 的 $p$ 分位数点之间的一个近似关系.

**2.34** 设 $\xi_n \sim \chi_n^2$, 证明: 当 $n \to \infty$ 时, $(\xi_n - n)/\sqrt{2n} \xrightarrow{\mathcal{L}} N(0,1)$. 利用这一事实给出 $\chi_n^2$ 的 $p$ 分位数点与 $N(0,1)$ 的 $p$ 分位数点之间的一个近似关系.

**中心极限定理**: 设 $\{X_i\}$ 为一列独立同分布的随机变量序列, 有共同的均值 $\mu$ 和方差 $0 < \sigma^2 < \infty$, 则有 $\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{\mathcal{L}} N(0,1)$, 其中 $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.

# 补充例题

设 $X_{(1)}, \cdots, X_{(n)}$ 为从均匀分布 $U(0,1)$ 中抽取的次序统计量

(1) 样本容量 $n$ 为多大时, 才能使 $\mathrm{P}(X_{(n)} \geq 0.99) \geq 0.95$?

(2) 求极差 $R_n = X_{(n)} - X_{(1)}$ 的密度函数;

(3) 证明统计量 $Z = 2n(1 - R_n)$ 极限分布为 $\chi_4^2$.

# 补充例题

设 $X_{(1)},\cdots,X_{(n)}$ 为从均匀分布 $U(0,1)$ 中抽取的次序统计量
(1) 样本容量 $n$ 为多大时, 才能使 $\mathrm{P}(X_{(n)}\geq 0.99)\geq 0.95$?
(2) 求极差 $R_n=X_{(n)}-X_{(1)}$ 的密度函数;
(3) 证明统计量 $Z=2n(1-R_n)$ 极限分布为 $\chi_4^2$.
**解** (1)$X_{(n)}$ 的分布函数为 $F_n(x)=(F(x))^n$. 为使得 $\mathrm{P}(X_{(n)}\geq 0.99)\geq 0.95$
即 $(F(0.99))^n\leq 0.05$. 由 $X_{(1)},\cdots,X_{(n)}$ 的分布有 $0.99^n\leq 0.05$, 故
$n\geq 298.07$, $n$ 取 299.

# 补充例题

设 $X_{(1)}, \cdots, X_{(n)}$ 为从均匀分布 $U(0,1)$ 中抽取的次序统计量
(1) 样本容量 $n$ 为多大时, 才能使 $\mathrm{P}(X_{(n)} \geq 0.99) \geq 0.95$?
(2) 求极差 $R_n = X_{(n)} - X_{(1)}$ 的密度函数;
(3) 证明统计量 $Z = 2n(1 - R_n)$ 极限分布为 $\chi_4^2$.
**解** (2) 作下列变换:

$$\begin{cases} R_n = X_{(n)} - X_{(1)}, \\ S = X_{(1)} \end{cases} \Longleftrightarrow \begin{cases} X_{(1)} = S, \\ X_{(n)} = R_n + S. \end{cases}$$

从而 Jacobi 行列式 $J = \left| \frac{\partial(X_{(i)}, X_{(j)})}{\partial R_n \partial S} \right| = 1$, 得到 $(R_n, S)$ 的联合密度

$$g_{1n}(r,s) = \begin{cases} n(n-1)r^{n-2}, & 0 < r < 1,\ 0 < s < 1,\ 0 < r+s < 1 \\ 0, & \text{其他}. \end{cases}$$

则 $R_n$ 的密度函数为

$$g(r) = \int_0^{1-r} n(n-1)r^{n-2} ds = n(n-1)r^{n-2}(1-r),\ 0 < r < 1.$$

# 补充例题

设 $X_{(1)}, \cdots, X_{(n)}$ 为从均匀分布 $U(0,1)$ 中抽取的次序统计量
(1) 样本容量 $n$ 为多大时, 才能使 $\mathrm{P}(X_{(n)} \geq 0.99) \geq 0.95$?
(2) 求极差 $R_n = X_{(n)} - X_{(1)}$ 的密度函数;
(3) 证明统计量 $Z = 2n(1 - R_n)$ 极限分布为 $\chi_4^2$.
**解** (3) 由 $Z = 2n(1 - R_n)$, $R_n = 1 - \frac{Z}{2n}$, $J = \frac{1}{2n}$, 从而 $Z$ 的密度函数为

$$f(z) = \frac{n-1}{4n} z(1 - \frac{z}{2n})^{n-2}, \ z > 0,$$

极限分布为

$$\lim_{n\to\infty} \frac{n-1}{4n} z(1 - \frac{z}{2n})^{n-2} = \lim_{n\to\infty} \frac{1}{4} z \left[(1 - \frac{z}{2n})^{-\frac{2(n-2)}{z}}\right]^{-\frac{z}{2}} = \frac{1}{4} z e^{-\frac{z}{2}},$$

即 $\chi_4^2$.

感谢观看！