

第 7 讲 统计量与抽样分布

知识梳理

· 事先说明：数理统计部分中的 X_1, \dots, X_n 都是独立同分布的随机变量

一 常见统计量

样本均值

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

样本方差

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

样本标准差

$$S = \sqrt{S^2}$$

样本 k 阶矩

$$A_k = \frac{1}{n} \sum_{i=1}^n X_i^k$$

样本 k 阶中心矩

$$B_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k$$

· 比较重要的是样本均值和样本方差

1. 样本均值的性质

样本均值的期望

$$E(\bar{X}) = E(X)$$

样本均值的方差

$$D(\bar{X}) = \frac{D(X)}{n}$$

2. 样本方差的性质

样本方差的期望

$$E(S^2) = D(X)$$

二 三个抽样分布

1. χ^2 分布

χ^2 分布

$$X_i \sim N(0, 1) \rightarrow Y = \sum_{i=1}^n X_i^2 \sim \chi^2(n)$$

① 期望和方差

χ^2 分布的期望

$$E(Y) = n$$

χ^2 分布的方差

$$D(Y) = 2n$$

② 性质

· 若 $Y_1 \sim \chi^2(m)$, $Y_2 \sim \chi^2(n)$ 且相互独立, 则 $Y_1 + Y_2 \sim \chi^2(m+n)$

2. t 分布

t 分布

$$X \sim N(0,1), Y \sim \chi^2(n) \text{ 且相互独立} \rightarrow t = \frac{X}{\sqrt{Y/n}} \sim t(n)$$

2. F 分布

t 分布

$$Y_1 \sim \chi^2(n_1), Y_2 \sim \chi^2(n_2) \text{ 且相互独立} \rightarrow F = \frac{Y_1/m}{Y_2/n} \sim F(m,n)$$

$$\cdot \frac{1}{F} \sim F(n,m), \text{ 若 } X \sim t(n), \text{ 则 } X^2 \sim F(1,n)$$

三 正态总体下抽样分布的特有性质

1. 均值和方差满足分布

$$\cdot \text{ 若 } X_i \sim N(\mu, \sigma^2), \text{ 则 } \bar{X} \text{ 和 } S^2 \text{ 相互独立, 且}$$

正态总体的样本均值

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

正态总体的样本方差

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

2. 正态总体的样本方差的方差

正态总体的样本方差的方差

$$D(S^2) = \frac{2\sigma^4}{n-1}$$

题型解析

十二 求解统计量的性质

1. 题型简述与解法

- 通常以填空题的形式考查由样本、样本均值、样本方差等构成的统计量的事件概率、期望、方差等
- 经常用到的知识点：
 - ① 期望、方差、协方差的运算性质
 - ② 正态分布的性质，以及正态变量的线性组合依然是正态变量
 - ③ 样本均值、样本方差的期望和方差，以及正态总体下的特殊性质
- 题干经常会有一大串的式子，其中一些实际上是样本方差的展开式，要能看出来
- 需要注意：不是所有的题目都是正态分布，要看清楚题干再使用正态总体下的结论

2. 历年考试典型例题

① 计算协方差或相关系数

例 1 (15-16 秋冬) 设总体 $X \sim N(0, 4)$, X_1, \dots, X_{100} 为来自 X 的简单随机样本, 则 $\sum_{i=1}^{60} X_i$ 与 $\sum_{i=51}^{100} |X_i|$ 的相关系数为_____。

解 根据协方差的性质:

$$\text{Cov}\left(\sum_{i=1}^{60} X_i, \sum_{i=51}^{100} |X_i|\right) = \sum_{i=1}^{60} \sum_{j=51}^{100} \text{Cov}(X_i, |X_j|)$$

由于 X_i 之间相互独立, 因此

$$\sum_{j=51}^{60} \text{Cov}(X_i, |X_i|)$$

由于

$$\text{Cov}(X_i, |X_i|) = E(X_i |X_i|) - E(X_i)E(|X_i|)$$

$$E(X_i |X_i|) = \int_{-\infty}^{+\infty} x |x| f(x) dx = 0 \quad (x |x| f(x) \text{ 为奇函数})$$

$$E(X_i) = 0$$

$$\text{因此 } \text{Cov}(X_i, |X_i|) = 0 \rightarrow \text{Cov}\left(\sum_{i=1}^{60} X_i, \sum_{i=51}^{100} |X_i|\right) = 0 \rightarrow \text{相关系数为 } 0$$

② 计算事件概率

例 2 (19-20 秋冬) 设总体 $X \sim N(\mu, 1)$, X_1, \dots, X_{16} 是 X 的简单随机样本, $\overline{Y}_1 = (X_1 + \dots + X_4)/4$, $\overline{Y}_2 = (X_5 + \dots + X_{16})/12$, 则 $P(|\overline{Y}_1 - \mu| < 1) =$ _____。

解 根据正态总体的性质, $\overline{Y}_1 \sim N(\mu, \frac{1}{4})$, 则由正态分布的性质, $\overline{Y}_1 - \mu \sim N(0, \frac{1}{4})$

$$\text{因此 } P(|\bar{Y}_1 - \mu| < 1) = P(|\frac{\bar{Y}_1 - \mu}{1/2}| < \frac{1}{1/2}) = 2\Phi(2) - 1$$

例 3 (16-17 春夏) 设总体 $X \sim N(\mu, \sigma^2)$, μ, σ^2 未知, X_1, \dots, X_{16} 为来自 X 的简单随机样本, \bar{X} 和 S^2 分别为样本均值与样本方差, 则 $P\left(\frac{X_1 + X_2}{2} - \bar{X} > \sigma\right) = \underline{\hspace{2cm}}$;

解 由于 \bar{X} 中包含了 X_1 和 X_2 , 需要注意减号左右两边不是相互独立的

$$\therefore \frac{X_1 + X_2}{2} \sim N(\mu, \frac{\sigma^2}{2}), \quad \bar{X} \sim N(\mu, \frac{\sigma^2}{16}), \quad \text{记 } Y = \frac{X_1 + X_2}{2} - \bar{X}$$

$$\therefore E(Y) = \mu - \mu = 0, \quad D(Y) = \frac{\sigma^2}{2} + \frac{\sigma^2}{16} - 2\text{Cov}(\frac{X_1 + X_2}{2}, \bar{X})$$

$$\therefore \text{Cov}(\frac{X_1 + X_2}{2}, \bar{X}) = \text{Cov}(\sum_{i=1}^2 \frac{1}{2} X_i, \sum_{j=1}^{16} \frac{1}{16} X_j) = \frac{1}{32} \text{Cov}(\sum_{i=1}^2 X_i, \sum_{j=1}^{16} X_j) = \frac{D(X_1) + D(X_2)}{32} = \frac{\sigma^2}{16}$$

$$\therefore D(Y) = \frac{\sigma^2}{2} + \frac{\sigma^2}{16} - 2 \cdot \frac{\sigma^2}{16} = \frac{7}{16} \sigma^2 \quad \therefore Y \sim N(0, \frac{7}{16} \sigma^2)$$

$$\therefore P(Y > \sigma) = 1 - \Phi(4/\sqrt{7}) = 1 - \Phi(1.51) = 0.07$$

② 计算均值和方差

例 4 (20-21 秋冬) 设总体 $X \sim N(\mu, \sigma^2)$, X_1, \dots, X_n 是 X 的简单随机样本, \bar{X} , S^2 分别是样本均值和样本方差, 若 $\mu = 0$, 则 $\text{Var}(S^2 - \bar{X}^2) = \underline{\hspace{2cm}}$.

解 根据正态总体的性质, S^2 与 \bar{X} 相互独立

$$\therefore \bar{X} \sim N(0, \frac{\sigma^2}{n}) \rightarrow \frac{\bar{X}}{\sigma/\sqrt{n}} \sim N(0, 1) \rightarrow \frac{n}{\sigma^2} \bar{X}^2 \sim \chi^2(1)$$

$$\therefore \text{Var}(\frac{n}{\sigma^2} \bar{X}^2) = 2 \rightarrow \frac{n^2}{\sigma^4} \text{Var}(\bar{X}^2) = 2 \rightarrow \text{Var}(\bar{X}^2) = \frac{2\sigma^4}{n^2}$$

$$\therefore \text{Var}(S^2) = \frac{2\sigma^4}{n-1} \rightarrow \text{Var}(S^2 - \bar{X}^2) = \text{Var}(S^2) + \text{Var}(\bar{X}^2) = \frac{2\sigma^4}{n-1} + \frac{2\sigma^4}{n^2}$$

例 5 (19-20 秋冬) 设总体 $X \sim N(\mu, 1)$, X_1, \dots, X_{16} 是 X 的简单随机样本, $\bar{Y}_1 = (X_1 + \dots + X_4)/4$,

$$\bar{Y}_2 = (X_5 + \dots + X_{16})/12, \quad \text{则 } \text{Var}\left[\sum_{i=1}^4 (X_i - \bar{Y}_1)^2 + \sum_{i=5}^{16} (X_i - \bar{Y}_2)^2\right] = \underline{\hspace{2cm}}.$$

解 注意到 $\sum_{i=1}^4 (X_i - \bar{Y}_1)^2 = (4-1)S_1^2$, $\sum_{i=5}^{16} (X_i - \bar{Y}_2)^2 = 11S_2^2$, 且两者相互独立 (没有重叠的 X_i)

由正态总体 $\text{Var}(S^2) = \frac{2\sigma^4}{n-1}$, 得

$$\begin{aligned} \text{Var}\left[\sum_{i=1}^4 (X_i - \bar{Y}_1)^2 + \sum_{i=5}^{16} (X_i - \bar{Y}_2)^2\right] &= \text{Var}\left[\sum_{i=1}^4 (X_i - \bar{Y}_1)^2\right] + \text{Var}\left[\sum_{i=5}^{16} (X_i - \bar{Y}_2)^2\right] \\ &= \text{Var}(3S_1^2) + \text{Var}(11S_2^2) = 3^2 \times \frac{2}{3} + 11^2 \times \frac{2}{11} = 28 \end{aligned}$$

十三 判断统计量函数的分布

1. 题型简述与解法

- 填空题, 判断统计量函数属于哪种抽样分布, 或已知含未知数的统计量函数满足某分布, 求该未知数
- 既然是判断抽样分布, 则总体一定是正态分布, 以下结论都是适用的:

常用结论 1

$$\frac{n(\bar{X} - \mu)^2}{\sigma^2} \sim \chi^2(1)$$

常用结论 2

$$\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \sim \chi^2(n)$$

常用结论 3

$$\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi^2(n-1)$$

常用结论 3 实际上是 S^2 的展开式, 因为题目通常给出方差的展开式而不是 S^2

- 如果 $\mu = 0$, 此时要意识到 X_i^2 可以除以 σ^2 变成 χ^2 分布 (或者用常用结论 2)
- 大多数题目会考察 F 分布, 即会把以上分布作比值处理, 此时 σ^2 、 n 等参数都会被合并或约掉, 使我们一眼看不出来, 这时就要根据以上结论将给的式子拆成两个 χ^2 分布

2. 历年考试典型例题

例 1 (15-16 春夏) 设总体 $X \sim N(\mu, \sigma^2)$, X_1, \dots, X_n 为来自 X 的简单随机样本, 则

$$\sum_{i=1}^3 \left(X_i - \frac{X_1 + X_2 + X_3}{3} \right)^2 / \sum_{i=4}^5 (X_i - \mu)^2 \sim \text{_____ 分布}.$$

解 $\sum_{i=1}^3 \left(X_i - \frac{X_1 + X_2 + X_3}{3} \right)^2$ 具有 3 个样本的方差形式, 根据常用结论 3:

$$\frac{1}{\sigma^2} \sum_{i=1}^3 \left(X_i - \frac{X_1 + X_2 + X_3}{3} \right)^2 \sim \chi^2(2)$$

根据常用结论 2:

$$\frac{1}{\sigma^2} \sum_{i=4}^5 (X_i - \mu)^2 \sim \chi^2(2)$$

两者相互独立, 因此

$$\sum_{i=1}^3 \left(X_i - \frac{X_1 + X_2 + X_3}{3} \right)^2 / \sum_{i=4}^5 (X_i - \mu)^2 = \frac{\frac{1}{\sigma^2} \sum_{i=1}^3 \left(X_i - \frac{X_1 + X_2 + X_3}{3} \right)^2 / 2}{\frac{1}{\sigma^2} \sum_{i=4}^5 (X_i - \mu)^2 / 2} \sim F(2, 2)$$

例 2 (15-16 秋冬) 设总体 $X \sim N(0, 4)$, X_1, \dots, X_{100} 为来自 X 的简单随机样本, 则 $\sum_{i=1}^{100} |X_i|$ 近似服从

分布, $\left(\sum_{i=1}^{50} X_i \right)^2 / \sum_{i=51}^{100} X_i^2 \sim \text{_____ 分布}.$

解 $\sum_{i=1}^{50} X_i$ 具有前 50 个样本的均值形式, 根据常用结论 1:

$$\frac{50(\frac{1}{50}\sum_{i=1}^{50}X_i)^2}{\sigma^2} = \frac{\frac{1}{50}(\sum_{i=1}^{50}X_i)^2}{\sigma^2} \sim \chi^2(1)$$

注意到 $\mu=0$ ，因此 $\sum_{i=51}^{100}X_i^2 = \sum_{i=51}^{100}(X_i-0)^2$ ，根据常用结论 2:

$$\frac{1}{\sigma^2}\sum_{i=51}^{100}X_i^2 \sim \chi^2(50)$$

$$\therefore \text{两者相互独立: } \left(\sum_{i=1}^{50}X_i\right)^2 / \sum_{i=51}^{100}X_i^2 = \frac{50 \cdot \frac{\frac{1}{50}(\sum_{i=1}^{50}X_i)^2}{\sigma^2}}{\frac{\frac{1}{\sigma^2}\sum_{i=51}^{100}X_i^2}{\frac{1}{\sigma^2}\sum_{i=51}^{100}X_i^2/50}} = \frac{\frac{\frac{1}{50}(\sum_{i=1}^{50}X_i)^2}{\sigma^2}}{\frac{1}{\sigma^2}\sum_{i=51}^{100}X_i^2/50} \sim F(1, 50)$$

例 3 (15-16 秋冬) 设总体 $X \sim N(\mu, 1)$ ， μ 未知， X_1, \dots, X_{25} 为来自 X 的简单随机样本， \bar{X} 是样本均值，若 $a(\bar{X} - X_1)^2 \sim \chi^2(1)$ ，则 $a =$ _____.

解 $\because \bar{X} \sim N(\mu, \frac{1}{25}) \quad X_1 \sim N(\mu, 1) \quad \text{Cov}(\bar{X}, X_1) = \frac{1}{25} D(X_1) = \frac{1}{25}$

$$\therefore D(\bar{X} - X_1) = \frac{1}{25} + 1 - 2 \times \frac{1}{25} = \frac{24}{25} \quad \rightarrow \quad \bar{X} - X_1 \sim N(0, \frac{24}{25})$$

$$\therefore \text{由常用结论 2, } \frac{1 \times (\bar{X} - X_1)^2}{24/25} \sim \chi^2(1), \text{ 比对得到 } a = \frac{25}{24}$$