概率论与数理统计　第 5–7 章复习

余启帆[1]

2022 年 12 月 11 日

在前

本次习题课主要对点估计和区间估计部分的重点内容进行梳理, 讲解作业题中的错题难题、小测题以及部分补充题.

1 知识梳理

1.1 统计学基本概念

1.1.1 统计量

1 样本均值

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{1.1}$$

2 样本方差

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \tag{1.2}$$

- 关于分母是 $n-1$ 的解释: 无偏性

3 样本矩

(1) k 阶原点矩

$$a_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k, \quad k = 1, 2, \cdots \tag{1.3}$$

(2) k 阶中心矩

$$m_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k, \quad k = 1, 2, \cdots \tag{1.4}$$

4 次序统计量　$X_{(1)} \leqslant X_{(2)} \leqslant \cdots \leqslant X_{(n)}$

1.1.2 三大分布

1 χ^2 分布　$X_1, \cdots, X_n\,\text{i. i. d.} \sim \mathcal{N}(0,1)$,

$$X = \sum_{i=1}^{n} X_i^2 \sim \chi_n^2 \tag{1.5}$$

是自由度为 n 的 χ^2 变量.

[1]闻道有先后, 讲义有疏漏. 发现错误欢迎联系我: qifan@mail.ustc.edu.cn
欢迎访问主页: http://home.ustc.edu.cn/~qifan/

(1) $\mathrm{E}\,X = n\,\mathrm{E}\,X_i^2 = n\,\mathrm{Var}\,X_i = n$

(2) $\mathrm{Var}\,X = n\,\mathrm{Var}\,X_i^2 = n(\mathrm{E}\,X_i^4 - 1) = 2n$

(3) X_1, X_2 独立, $X_1 \sim \chi_m^2$, $X_2 \sim \chi_n^2$, 则 $X_1 + X_2 \sim \chi_{m+n}^2$

2 t 分布　$X \sim \mathcal{N}(0,1)$, $Y \sim \chi_n^2$, X, Y 独立,

$$T = \frac{X}{\sqrt{\frac{Y}{n}}} \sim t_n \tag{1.6}$$

是自由度为 n 的 t 变量.

(1) 密度函数与正态分布相似, 但更扁平

(2) 期望与方差:

(a) $n = 1$, $\mathrm{E}\,T$ 不存在

(b) $n \geqslant 2$, $\mathrm{E}\,T = 0$

(c) $n \geqslant 3$, $\mathrm{Var}\,T = \dfrac{n}{n-2}$

3 F 分布　$X \sim \chi_m^2$, $Y \sim \chi_n^2$, X, Y 独立,

$$F = \frac{X/m}{Y/n} \sim F_{m,n} \tag{1.7}$$

是自由度为 (m, n) 的 F 变量.

(1) $F \sim F_{m,n} \implies \dfrac{1}{F} \sim F_{n,m}$

(2) $T \sim t_n \implies T^2 \sim F_{1,n}$

(3) $F_{n,m}(1-\alpha) = \dfrac{1}{F_{m,n}(\alpha)}$, $F_{m,n}(\alpha)$: 上 α 分位点

4 正态总体的分布　$X_1, \cdots, X_n$ i. i. d. $\sim \mathcal{N}(\mu, \sigma^2)$,

$$\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i, \quad S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2 \tag{1.8}$$

(1) $\overline{X} \sim \mathcal{N}\left(\mu, \dfrac{1}{n}\sigma^2\right)$

(2) $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$

(3) $\overline{X}$ 和 S^2 独立

(4)

$$T = \frac{\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}\Big/(n-1)}} = \frac{\sqrt{n}(\overline{X}-\mu)}{S} \sim t_{n-1} \tag{1.9}$$

5 混合样本　$X_1, \cdots, X_m$ i. i. d. $\sim \mathcal{N}(\mu_1, \sigma_1^2)$, $Y_1, \cdots, Y_n$ i. i. d. $\sim \mathcal{N}(\mu_2, \sigma_2^2)$ 且两者独立. 混合样本方差

$$S_w^2 := \frac{1}{m+n-2}[(m-1)S_1^2 + (n-1)S_2^2] \tag{1.10}$$

(1) 若两者方差相等但未知, 即 $\sigma_1^2 = \sigma_2^2 := \sigma^2$, 则

$$\frac{(m-1)S_1^2 + (n-1)S_2^2}{\sigma^2} \sim \chi_{m+n-2}^2 \tag{1.11}$$

$$T = \sqrt{\frac{mn}{m+n}} \cdot \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{S_w} \sim t_{m+n-2} \tag{1.12}$$

(2)

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{m-1,n-1} \tag{1.13}$$

1.2 点估计

1.2.1 矩估计

$$\hat{\theta}_i = \hat{\theta}_i(X_1, \cdots, X_n) = g_i(a_1, \cdots, a_k), \quad i = 1, 2, \cdots, k \tag{1.14}$$

- 参数个数 = 矩的个数

1.2.2 最大似然估计

$$\hat{\theta} = \arg\max L(\theta) \tag{1.15}$$

- 最终目的是为了求似然函数 $L(\theta)$ 的最大值, 因此 $\left.\dfrac{\partial L}{\partial \theta}\right|_{\hat{\theta}} = 0$ 或 $\hat{\theta}$ 为区间端点

1.2.3 点估计的优良准则

1 无偏性 $\leftrightarrow$ 期望

$$\mathrm{E}\,\hat{g}(\theta) = \hat{g}(X_1, \cdots, X_n) = g(\theta) \tag{1.16}$$

2 有效性 $\leftrightarrow$ 方差

$$\mathrm{Var}\,\hat{\theta}_1 \leqslant \mathrm{Var}\,\hat{\theta}_2 \tag{1.17}$$

则 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 更有效.

- Cramer-Rao 不等式

$$\mathrm{Var}\,\hat{\theta} \geqslant \frac{1}{nI(\theta)}, \quad I(\theta) = -\mathrm{E}\left[\frac{\partial^2 \ln f(x;\theta)}{\partial^2 \theta}\right] = \mathrm{E}\left[\frac{\partial \ln f(x;\theta)}{\partial \theta}\right]^2 \tag{1.18}$$

3 相合性 $\leftrightarrow$ 大数律

4 渐近正态性 $\leftrightarrow$ 中心极限定理

1.3 区间估计

1.3.1 枢轴变量法

利用三大分布及统计量的性质

1.3.2 大样本法

利用中心极限定理

2 例题

2.1 作业题

2.1.1 (4.45)

答案 由 Chebyshev 不等式求得概率 $p_1 = 0.8$, 由中心极限定理求得概率 $p_2 = 0.97$.

2.1.2 (4.49)

答案 1167.

2.1.3 (4.52)

引理 设 $X_1, \cdots, X_n$ 独立同分布, $N \sim \mathrm{Poi}(\lambda)$, 则 $X = \sum_{i=1}^{N} X_i$ 服从参数为 λ 的复合 Poisson 分布,

$$\mathrm{E}\, S = \mathrm{E}\, X_i \,\mathrm{E}\, N = \lambda \,\mathrm{E}\, X, \quad \mathrm{Var}\, S = \lambda \,\mathrm{E}\, X_i^2.$$

解 记第 i 辆车的索赔额度为 Z_i $(i = 1, 2, \cdots, 2400)$, 发生 X_i 次事故, 每次事故的索赔额度为 Y_{ij}, 则 $Z_i = \sum_{j=1}^{X_i} Y_{ij}$, 服从复合 Poisson 分布.

$$\mathrm{E}\, Z_i = \mathrm{E}\, X_i \,\mathrm{E}\, Y_{ij} = 2 \times 3000 = 6000,$$

$$\mathrm{Var}\, Z_i = \lambda \,\mathrm{E}\, Y_{ij}^2 = 2 \int_{1000}^{5000} \frac{1}{4000} t^2 \,\mathrm{d}t = 20.67 \times 10^6,$$

由中心极限定理, 盈利 2×10^6 的概率为

$$\begin{aligned} P\left(\sum_{i=1}^{2400} Z_i \leqslant 2400 \times 5000 - 2 \times 10^6\right) &= P\left(\frac{\sum_{i=1}^{2400} Z_i - 2400\,\mathrm{E}\, Z_i}{\sqrt{2400\,\mathrm{Var}\, Z_i}} \leqslant \frac{10^7 - 2400\,\mathrm{E}\, Z_i}{\sqrt{2400\,\mathrm{Var}\, Z_i}}\right) \\ &= \Phi(-19.757). \end{aligned}$$

□

2.1.4 (5.8)

答案 样本空间: $\{(x_1, \cdots, x_5) : x_i = 0, 1,\ 1 \leqslant i \leqslant 5\}$
抽样分布 $P(X_1 = x_1, \cdots, X_5 = x_5) = p^{\sum x_i}(1-p)^{5-\sum x_i}$
$X_1 + X_2, \min X_i$ 是统计量, 其余与参数 p 有关, 不是统计量.

2.1.5 (5.14)

解 $(X_1, X_2) \to \frac{1}{\sqrt{2}}(X_1 - X_2, X_1 + X_2)$ 是一个正交变换, 计算其联合密度函数可知两者独立且服从标准正态分布, 从而 $Y = \dfrac{\left(\frac{X_1 - X_2}{\sqrt{2}}\right)^2}{\left(\frac{X_1 + X_2}{\sqrt{2}}\right)^2}$ 的分子分母独立同分布 $\sim \chi_1^2$, 因此 $Y \sim F_{1,1}$. □

2.1.6　(5.17)

解　与上题同理, $Y \sim F_{10,5}$. □

2.1.7　(5.19)

解　$S_n^2 = \frac{n-1}{n} S^2 \implies \operatorname{E} S_n^2 = \frac{n-1}{n} \operatorname{E} S^2 = \frac{n-1}{n} p(1-p)$. □

2.1.8　(5.22)

解　运用中心极限定理.

(1) 299

(2) $\frac{n-1}{n+1}$

□

2.1.9　(6.3)

解　(2) 由分布律得 $X \sim B(m, \theta)$, 因此 $\overline{X} \sim \operatorname{E} X = m\theta$, $S^2 \sim \operatorname{Var} X = m\theta(1-\theta)$, 可得 $\hat{\theta} = 1 - \frac{S^2}{\overline{X}}$.

(3) $\operatorname{E} X = \frac{2}{\theta} \implies \hat{\theta} = \frac{2}{\overline{X}}$. □

注意　第 2 问包含两个参数 m, θ, 因此需要考虑二阶矩才能得出参数估计. 一般情况下, 除非题目特别说明, 只有样本量 n 是已知的参数, 其余参数均未知.

2.1.10　(6.4)

解　(1) $\operatorname{E} X = \frac{\theta}{3} \implies \hat{\theta} = 3\overline{X}$;

(5) $\operatorname{E} X = \frac{\theta}{2} \implies \hat{\theta} = 2\overline{X}$. □

2.1.11　(6.25)

解　(2) $\hat{\theta} = \overline{X}$; (3) $\hat{\theta} = \frac{2}{\overline{X}}$. □

2.1.12　(6.26)

解　(1) $\hat{\theta}$ 是方程 $\sum_i \frac{1}{\theta - x_i} = \frac{2n}{\theta}$ 的根; (5) $\hat{\theta}$ 是方程 $\sum_i \frac{1}{\theta - x_i} = \frac{3n}{\theta}$ 的根. □

2.1.13　(6.29)

解　(1) $\hat{\theta}_M = 2\overline{X}$, $\hat{\theta}_L = \min X_i$;

(2) $\hat{\theta}_M = \frac{2}{3}\overline{X}$, $\hat{\theta}_L = \frac{1}{2} \max X_i$.

□

2.1.14　(6.33)

解　由参数函数的估计得: $\hat{g}(p) = \overline{X}(1-\overline{X})^2$. □

2.1.15　(6.39)

解　(1) $\frac{\sqrt{\pi\theta}}{2}$

(2) $\frac{1}{n}\sum_i X_i^2$

(3) 取 $a=\theta$

□

2.1.16　(6.41)

解　(1) $\hat{\theta}_l=\frac{1}{c}\max X_i$

(2) $\hat{\theta}_m=\frac{2\overline{X}}{c+1}$, 是无偏估计.

□

2.1.17　(6.44)

解　(1) $\hat{\theta}_1=\overline{X}-\sigma$, $\hat{\theta}_2=\min X_i$

(2) $\tilde{\theta}_1=\hat{\theta}_1$, $\tilde{\theta}_2=\hat{\theta}_2-\frac{\sigma}{n}$ 是修正后的无偏估计

(3) $\operatorname{Var}\tilde{\theta}_1=\frac{\sigma^2}{n}$, $\operatorname{Var}\tilde{\theta}_2=\frac{\sigma^2}{n^2}$, 因此 $\tilde{\theta}_2$ 更优.

□

注意　需要算出两者方差的具体值.

2.1.18　(6.47)

解　$\lambda_M=\left(\frac{\Gamma\left(1+\frac{1}{\alpha}\right)}{\overline{X}}\right)^\alpha$, $\lambda_L=\frac{n}{\sum X_i^\alpha}$　□

2.1.19　(6.56)

解　(1) $\hat{\theta}_M=\frac{3-\overline{X}}{5}=\frac{2}{5}$, $\hat{\theta}_L=\frac{n-n_3}{3n}=\frac{4}{15}$;

(2) 计算期望可得其均为无偏估计;

(3) $\operatorname{Var}\hat{\theta}_M=\frac{\theta(2-5\theta)}{50}>\operatorname{Var}\hat{\theta}_L=\frac{\theta(1-3\theta)}{30}$, 因此 $\hat{\theta}_L$ 更有效.

□

2.1.20　(6.60)

解　由 Cramer-Rao 不等式可计算得方差的下界 $\frac{1}{nI(\lambda)}=\frac{\lambda^2}{n}$, 取 $\hat{\lambda}=\frac{n-1}{n\overline{X}}$ 是修正后的矩估计, 为无偏估计, 其方差 $\operatorname{Var}\hat{\lambda}=\frac{\lambda^2}{n-2}>\frac{\lambda^2}{n}$, 因此不是 MVUE.　□

说明　本题分为两步, (1) 计算方差的下界; (2) 选取一个无偏估计量, 计算其方差并与下界比较.

2.1.21　(6.61)

说明　先统一约定题目的含义: 先所有人混检, 混检异常后该管中所有人单检.

解　直接考虑最一般情况. 假设总共 N 人, 分为 m 组, 每组 n 人, $mn=N$, 人群中阳性的概率为 p. 则混管正常的概率为 $(1-p)^n$, 异常概率为 $1-(1-p)^n$, 从而平均检测次数为混检次数 + 单检次数:

$$\operatorname{E}X=m+m(1-(1-p)^n)\times n=mn(1-(1-p)^n)+m=N(1-(1-p)^n)+\frac{N}{n}:=f(n),$$

求 $f(n)$ 的最值即可. 对于本题而言, $n=6$ 时取得最小值. □

注意 以下涉及区间估计的题目都应当写出枢轴变量及其服从的分布, 区间端点的结果要计算出具体的数值, 一般无特别说明的情况下, 保留三位小数. 此外, 在中间步骤运算过程中, 应当保留完整结果, 避免截断误差. 枢轴变量的选取请参照**知识梳理**部分.

2.1.22 (7.4)

答案 (1) $[2.127, 2.137]$; (2) $[2.120, 2.145]$.

2.1.23 (7.5)

答案 (1) $[119.80, 124.54]$; (2) $[118.69, 127.53]$; (3) $[118.43, 124.01]$;

2.1.24 (7.11)

答案 (1) $[71.03, 110.97]$; (2) $[72.22, 114.96]$; (3) $[-10.06, 4.88]$ 或 $[-4.88, 10.06]$.

2.1.25 (7.15)

答案 (1) $[3.187, 3.213]$; (2) $[0.024, 0.043]$.

2.1.26 (7.18)

答案 $\sigma : [8.86, 18.57]$, $\sigma^2 : [78.58, 344.93]$

2.2 小测题

2.2.1 设 $X_1, X_2, \cdots, X_n$ 是来自于正态总体 X 的一组简单随机样本, 记 $\overline{X}$ 和 S^2 分别为样本均值和样本方差, 若 $\operatorname{Var} X = \sigma^2$, 则 ________.

(A) S 是 σ 的无偏估计量

(B) S 是 σ 的极大似然估计量

(C) S 与 $\overline{X}$ 相互独立

(D) 以上都不对

答案 C. 由正态总体的性质可知.

2.2.2 对于一正态分布总体 $\mathcal{N}(\mu, 100)$ 的均值 μ 求置信水平为 $95\,\%$ 的置信区间, 若要求其区间长度不大于 4, 则样本容量 n 至少应取多少?

解 构造枢轴变量

$$N = \frac{\sqrt{n}(\overline{X}-\mu)}{\sigma} \sim \mathcal{N}(0,1),$$

因此置信区间为 $\left[\overline{X} - \frac{\sigma}{\sqrt{n}} u_{0.025}, \overline{X} + \frac{\sigma}{\sqrt{n}} u_{0.025}\right]$, 长度为

$$\frac{2\sigma}{\sqrt{n}} u_{0.025} \leqslant 4 \implies n \geqslant \left(\frac{\sigma}{2} u_{0.025}\right)^2 = 96.04 \implies n \geqslant 97.$$

□

2.2.3 设总体 X 的密度函数为 $f(x)=\dfrac{2x}{a^2}$ $(0\leqslant x\leqslant a)$, 其中 a 为未知参数, 而 $X_1,X_2,\cdots,X_n$ 是来自于该总体的一组简单随机样本.

(1) 求 a 的矩估计量 $\hat{a}_1$ 和极大似然估计量 $\hat{a}_2$;

(2) 求 $p=P(0<X<\sqrt{a})$ 的极大似然估计量 $\hat{p}$;

(3) 问 $\hat{a}_1,\hat{a}_2$ 是否为无偏估计? 若是, 请证明你的结论; 若不是, 请修正之;

(4) 问 (3) 中所得的无偏估计, 哪个更有效?

解 (1) 计算 $\mathrm{E}\,X=\dfrac{2}{3}a \implies \hat{a}_1=\dfrac{3}{2}\overline{X}$;

似然函数 $L(x;a)=\dfrac{2^n}{a^{2n}}\displaystyle\prod_{i=1}^n x_i$ 随 a 单调递减, 因此 $\hat{a}_2=X_{(n)}=\max\limits_i X_i$.

(2) $p=\dfrac{1}{a} \implies \hat{p}=\dfrac{1}{X_{(n)}}$.

(3) $\mathrm{E}\,\hat{a}_1=\dfrac{3}{2}\mathrm{E}\,X=a$, 因此 $\hat{a}_1$ 是无偏的;

$X_{(n)}$ 的密度函数

$$h(x)=n[F(x)]^{n-1}f(x)=\frac{2n}{a^{2n}}x^{2n-1},\quad 0<x<a,$$

可得 $\mathrm{E}\,\hat{a}_2=\dfrac{2n}{2n+1}a$, $\hat{a}_2$ 不是无偏估计, 可以修正为 $\hat{a}_2^*=\dfrac{2n+1}{2n}X_{(n)}$.

(4) 计算其方差

$$\operatorname{Var}\hat{a}_1=\frac{1}{8n}a^2,\quad \operatorname{Var}\hat{a}_2^*=\frac{1}{4n(n+1)}a^2,$$

$\operatorname{Var}\hat{a}_1\geqslant\operatorname{Var}\hat{a}_2^*$, 因此 $\hat{a}_2^*$ 更有效. □

2.2.4 设总体 X 的概率密度函数为 $f(x)=(\theta+1)x^\theta$ $(0\leqslant x\leqslant 1)$, 其中 $\theta>-1$ 为未知参数, 而 $X_1,X_2,\cdots,X_n$ 是来自于该总体的一组简单随机样本.

(1) 求 θ 的矩估计量 $\hat{\theta}$;

(2) 求 $g(\theta)=\dfrac{1}{\theta}$ 的极大似然估计量 $\hat{g}$;

(3) 问 g 是否为无偏估计? 若是, 请证明你的结论;

(4) 求常数 b, 使得对于任意使得实数 x, 都有 $\lim\limits_{n\to\infty}P\left(\dfrac{\sqrt{n}(\hat{g}-g(\theta))}{b}\leqslant x\right)=\Phi(x)$ 成立, 其中 $\Phi(x)$ 为标准正态分布函数.

解 (1) 由 $\mathrm{E}\,X=\dfrac{\theta+1}{\theta+2} \implies \hat{\theta}=\dfrac{2\overline{X}-1}{1-\overline{X}}$;

(2) 似然函数 $L(\theta)=(\theta+1)^n\displaystyle\prod_{i=1}^n x_i^\theta$, 从而对数似然函数

$$l(\theta)=\ln L(\theta)=n\ln(\theta+1)+\theta\sum_{i=1}^n\ln x_i,$$

令

$$\frac{\mathrm{d}l(\theta)}{\mathrm{d}\theta}=0 \implies \frac{n}{\theta+1}+\sum_{i=1}^n\ln x_i=0 \implies \hat{g}(\theta)=-\frac{1}{n}\sum_{i=1}^n\ln x_i.$$

(3) 记 $Y_i=-\ln X_i$, 则 $Y_i\,\mathrm{i.\,i.\,d.}\sim\mathrm{Exp}(\theta+1)$, 从而 $\mathrm{E}\,\hat{g}=\mathrm{E}\,Y_i=\dfrac{1}{\theta+1}$.

(4) 由中心极限定理知, $b=\sqrt{\operatorname{Var} Y_i}=\dfrac{1}{\theta+1}$.

□

2.2.5 为比较两种型号步枪子弹的枪口速度, 随机抽取 A 型子弹 10 发, 得到枪口速度的平均值 $\overline{x}=500\,\mathrm{m/s}$, 样本标准差 $s_1=1.10\,\mathrm{m/s}$, 随机抽取 B 型子弹 20 发, 得到枪口速度的平均值 $\overline{x}=496\,\mathrm{m/s}$, 样本标准差 $s_1=1.20\,\mathrm{m/s}$. 假设 A 和 B 型子弹的枪口速度分别近似服从 $\mathcal{N}(\mu_1,\sigma_1^2),\mathcal{N}(\mu_2,\sigma_2^2)$, 试求参数 $\mu_1-\mu_2$ 和 $\dfrac{\sigma_1^2}{\sigma_2^2}$ 的 $1-\alpha$ 置信区间.

解 (1) 设两者方差相等, 则

$$T=\sqrt{\frac{mn}{m+n}}\cdot\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w}\sim t_{m+n-2},$$

其中

$$S_w^2:=\frac{1}{m+n-2}[(m-1)S_1^2+(n-1)S_2^2],\quad m=10,\ n=20,$$

因此 $\mu_1-\mu_2$ 的置信区间为

$$\left[\overline{X}-\overline{Y}-\sqrt{\frac{m+n}{mn}}S_w t_{m+n-2}\left(\frac{\alpha}{2}\right),\overline{X}-\overline{Y}+\sqrt{\frac{m+n}{mn}}S_w t_{m+n-2}\left(\frac{\alpha}{2}\right)\right]$$

(2) 统计量

$$F=\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\sim F_{m-1,n-1},$$

因此置信区间为

$$\left[\frac{S_1^2}{S_2^2}\cdot\frac{1}{F_{m-1,n-1}\left(\frac{\alpha}{2}\right)},\frac{S_1^2}{S_2^2}F_{n-1,m-1}\left(\frac{\alpha}{2}\right)\right]$$

□

2.3 补充题

2.3.1 设随机变量 X 和 Y 都服从标准正态分布，则下述正确的是 ________.

(A) $X+Y$ 服从正态分布

(B) X^2+Y^2 服从 χ^2 分布

(C) X^2 和 Y^2 都服从 χ^2 分布

(D) $\dfrac{X^2}{Y^2}$ 服从 F 分布

答案 C. 其余选项成立均要求 X,Y 独立.

2.3.2 设 $X_1,X_2,\cdots,X_n$ 是来自正态总体样本 $\mathcal{N}(\mu,\sigma^2)$ 的简单随机样本, 其中 μ 为已知常数, 记 $\overline{X}$ 和 S^2 分别为样本均值和样本方差, 则下列统计量与 $\overline{X}$ 不独立的是 ________.

(A) 样本标准差 S　(B) $\dfrac{1}{n}\displaystyle\sum_{i=1}^n(X_i-\overline{X})^2$　(C) $\dfrac{1}{n}\displaystyle\sum_{i=1}^n(X_i-\mu)^2$　(D) X_1-X_2

答案 C. 该统计量展开后含有 $\overline{X}$ 项.

2.3.3　设正态总体 $X \sim \mathcal{N}(\mu, \sigma^2)$, 其中参数 μ, σ^2 均未知. 若样本容量 n 和置信系数 $1-\alpha$ 均保持不变, 对于不同的样本观测值, 总体均值 μ 的置信区间长度 ________.

(A) 始终保持不变　(B) 与 μ 的真值有关　(C) 与样本均值有关　(D) 不固定

答案　D. 易知区间长度为 $\dfrac{2S}{\sqrt{n}} t_{n-1}\left(\dfrac{\alpha}{2}\right)$, 与实测的样本方差有关, 因此不固定.

2.3.4　设 $X_1, X_2, \cdots, X_n$ $(n>2)$ 是来自正态总体 $\mathcal{N}(\mu, \sigma^2)$ 的简单随机样本, 且 $\overline{X}$ 为样本均值. 若统计量 $T = c(X_1 + X_n - 2\overline{X})^2$ 为 σ^2 的无偏估计, 则常数 $c =$ ________.

解　将 $X_1 + X_n - 2\overline{X}$ 看作 $X_1, X_2, \cdots, X_n$ 的线性组合:

$$X_1 + X_n - 2\overline{X} = \frac{n-2}{n} X_1 + \frac{n-2}{n} X_n - \frac{2}{n}(X_2 + \cdots + X_{n-1}),$$

从而

$$\begin{aligned}
\mathrm{E}(X_1 + X_n - 2\overline{X})^2 &= \mathrm{Var}(X_1 + X_n - 2\overline{X}) + \left[\mathrm{E}(X_1 + X_n - 2\overline{X})\right]^2 \\
&= \left(\left(\frac{n-2}{n}\right)^2 \times 2 + \frac{4}{n^2} \times (n-2)\right)\sigma^2 \\
&= \frac{2(n-2)}{n}\sigma^2,
\end{aligned}$$

因此 $c = \dfrac{n}{2(n-2)}$. □

2.3.5　设 $X_1, X_2, \cdots, X_n$ 是来自二项总体 $B(n,p)$ 的一组简单随机样本, 且记 $\overline{X}$ 和 S^2 分别为样本均值和样本方差. 若 $\overline{X} + kS^2$ 为 np^2 的一个无偏估计, 则常数 $k =$ ________.

解　由于 S^2 是 σ^2 的无偏估计, $\mathrm{E}\left[\overline{X} + kS^2\right] = np + k \cdot np(1-p) = np^2 \implies k = -1$. □

2.3.6　某药厂试制了一种新药, 声称对贫血患者的治疗有效率达到 80%. 医药监管部门随机抽取 200 个贫血患者进行此药的临床试验, 若至少有 152 人用药有效, 就批准此药的生产. 试用中心极限定理, 求解如下问题:

(1) 若该药的有效率确实达到 80%, 此药被批准生产的概率大约是多少?

(2) 若监管部门的方案是 200 个人中要有 160 人用药有效才批准, 这对药厂是否公平? 请说明理由.

提示　直接利用中心极限定理即可. 可求得第 2 问的概率为 0.5, 这相当于抛硬币, 因此对药厂不公平.