



中国人民大学夏令营试题汇编

数学学院励志书院数学试验班师生联合党支部 整理人：李怡龙 (25)



中国人民大学统计学院
2022 年夏令营招生考试试题（统计学）

1. 计算定积分 $\int_0^1 x \ln(1+x) dx$.
2. 设数列 a_n 满足条件 $a_0 = 1, a_1 = 1, a_n = a_{n-1} + a_{n-2} (n = 2, 3, \dots)$. 判断级数 $\sum_{n=1}^{\infty} \frac{1}{a_n}$ 的收敛性.
3. 设矩阵 $\mathbf{A} = (\alpha_1, \alpha_2, \alpha_3)$ 有三个不同的特征值, 其中 $\alpha_1, \alpha_2, \alpha_3$ 为 $\mathbf{A}$ 的列向量且满足 $\alpha_3 = \alpha_1 + 2\alpha_2$.
 - (1) 证明: $\text{rank}(\mathbf{A}) = 2$.
 - (1) 已知向量 $\beta = \alpha_1 + \alpha_2 - 3\alpha_3$, 求解线性方程组 $\mathbf{A}x = \beta$.
4. 现有两枚硬币, 硬币 A 正面和反面出现的概率均为 $\frac{1}{2}$, 硬币 B 正面和反面出现的概率分别为 $\frac{3}{4}$ 和 $\frac{1}{4}$. 随机选择一枚硬币进行抛掷, 若前三次结果均为正面, 求第四次抛掷出现正面的概率.
5. 有 m 种不同的卡片, 每次抽到第 i 种卡片的概率为 p_i , 且满足 $\sum_{i=1}^m p_i = 1$. 现有放回地抽取 n 次, 定义随机变量 X 为抽到的不同卡片种类数, 求 $\mathbb{E}(X)$.
6. 设总体 X 服从均匀分布 $U\left(\theta - \frac{1}{2}, \theta + \frac{1}{2}\right)$, 令 $X_1, \dots, X_n$ 为从总体 X 中抽取的简单样本.
 - (1) 证明: 均值 $\hat{\theta}_1 = \bar{X}$ 和 $\hat{\theta}_2 = \frac{1}{2}(X_{(1)} + X_{(n)})$ 均为 θ 的无偏估计.
 - (1) 求 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 的方差, 并说明哪一个更有效.



中国人民大学统计学院
2022 年夏令营招生考试试题（应用统计）

1. 计算积分 $\int_a^{+\infty} 4x^2 e^{-2x} dx$.
2. 计算二重积分 $\iint_D y dx dy$, 其中积分区域 D 由直线 $x = -2$, $y = 2$, $y = 0$ 和半圆 $x^2 + (y-1)^2 = 1$ ($x < 0$) 共同围成.
3. 设矩阵 $A = \begin{pmatrix} 2 & 2 & -2 \\ 2 & 5 & -4 \\ -2 & -4 & a \end{pmatrix}$ 有一个特征值为 10.
 - (1) 求参数 a 的值.
 - (1) 求 A 的全部特征值.
 - (1) 求正交矩阵 P 使得 $P^{-1}AP$ 为对角矩阵.
4. 现有两枚硬币, 硬币 A 正面和反面出现的概率均为 $\frac{1}{2}$, 硬币 B 正面和反面出现的概率分别为 $\frac{3}{4}$ 和 $\frac{1}{4}$. 随机选择一枚硬币进行抛掷, 若前三次结果均为正面, 求第四次抛掷出现正面的概率.
5. 有 m 种不同的卡片, 每次抽到第 i 种卡片的概率为 p_i , 且满足 $\sum_{i=1}^m p_i = 1$. 现有放回地抽取 n 次, 定义随机变量 X 为抽到的不同卡片种类数, 求 $E(X)$.
6. 设总体 X 服从均匀分布 $U\left(\theta - \frac{1}{2}, \theta + \frac{1}{2}\right)$, 令 $X_1, \dots, X_n$ 为从总体 X 中抽取的简单样本.
 - (1) 证明: 均值 $\hat{\theta}_1 = \bar{X}$ 和 $\hat{\theta}_2 = \frac{1}{2}(X_{(1)} + X_{(n)})$ 均为 θ 的无偏估计.
 - (1) 求 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 的方差, 并说明哪一个更有效.



中国人民大学统计学院
2023 年夏令营招生考试试题（应用统计）

注：考试时间 120 分钟。

1. 求极限 $\lim_{x \rightarrow 0} \left(\frac{e^x + e^{2x} + \cdots + e^{nx}}{n} \right)^{\frac{e}{x}}$ ，其中 n 是给定的正整数。

2. 求幂级数 $\sum_{n=0}^{\infty} \frac{x^{2n+2}}{(n+1)(2n+1)}$ 的和函数。

3. 已知矩阵

$$A = \begin{pmatrix} 1 & -a & -a \\ -a & 1 & -a \\ -a & -a & 1 \end{pmatrix}.$$

(1) 求 A 的特征值和特征向量。

(1) 当 $a = 0.6$ 时，判断是否存在随机向量 (X, Y, Z) 使其协方差矩阵恰好为 A 。

4. 设矩阵 A 结构如下

$$A = \begin{pmatrix} 1 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 1 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 & 1 \\ x_{100} & x_{99} & x_{98} & \cdots & x_2 & x_1 \end{pmatrix},$$

项目的收益 $Y = |A|$ ，随机变量 X_i ($i = 1, 2, \dots, 100$) 独立同分布于标准正态分布 $N(0, 1)$ 。现需确定储备资金 M 的最小值，使得项目不破产的概率满足 $\mathbb{P}(Y + M \geq 0) \geq 0.95$ ，求满足上述条件的 M 的最小值（标准正态分布的 0.95 分位数取 1.645）。

5. 在一个有 n 个人参加的晚会上，每个人带了一件礼物，且假定各人带的礼物都不相同。晚会期间各人从放在一起的 n 件礼物中随机抽取一件，试求抽中自己礼品的人数 X 的均值和方差。

6. 设某场景中有两种型号的电池，其使用时长分别服从参数为 λ_1 和 λ_2 的指数分布。随机选取电池时，选中型号 1 的概率为 p_1 ，选中型号 2 的概率为 p_2 （满足 $p_1 + p_2 = 1$ ）。已知在时刻 t 电池仍能正常工作，求该电池后续能继续使用至少 s 小时的条件概率。

7. 设随机变量 X 服从对数正态分布，即其对数变换 $Z = \ln X$ 服从正态分布 $N(\mu, \sigma^2)$ 。

(1) 计算均值 $\mathbb{E}(X)$ 。

(1) 已知 $x_1, x_2, \dots, x_n$ 是从该对数正态分布中抽取的独立同分布样本，求 μ 和 σ^2 的最大似然估计。

(1) 当 $n = 20$ 时，给定具体样本值

52	24	15	67	15	22	63	26	16	32
7	33	28	14	7	29	10	6	59	30

求均值 $\mathbb{E}(X)$ 的最大似然估计。



中国人民大学统计与大数据研究院
2023 年夏令营招生考试试题（应用统计）

1. (1) 证明: $\left(\int_a^b f(x)dx\right)^2 = 2 \int_a^b f(x) \left(\int_a^x f(t)dt\right) dx$.
 (1) 判断 $f(x) = \frac{x^3}{1+x^2}$ 在 $\mathbb{R}$ 上的连续性和一致连续性.
 (1) 判断下列命题的正确性, 并给出证明或反例.
 - i. 若方阵 A 满足 $A^2 = O$, 则 $A = O$.
 - i. 若矩阵 A, B, C 满足 $AB = AC$, $A \neq O$, 则 $B = C$.
2. 设 $\Phi(x)$ 和 $\varphi(x)$ 分别为标准正态分布 $N(0, 1)$ 的分布函数和密度函数. 证明下列命题:
 - (1) $\varphi'(x) + x\varphi(x) = 0$.
 - (1) 当 $x > 0$ 时, $\varphi(x) \left(\frac{1}{x} - \frac{1}{x^3}\right) \leq 1 - \Phi(x) \leq \varphi(x) \left(\frac{1}{x} - \frac{1}{x^3} + \frac{3}{x^5}\right)$.
3. f 为正定二次函数 $f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^\top \mathbf{A} \mathbf{x} + \mathbf{b}^\top \mathbf{x}$, $\mathbf{d}^k$ 为下降方向, $\mathbf{x}^k$ 为当前迭代点.
 - (1) 试求出精确线搜索步长 $\alpha_k = \arg \min_{\alpha > 0} f(\mathbf{x}^k + \alpha \mathbf{d}^k)$.
 - (1) 证明: 最速下降法的步长 $\alpha_k = \frac{\|\nabla f(\mathbf{x}^k)\|^2}{\nabla f(\mathbf{x}^k)^\top \mathbf{A} \nabla f(\mathbf{x}^k)}$.
4. 设样本 $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} B(1, p_1)$, $Y_1, Y_2, \dots, Y_m \stackrel{\text{i.i.d.}}{\sim} B(1, p_2)$, 其中 $B(1, p)$ 表示 0-1 二项分布. 定义统计量

$$n_1 = \sum_{i=1}^n \mathbb{I}\{X_i = 1\}, \quad n_0 = \sum_{i=1}^n \mathbb{I}\{X_i = 0\}, \quad m_1 = \sum_{j=1}^m \mathbb{I}\{Y_j = 1\}, \quad m_0 = \sum_{j=1}^m \mathbb{I}\{Y_j = 0\}.$$

- (1) 设随机变量序列 $\{X_n\}$ 满足

$$\sqrt{n}(X_n - \mu) \xrightarrow{\mathcal{L}} N(0, \sigma^2),$$

其中 $\sigma^2 > 0$ 和 μ 均为有限数. $g(x)$ 的一阶导函数 $g'(x)$ 在 $x = \mu$ 处连续且非零, 证明:

$$\sqrt{n}[g(X_n) - g(\mu)] \xrightarrow{\mathcal{L}} N(0, [g'(\mu)]^2 \sigma^2).$$

- (1) 针对以下两个参数, 构造其点估计并给出统计量的极限分布.

$$OR = \ln \left(\frac{p_1}{1-p_1} \right) - \ln \left(\frac{p_2}{1-p_2} \right), \quad RR = \ln \left(\frac{p_1}{p_2} \right).$$

- (1) 基于 (2) 中结果求 OR 和 RR 的置信水平为 95% 的置信区间.



中国人民大学统计与大数据研究院
2024 年夏令营招生考试试题（统计学）

注：考试时间 120 分钟。

1. (1) 求极限

$$\lim_{x \rightarrow +\infty} \left[\sqrt[3]{x^3 + x^2 + x + 2025} - \sqrt{x^2 + x + 2024} \cdot \frac{\ln(e^x + x)}{x} \right].$$

(1) 计算二重积分 $\iint_{|x|+|y|\leq 2} f(x, y) dx dy$, 其中函数

$$f(x, y) = \begin{cases} x^2, & |x| + |y| \leq 1 \\ \frac{1}{\sqrt{x^2 + y^2}}, & 1 < |x| + |y| \leq 2 \end{cases}.$$

(1) 设 $f(x)$ 在 $[0, +\infty)$ 内连续, 证明:

$$\int_0^x f(t)(x-t)dt = \int_0^x \left(\int_0^t f(u)du \right) dt.$$

(1) 证明: 行满秩矩阵具有右消去律, 即若矩阵 A 行满秩且 $AC = C$, 则 $A = I$ 。

2. 考虑线性混合模型

$$y_{ij} = \mu + \alpha_j + \beta x_{ij} + \varepsilon_{ij}, \quad \alpha_j \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma_\alpha^2), \quad \varepsilon_{ij} \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma_\varepsilon^2)$$

其中所有 α_j 与 ε_{ij} 相互独立, $i = 1, 2, \dots, n, j = 1, 2, \dots, m$ 。

(1) 给出 $\mathbf{y}_j = (y_{1j}, y_{2j}, \dots, y_{nj})^\top$ 的分布。

(1) 若 σ_ε^2 已知, 求 $\boldsymbol{\theta} = (\mu, \beta)^\top$ 的最大似然估计。

(1) 若 σ_ε^2 未知, 写出给定 $\mathbf{y}_j$ 时 α_j 的条件分布, 并求 $\mathbb{E}(\alpha_j | \mathbf{y}_j)$ 和 $\text{Var}(\alpha_j | \mathbf{y}_j)$ 。

(1) 若 σ_ε^2 未知, 设计算法估计 β 和 σ_ε^2 的值。

3. 设 $X_1, X_2, \dots, X_n$ 为从总体 X 中抽取的简单样本, 定义统计量

$$Y_k = \frac{X_1 + \dots + X_k}{X_1 + \dots + X_n}, \quad k = 1, 2, \dots, n.$$

另设 $U_{(1)} \leq U_{(2)} \leq \dots \leq U_{(n)}$ 是来自均匀分布 $U(0, 1)$ 总体的次序统计量。

(1) 证明: $\mathbb{E}(Y_k) = \frac{k}{n}$ 。

(1) 设总体分布为指数分布 $\text{Exp}(1)$, 判断 $U_{(k)}$ 与 Y_k 是否具有相同的分布。

(1) 设总体分布为指数分布 $\text{Exp}(1)$, 证明: Y_k 与 $\sum_{i=1}^n X_i$ 相互独立。

(1) 证明: $n(1 - U_{(n)}) \xrightarrow{L} \text{Exp}(1)$ 。



中国人民大学统计与大数据研究院
2024 年夏令营招生考试试题（应用统计）

注：考试时间 120 分钟。

1. (1) 求极限 $\lim_{n \rightarrow \infty} \frac{n}{n^2 + 1} + \frac{n}{n^2 + 2^2} + \cdots + \frac{n}{n^2 + n^2}$.
- (1) 证明：函数 $f(x) = \frac{|\sin x|}{x}$ 在区间 $[-\pi, 0)$ 和 $(0, \pi]$ 均为一致连续，但在 $[-\pi, 0) \cup (0, \pi]$ 上不一致连续.

2. 设随机变量 X 服从 Pareto 分布，概率密度函数

$$f(x) = \frac{\beta \alpha^\beta}{x^{\beta+1}} \mathbb{I}\{x \geq \alpha\} \quad (\alpha > 0, \beta > 2).$$

- (1) 计算 $\mathbb{E}(X)$ 和 $\text{Var}(X)$.
- (1) 求随机变量 $Y = 2\beta(\ln X - \ln \alpha)$ 的分布。

3. 设独立观测样本满足以下条件：

- (a) $X_i^{(1)}$ 和 $X_i^{(0)}$ 分别服从参数为 λ_1, λ_0 的 Poisson 分布 $P(\lambda_1), P(\lambda_0)$.
- (b) T_i 服从两点分布 $B(1, p)$ ，且与 $X_i^{(1)}, X_i^{(0)}$ 独立。

观测值为 $X_i = T_i X_i^{(1)} + (1 - T_i) X_i^{(0)}$.

- (1) 求 λ_0 和 λ_1 的无偏估计量 $\hat{\lambda}_0$ 和 $\hat{\lambda}_1$ ，并证明其无偏性。
- (1) 推导 $\hat{\lambda}_1 - \hat{\lambda}_0$ 和 $\frac{\hat{\lambda}_1}{\hat{\lambda}_0}$ 的渐近分布.
- (1) 确定 p 的取值，使得 $\hat{\lambda}_1 - \hat{\lambda}_0$ 和 $\frac{\hat{\lambda}_1}{\hat{\lambda}_0}$ 的渐近方差最小.
- (1) 构造 $\hat{\lambda}_1 - \hat{\lambda}_0$ 和 $\frac{\hat{\lambda}_1}{\hat{\lambda}_0}$ 的置信水平为 95% 的近似置信区间.