

第一章 概率论

1.2 随机变量

在随机试验中, 相较试验的具体结果而言, 我们通常对结果的某些函数更感兴趣. 例如, 在同时扔两颗骰子的试验中, 相比于它们各自的点数组合, 我们往往更关心它们的点数和, 就是说, 我们关心其点数和为 7, 并不关心实际结果是 $(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)$ 中的哪一个; 对于一款电池, 我们往往更关心“它的寿命是否超过两年”而非一个确切的寿命值. 我们所关注的这些量被称为随机变量, 其具体定义如下:

定义 1.2.1 (随机变量). 定义在样本空间 Ω 上的一个实值函数 X 被称为**随机变量**, 它使得 Ω 中的每个样本点 ω 都有唯一的实数 $X(\omega)$ 与之对应.

若随机变量只能取有限个或可列个¹, 称这样的随机变量为**离散型随机变量**.

实际上, 还有大量问题中的随机变量不是离散型的, 它们的取值往往充满整个区间 I 或整个实数集 $\mathbb{R}$, 但单取一点的概率为 0, 称这样的随机变量为**连续型随机变量**.

对于离散型随机变量, 通常使用分布律来进行描述, 具体如下

定义 1.2.2 (离散型随机变量的分布律). 已知随机变量 X 的取值为 $x_1, x_2, \dots$, 满足

$$P(X = x_i) = p_i, \quad i = 1, 2, \dots$$

其中 $p_i \geq 0, \sum_i p_i = 1$. 称上式为离散型随机变量 X 的**分布律**, 通常用表格表示:

X	x_1	x_2	x_3	$\dots$
P	p_1	p_2	p_3	$\dots$

一般通过演算所有概率之和是否为 1 来初步验证分布律是否列对.

定义 1.2.3 (分布函数). 设 X 是随机变量, 称函数 $F_X(x) = P(X \leq x), x \in \mathbb{R}$ 为 X 的**分布函数**.

分布函数具有以下性质:

- (1) $0 \leq F_X(x) \leq 1$;
- (2) $F_X(x)$ 单调不减;
- (3) $\lim_{x \rightarrow -\infty} F_X(x) = 0, \lim_{x \rightarrow +\infty} F_X(x) = 1$.
- (4) $P(a < X \leq b) = F_X(b) - F_X(a)$.

对于连续型随机变量, 通常使用概率密度函数来进行描述, 具体如下

¹可列指的是 X 的所有可能取值能按一定的顺序排列, 例如: 若随机变量 X 的取值为 $(0, 1)$ 内的所有实数, 那么 X 不是离散型随机变量, 因为我们无法指出 0.5 后 X 取哪个值.

定义 1.2.4 (概率密度函数). 若存在非负函数 $p(x)$, 使得

$$P(a \leq X \leq b) = \int_a^b p(x) dx$$

其中 $p(x)$ 满足 $\int_{-\infty}^{+\infty} p(x) dx = 1$, 则称 X 是连续型随机变量, $p(x)$ 为 X 的**概率密度函数**.

分布函数与概率密度之间的关系为:

$$F_X(x) = \int_{-\infty}^x p(t) dt, \quad p(x) = F'_X(x).$$

需要强调:

$$P(X = x) = 0,$$

但这并不意味着连续型随机变量不会取到某个确定的数, 而是单点对应的区间长度为零.

1.2.1 随机变量的数字特征

定义 1.2.5 (离散型随机变量的数学期望). 设离散型随机变量 X 的分布律为 $P(X = x_i) = p_i, i = 1, 2, \dots$, 若级数 $\sum_{i=1}^{\infty} |x_i| p_i$ 收敛, 则称 $\sum_{i=1}^{\infty} x_i p_i$ 为 X 的**数学期望**或**均值**², 记为 EX . 若 $\sum_{i=1}^{\infty} |x_i| p_i$ 发散, 则称随机变量 X 不存在数学期望.

问题 1.1 (离散型随机变量期望的性质) 若 X, Y 是随机变量且均存在数学期望, a, b 是实常数, 证明:

- (1) $E(a) = a$;
- (2) $E(aX + b) = aEX + b$;
- (3) (**线性可加性**) $E(X + Y) = EX + EY$;
- (4) (**独立可乘性**)³若随机变量 X, Y 相互独立, 则 $E(XY) = EXEY$

一般地, 对任意实数 $a_1, a_2, \dots, a_n$ 和期望有限的随机变量 $X_1, X_2, \dots, X_n$, 有

$$E(a_1X_1 + a_2X_2 + \dots + a_nX_n) = a_1EX_1 + a_2EX_2 + \dots + a_nEX_n$$

定义 1.2.6 (离散型随机变量的方差). 设 X 是随机变量, 若 $EX^2 < +\infty$, 则称 $E(X - EX)^2$ 为随机变量 X 的**方差**, 记为 DX . 称 $\sqrt{DX}$ 为 X 的**标准差**.

² X 的期望是 X 所有可能取值的加权平均, 权重为 X 取此值的概率.

³ 高中阶段不要求掌握.

问题 1.2 (离散型随机变量方差的性质) 解答下述问题:

1. 若随机变量 X 存在方差 DX , 证明: $DX = EX^2 - (EX)^2$;
2. 已知 X, Y 是离散型随机变量且均存在方差, a, b 是实常数, 证明:

$$(1) D(a) = 0;$$

$$(2) D(aX + b) = a^2 DX;$$

$$(3) D(X \pm Y) = DX + DY \pm 2[E(XY) - EXEY]$$

特别地, 当 X 和 Y 独立时, 有 $D(X \pm Y) = DX + DY$

1.2.2 随机变量的概率分布

接下来研究几类典型的概率分布.

二项分布 (Binomial Distribution) 若随机变量 X 的分布律满足

$$P(X = k) = C_n^k p^k (1-p)^{n-k}, k = 0, 1, \dots, n$$

则称 X 服从参数为 n, p 的二项分布, 记作 $X \sim B(n, p)$.

一类典型的二项分布是: 进行 n 重伯努利试验, 设每次试验的成功概率为 p , 记随机变量 X 表示成功次数, 可知 X 服从参数为 n, p 二项分布.

问题 1.3 设 $X \sim B(n, p)$,

1. 使用组合方法求 EX, DX ;
2. 借助期望可加性与方差独立可加性求 EX, DX ;
3. 求 $f(k) = P(X = k)$ 的最大值点;
4. 记 X 等于奇数的概率为 $f_a(n)$, X 等于偶数的概率为 $f_b(n)$, 其中 $n = 1, 2, \dots$,

$$(1) \text{ 证明: 当 } p = \frac{1}{2} \text{ 时, } f_a(n) = f_b(n);$$

$$(2) \text{ 讨论 } f_a(n) \text{ 的单调性.}$$

*** 泊松分布 (Poisson Distribution)** ⁴19 世纪, 法国数学家泊松 (Siméon Denis Poisson) 在研究二项分布 $B(n, p)$ 时, 思考了一个问题: 如果试验次数 n 非常大, 而单次发生的概率 p 非常小, 但它们的乘积 $np = \lambda$ (平均发生次数) 保持稳定, 那么这个概率公式呈现出何种形式? 我们尝试顺着泊松的假设推导下去.

二项分布的概率公式为: $P(X = k) = C_n^k p^k (1-p)^{n-k}$, 将 $p = \frac{\lambda}{n}$ 代入, 得

⁴泊松分布不属于高中学习范畴.

$$\begin{aligned}
 P(X=k) &= \frac{n(n-1)\dots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} \\
 &= \frac{\lambda^k}{k!} \cdot \underbrace{\frac{n(n-1)\dots(n-k+1)}{n^k}}_A \cdot \underbrace{\left(1-\frac{\lambda}{n}\right)^n}_B \cdot \underbrace{\left(1-\frac{\lambda}{n}\right)^{-k}}_C
 \end{aligned}$$

保持 λ 不变, 让 $n \rightarrow +\infty$ (相应的, $p \rightarrow 0^+$).

问题 1.4 计算这三部分的极限.

最终得到

$$P(X=k) = \frac{\lambda^k e^{-\lambda}}{k!}, k=0, 1, 2, \dots$$

这就是泊松分布的分布律.

在计算机普及之前, 当二项分布的 n 较大时, 计算 C_n^k 是很麻烦的. 泊松这一发现表明, 在 n 较大, p 较小, $\lambda = np$ 适中时, 可以使用泊松分布近似二项分布, 受这一前提限制, 泊松分布多被用来描述“稀有事件在连续时间或空间中发生的次数”, 例如:

- (1) 一本书里每一页的印刷错误.
- (2) 放射性物质在 1 秒内衰变的原子数.
- (3) 普鲁士骑兵军团中每年被马踢死的人数⁵

*** 几何分布 (Geometric Distribution)** ⁶ 若随机变量 X 的分布律满足

$$P(X=k) = (1-p)^{k-1}p, k=1, 2, \dots$$

其中 $0 < p < 1$ 为常数, 则称 X 服从参数为 p 的几何分布, 记为 $X \sim g(p)$

一类典型的几何分布是: 进行若干重伯努利试验, 设每次试验的成功概率为 p , 当首次试验成功时终止操作, 记随机变量 X 为终止操作时的试验总次数, 可知 X 服从参数为 p 的几何分布.

问题 1.5 解答下述问题

1. 若随机变量 $X \sim g(p)$, 求 EX, DX ;
2. 证明几何分布具有无记忆性: 若随机变量 $X \sim g(p)$, 则

$$P(X > s+t \mid X > s) = P(X > t)(s, t \in \mathbb{N})$$

⁵这是泊松分布在历史上最著名的应用案例.

⁶几何分布不属于高中学习范畴.

超几何分布 (Hypergeometric Distribution) 若随机变量 X 的分布律满足

$$P(X = k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n} (\max\{0, n - N + M\} \leq k \leq \min\{n, M\})$$

其中 N, M 是给定的常数, 且 $N \geq M$, 则称 X 服从参数为 N, M, n 的超几何分布, 记为 $X \sim H(N, M, n)$.

一类典型的超几何分布是: N 件产品中有 M 件次品. 若一次性从中取出 n 件产品, 记随机变量 X 为这 n 件产品中次品个数, 则 X 服从参数为 N, M, n 超几何分布.

问题 1.6 假设随机变量 $X \sim H(N, M, n)$,

1. 用组合法证明:

$$EX = \frac{nM}{N}, DX = \frac{nM(N-M)(N-n)}{N^2(N-1)}$$

2. 利用期望可加性证明: $EX = \frac{nM}{N}$

正态分布 (Normal Distribution) 若随机变量 X 的概率密度函数 $p(x)$ 满足

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}$$

则称 X 服从参数为 μ, σ^2 的正态分布, 记为 $X \sim N(\mu, \sigma^2)$. 当 $\mu = 0, \sigma^2 = 1$ 时, 称 X 服从**标准正态分布**, 其概率密度函数记为 $\varphi(x)$, 分布函数记为 $\Phi(x)$.

利用微积分的相关知识可以证明: 若 $X \sim N(\mu, \sigma^2)$, 则 (1) $aX + b \sim N(a\mu + b, a^2\sigma^2)$

(2) 由第 1 条, 若 $X \sim N(\mu, \sigma^2)$, 则 $\frac{X - \mu}{\sigma} \sim N(0, 1)$

(3) $EX = \mu, DX = \sigma^2$

正态分布密度函数 $p(x)$ 的图象如下图所示, 其形状呈钟形, 具有如下特点:

(1) 关于直线 $x = \mu$ 对称;

(2) 当 $x \rightarrow \pm\infty$ 时, $p(x) \rightarrow 0$, 曲线以 x 轴为渐近线;

(3) $p(x)$ 在 $x = \mu$ 处取得最大值 $\frac{1}{\sqrt{2\pi}\sigma}$, 由于密度图形覆盖面积总为 1, 因此固定 μ 时, σ 越小, 最大值越大, 图形越高越陡峭; σ 越大, 最大值越小, 图形越低越平缓.

正态分布是概率统计中最重要的分布之一, 很多自然和社会现象中的随机变量都服从正态分布, 如某地区的成年男子的身高、一批工件的直径、测量误差等.⁷正态分布的概率计算常通过查表实现, 即将一般的正态分布化为标准正态分布, 然后查标准正态分布的分布函数表 $\Phi(x)$, 得到所求概率.

问题 1.7 已知 $X \sim N(\mu, \sigma^2)$, 用含 $\Phi(k)$ 的式子表示 $P(\mu - k\sigma \leq X \leq \mu + k\sigma)$

正态分布的随机变量 $X \sim N(\mu, \sigma^2)$, 虽然其理论取值范围是 $(-\infty, +\infty)$, 但是在区间 $[\mu - 3\sigma, \mu + 3\sigma]$ 内取值的可能性超过 99.7%, 在这之外取值的可能性不足千分之三. 这表明在一次随机试验中, X 的取值几乎可以肯定落入 $[\mu - 3\sigma, \mu + 3\sigma]$ 中, 这称为“3 σ 法则”. 该法则在针对实际问题, 特别是产品质量问题的统计推断中有着重要的应用.

⁷详见大数定律与中心极限定理这一节.

正文问题答案

问题 9.2.6.

(1) 参考

(2) 对于取 n 件产品的一次操作, 定义随机变量 $X_i = \begin{cases} 1, & \text{第 } i \text{ 件次品被取出} \\ 0, & \text{第 } i \text{ 件次品没有被取出} \end{cases}$, 于是该轮操作取出的次品数这一随机变量 X 可被表示成 $X_1, X_2, \dots, X_M$, 而 $EX_i = P(X_i = 1) = \frac{M}{N}$, 于是由期望的线性可加性得到 $EX = \sum_{i=1}^n EX_i = \frac{nM}{N}$ 。

下面计算方差。由 $D(X) = E(X^2) - (EX)^2$, 于是只需要再求出 $E(X^2)$ 。因为 $X = X_1 + X_2 + \dots + X_n$, 所以

$$E(X^2) = E[(X_1 + X_2 + \dots + X_n)^2] = \sum_{i=1}^n E(X_i^2) + 2 \sum_{1 \leq i < j \leq n} E(X_i X_j).$$

由于 X_i 的所有可能取值是 0 和 1, 所以 $X_i^2 = X_i$ 。而当 $i \neq j$ 时,

$$E(X_i X_j) = P(X_i = 1, X_j = 1) = \frac{M(M-1)}{N(N-1)},$$

由期望的线性可加性可得

$$E(X^2) = \frac{nM}{N} + n(n-1) \cdot \frac{M(M-1)}{N(N-1)},$$

于是

$$D(X) = E(X^2) - (EX)^2 = \frac{nM}{N} + n(n-1) \cdot \frac{M(M-1)}{N(N-1)} - \left(\frac{nM}{N}\right)^2.$$

第一章 概率论

1.2 随机变量

习 题

A 组

1. 已知随机变量 X, Y 满足 $X \sim N(1, 2), Y \sim N(2, 1)$, 则

(A) $P(X < 0) = P(X > 2)$

(B) $P(X > 3) = P(Y > 3)$

(C) $P(X < 2) > P(Y > 2)$

(D) $P(X < 2) < P(Y > 1)$

2. 【2015 湖北 4】设 $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$. 这两个正态分布密度曲线如图所示, 下列结论中正确的是

(A) $P(Y \geq \mu_2) \geq P(Y \geq \mu_1)$

(B) $P(X \leq \sigma_2) \leq P(X \leq \sigma_1)$

(C) 对任意正数 $t, P(X \leq t) \geq P(Y \leq t)$

(D) 对任意正数 $t, P(X \geq t) \geq P(Y \geq t)$

3. 【2007 安徽理 10】用 $\Phi(x)$ 表示标准正态总体在区间 $(-\infty, x)$ 内取值的概率, 若随机变量 ξ 服从正态分布 $N(\mu, \sigma^2)$, 则 $P(|\xi - \mu| < \sigma) =$

(A) $\Phi(\mu + \sigma) - \Phi(\mu - \sigma)$

(B) $\Phi(1) - \Phi(-1)$

(C) $\Phi\left(\frac{1-\mu}{\sigma}\right)$

(D) $2\Phi(\mu + \sigma)$

4. 【2026 考研数学一 10】设随机变量 X 的概率分布为 $P\{X = k\} = \frac{1}{2^{k+1}} + \frac{1}{3^k}$ ($k = 1, 2, \dots$), 则对于正整数 m, n , 有

(A) $P\{X > m+n | X > m\} = P\{X > m\}$

(B) $P\{X > m+n | X > m\} = P\{X > n\}$

(C) $P\{X > m+n | X > m\} > P\{X > m\}$

(D) $P\{X > m+n | X > m\} > P\{X > n\}$

5. 【2025 炎奖杯 11】(多选) 设某离散型随机变量 X 的数学期望为 1, 则

(A) $E[(X-1)^2] < EX^2$

(B) $E(|X-1|) < EX$

(C) $D[(X-1)^2] \leq DX^2$

(D) $D(|X-1|) \leq DX$

6. 设随机变量 $X \sim B(2, \frac{1}{2}), Y \sim B(3, \frac{1}{2})$, 求 $P(X+Y > 1)$.

7. 袋中有相同大小的 5 个白球和 5 个黑球, 从中育放回地随机取 5 次, 每次取 1 个球, 记随机变量 X 为取出白球的次数, 求 $P(X=5 | X \geq 3)$.

8. 【2026 考研数学一 8】设随机变量 $X \sim N(1, 2)$, 令 $f(t) = E[(X+t)^2]$, 求 $f(t)$ 的最小值点与最小值.

9. 【2020 浙江 16】盒中有 4 个球, 其中 1 个红球, 1 个绿球, 2 个黄球. 从盒中不放回地随机取球, 每次取 1 个, 直到取出红球为止. 设此过程中取到黄球的个数为 ξ , 求 ξ 的分布列与数学期望.

10. 【2015 上海理 12】某种赌博每局的规则是：赌客先在标记有 1,2,3,4,5 的卡片中随机摸取一张，将卡片上的数字作为其赌金，随后放回该卡片，再随机摸取两张，将这两张卡片上数字之差的绝对值的 1.4 倍作为其奖金。若随机变量 ξ_1 和 ξ_2 分别表示赌客在一局赌博中的赌金和奖金，求 $E(\xi_1 - \xi_2)$ 。

11. 解答下述问题：

- (1) n 个人将帽子扔到房间中央. 帽子混杂后, 每人随机从中取一个, 求取到自己帽子的人数的数学期望;
- (2) 【2024 新高考 I 卷 14】甲, 乙两人各有 4 张卡片, 每张卡片上标有一个数字, 甲的卡片上分别标有数字 1,3,5,7, 乙的卡片上分别标有数字 2,4,6,8. 两人进行四轮比赛, 在每轮比赛中, 两人各自从自己持有的卡片中随机选择一张, 并比较所选卡片上的数字大小, 数字大的人得 1 分, 数字小的人不得分, 然后各自弃置本轮所选的卡片 (弃置后的卡片在此后的轮次中不能使用), 求四轮比赛后, 甲的总得分不小于 2 的概率;
- (3) 【2025 新高考 I 卷 14】一个箱子里有 5 个球, 分别以 1,2,3,4,5 编号, 若有放回地从中随机取 3 次球, 每次取 1 个球, 记至少被取出一次的球的个数为 X , 求 X 的数学期望;
- (4) 【2026 新高考 I 卷 8】设 $U = \{(x_1, x_2, x_3) \mid x_i \in \{-2, -1, 1, 2\}, i = 1, 2, 3\}$, 记 $P(1, 1, 1)$, 样本空间 $\Omega = \mathcal{C}_U\{P\}$, 从 Ω 中随机取一个点, 定义随机变量 X : 对 Ω 中每一个点 $A(x_1, x_2, x_3)$, 令 $X(A) = x_1 + x_2 + x_3$, 求 X 的数学期望.
12. 【2012 上海 17】设 $10 \leq x_1 < x_2 < x_3 < x_4 \leq 10^4, x_5 = 10^5$, 随机变量 ξ_1 取值 x_1, x_2, x_3, x_4, x_5 的概率均为 0.2, 随机变量 ξ_2 取值 $\frac{x_1 + x_2}{2}, \frac{x_2 + x_3}{2}, \frac{x_3 + x_4}{2}, \frac{x_4 + x_5}{2}, \frac{x_5 + x_1}{2}$ 的概率也均为 0.2. 记 $D\xi_1, D\xi_2$ 分别为 ξ_1, ξ_2 的方差, 比较 $D\xi_1$ 与 $D\xi_2$ 的大小.
13. 【2014 浙江理 9】已知甲盒中有 1 个红球, 乙盒中有 m 个红球和 n 个蓝球 ($m, n \geq 3$), 从乙盒中随机取 i 个球放入甲盒. 记此时甲盒中有 ξ_i 个红球, 从甲盒中随机取 1 个球, 取出红球的概率为 p_i , 比较 p_1 与 p_2 的大小以及 $E\xi_1$ 与 $E\xi_2$ 的大小,
14. 【2025 “数海漫游二模” (网络联考) 8】将 3 个箱子编号为 1,2,3, 把 p 个相同的小球依次等可能地放入这 3 个箱子, 最终第 i 个箱子中有 x_i 个小球, 记 $X_{k,p} = |x_1 - x_2|^k + |x_2 - x_3|^k + |x_3 - x_1|^k$, 求 $EX_{2,3}$ 与 $EX_{1,4}$.
15. 【2025 温州一模 8】一款飞行棋的规则是: 掷一枚均匀的标准骰子, 正面朝上的点数是棋子向前走的步数, 如果步数恰好为到达终点所需的步数, 则可以到达终点, 如果超出到达终点所需的步数, 则需往回走超出点数对应的步数, 若当前棋子离终点还需走 3 步, 设从现在起到该棋子到达终点时所掷骰子的次数为 X , 求 X 的数学期望.
16. 【2024 成都零诊 14 改编】袋中装有三个相同的小球, 依次编号为 1,2,3. 每次从袋中随机取一球, 记下编号后放回袋中. 设第 n 次取得的球编号为 $a_n (n = 1, 2, \dots, 7)$, 已知

$$E_k: \frac{x^2}{a_k} + \frac{y^2}{a_{k+1}} = 1 (k = 1, 2, \dots, 6)$$

设 $E_1, E_2, \dots, E_6$ 中, 圆的个数为 X , 半径为 2 的圆的个数为 Y , 焦点在 x 轴上的椭圆的个数为 Z .

- (1) 分别求 X, Y, Z 的均值;

- (2) 求 $E_1, E_2, \dots, E_6$ 中有且只有 E_6 是焦点在 x 轴上的椭圆的概率;
- (3) 判断 X 是否服从二项分布, 并说明理由;
- (4) 判断 Y 是否服从二项分布, 并说明理由;
- (5) 判断 Z 是否服从二项分布, 并说明理由.
17. 【2021 八省联考 16】对一个物理量做 n 次测试, 并以测量结果的平均值作为该物理量的最后结果, 已知最后结果的误差 $\varepsilon_n \sim N(0, \frac{2}{n})$, 为使误差 ε_n 在 $(-0.5, 0.5)$ 的概率不小于 0.9545, 求至少要测量的次数.
18. 【2018 全国 I 卷理 20】某工厂的某种产品成箱包装, 每箱 200 件, 每一箱产品在交付用户之前要对产品作检验, 如检验出不合格品, 则更换为合格品. 检验时, 先从这箱产品中任取 20 件作检验, 再根据检验结果决定是否对余下的所有产品作检验. 设每件产品为不合格品的概率都为 $p(0 < p < 1)$, 且各件产品是否为不合格品相互独立.
- (1) 记 20 件产品中恰有 2 件不合格品的概率为 $f(p)$, 求 $f(p)$ 的最大值点 p_0 .
- (2) 现对一箱产品检验了 20 件, 结果恰有 2 件不合格品, 以 (1) 中确定的 p_0 作为 p 的值. 已知每件产品的检验费用为 2 元, 若有不合格品进入用户手中, 则工厂对每件不合格品支付 25 元的赔偿费用.
- i. 若不对该箱余下的产品作检验, 这一箱产品的检验费用与赔偿费用的和记为 X , 求 EX .
- ii. 以检验费用与赔偿费用的和的期望值为决策依据, 是否该对这箱余下的所有产品作检验?
19. 【2023 四省联考 20】一个池塘里的鱼的数目记为 N , 从池塘里捞出 200 尾鱼, 给鱼作上标识后把鱼放回池塘里, 过一段时间后再从池塘里捞出 500 尾鱼, 记捞出的 500 尾鱼中有标识的鱼的数目为 X .
- (1) 若 $N = 5000$, 求 X 的数学期望;
- (2) 已知捞出的 500 尾鱼中 15 尾有标识, 试给出 N 的估计值 (以使得 $P(X = 15)$ 最大的 N 的值作为 N 的估计值)
20. 【2026 长沙适应性考试 18】已知集合 U 满足 $|U| = n(n \geq 2)$, 先后两次随机、独立地选取集合 U 的两个子集, 分别记为 A, B , 设 $X = |A \cup B|$.
- (1) 若 $U = \{1, 2\}$, $X = 1$, 列举所有满足条件的 A, B ;
- (2) 求随机变量 X 的数学期望;
- (3) 设 $f(k) = P(X = k)$ 的最大值点为 m , 建立 m 与 EX 的函数关系.
21. 十条横路与十条纵路交叉形成 100 个路口, 每个路口上有一盏灯和一面旗帜. 每条横路独立地以概率 p 同时点亮该路上的 10 盏灯, 每条纵路独立地以概率 q 同时升起该路上的 10 面旗帜, 其中 $p, q \in (0, 1)$.
- (1) 对某个固定的路口, 事件“该路口开灯”与“该路口升旗”是否独立?
- (2) 记事件 A_{ij} 为第 i 条横路与第 j 条纵路交叉口既亮灯又升旗, (i, j) 与 (i', j') 满足什么条件时, 有 A_{ij} 与 $A_{i'j'}$ 独立?

(3) 记 $X_{ij} = \begin{cases} 1, & \text{事件 } A_{ij} \\ 0, & \text{事件 } \bar{A}_{ij} \end{cases}$, $\sum_{i=1}^{10} X_{ii}$ 是否服从二项分布?

(4) 又亮灯又升旗的路口总数是否服从二项分布?

22. 混管病毒检测是为了应对单管病毒检测效率低下的问题出现的一个创新型病毒检测策略, 混管检测结果为阴性, 则参与该混管检测的所有人均均为阴性, 混管检测结果为阳性, 则参与该混管检测的人中至少有一人阳性. 假设一组样本有 N 个人, 每个人感染病毒的概率相互独立且均为 $p(0 < p < 1)$. 目前, 我们采用 K 人混管病毒检测, 定义成本函数 $f(X) = \frac{N}{K} + KX$, 这里 X 指该组样本 N 个人中感染病毒的人数.

(1) 证明: $E[f(X)] \geq 2\sqrt{p} \cdot N$;

(2) 若 $0 < p < 10^{-4}$, $10 \leq K \leq 20$, 证明: 某混管检测结果为阳性, 则参与该混管检测的人中大概率恰有一人为阳性

23. 【2024 宁波一模 18 节选】有 n 根绳子, 每根绳子有 2 个绳头, 每次随机从未被打结的绳头中选定两个, 将这两个绳头打结, 直到所有绳头打结完.

(1) 当 $n = 3$ 时, 记随机变量 X 为绳子围成的圆的个数, 求 X 的分布列与数学期望;

(2) 求证: 这 n 根绳子打结完恰好围成一个圆的概率为 $\frac{2^{2n-1} \cdot n!(n-1)!}{(2n)!}$

24. 编号为 $1, 2, \dots, n(n \in \mathbb{N}^*, n \geq 3)$ 的 n 位球员围成一个圆, DM 站在圆心位置手持球, 按 $1 \rightarrow 2 \rightarrow \dots \rightarrow n$ 的顺序依次把球传给球员. 每次传球时, DM 只从还没有接到过球的球员中随机选择一人传球. 若选中的人正好是当前应传给的目标球员, 那么这次传球成功, 该球员接球后把球传给 DM, 并且之后不再参加传球; 若选中的不是当前目标球员, 那么这次传球失败, 该球员把球传回 DM, DM 记下他的编号, 待其成为目标球员时直接传球给他, 在那之前, DM 不会再向他传球; 传球无论成功与否, 都算 1 次传球, 当第 n 号球员接球后传回 DM 时, 游戏结束.

(1) 当 $n = 5$ 时,

i. 求 DM 第 3 次恰好成功把球传给 2 号球员的概率;

ii. 设 X 为完成全过程所需的传球总次数, 求 X 的分布列及数学期望;

(2) 设 X_n 为完成全过程所需的传球总次数, 证明: $E(X_n) = 2n - (1 + \frac{1}{2} + \dots + \frac{1}{n})$

1.

2. 新答案 (来源: 2024 上海, 13 (统计) .md): C

【解题思路】 x, y 负相关, y, z 正相关, 则 x, z 负相关.

3.

4.

5.

6.

7.

8.

9.

10.

11. (1)

(2)

(3) 61/25. 方法一: 直接计算分布列.

事件 $\{X=1\}$ 发生意味着同一个球被取出 3 次, 该球可以是 5 个球中的任意一个; 即从 5 个球中任意挑选一个, 3 次都取出该球. 因此, $P(X=1) = C_5^1 \left(\frac{1}{5}\right)^3 = \frac{1}{25}$.

事件 $\{X=2\}$ 发生意味着有放回地随机取 3 次球后, 至少被取出 1 次的球的个数为 2, 也就是说, 从 5 个球中先任意挑选 2 个, 然后 3 次取球都从这 2 个球中取, 这 2 个球中的一个被取出 1 次, 另一个被取出 2 次. 因此 $P(X=2) = C_5^2 C_2^1 C_3^1 \left(\frac{1}{5}\right)^3 = \frac{12}{25}$.

事件 $\{X=3\}$ 发生意味着从 5 个球中先任意挑选 3 个, 这 3 个球各被取出 1 次. 因此 $P(X=3) = C_5^3 A_3^3 \left(\frac{1}{5}\right)^3 = \frac{12}{25}$.

于是计算出期望 $EX = \frac{61}{25}$.

注: 也可先计算 $P(X=3), P(X=1)$, 再由概率归一性得到 $P(X=2)$.

方法二: 利用期望可加性, 构造随机变量

$$X_i = \begin{cases} 1, & i \text{ 号球至少被取出 1 次, } (i=1, 2, \dots, 5), \\ 0, & i \text{ 号球未被取出过} \end{cases}$$

可知 X_i 服从两点分布, 且 $P(X_i=0) = \left(\frac{4}{5}\right)^3 = \frac{64}{125}$, $P(X_i=1) = p_i = 1 - P(X_i=0) = \frac{61}{125}$. 随机变量 X 描述了取 3 次球之后, 1 号球至 5 号球中至少被取出 1 次的球的个数, 而 $X_i=1$ 代表 i 号球至少被取出 1 次, $X_i=0$ 代表 i 号球未被取出过, 因此 $X = \sum_{i=1}^5 X_i$. 进而

$$E\left(\sum_{i=1}^5 X_i\right) = \sum_{i=1}^5 p_i = 5 \times \frac{61}{125} = \frac{61}{25}$$

(4) $-1/21$.

评: 其余可利用期望可加性解决的高考题:

【2011 浙江 15】某毕业生参加人才招聘会, 分别向甲、乙、丙三个公司投递了个人简历. 假定该毕业生得到甲公司面试的概率为 $\frac{2}{3}$, 得到乙、丙两公司面试的概率均为 p , 且三个公司是否让其面试是相互独立的. 记 X 为该毕业生得到面试的公司个数. 若 $P(X=0) = \frac{1}{12}$, 求随机变量 X 的数学期望. $\boxed{5/3}$.

【2012 浙江 19 (2)】已知箱中装有 4 个白球和 5 个黑球, 且规定: 取出一个白球得 2 分, 取出一个黑球得 1 分. 现从该箱中任取 (无放回, 且每球取到的机会均等) 3 个球, 记随机变量 X 为取出 3 球所得分数之和. 求 X 的数学期望. $\boxed{13/3}$.

12. 方法一: 两个随机变量的均值都是 $\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5}$. 从直观上看, ξ_1 只有 x_5 偏离 $\bar{x}$ 很多, 而 ξ_2 有 $\frac{x_4 + x_5}{2}$ 、 $\frac{x_5 + x_1}{2}$ 偏离 $\bar{x}$ 很多, 计算方差的时候只需要直观上看“大头”, 注意

$$\begin{aligned}(x_5 - \bar{x})^2 &\approx \frac{9}{16}x_5^2, \\ \left(\frac{x_4 + x_5}{2} - \bar{x}\right)^2 + \left(\frac{x_5 + x_1}{2} - \bar{x}\right)^2 &\approx \frac{9}{50}x_5^2,\end{aligned}$$

所以 $D\xi_1 > D\xi_2$.方法二: 把方法一严格化, 证明如下: 由于 $10 \leq x_1 < x_2 < x_3 < x_4 \leq 10^4$, $x_5 = 10^5$, 所以

$$2 \times 10^4 < \bar{x} < 2.8 \times 10^4.$$

从而对 $k=1, 2, 3, 4$, $1 \times 10^4 \leq |x_k - \bar{x}| \leq 1.8 \times 10^4$; 并且

$$7.2 \times 10^4 \leq |x_5 - \bar{x}| \leq 8 \times 10^4.$$

所以

$$\left|\frac{x_k + x_5}{2} - \bar{x}\right| < 3.5 \times 10^4.$$

$$D\xi_1 = \frac{1}{5} \sum_{k=1}^5 (x_k - \bar{x})^2 \geq \frac{1}{5} (x_5 - \bar{x})^2 = 10.368 \times 10^8,$$

$$D\xi_2 = \frac{1}{5} \sum_{k=1}^5 \left(\frac{x_k + x_{k+1}}{2} - \bar{x}\right)^2 (x_6 = x_1) \leq \frac{1}{5} \left[3 \times (1.8 \times 10^4)^2 + 2 \times (3.5 \times 10^4)^2\right] = 6.844 \times 10^8.$$

所以 $D\xi_1 > D\xi_2$.可知 $E\xi_1 = E\xi_2$, 下比较 $E\xi_1^2$ 和 $E\xi_2^2$.由不等式 $\left(\frac{u+v}{2}\right)^2 \leq \frac{u^2+v^2}{2}$ 可得

$$y_1^2 + y_2^2 + y_3^2 + y_4^2 + y_5^2 \leq x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2.$$

于是 $E(\xi_2^2) \leq E(\xi_1^2)$, 从而

$$D\xi_2 \leq D\xi_1.$$

当 $x_1 = x_2 = x_3 = x_4 = x_5$ 时等号成立, 显然不可能, 故 $D\xi_1 > D\xi_2$.

13.

14.

15.

16.

17. 由题意, $P\left(|X| < 2\sqrt{\frac{2}{n}}\right) = 0.9545 \leq P(|X| < 0.5)$, 因此集合 $\left\{x \mid |x| < 2\sqrt{\frac{2}{n}}\right\} \subseteq \{x \mid |x| < 0.5\}$, 故只需 $2\sqrt{\frac{2}{n}} \leq 0.5$, 解得 $n \geq 32$. 因此至少要测量 32 次.

18. (1) $f(p) = C_{20}^2 p^2 (1-p)^{18}$, $f'(p) = 2C_{20}^2 p(1-p)^{17}(1-10p)$. 令 $f'(p) = 0$ 得 $p = 0.1$,

从而 $p \in (0, 0.1)$ 时 $f'(p) > 0$, $p \in (0.1, 1)$ 时 $f'(p) < 0$. 则 $f(p)$ 最大值点是 $p_0 = 0.1$.

(2) 由 (1) 可知, $p = 0.1$.

(i) 令 Y 表示余下的 180 件产品中的不合格品件数, 依题意可知 $Y \sim B(180, 0.1)$, $EY = 180 \times 0.1 = 18$. 则 $X = 20 \times 2 + 25Y = 40 + 25Y$, 所以 $EX = E(40 + 25Y) = 40 + 25EY = 490$.

(ii) 如果对余下的产品作检验, 则这一箱产品所需要的检验费为 400 元. 而 $EX > 400$, 故应该对余下的产品作检验.

19. (1) X 的所有可能取值是 $0, 1, \dots, 200$, 并且分布列为

$$P(X = k) = \frac{C_{200}^k C_{N-200}^{500-k}}{C_N^{500}}, k = 0, 1, 2, \dots, 200.$$

因此 X 服从超几何分布, 且 $N = 5000$, $M = 200$, $n = 500$, 从而 $E(X) = 500 \times \frac{200}{5000} = 20$.

(2) 当 $N < 685$ 时, 捞出的 500 尾鱼至少有 16 尾有标识, 因此 $P(X = 15) = 0$.

当 $N \geq 685$ 时, $P(X = 15) = \frac{C_{200}^{15} C_{N-200}^{485}}{C_N^{500}}$. 记 $a(N) = \frac{C_{200}^{15} C_{N-200}^{485}}{C_N^{500}}$. 根据

$$C_{n+1}^k = \frac{(n+1)!}{k!(n+1-k)!} = \frac{n+1}{n+1-k} \cdot \frac{n!}{k!(n-k)!} = \frac{n+1}{n+1-k} C_n^k,$$

可知

$$\begin{aligned} \frac{a(N+1)}{a(N)} &= \frac{C_{N+1-200}^{485} C_N^{500}}{C_{N+1}^{500} C_{N-200}^{485}} \\ &= \frac{(N+1-500)(N+1-200)}{(N+1)(N+1-200-485)} \\ &= \frac{(N-499)(N-199)}{(N+1)(N-684)} \\ &= \frac{N^2 - 698N + 499 \times 199}{N^2 - 683N - 684}. \end{aligned}$$

由 $N^2 - 698N + 499 \times 199 > N^2 - 683N - 684$ 当且仅当 $N < \frac{499 \times 199 + 684}{15} \approx 6665.7$, 可知当 $685 \leq N \leq 6665$ 时, $a(N+1) > a(N)$; 当 $N \geq 6666$ 时, $a(N+1) < a(N)$, 故 $N = 6666$ 时 $a(N)$ 最大, 所以 N 的估计值为 6666.

评: 此题背景为极大似然估计.

20.

21.

22.

23.

24.

第一章 概率论

1.2 随机变量

B 组

1. 现有一种不断分裂的细胞 X, 每个时间周期 T 内分裂一次, 一个 X 细胞每次分裂分别能生成一个或两个新的 X 细胞, 每次分裂后原 X 细胞消失, 设每次分裂成一个新 X 细胞的概率为 p , 分裂成两个新 X 细胞的概率为 $1-p$, 新细胞在下一个周期 T 内可以继续分裂, 每个细胞间相互独立, 设有一个初始的 X 细胞, 在第一个周期 T 中开始分裂, 其中 $p \in (\frac{1}{2}, 1)$.

(1) 设 $2T$ 结束后, X 细胞的数量为 ξ , 求 ξ 的分布列和数学期望;

(2) 设 $nT (n \in \mathbb{N}^*)$ 结束后, X 细胞数量为 m 的概率为 $P_m(n)$

i. 求 $P_2(n)$;

ii. 证明: $P_3(n) < \frac{8}{27p^2}$

2. 【2025 山西三重教育 12 月联考 19】已知集合 $S = \{0, 1, \dots, n\} (n \geq 3, n \in \mathbb{N}^*)$, 设 S 的所有元素按一定顺序排列得到数列 $X: x_0, x_1, \dots, x_n$ 和 $Y: y_0, y_1, \dots, y_n$, 若 X 和 Y 满足 $\{\varepsilon | \varepsilon = |x_i - y_i|, i = 0, 1, \dots, n\} = S$, 则称 X 和 Y 关于 S 全封闭, 否则称 X 和 Y 关于 S 半封闭.

(1) 若 $n = 3, X: 2, 0, 3, 1$, 写出两个不同的 Y , 使得 X 和 Y 关于 S 全封闭;

(2) 设 $i, j \in S$, 且 $i \neq j$, 证明: 若 $x_i + x_j = y_i + y_j$, 则 X 和 Y 关于 S 半封闭;

(3) 设数列 $Z: 0, 1, \dots, n$, 随机变量 ξ, η 分别服从 $P(\xi = i) = \frac{2x_i}{n(n+1)}, P(\eta = i) = \frac{2y_i}{n(n+1)}, i = 0, 1, \dots, n$, 证明: 若 X 和 Z 关于 S 全封闭, 则存在 $X \neq Y$, 使得 $E(\xi) = E(\eta) > \frac{\sum_{i=0}^n x_i y_i}{\sum_{i=0}^n (x_i + y_i)}$.

3. 数列 $a_1, a_2, \dots, a_{2n} (n \geq 2, n \in \mathbb{Z})$ 是公差为 $d (d > 0)$ 的等差数列, 现将这 $2n$ 项平均分为 A, B 两组, 从 A 中随机挑选一项 a , 从 B 中随机挑选一项 b , 记 $X = |a - b|$, X 的期望为 $E(X)$.

(1) 若 $n = 2$, 求 $P(X = d)$ 与 $E(X)$ 的所有可能值;

(2) 求 $P(X = d)$ 的最大值;

(3) 证明: $E(X) \leq nd$.

1.

2.

3.

第一章 概率论

1.2 随机变量

杂 题

1. 解答下述问题

(1) 已知随机事件 A, B , 证明: $|P(AB) - P(A)P(B)| \leq \frac{1}{4}$;

(2) 已知随机事件 A, B, C 发生的概率均为 $\frac{2}{3}$, 且这三个事件两两独立。求 $P(ABC)$ 的取值范围。

2. 6 个相同的球分别标有数字 1,2,3,4,5,6, 从中不放回地随机取球, 在某轮取完球后进行判定, 若存在正整数 $1 \leq k \leq 3$, 使得标有数字 $2k-1$ 和 $2k$ 的球均已被取出, 则终止操作。记终止操作时已经取出了 X 个球, 求 X 的数学期望。

1.

2.

第一章 概率论

1.2 随机变量