

一、填空题

1. 设总体共有21个个体，按照1 ~ 21进行编号。若使用系统抽样法抽取容量为5的样本，且第一个被抽中的样本编号为11，则其余被抽中的样本编号依次为_____。
2. 假设抛一枚正面向上概率为3/4的硬币1000次，随机变量 X 定义为硬币正面向上的次数，使用 Chebyshev 不等式估算 $X > 900$ 的概率上界为_____。
3. 假设抛一枚正面向上概率为1/5的硬币800次，随机变量 X 定义为硬币正面向上的次数，使用 Chernoff 不等式估算 $X < 40$ 的概率上界为_____。
4. 当哈希函数 $h(x) = (3x + 1) \bmod 5$ 被用于行排列变换时，集合 $A = \{0,1,4\}$ 和 $B = \{2,3,4\}$ 的最小哈希值分别为_____。
5. 设 $k = 3$ ，使用 Misra Gries 算法求得输入数据流 $\langle a, b, b, c, c, a, a, d \rangle$ 中的频繁元素为_____。
6. 对于输入数据流 $\langle 0,1,1,2,3,3 \rangle$ ，假设给定哈希函数 $h_1(x) = (2x + 1) \bmod 3$ 和 $h_2(x) = (x + 1) \bmod 3$ ，用 CM Sketch 估计元素 0 的频度为_____。
7. 对于数据流 $\langle 0,0,1,2,2,3,3,3 \rangle$ ，假设给定哈希函数 $h(x) = (7x + 2) \bmod 3$ 和 $g(x) = \begin{cases} +1, & \text{if } x \bmod 2 = 0 \\ -1, & \text{if } x \bmod 2 = 1 \end{cases}$ ，用 Count Sketch 估计元素 1 的频度为_____。
8. 对于转移概率矩阵为 $\mathbf{P} = \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 0 & 1/4 & 3/4 \\ p & 1-p & 0 \end{pmatrix}$ 的马尔可夫链，已知其平稳分布为 $(1/3, 1/3, 1/3)$ ，则参数 p 的取值是_____。

二、设一组独立随机变量 $x_{ij} (i = 1, \dots, k; j = 1, \dots, n)$ 服从参数为 p 的伯努利分布。定义随机变量 $X_i = \frac{1}{n} \sum_{j=1}^n x_{ij} \quad (i = 1, \dots, k)$ 。设随机变量 $Y = \min_{1 \leq i \leq k} X_i$ ，计算事件 $Y > (1 + \varepsilon)p$ 的概率上界；

三、给定两个集合 A, B 各包含 50 亿个元素，每个元素占用 64B。当内存使用被限制在 4×10^9 B 时，设计恰当的方案计算集合 A 和 B 的交集，并分析方案的误差。

四、存在四个网站 A, B, C, D ，其链接关系如图 2 所示。试判断基于该链接关系的随机游走是否存在平稳分布。若存在，则计算其平稳分布；若不存在，请提出方案改进其概率转移矩阵，使改进后的随机游走存在平稳分布。

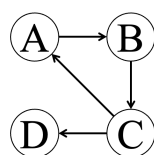


图 2