

作业 3

1. 教材习题二的1、3、19以及习题三的3.

2. (蓄水池抽样) 有一系列的数据流经某系统。我们希望对该数据流进行采样, 希望能够从该数据流中采样一个数据, 使得该数据等可能地为所有已经流经该系统的数据中的一个。假设我们不知道数据流中数据的个数, 同时也不保存已经流经系统的数据。

考虑如下算法: 当第一个数据经过时, 我们将其存放在内存中。当第 k 个数据经过时, 我们以 $1/k$ 的概率用其替代内存中的数据。

- 证明该算法的有效性。
- 如果替代的概率为 $1/2$, 求内存中存放的是哪一个数据的分布。

3. 考虑取值为正整数的随机变量 X , 其分布律为 $P(X = i) = \frac{6}{\pi^2} i^{-2}$ (注意 $\sum_{i=1}^{\infty} i^{-2} = \pi^2/6$)。分析 X 的期望值。

4. 考虑 n 个人玩一个 n 轮的游戏。在第 i ($i = 1, 2, \dots, n$) 轮游戏中, 从剩下的 $2n - 2(i - 1)$ 只手中随机挑选2只进行牵手。求牵手形成环的个数的期望值(某人左右手牵在一起也算形成一个环)。

5. (Jensen不等式) 设 f 为下凸函数, 即对于任意 x_1, x_2 且 $0 \leq \lambda \leq 1$,

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2).$$

假设 X 为只有有限个取值的离散型随机变量, 求证

$$E[f(X)] \geq f(E[X])$$

(注: 该不等式对连续型随机变量同样适用)