

10245501435 杨茜雅 Dase算法 1st

习题 2 p28-29

1. 1-160 灯泡

$$k = \frac{N}{n} = \frac{160}{20} = 8 \quad \text{直线等距抽样}$$

设第一次抽的是第 r 个个体, 以 r 为起始位置

$$\begin{array}{ccccccc} r & r+8 & r+8 \times 2 & \cdots & r+8 \times 15 = 126 \\ & 2 & 3 & & 16 \end{array}$$

$$r=6$$

\therefore 起始抽取编号为 6

4. $k = \frac{N}{n} = \frac{14}{4} = 3.5$

\therefore 采用圆形等距抽样 $k = \lfloor 3.5 \rfloor = 3$

$r=1$ 为起始位置

$$\begin{array}{ccc} r+3 \times 1 & r+3 \times 2 & r+3 \times 3 \\ 4 & 7 & 10 \end{array}$$

\therefore 其余 3 个样本编号为 4, 7, 10

5. 采用分层抽样

样本容量与总体容量之比 $100:500 = 1:5$

$$A: \frac{280}{5} = 56 \quad B: \frac{95}{5} = 19 \quad C: \frac{125}{5} = 25$$

A 抽取 56 件 B 抽取 19 件 C 抽取 25 件

8. 分层抽样 按比例分配法

$$20:500 = 1:25 \quad O: \frac{200}{25} = 8 \quad A: \frac{125}{25} = 5 \quad B: \frac{125}{25} = 5$$

$$AB: \frac{50}{25} = 2 \quad \therefore O \text{ 型 } 8 \text{ 人}, A \text{ 型 } 5 \text{ 人}, B \text{ 型 } 5 \text{ 人}, AB \text{ 型 } 2 \text{ 人}$$



13.

- - - N

采用水库抽样算法

该链表可被看作一个数据规模大小未知的数据流

只能遍历链表一次 \rightarrow 即数据又能被访问一次

每个元素被抽中的概率相等

用水库抽样算法

① 先将链表中的前 k 个元素保留下来，构建一个大小为 k 的水库



1 2 3 ... k ... N

② 对于第 m 条记录 ($m > k$)，以 $\frac{k}{m}$ 的概率决定是否由这条记录替换水库中的一条记录

③ 循环这一步

④ 直到遍历链表

当读到第 i 个元素时，产生 $1 \sim i$ 的均匀随机数 r 。if $r \leq k$,

则当前元素替换水库中第 r 个元素，发生概率为 $\frac{k}{i}$

证明：等概率且为 $\frac{k}{N}$ ① $N = k$ 时 $prob = \frac{k}{k} = 1$

② $N = m (> k)$ 时，假设 $prob = \frac{k}{m}$

③ 当 $N = m+1 (> k)$ 时 $prob = P(\text{替换不发生}) + P(\text{替换但不替换 } t, \text{ } t \text{ 是一直保留在水库中的元素})$

$$P(\text{替换不发生}) = 1 - \frac{k}{m+1} \quad \text{从 } k \text{ 个数中选中 } t$$

$$P(\text{替换但不替换 } t) = \frac{k}{m+1} \times (1 - P(\text{替换 } t)) = \frac{k}{m+1} \times (1 - \frac{1}{k})$$

$$\therefore P(t \text{ 一直保留}) = \frac{k}{m} \times (\frac{k}{m+1} \times \frac{k-1}{k} + \frac{m+1-k}{m+1}) = \frac{k}{m+1}$$



14.

———
|
———

假设每个元素被赋予的随机数是 无重复的, 则是等概率抽样。

证明:

① 当 $N=k$ 时, 则选取最大的 k 个数时 (Top- k), 则会选出所有数, $\text{prob} = 1 = \frac{k}{N}$

② 当 $N=m$ ($m > k$) 时, 假设, 水库中任意一条记录被抽取到的概率为 $\frac{k}{m}$

③ 假设记录 t 是水库中的记录, 在上一轮抽样中, 它以 $\frac{k}{m}$ 的概率保存在水库中, 当 $N=m+1$ 时, t 留下的情况有二

a. 等 $m+1$ 条记录被赋予的随机数不在 top- k 个中

$$\text{prob} = 1 - \frac{k}{m+1}$$

b. 第 $m+1$ 条记录被赋予的随机数在 top- k 个中但不替代 t

$$\frac{k}{m+1} \left(1 - \frac{1}{k} \right)$$

$$\therefore \text{prob} = \frac{k}{m} \times \left(\frac{m+1-k}{m+1} + \frac{k}{m+1} \times \frac{k-1}{k} \right)$$

$$= \frac{k}{m} \times \frac{m}{m+1} = \frac{k}{m+1}$$

为等概率抽样

等 top k

1	2	3	...	t	t+1	...	m	m+1
1	5	3	6	12	11	...	Random	

$m+1$ 个数

($m+1$) 被赋予的随机数不在 top- k 个之间的 prob $1 - \frac{k}{m+1}$

在 top- k 个之间的 prob $\frac{k}{m+1}$ 且正好取代

t , 则证明 t 当初被赋予的随机数是 top- k 中 最小的

$$\text{prob} = \frac{1}{k}$$

$$\therefore \frac{k}{m+1} \left(1 - \frac{1}{k} \right)$$

