第 8 讲: 概率分析与随机算法

姓名: 朱宇博　　**学号:** 191220186

评分: ______　　**评阅:** ______

2020 年 4 月 22 日

请独立完成作业，不得抄袭。
若得到他人帮助, 请致谢。
若参考了其它资料，请给出引用。
鼓励讨论，但需独立书写解题过程。

1 作业（必做部分）

题目 1 (CS 5.6-4)
In a card game, you remove the jacks, queens, kings, and aces from an ordinary deck of cards and shuffle them. You draw a card. If it is an ace, you are paid \$1.00, and the game is repeated. If it is a jack, you are paid \$2.00, and the game ends. If it is a queen, you are paid \$3.00, and the game ends. If it is a king, you are paid \$4.00, and the game ends. What is the maximum amount of money a rational person would pay to play this game?

解答:
设 X 为游戏收益
$E(X) = \frac{1}{4}(1 + E(X)) + \frac{1}{4}2 + \frac{1}{4}3 + \frac{1}{4}4$
$E(X) = \frac{10}{3}$.
Therefore, the maximum amount of money a rational person would pay to play this game is $\frac{10}{3}$\$.

题目 2 (CS 5.6-8)
Prove **Theorem 5.23**: Let X be a random variable defined on a sample space S and let $F_1, F_2, \ldots F_n$ be disjoint events whose union is S (i.e., a partition of S). Then

$$E(X) = \sum_{i=1}^{n} E\left(X|F_i\right) P\left(F_i\right)$$

解答:

$$\begin{aligned}
E(X) &= \sum_{s:s\in S} P(s)X(s) \\
&= \sum_{i=1}^{n} \sum_{s:s\in F_i} X(S)P(s) \\
&= \sum_{i=1}^{n} \sum_{s:s\in F_i} X(S)\frac{P(s)}{P(F_i)}P(F_i) \\
&= \sum_{i=1}^{n} E\left(X|F_i\right) P\left(F_i\right)
\end{aligned}$$

题目 3 (CS 5.7-2)
In Problem 1, let X_i be the number of correct answers the student gets on Question i, that is, X_i is either 0 or 1. What is the expected value of X_i ? What is the variance of X_i ? How does the sum of the variances of X_1 through X_5 relate to the variance of X for Problem 1?

解答:
$E(X_i) = 0.6 * 1 + 0.4 * 0 = 0.6$
$V(X_i) = 0.6 * (1 - 0.6) = 0.24$
$V(X_1) + V(X_2) + V(X_3) + V(x_4) + V(x_5) = V(\sum_{i=1}^{5} X_i) = V(X)$

题目 4 (CS 5.7-12)
How many questions need to be on a short-answer test for you to be 95% sure that someone who knows 80% of the course material gets a grade between 75% and 85%?

解答:
对于 1 个问题，其方差为 0.16.
则对于 n 个问题，其期望为 0.8n，方差为 0.16n，标准差为 $0.4\sqrt{n}$.
95% 近似对应两个标准差。
$2 \times 0.4\sqrt{n} = 0.05n$
解得 $n = 256$

题目 5 (TC 5.2-4)
Use indicator random variables to solve the following problem, which is known as the hat-check problem. Each of n customers gives a hat to a hat-check person at a restaurant. The hat-check person gives the hats back to the customers in a random order. What is the expected number of customers who get back their own hat?

解答:
对于每个人而言，其能拿到自己的帽子的概率为 $\frac{(n-1)!}{n!} = \frac{1}{n}$.
则期望人数为: $n \cdot 1 \cdot \frac{1}{n} = 1$

题目 6 (TC 5.2-5)
Let $[1..n]$ be an array of n distinct numbers. If $i < j$ and $A[i] > A[j]$, then the pair (i, j) is called an inversion of A. (See Problem 2-4 for more on inversions.) Suppose that the elements of A form a uniform random permutation of $< 1, 2, ..., n >$. Use indicator random variables to compute the expected number of inversions.

.

解答:
令 $A_{i,j}(i < j)$ 表示在排列 A 中，$A[i]$ 与 $A[j]$ 构成逆序对
X
$X_{i,j} = I(A_{i,j}) = \begin{cases} 1 & \text{A} \\ 0 & \overline{A} \end{cases}$

$Pr(I(A_{i,j})) = \frac{1}{2}, Pr(I(\overline{A_{i,j}})) = \frac{1}{2}$

$X = \sum\limits_{i=1}^{n-1} \sum\limits_{j=i+1}^{n} X_{i,j}$

$\forall 1 < i < j < n,\ E(X_{i,j}) = 1 \cdot Pr(A) + 0 \cdot Pr(\overline{A}) = \frac{1}{2}$

$$
\begin{aligned}
E(X) &= E(\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{i,j}) \\
&= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E(X_{i,j}) \\
&= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{2} \\
&= \frac{n(n-1)}{4}
\end{aligned}
$$

题目 7 (TC 5.3-3)

Suppose that instead of swapping element $A[i]$ with a random element from the subarray $A[i..n]$, we swapped it with a random element from anywhere in the array:

PERMUTE-WITH-ALL(A)

```
n = A.length
  for i = 1 to n
    swap A[i] with A[RANDOM(1, n)]
```

Does this code produce a uniform random permutation? Why or why not?

解答:

不能。因为生成每种排列的概率不相同。

例如, 在 $n = 3$ 时

设生存排列 k 的概率为 $p(k)$

$p((1, 2, 3)) = \frac{4}{27}$

$p((1, 3, 2)) = \frac{5}{27}$

$p((2, 1, 3)) = \frac{5}{27}$

$p((2, 3, 1)) = \frac{5}{27}$

$p((3, 1, 2)) = \frac{4}{27}$

$p((3, 2, 1)) = \frac{4}{27}$

题目 8 (TC 5.3-4)

Professor Armstrong suggests the following procedure for generating a uniform random permutation:

PERMUTE-B Y-CYCLIC (A)

```
  n = A.length
let B[1...n] be a new array
offset = RANDOM (1, n)
for i = 1 to n
  dest = i + offset
  if dest > n
   dest = dest - n
 B[ dest ] = A[i]
return B
```

Show that each element $A[i]$ has a $1/n$ probability of winding up in any particular position in B. Then show that Professor Armstrong is mistaken by showing that the resulting permutation is not uniformly random.

解答:
(1)
每次赋值为 $B[dest] \leftarrow A[i]$, 其中 $dest = (i+\text{offset}-1) \mod n + 1$。
因为 offset 为 1 到 n 之间的随机整数，即 offset 取到每一个值的概率为 $\frac{1}{n}$，对于确定的 i，dest 取到每一个值的概率都为 $\frac{1}{n}$.
因此，$A[i]$ 分到任何一个 position 的概率都为 $\frac{1}{n}$.
(2)
对于确定的 offset，算法的实质为将这个排列在取模意义下向后移动若干位。
offset 从 1 到 n，仅对应了 n 个排列
因此，该算法不能实现生成 uniform random permutation。

题目 9
Now consider a deterministic linear search algorithm, which we refer to as DETERMINISTIC-SEARCH. Specifically, the algorithm searches A for x in order, considering $A[1], A[2], A[3], \ldots, A[n]$ until either it finds $A[i] = x$ or it reaches the end of the array. Assume that all possible permutations of the input array are equally likely.
e. Suppose that there is exactly one index i such that $A[i] = x$. What is the average-case running time of DETERMINISTIC-SEARCH? What is the worstcase running time of DETERMINISTIC-SEARCH?
f. Generalizing your solution to part (e), suppose that there are $k \geq 1$ indices i such that $A[i] = x$. What is the average-case running time of DETERMINISTICSEARCH? What is the worst-case running time of DETERMINISTIC-SEARCH? Your answer should be a function of n and k
g. Suppose that there are no indices i such that $A[i] = x$. What is the average-case running time of DETERMINISTIC-SEARCH? What is the worst-case running time of DETERMINISTIC-SEARCH?

解答:
(e)
设 X 代表搜索的次数
$E(X) = \sum\limits_{i=1}^{n} \frac{1}{n} i = \frac{n+1}{2}$
则平均情况的运行时间为 $\frac{n+1}{2}$.
在最坏情况下，x 位于第 n 位，则需要 n 次
(f)
设 X 代表搜索的次数
根据题意, 对于第 i 个位置，$A[i] = x$ 的概率为 $\frac{C_{n-i}^{k-1}}{C_n^k}$

$$
\begin{aligned}
E(X) &= \sum_{i=1}^{n-k+1} \frac{C_{n-i}^{k-1}}{C_n^k} i \\
&= \frac{C_{n+1}^{k+1}}{C_n^k} \\
&= \frac{n+1}{k+1}
\end{aligned}
$$

在最坏情况下，后 k 个位置全为 x，则需要 $n - k + 1$ 次
(g)
此时，无论如何，每个元素都会被依次检索。
平均情况和最坏情况的时间都为 n

---

2 作业（选做部分）

题目 1 (The Coin Problem (Provided by Pei))
Suppose you have a fair coin. What is the expected number of tosses to get 3 Heads in a row (连续三次正面朝上)? What about n Heads in a row?

解答：

---

3 Open Topics

Open Topics 1 (Average-case Analysis of *FindMax*)
学习如下讲座视频，讲解其中的算法分析过程。
参考资料：

- The Analysis of Algorithms.mp4 by Donald Knuth @ Stanford Lecture

Open Topics 2 (Average-case Analysis of Binary Search)
分析 Binary Search 的平均情况时间复杂度。
参考资料：

- Section 6.2.1 of "The Art of Computer Programming (Vol 3; 2nd Edition)" by Donald Knuth. ("Fibonaccian search" 部分可选)

4 反馈