

Introduction to Algorithms
CLRS
3rd Edition

Md. Mohsin Ali Khan

November 10, 2016

5 Probabilistic Analysis and Randomized Algorithms

5.2 Indicator random variables

- 5.2-1 In HIRE -ASSISTANT, assuming that the candidates are presented in a random order, what is the probability that you hire exactly one time? What is the probability that you hire exactly n times? **Solution:** Exactly one time is $\frac{(n-1)!}{n!} = \frac{1}{n}$ and exactly n times is $\frac{1}{n!}$
- 5.2-2 In HIRE -A SSISTANT , assuming that the candidates are presented in a random order, what is the probability that you hire exactly twice? **Solution:** Let i be the index where the best candidate comes in for the interview. Then there are $x = \sum_{i=2}^n \binom{n-1}{i-1} (i-2)!(n-i)!$ ways the candidates can be permuted so that exactly 2 candidates are hired. And there are total $n!$ possible ways that the candidates can be permuted for the

interview. So, the probability in question is

$$\begin{aligned}
 p &= \frac{x}{n!} \\
 &= \frac{\sum_{i=2}^n \binom{n-1}{i-1} (i-2)! (n-i)!}{n!} \\
 &= \sum_{i=2}^n \frac{(n-1)! (i-2)! (n-i)!}{(n-1-i+1)! (i-1)! n!} \\
 &= \sum_{i=2}^n \frac{(n-1)! (i-2)!}{(i-1)! n!} \\
 &= \sum_{i=2}^n \frac{1}{(i-1)n} \\
 &= \frac{1}{n} \left(1 + \frac{1}{2} + \cdots + \frac{1}{n-1} \right) \\
 &\leq \frac{1}{n} \ln(n-1) + 1
 \end{aligned}$$

5.2-3 Use indicator random variables to compute the expected value of the sum of n dice. **Solution:** Let us consider random variables $D_{(i,j)}$ which take values 1 when the outcome of the i -th dice is j and 0 otherwise. Let n_j is the number of dices that produces outcome j . Then we have $n_j =$

$\sum_{i=1}^n D_{(i,j)}$. Let X be the sum of the outcome of the dices. Then

$$\begin{aligned}
X &= \sum_{j=1}^6 j \cdot n_j \\
&= \sum_{j=1}^6 j \cdot \sum_{i=1}^n D_{(i,j)} \\
E[X] &= E \left[\sum_{j=1}^6 j \cdot \sum_{i=1}^n D_{(i,j)} \right] \\
&= \sum_{j=1}^6 E \left[j \cdot \sum_{i=1}^n D_{(i,j)} \right] \\
&= \sum_{j=1}^6 j \cdot E \left[\sum_{i=1}^n D_{(i,j)} \right] \\
&= \sum_{j=1}^6 j \sum_{i=1}^n E[D_{(i,j)}] \\
&= \sum_{j=1}^6 j \sum_{i=1}^n \Pr(D_{(i,j)} = 1) \cdot 1 + \Pr(D_{(i,j)} = 0) \cdot 0 \\
&= \sum_{j=1}^6 j \sum_{i=1}^n \Pr(D_{(i,j)} = 1) \\
&= \sum_{j=1}^6 j \sum_{i=1}^n \frac{1}{6} \\
&= \sum_{j=1}^6 j \cdot \frac{n}{6} \\
&= \frac{n}{6} \sum_{j=1}^6 j \\
&= \frac{n}{6} \cdot \frac{6(6+1)}{2} \\
&= \frac{7}{2}n
\end{aligned}$$

5.2-3 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? **Solution:** Let C_i be the indicator

random variables which take values 1 if the i -th customer gets his cap back and 0 otherwise. We claim that the probability, $Pr(C_i = 1) = \frac{1}{n}$. This is because the probability that the i -th customer has picked up his cap is the probability that none of the customers before the i -th customer has taken the i -th customer's cap and the i -th customer has taken his own cap. This probability p can be expressed in the following form

$$\begin{aligned} p &= \left(1 - \frac{1}{n}\right) \left(1 - \frac{1}{n-1}\right) \cdots \left(1 - \frac{1}{n-i+2}\right) \left(\frac{1}{n-i+1}\right) \\ &= \left(\frac{n-1}{n}\right) \left(\frac{n-2}{n-1}\right) \cdots \left(\frac{n-i+1}{n-i+2}\right) \left(\frac{1}{n-i+1}\right) \\ &= \frac{1}{n} \end{aligned}$$

Let X be the total number of customers who get back their caps. Then we have

$$\begin{aligned} X &= \sum_{i=1}^n C_i \\ E[X] &= E\left[\sum_{i=1}^n C_i\right] \\ &= \sum_{i=1}^n E[C_i] \\ &= \sum_{i=1}^n Pr(C_i = 1) \cdot 1 + Pr(C_i = 0) \cdot 0 \\ &= \sum_{i=1}^n Pr(C_i = 1) \\ &= \sum_{i=1}^n \frac{1}{n} \\ &= 1 \end{aligned}$$

5.2-4 Let $A[1 \cdots n]$ 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 $\langle 1, 2, \dots, n \rangle$. Use indicator random variables to compute the expected number of inversions. **Solution:** We can obtain $2\binom{n}{2}$ number of pairs from a set of size n . Observe that for every pair of elements a and b there are two distinct pairs (a, b) and (b, a) . As every element in the set is a distinct number and numbers form a total order, for exactly one of these pairs the first element is bigger than the second element. Which means, if you choose arbitrarily any pair of elements in the given array, the first element being bigger than the second element is exactly $\frac{1}{2}$.

For all $i < j$, let, $I_{(i,j)}$ be the indicator random variables which take value 1 when the pair (i, j) is an inversion of A . They take the value 0 otherwise. Let X be the total number of inversions of A . Then we have

$$\begin{aligned}
X &= \sum_{i=1}^{n-1} \sum_{j=i+1}^n I_{(i,j)} \\
E[X] &= E \left[\sum_{i=1}^{n-1} \sum_{j=i+1}^n I_{(i,j)} \right] \\
&= \sum_{i=1}^{n-1} E \left[\sum_{j=i+1}^n I_{(i,j)} \right] \\
&= \sum_{i=1}^{n-1} \sum_{j=i+1}^n E [I_{(i,j)}] \\
&= \sum_{i=1}^{n-1} \sum_{j=i+1}^n Pr(I_{(i,j)} = 1) \cdot 1 + Pr(I_{(i,j)} = 0) \cdot 0 \\
&= \sum_{i=1}^{n-1} \sum_{j=i+1}^n Pr(I_{(i,j)} = 1) \\
&= \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{1}{2} \\
&= \sum_{i=1}^{n-1} \frac{(n-j+1)}{2} \\
&= \sum_{i=1}^{n-1} \frac{(n+1)}{2} - \sum_{i=1}^{n-1} \frac{j}{2} \\
&= \frac{(n-1)(n+1)}{2} - \frac{1}{2} \cdot \frac{(n-1)n}{2} \\
&= \frac{(n-1)(n+1)}{2} - \frac{(n-1)n}{4} \\
&= \frac{2(n-1)(n+1) - n(n-1)}{4} \\
&= \frac{(n-1)(2n+2-n)}{4} \\
&= \frac{(n-1)(n+2)}{4}
\end{aligned}$$

5 Probabilistic Analysis and Randomized Algorithms

5.4 Probabilistic analysis and further uses of indicator random variables

- 5.4-1 How many people must there be in a room before the probability that someone has the same birthday as you do is at least $1/2$? How many people must there be before the probability that at least two people have a birthdays on July 4 is greater than $1/2$? **Solution:** The number of people required to find the first person who has the same same birthday as mine is geometrically distributed where the success probability is $p = 1/365$. Let's assume that the probability in question will be at least $1/2$ when the number of people become n . Then the cumulative mass function of geometric distribution outputs at least $1/2$ when k and p are plugged in it. The pmf is $F(n|p) = 1 - (1 - p)^k \geq 1/2$. So, we set

$$\begin{aligned} 1 - (1 - p)^n &\geq 1/2 \\ (1 - p)^n &\leq 1/2 \end{aligned}$$

Taking logarithm on both side we have

$$\begin{aligned} \log(1 - p)^n &\leq \log 2^{-1} \\ n \log(1 - p) &\leq -1 \\ n &\geq \frac{-1}{\log(1 - p)} \text{ (as } \log(1-p) \text{ is less than zero)} \\ n &\geq \frac{-1}{\log(1 - 1/365)} \\ n &\geq \frac{-1}{\log(364/365)} \\ n &\geq \frac{-1}{-.003958} \\ n &\geq 253 \end{aligned}$$

So, including me there shall be 254 people. The other part of the problem can be solved using binomial distribution. The number of people who have birthdays on July 4 are binomially distributed. Lets say we need n many people in the room to have the probability in question to be at least $1/2$. Which means the cumulative mass function of binomial distribution F will output less than $1/2$ for the value $k = 1$ when the success probability p

and n are plugged in it. So we have

$$\begin{aligned}
F(k=1|p, n) &\leq 1/2 \\
\sum_{i=0} \binom{n}{i} p^i (1-p)^{n-i} &\leq 1/2 \\
\binom{n}{0} p^0 (1-p)^n + \binom{n}{1} p (1-p)^{n-1} &\leq 1/2 \\
(1-p)^n + np(1-p) &\leq 1/2 \\
(364/365)^n + \frac{n}{365} (364/365)^{n-1} &\leq 1/2
\end{aligned}$$

Maybe now run a program to compute the smallest value of n so that

- 5.4-2 Suppose that we toss balls into b bins until some bin contains two balls. Each toss is independent, and each ball is equally likely to end up in any bin. What is the expected number of ball tosses? **Solution:** Let X_i be a random variable that takes value 1 if all the ball toss before i -th toss and also the i -th toss fails to get at least two balls in one bin. X_i takes value 0 otherwise. Let X be the number of toss required to get the first bin that receives two balls. Then

$$\begin{aligned}
X &= 1 + X_1 + \cdots + X_b \\
E[X] &= E[1 + X_1 + \cdots + X_b] \\
&= 1 + E[X_1] + \cdots + E[X_b] \\
&= 1 + \sum_{i=1}^b E[X_i] \\
&= 1 + \sum_{i=1}^b P(X_i = 1) \cdot 1 + P(X_i = 0) \cdot 0 \\
&= 1 + \sum_{i=1}^b P(X_i = 1) \\
&= 1 + \sum_{i=1}^b \frac{b(b-1) \cdots (b-i+1)}{b^i} \\
&= 1 + \sum_{i=1}^b \frac{b!}{(b-i)!b^i}
\end{aligned}$$

- 5.4-3 For the analysis of the birthday paradox, is it important that the birthdays be mutually independent, or is pairwise independence sufficient? Justify your answer. **Solution:** They has to be mutually independent. Because When there are three persons in the room, the fact that the first two persons have certain birthday doesn't impact the third person to be biased

to take a certain birthday. The third person still uniformly likely to take any of the possible 365 birth days. And the analysis considers this fact when it argues that the third person would have 363 possible birth days so that it doesn't have the same birth day with any of the first two. And in the analysis this probability is computed as $\frac{1}{365} \cdot 363$. That is, it considers the identical and uniform probability for each possible birth day. The same argument applies for all the subsequent persons.

- 5.4-4 How many people should be invited to a party in order to make it likely that there are three people with the same birthday? **Solution:** Let us consider Indicator random variables $I_{(i,j,k)}$ such that they take values 1 when the persons i, j and k have the same birthdays. Otherwise they take values 0. Let us assume that the total number of 3-subsets that have the same birthdays is X . Then

$$\begin{aligned}
 X &= \sum_{i,j,k} I_{(i,j,k)} \\
 E[X] &= E \left[\sum_{i,j,k} I_{(i,j,k)} \right] \\
 &= \sum_{i,j,k} E[I_{(i,j,k)}] \\
 &= \sum_{i,j,k} Pr(I_{(i,j,k)} = 1) \cdot 1 + Pr(I_{(i,j,k)} = 0) \cdot 0 \\
 &= \sum_{i,j,k} Pr(I_{(i,j,k)} = 1) \\
 &= \sum_{i,j,k} \binom{m}{3} \frac{1}{365^3} \cdot 365 \\
 &= \sum_{i,j,k} \binom{m}{3} \frac{1}{365^2} \\
 &\quad (m \text{ is total number of people present in the room}) \\
 &= \frac{m(m-1)(m-2)}{6} \cdot \frac{1}{365^2}
 \end{aligned}$$

Now if we set $E[X] = 1$ then we have $\frac{m(m-1)(m-2)}{6} \cdot \frac{1}{365^2} = 1$ Using wolfram alpha, I get $m = 93.810 \approx 94$

- 5.4-6 Suppose that n balls are tossed into n bins, where each toss is independent and the ball is equally likely to end up in any bin. What is the expected number of empty bins? What is the expected number of bins with exactly one ball? **Solution:** Let us consider the indicator random variables B_i which take value 1 if the bin i is empty after tossing all the balls and 0 otherwise. And let X be the number of empty bins after tossing all the

balls. Then we have

$$\begin{aligned}
 X &= \sum_{i=1}^n B_i \\
 E[X] &= E\left[\sum_{i=1}^n B_i\right] \\
 &= \sum_{i=1}^n E[B_i] \\
 &= \sum_{i=1}^n Pr(B_i = 1) \cdot 1 + Pr(B_i = 0) \cdot 0 \\
 &= \sum_{i=1}^n Pr(B_i = 1) \\
 &= \sum_{i=1}^n \left(\frac{n-1}{n}\right)^n \\
 &= n \left(\frac{n-1}{n}\right)^n \\
 &= \frac{(n-1)^n}{n^{n-1}}
 \end{aligned}$$

Solution of the other part: Let us consider the indicator random variables B_i which take value 1 if the bin i has exactly one ball after tossing all the balls and 0 otherwise. And let X be the number of bins with exactly

one ball after tossing all the balls. Then we have

$$\begin{aligned}
X &= \sum_{i=1}^n nB_i \\
E[X] &= E\left[\sum_{i=1}^n B_i\right] \\
&= \sum_{i=1}^n E[B_i] \\
&= \sum_{i=1}^n Pr(B_i = 1) \cdot 1 + Pr(B_i = 0) \cdot 0 \\
&= \sum_{i=1}^n Pr(B_i = 1) \\
&= \sum_{i=1}^n \binom{n}{1} \left(\frac{1}{n}\right)^1 \left(1 - \frac{1}{n}\right)^{n-1} \\
&= \sum_{i=1}^n n \cdot \frac{1}{n} \cdot \left(\frac{n-1}{n}\right)^{n-1} \\
&= \sum_{i=1}^n \left(\frac{n-1}{n}\right)^{n-1} \\
&= n \left(\frac{n-1}{n}\right)^{n-1} \\
&= \frac{(n-1)^{n-1}}{n^{n-2}}
\end{aligned}$$

6 Heapsort

6.1 Heaps

6.1-1 **What are the minimum and maximum numbers of elements in a heap of height h ?** **Solution:** We claim that a full binary tree of height h has 2^h leaves. Let's show the claim by induction. Base case $h = 0$ is immediate. Let's assume it is true for h . The tree with height $h + 1$ will have $2 \times 2^h = 2^{h+1}$ leaves. So, it is true for $h + 1$. Consequently the claim is true.

The maximum number of elements in a heap of height h is the total number of elements in a full binary tree of height h . The total number of elements m in a full binary tree is the sum of number of leaves in full binary trees of heights $h \in \{0, 1, \dots, h\}$. So,

$$\begin{aligned}
m &= 2^0 + 2^1 + \cdots + 2^h \\
&= 2^{h+1} - 1
\end{aligned}$$

And the minimum number of elements in a heap of height h is the total number of elements in a heap of height $h - 1$ plus 1. Which is $2^{h-1+1} - 1 + 1 = 2^h$

6.1-2 **Show that an n -element heap has height $\lfloor \log_2 n \rfloor$**

Solution 1: Let h be the height of the heap. From the first problem of this section we know

$$\begin{aligned}
2^h &\leq n < 2^{h+1} \\
\log_2 2^h &\leq \log_2(n+1) < \log_2 2^{h+1} \\
h &\leq \log_2 n < h+1
\end{aligned}$$

As h is an integer, we have $\lfloor \log_2 n \rfloor = h$

□

Solution 2: Let us consider a function $h : \mathbb{N} \rightarrow \mathbb{N}_0$ that takes the number of elements of a heap as input and outputs the height of the heap. We show the claim in question is true by induction.

- (a) The base case is $n = 1$. The height of the heap with 1 element is $\lfloor \log_2 1 \rfloor = 0$. Which satisfy the base case because when there is only one element in the heap, the height of the heap is 0.
- (b) Let us assume that the claim is true for n . That is $h(n) = \lfloor \log_2 n \rfloor$.
- (c) We now show that the claim is also true for $n + 1$. That is $h(n + 1) = \lfloor \log_2(n + 1) \rfloor$. As a heap is a complete binary tree, we can write $h(n + 1) = h(n) + a$ where a is the increase of the height of the heap because of the additional one element. There are two cases to consider now. One is the case in which the heap is a full binary tree and in the other, it is not full.

Full: If the heap with n element is a full binary tree, that means the heap contains the maximum number of elements possible with its current height. So, $a = 1$ and we have,

$$h(n + 1) = h(n) + a = \lfloor \log_2 n \rfloor + 1 \quad (1)$$

Observe that when the heap is a full binary tree, we have

$$\begin{aligned}
n &= 2^{\lfloor \log_2 n \rfloor + 1} - 1 \\
n + 1 &= 2^{\lfloor \log_2 n \rfloor + 1} \\
\log_2(n + 1) &= \log_2 2^{\lfloor \log_2 n \rfloor + 1} \\
\log_2(n + 1) &= \lfloor \log_2 n \rfloor + 1 \\
\lfloor \log_2(n + 1) \rfloor &= h(n + 1) \text{ [Using equation ??]} \\
h(n + 1) &= \lfloor \log_2(n + 1) \rfloor
\end{aligned}$$

Not full: Now consider the case when the heap is not a full binary tree. In this case $a = 0$. So we have

$$h(n+1) = h(n) + a = \lfloor \log_2 n \rfloor \quad (2)$$

Observe that when the heap is not a full binary tree, we have

$$\begin{aligned} 2^{\lfloor \log_2 n \rfloor} &\leq n < 2^{\lfloor \log_2 n \rfloor + 1} - 1 \\ 2^{\lfloor \log_2 n \rfloor} &\leq n + 1 < 2^{\lfloor \log_2 n \rfloor + 1} \\ \log_2 2^{\lfloor \log_2 n \rfloor} &\leq \log_2(n+1) < \log_2 2^{\lfloor \log_2 n \rfloor + 1} \\ \lfloor \log_2 n \rfloor &\leq \log_2(n+1) < \lfloor \log_2 n \rfloor + 1 \end{aligned}$$

It is immediate to observe from the last inequalities that $\lfloor \log_2(n) \rfloor = \lfloor \log_2(n+1) \rfloor$. Using this equality in equation ??, we have $h(n+1) = \lfloor \log_2(n+1) \rfloor$ \square

- 6.1-3 **Show that in any subtree of a max-heap, the root of the subtree contains the largest value occurring anywhere in that subtree.** **Solution:** We show the claim by induction. All the children of the root have smaller or equal values than that of the root by the heap property of max-heap. So, the base case when the height of the subtree is 1 is true. We assume that the claim is true when the height of the subtree is h . Then the value of every leaf of the subtree of height $h+1$ is smaller than its parent. And the parent is a leaf of the subtree of length h , which is smaller than the root by induction hypothesis. As a result, by the transitivity of the order relation (we assume the order relation is transitive), every leaf of the subtree of length $h+1$ is smaller or equal than the root. Consequently the claim in question is true. \square
- 6.1-4 **Where in a max-heap might the smallest element reside, assuming that all elements are distinct?** **Solution:** The smallest element will be one of the leaves of the underlying complete binary tree of the heap.
- 6.1-6 **Is an array that is in sorted order a min-heap?** **Solution:** Yes, because in the underlying binary tree, the children of a parent node are elements which are on the right side of the corresponding element of the parent node in the array. As the array is sorted, any element is smaller or equal than any other elements on its right. Consequently, the value of any node in the tree is smaller than that of any children.
- 6.1-6 **Is the array with values $\langle 23, 17, 14, 6, 13, 10, 1, 5, 7, 12 \rangle$ is a max-heap??** **Solution:** No, because in the underlying binary tree, 7 becomes a child of 6. But in max heap the value stored in a node must be greater or equal than the values of both of the left and right children individually.
- 6.1-7 **Show that, with the array representation for storing an n -element heap, the leaves are the nodes indexed by $\lfloor \frac{n}{2} \rfloor + 1, \lfloor \frac{n}{2} \rfloor + 2, \dots, n$** **Solution:** We know the element in the array at index i has a left child at position $2i$

and right child at index $2i + 1$. Observe that any element at index less or equal to $\lfloor \frac{n}{2} \rfloor$ surely has a left child because $2\lfloor \frac{n}{2} \rfloor \leq n$. So any element in the array at index less or equal to $\lfloor \frac{n}{2} \rfloor$ can't be a leaf. On the other hand every element greater than $\lfloor \frac{n}{2} \rfloor$ can't have either left or right child because $2(\lfloor \frac{n}{2} \rfloor + 1) > n$ and also $2(\lfloor \frac{n}{2} \rfloor + 1) + 1 > n$. As a result all the elements in the array at indices mentioned in the question are leaves of the heap and there are no other leaves.