CS 600 A - Advanced Algorithms – Homework 10

Name: Deepali Nagwade | 20013393

Question 19.8.3: Suppose a certain birth defect occurs independently at random with probability p = 0.02 in any live birth. Use a Chernoff bound to bound the probability that more than 4% of the 1 million children born in a given large city have this birth defect.

Answer:

μ= E[X] = $\sum_1^n$ p = 0.02 + 0.02 + ... + 0.02 = 20000

Where X = $X_1 + X_2 ...+ X_n$ is an independent random variable that denotes live birth defects among n births.

We have to bound the probability that 4% of the 1 million children born in a given large city have this birth defect and 4% of 1 million equals to 40000.

Therefore we have to bound the probability for X > 4% or X > 40000

By Chernoff Bounds, For ∂ = 1 the upper bound is

$P(x >= (1+\partial)\ \mu) = P(X >= 40000) <= [e^{\partial} / (1 + \partial)^{1+\partial}] <= [e / 4]^{20000}$

Question 19.8.18: Consider a modification of the Fisher-Yates random shuffling algorithm where we replace the call to random (k + 1) with random (n) , and take the for-loop down to 0, so that the algorithm now swaps each element with another element in the array, with each cell in the array having an equal likelihood of being the swap location. Show that this algorithm does not generate every permutation with equal probability. Hint: Consider the case when n = 3.

Answer: Consider modifying the Fisher-Yates random shuffling algorithm by replacing the call to random(k+1) with random(n). It is critical to show that this modified algorithm does not create every permutation with equal probability.

Let's look at the permutation (1,2,3) to see how this works. The amount of ways a certain permutation can be created through element swaps influences the likelihood of obtaining it. For example, (1,2,3) can be generated by switching the first and second or first and third elements. (1,3,2), on the other hand, can only be formed by exchanging the first and third elements. As a result, (1,2,3) has more possibilities, making it more likely.

When all potential permutations are considered, each with a probability of 1/27 (there are 3! = 6 possible permutations), the distribution becomes 1/27 for each of the remaining five permutations: (1,3,2), (2,1,3), (2,3,1), (3,1,2), and (3,2,1). When we add these probabilities together, we get 1/27 + 1/27 + 1/27 + 1/27 + 1/27 + 1/27 + 1/27 = 6/27. This disparity suggests that, under this updated approach, not all permutations are equally likely.

Question 19.8.35 In a famous experiment, Stanley Milgram told a group of people in Kansas and Nebraska to each send a postcard to a lawyer in Boston, but they had to do it by forwarding it to someone that they knew, who had to forward it to someone that they knew, and so on. Most of the postcards that were successfully forwarded made it in 6 hops, which gave rise to the saying that everyone in America is separated by "six degrees of separation." The idea behind this experiment is also behind a technique, called probabilistic packet marking, for doing traceback during a distributed denial-of-service attack, where a website is bombarded by connection requests. In implementing the probabilistic packet marking strategy, a router, R, will, with some probability, p ≤ 1 / 2 , replace some seldom-used parts of a packet it is processing with the IP address for R , to enable tracing back the attack to the sender. It is as if, in the Milgram experiment, there is just one sender, who is mailing multiple postcards, and each person forwarding a postcard would, with probability, p , erase the return address and replace it with his own. Suppose that an attacker is sending a large number of packets in a denial-of-service attack to some recipient, and every one of the d routers in the path from the sender to the recipient is performing probabilistic packet marking.

(a) What is the probability that the router farthest from the recipient will mark a packet and this mark will survive all the way to the recipient?

(b) Derive a good upper bound on the expected number of packets that the recipient needs to collect to identify all the routers along the path from the sender to the recipient.

Answer:

(a) The likelihood that a router will execute probabilistic packet marking is given as p <. Now, for the packet to survive with a mark created by the furthest router, other routers must not execute probabilistic packet marketing on that packet, and the chance of this is given by (1 - p). As a result, the overall likelihood that the router farthest away from the destination will mark a packet and that this mark will survive all the way to the recipient is $p\ (1 - p)^{d-1}$, where d is the total number of routers.

(b) The expected number of packets that the recipient needs to collect to identify all the routers along the path from the sender to the recipient is at most d/(1 - p).

The probability that a packet will be marked by the farthest router and that this mark will survive all the way to the recipient is (1 - p)^d.

This is because the probability that a router will mark a packet is p, and the probability that a packet will be forwarded to the next router is (1-p). Therefore, the probability that a packet will be marked by the farthest router and that this mark will survive all the way to the recipient is p(1-p)^d.

The expected number of packets that the recipient needs to collect to identify all the routers along the path from the sender to the recipient is at most d/(1 - p).

This is because the probability that a router will mark a packet is p, and the probability that a packet will be forwarded to the next router is (1-p). Therefore, the expected number of packets that the recipient needs to collect to identify all the routers along the path from the sender to the recipient is at most d/(1 - p).

The number of packets that the recipient needs to collect in order to identify all the routers along the path from the sender to the recipient is at most d/(1 - p).