Romil V. Shah
20008692

CS / CPE 600
Prof. Reza Peyrovian

Homework Assignment 10
Submission Date: 12 / 04 / 2022

Q1. No. 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.

Sol.

For i = 1, …, 10^6, Compute μ

μ = E[X]= ∑(i=1)^1000000 E[x_i] = ∑(i=1)^1000000 0.02 = 20000

4% of 1 million children would be = 0.04 * 1000000 = 40000

By Chernoff bounds, for δ = 1, upper bound is

$Pr(X \geq (1+\delta)\mu) = P(X \geq 40000) <$ [[e^δ] / (1+δ)^(1+δ)] ^ μ = [e / 4]^20000

Q2. No. 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 $n = 3$.

Sol.

There are n^n possibilities how this algorithm chooses random numbers.
There is n! permutations, it is not possible to divide n^n by n!

This algorithm does not generate every permutation with equal probability.
For example, when n = 4, the permutation (1, 2, 3 ,4) is twice as likely to be generated as the permutation (1, 4, 3, 2).

Let us take an example,
$n = 4$, $n^n = 4^4 = 256$ possible choices for random numbers.
$n! = 4! = 24$ possible permutations.
256/24 = 10.67 which is not possible.
The possibility of permutations to be more likely is more.

Q3. No. 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, $p \leq 1/2$, 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.

Sol.

a.

For implementing probabilistic packet marking strategy, a router R with some probability (p <= (1 / 2)).

The probability of the packet received by the recipient i.e. marked by the i^{th} (1<= i <= d) router along the attack path is = $p\ (1 - p)^{(d - i)}$ where d is total number of routers.

b.

Above problem is same as coupon collector problem. The recipient must collect d routers by visiting series of routers.

Let X be the random variable representing number of times to visit for d routers:
X can be written as

$$X = X_1 + X_2 + X_3 + \dots X_d$$

Let Xi be the number of trips recipient must made in order to go from having i-1 distinct routers. Got i-1 distinct coupons, probability of getting new router will be

$$p_i = (d - (i - 1)) / d$$

Since there are d routers, and d - (i - 1) we don't have. By the linearity of expectation

$$\begin{aligned} E[X] &= E[X_1] + E[X_2] + E[X_3] + \dots E[X_d] \\ &= 1/p_1 + 1/p_2 + 1/p_3 + \dots + 1/p_d \\ &= dH_d \qquad \text{(computed in book)} \end{aligned}$$

where H_d is the harmonic number and can be approximated as $\ln d < H_d < \ln d+1$
Now according to tail estimate, recipient must make more than d ln d traceback to get all d routers.