CS 600 HW 10
SHUCHI PARAGBHAI MEHTA
CWID: 20009083

- 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.

For i = 1,..,10^6, Compute μ

$\mu = E[X] = \sum_{i=1}^{1000000} E[x_i] = \sum_{i=1}^{1000000} 0.02 = 20000$

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

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

$Pr(X \geq (1+\delta)\mu) = P(X \geq 40000) < [\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}]^{\mu} = \left[\frac{e}{4}\right]^{20000}$

- 19.8.18

Consider a modification of the Fisher-Yates random shuffling algorithm where we replace the call to $\texttt{random}(k+1)$ with $\texttt{random}(n)$, 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 $n = 3$.

n^n different ways exist for an algorithm to select random integers. Since there are n! permutations, n^n cannot be divided by n!.

Example:

n=3, $n^n = 3^3 = 27$ possible choices for random numbers.

n! = 3! = 6 possible permutations

27/6 = 4.5 which is not possible.

The 4 and 5 random number selections result in three different permutations.
There is a greater potential for permutations to occur.

- 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.

a) Using a router R with some probability $p \leq ½$, design a probabilistic packet marking scheme.
The likelihood that a packet will be received by the recipient, identified as i^{th} ($1 \leq i \leq d$) router along the attack path,

$= p(1-p)^{d-i}$, Where d is total number of routers.

b) The mentioned issue is the same as the coupon collector problem. The recipient must visit a series of routers in order to obtain D routers.
Let X be the random variable that represents the number of times that d routers will need to visit:
X may be expressed as,

$X = X_1 + X_2 + X_3 + ... X_d$

Let Xi represent the number of trips the recipient must make to transition from having i-1 unique routers. received i-1 different coupons, increasing the chances of getting a new router, which will be,

$$p_i = \frac{d-(i-1)}{d}$$

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

$E[X] = E[X_1] + E[X_2] + E[X_3] + ... E[X_d]$

$= 1/p_1 + 1/p_2 + 1/p_3 + ...+ 1/p_d$

$= dH_d$ (computed in book)

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