Problem Set 11

Due: Friday, December 3 at 5 PM

Problem 1. Your TA has invented a random number generator using a bunch of red balls, a bunch of green balls, and a bin. This is how it works.

Initially, there is a red ball and a green ball in the bin. Your TA randomly removes one ball from the bin, and then puts back into the bin two balls of the same color as the one removed. He/she repeats this process until there are exactly $n$ balls in the bin. Define the random variable

$$R_n = \text{number of red balls in the bin when the bin has } n \text{ balls.}$$

Your TA observes that $R_n$ is a random variable taking integer values from 1 to $n - 1$ with a uniform distribution. Prove this. Hint: Induction.

Problem 2. A biased coin will be flipped $n$ times. The probability of Heads on each flip is $h$, and the flips are independent. Write simple formulas for:

(a) The probability of $k$ Heads.

(b) The probability of at least one Head.

(c) The expected number of flips before the first Head.

Problem 3. Suppose $n$ balls are thrown randomly into $n$ boxes, so each ball lands in each box with uniform probability. Also, suppose the outcome of each throw is independent of all the other throws.

(a) Let $X_i$ be an indicator random variable whose value is 1 if box $i$ is empty and 0 otherwise. Write a simple closed form expression for the probability distribution of $X_i$. Are $X_1, X_2, \ldots, X_n$ independent random variables?

(b) Find a constant, $c$, such that the expected number of empty boxes is asymptotically equal ($\sim$) to $cn$.

(c) Show that

$$\Pr(\text{at least } k \text{ balls fall in the first box}) \leq {n \choose k} \left( \frac{1}{n} \right)^k .$$
(d) Let \( R \) be the maximum of the numbers of balls that land in each of the boxes. Conclude from the previous parts that
\[
\Pr (R \geq k) \leq \frac{n}{k!}.
\]

(e) Conclude that
\[
\Pr (R \geq n\epsilon) \sim 0
\]

for all \( \epsilon > 0 \).

**Problem 4.** We are given a set of \( n \) positive integers. We then determine the maximum of these numbers by the following procedure:

Randomly arrange the numbers in a sequence.

Let the “current maximum” initially be the first number in the sequence and the “current element” be the second element of the sequence. If the current element is greater than the current maximum, update the current maximum to be the current element, and let the current element be the next element of the sequence. Repeat this process until there is no next element.

Prove that the expected number of updates is \( \sim \ln n \).

*Hint:* Let \( M_i \) be the indicator variable for the event that the \( i \)th element of the sequence is bigger than all the previous elements in the sequence.

**Problem 5.** A true story from World War II:

The army needs to identify soldiers with a disease called “klep”. There is a way to test blood to determine whether it came from someone with klep. The straightforward approach is to test each soldier individually. This requires \( n \) tests, where \( n \) is the number of soldiers. A better approach is the following: group the soldiers into groups of \( k \). Blend the blood samples of each group and apply the test once to each blended sample. If the group-blend doesn’t have klep, we are done with that group after one test. If the group-blend fails the test, then someone in the group has klep, and we individually test all the soldiers in the group.

Assume each soldier has klep with probability, \( p \), independently of all the other soldiers.

(a) What is the expected number of tests as a function of \( n \), \( p \), and \( k \)?

(b) How should \( k \) be chosen to minimize the expected number of test performed, and what is the resulting expectation?

(c) What fraction of the work does the grouping method expect to save over the straightforward approach in a million-strong army where 1% have klep?

**Problem 6.** A tournament is a digraph in which there is exactly one edge (in one direction or the other) between two vertices. The vertices are called “players,” and if there is a edge from player \( A \) to player \( B \), then player \( A \) is said to have beaten player \( B \). A ranking in the tournament is a path visiting every player exactly once. (That is, the first player on the path beat the second, the second beat the third,\ldots, the next-to-last beat the last. However, it need not be the case that the first player beat the third player, or more generally, that a player beats any player other than the next one in the ranking.)
(a) Describe a tournament of 9 players that has $3^3 = 27$ rankings.

(b) Prove that there is a tournament with 9 players that has over 1000 distinct rankings. *Hint:* If each players is equally likely to beat any other player, what is the expected number of rankings?

(This problem is an instance of the *probabilistic method.* It uses probability to prove the existence of an object without constructing it.)

**Problem 7.** In this problem you will check a proof of:

**Theorem.** Let $E_1, E_2, \ldots, E_n$ be a sequence of mutually independent events, and let $K$ be the random variable equal to the number of these events that occur. The probability that none of the events occur is at most $e^{-E[K]}$.

To prove the Theorem, let $K_i$ be the indicator variable for the event $E_i$. Justify each line in the following derivation:

**Proof.**

$$
Pr(K = 0) = Pr\left(\bigcup_{i=1}^{n} E_i\right)
= Pr\left(\bigcap_{i=1}^{n} \overline{E_i}\right)
= \prod_{i=1}^{n} (1 - Pr(E_i))
\leq \prod_{i=1}^{n} e^{-Pr(E_i)}
= e^{-\sum_{i=1}^{n} Pr(E_i)}
= e^{-\sum_{i=1}^{n} E[K_i]}
= e^{-E[K]}.
$$