UC Berkeley
Department of Electrical Engineering and Computer Sciences

EE126: Probability and Random Processes

Problem Set 2

Spring 2016

Issued: Thursday, January 28, 2016 **Due:** 9am, Thursday, February 4, 2016

Problem 1. N couples enter a casino. After two hours, N of the original $2N$ people remain (the rest have left). Each person decides to leave with probability p independent of others' decisions. What is the expected number of couples still in the casino at the end of two hours?

Problem 2. A bin contains balls numbered $1, 2, \dots, n$. You reach in and select k balls at random. Let T be the sum of the numbers on the balls you picked.

(a) Say $k = 1$, what is $E[T]$

(b) Find $E[T]$ for general values of k.

(c) What is $\text{Var}(T)$?

Problem 3. Consider a binary tree with n levels as shown in Figure 1. Suppose that each link in this tree works with probability p, and is defective with probability $1 - p$, independently of other links. Find the probability that there is a working path from the root node (R) to a leaf. You only need to find a recursive formula for this probability.

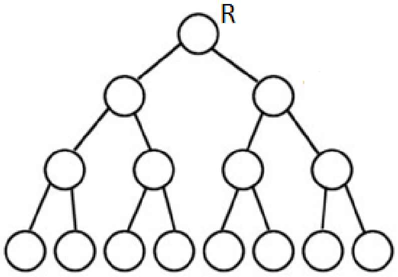

Figure 1: reliability graph for a binary tree with $n = 3$.

Problem 4. Consider the following scenario: two envelopes are placed in front of you, each of which contains a different positive, integer amount. You randomly select one of the two envelopes and peek inside so that you see the amount in that envelope. You may either keep the amount in this envelope, or switch to the other envelope, but at that point your choice is fixed. Your friend tells you that if you toss a coin until you see a heads and add $\frac{1}{2}$ to the number tosses it took to see a heads *and* that number is greater than the number in the envelope, you should switch and select the other envelope. Is your friend correct?

Problem 5. This problem will explore an important probabilistic concept of clustering that is widely used in machine learning applications today. Consider n students. For each pair of students, say student i and student j, they are friends with probability p, independently of other pairs. Here we assume that friendship is mutual, then we can see that the friendship among the n students can be represented by an undirected graph G. Let $N(i)$ be the number of friends of student i and $T(i)$ be the number of triangles attached to student i. We define the *clustering coefficient* $C(i)$ for student i as follows:

$$C(i) = \frac{T(i)}{\binom{N(i)}{2}}.$$

Clustering coefficient is not defined for the students who have no friends. An example is shown in Figure 2. Student 3 has 4 friends: 1,2,4,5, and there are two triangles attached to student 3, i.e., triangle 1-2-3 and triangle 2-3-4. Therefore $C(3) = \frac{2}{\binom{4}{2}} = \frac{1}{3}$. Find $\mathbb{E}[C(i)|N(i) \geq 2]$.

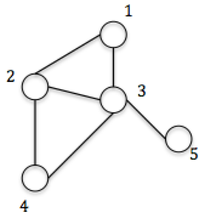

Figure 2: Friendship and clustering coefficient.

Problem 6. Consider a balls-and-bins model with K balls and M bins as shown in Figure 2. For each pair of ball and bin (e.g. the ith ball and the jth bin), the ball is thrown into the bin with probability p, independently from other pairs. The balls-and-bins model can also be viewed as a random graph. As we can see, the degree of the bins (number of balls thrown in a bin) is a random variable.

(a) Find the distribution of the degree of a bin.

(b) What is the expected degree of a bin?

(c) Now suppose you pick a random edge in the graph. What is the distribution of the degree of the right node connected to this edge? Is it the same as part (a)?

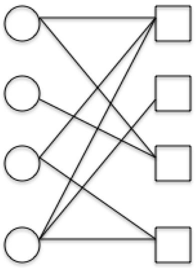

Figure 3: Balls-and-bins model.

(d) We call a bin with degree 1 a *singleton*. What is the average number of singletons in a random balls-and-bins model?

(e) Find the average number of balls that are connected to at least one singleton.

Problem 7. (Optional: this problem is of a more theoretical nature and will not be covered in an exam)

Jeff and Lester are bored at work and are trying to compress their messages to each other as much as possible. The messages they represent are modeled as bit strings of length n, where the bits are iid random variables such that $X_i = 1$ with probability p and $X_i = 0$ with probability $1 - p$.

(a) Let N be the number of 1's in the sequence. Find $P(N = k)$ and $E[N]$

In the next few parts, we will show that:

$$P(N = k) = C_{n,k}2^{-nD\left((\frac{k}{n},\frac{n-k}{n})||(p,1-p))\right)}$$

where $C_{n,k}$ is a term involving k and n that is on the order of $\sqrt{n}$ and

$$D\big((\frac{k}{n},\frac{n-k}{n})||(p,1-p))\big) = -\frac{k}{n}\log\frac{p}{\frac{k}{n}} - \frac{n-k}{n}\log\frac{1-p}{\frac{n-k}{n}}$$

$D(p||q) = \sum_i p(i)\log\frac{p(i)}{q(i)}$ is called the Kullback-Leibler divergence and measures the distance between two probability distributions. We are trying to show that as $\frac{k}{n}$ moves away from p, the probability of a sequence containing k ones drops off exponentially fast.

(b) Show that $\binom{n}{k} = C_{n,k}2^{nH(\frac{k}{n},\frac{n-k}{n})}$. Where $H(\frac{k}{n},\frac{n-k}{n}) = -\frac{k}{n}\log\frac{k}{n} - \frac{n-k}{n}\log\frac{n-k}{n}$ and is called the *entropy* and $C_{n,k} = \frac{\sqrt{2\pi n}}{\sqrt{2\pi k}\sqrt{2\pi(n-k)}}$.

Hint: Use $n! \approx (\frac{n}{e})^n\sqrt{2\pi n}$.

(c) Show that

$$P(N = k) \approx C_{n,k}2^{-nD\left((\frac{k}{n},\frac{n-k}{n})||(p,1-p))\right)}$$

As $n \to \infty$, what happens? Use this to devise a simple compression scheme.

(d) Consider $n = 10000, p = 0.7$. What is the sequence with the highest probability? Is this sequence considered in your compression scheme?