Final Review
CS244

I. Sample spaces, probability functions, Random Variables PMF, PDF, CDF

(1) Three cards are drawn from a standard deck. What is the probability that they are all 3's?
This problem has two different interpretations, leading to two different answers, depending on whether or not you replace a card after each draw. In both cases, identify the relevant sample space and probability function, carefully define the event in question, and compute the probability of the event.

(2) Three dice are rolled. Let X be the random variable denoting the number of distinct values that appear. For instance, if you roll a 1, a 3, and then another 3, X has the value 2. Calculate the PMF of X.

(3) *The billion-dollar bracket*. You could have won a billion dollars had you correctly predicted the participants and outcome of all 63 games in the NCAA College basketball tournament. Let us say that you are very good at predicting the outcome of games, and have a 65% success rate for each game. What is the probability that you will win at this challenge? Suppose all 300 million people in the United States, each with the same prediction ability as you, enter the contest. What is the probability that someone will win?

(4)(a) A spinner, giving a value in the range 0 to 1 is spun, with result X. Find the CDF and PDF of the random variable $X^{2.}$

(b) Two spinners, giving values in the range 0 to 1, are spun, with results X and Y. Let Z be the random variable $X^2 + Y^2$. Let $F_Z(x)$ be the CDF of Z and $p_Z(x)$ the PDF.

*Find and plot $F_Z(x)$ and $p_Z(x)$ first for x between 0 and 1, and then try to plot these functions for x > 1 (which is a much harder problem, mostly a calculus problem, but see the related problems below).

II. Combinatorics

5. Find $\binom{100}{0}$ and $\binom{100}{99}$ without lifting a finger. Find $\binom{100}{2}$ without breaking a sweat.

6. There are 100 bottles of beer on the pantry shelf. 40 of them are India Pale Ale, 30 are Pilseners, and 30 are Stout. You root around in the dark to find six bottles for you and your friends. What is the probability that you get at least one of each kind of beer?

III. Independence, Conditional Probability and Bayes's Theorem

7. An experiment consists of drawing 2 cards from a deck. Let the random variables X and Y denote, respectively, the number of 3's drawn and the number of hearts drawn. Show by a careful calculation that X and Y are not independent. (This should be obvious on an intuitive basis, but show me the math.)

8. A continuous version of the preceding problem: Choose a point uniformly and at random in the interior of the unit circle (the circle of radius 1 in the plane centered at the origin). Let X and Y be the x- and y-coordinates of the selected points. Show that the random variables X and Y are not independent.

9. Let's redo Problem 6 with a different story: There are two pantries, each with 100 bottles of beer. Pantry One has the same distribution of beers as in Problem 6, and Pantry 2 has 25 Pale Ales, 35 Pilseners and 40 Stouts. You open one of the two pantry doors, pull out three beers, and find they are all different. What is the probability that the door you opened was to Pantry One?

IV. Expected Value

10. Find the expected value of the random variables X and Z of Problem 4. (For part (b) this is easier than Problem 4 itself, which you don't need to solve completely in order to do this one.)

11. Roll a die repeatedly until a value repeats, and let X be the random variable giving the number of times you roll. For example, if you roll 1,3, 4, 3, the game lasts 4 rounds, so X has the value 4. What is E(X)?

V. Variance and Standard Deviation

12. Find the variance and standard deviation of the random variable Z of Problems 4 and 10.

13. Find the variance and standard deviation of the random variable X of Problem 11.

VI Special Distributions

14.You go outside during a light rainfall and set up instruments to count the number of raindrops falling on a small area of sidewalk during one minute. You divide the minute into 60 one-second periods, and find that at least one raindrop hits the sidewalk in 20 of these periods.

What is the average number of raindrops per second? (It is not 1/3.) What is the average wait time until the next raindrop? (It is not 3 seconds.) What is the probability that at least two raindrops will fall in a given one-second period?

15. Spin a spinner 100 times and take the sum of the squares of the values. Use the Central Limit Theorem to estimate the probability that this sum is greater than 30.

VII. Eigenvalues and Principal Component Analysis

16. Let

$$X = \begin{pmatrix} 1 & 3 & -2 & 10 \\ 2 & 5 & -1 & 6 \\ -3 & 7 & 4 & 4 \end{pmatrix}$$

You can think of the columns as samples drawn from some distribution.

(a) Compute the mean value **m** of the four columns, and replace X by a matrix Y in which all the columns are translated by **m** so that the mean value for Y is the all-zero vector.

(b) Compute YY^T by hand, and then compute cov(X) by computer. You should see, and be able to explain, a close relationship between these two matrices.

(c) Use the computer to find a diagonal matrix D and an orthogonal matrix U such that

cov(X) = UDU^T

and verify that U is indeed orthogonal, and that its columns are eigenvectors of cov(X) associated with eigenvalues given by the diagonal entries of D.

(d) Compute the projections of the columns of Y onto the directions given by the eigenvectors of cov(X) associated with the two largest eigenvalues. Use this and the mean **m** to approximately reconstitute the original matrix X and compare to the original. (The idea of this problem is that even though the original data is 3-dimensional, it is 'almost' two-dimensional, and we can partially recover it from the two original components.)

VIII. Markov Chains

17. Toss a coin repeatedly until the pattern HTH appears. What is the expected number of tosses? To solve this problem, model the experiment as a Markov chain with 4 states {none,H,HT,HTH}. The state is the longest suffix of sequence of tosses seen so far that is also a prefix of HTH, so that it records our progress towards reaching the desired pattern. (For example, if we toss HTTHHTH we visit the sequence of states none,H,HT,none,H,H,HT,HTH.) The state HTH is an absorbing

state. Draw the state diagram and write down the transition matrix, then use this matrix to compute the expected number of tosses until absorption.

18. (PageRank for a very tiny web.) Web page 1 links to pages 2 and 3, web page 2 links to page 3, and page 3 links back to page 2. A random web surfer starts on any of the three pages. With probability 0.8 he follows a random link on the current page; with probability 0.2 he moves to one of the three pages selected at random.

Determine the probabilities that after a long session of surfing, he winds up on pages1,2, or 3. Do this by computing the transition matrix for the underlying Markov chain, and a left eigenvector associated with the eigenvalue 1 giving the stationary distribution.

IX Simulations

With the exception of problems 3,5 and 16, you can reality-check all of your answers by writing simulations. Do a few of these and see what you come up with.