University of Illinois **Fall 2012**

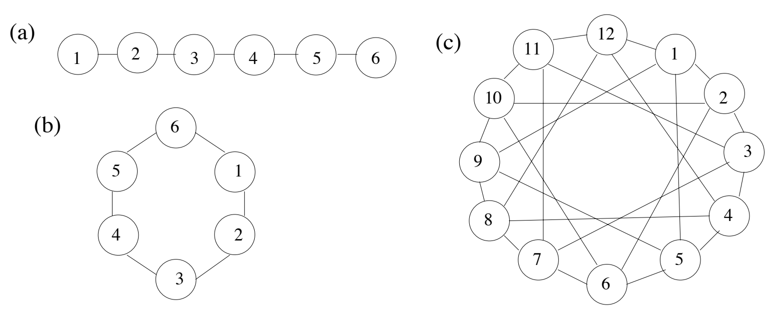
ECE 313: Problem Set 2
Discrete random variables

Due: Wednesday, September 12 at 4 p.m.
Reading: *ECE 313 Course Notes,* Sections 2.1-2.2

Reminder on turning in homework: Place into the ECE 313 drop box in the basement of Everitt Lab by the deadline. All assignments must be submitted stapled if they consist of more than one sheet with the following heading in block capital letters with at least 12pt size font or equivalent neat handwriting, on the top right corner of the first page:

NAME AS IT APPEARS ON COMPASS
NETID
SECTION
PROBLEM SET #

1. **[Distance between two randomly selected vertices]**
Solve the following problem for each of the three undirected graphs below. For a given graph, two verticies, i and j, are selected at random, with all possible values of (i, j) having equal probability, including the cases with $i = j$. Let D denote the distance between i and j, which is the minimum number of edges that must be crossed to walk in the graph from i to j. If $i = j$ then $D = 0$. Find and sketch the pmf of D, and find its mean and variance. (Hint: For (b) and (c), by symmetry, it can be assumed that $i = 1$ and only j is selected at random.)



2. **[Up to five rounds of double or nothing]**
A gambler initially having one chip participates in up to five rounds of gambling. In each round, the gambler bets all of her chips, and with probability one half, she wins, thereby doubling the number of chips she has, and with probability one half, she loses all her chips. Let X denote the number of chips the gambler has after five rounds, and let Y denote the maximum number of chips the gambler ever has (with stopping after five rounds). For example, if the gambler wins in the first three rounds and loses in the fourth, $Y = 8$. If the gambler loses in the first round, $Y = 1$.

   (a) Find the pmf, mean, and variance of X.
   (b) Find the pmf, mean, and variance of Y.

3. **[Selecting supply for a random demand]**
A reseller reserves and prepays for L rooms at a luxury hotel for a special event, at a price of a dollars per room, and the reseller sells the rooms for b dollars per room, for some known, fixed values a and b with $0 < a \leq b$. Letting U denote the number of potential buyers, the actual number of buyers is $\min\{U, L\}$ and the profit of the reseller is $b \min\{U, L\} - aL$. The reseller must declare L before observing U, but when L is selected the reseller assumes U takes on the possible values $1, 2, \cdots, M$, each with probability $\frac{1}{M}$ for some known $M \geq 1$.

   (a) Express the expected profit of the reseller in terms of M, L, a, and b. Simplify your answer as much as possible. For simplicity, without loss of generality, assume $0 \leq L \leq M$. (Hint: $1 + \cdots + L = \frac{L(L+1)}{2}$. Your answer should be valid whenever $0 \leq L \leq M$; for $L = 3$ and $M = 5$ it should reduce to $E[\text{profit}] = \frac{12b}{5} - 3a$.)

   (b) Determine the value of L as a function of a, b and M that maximizes the expected profit.

4. **[First and second moments of a ternary random variable]**
This problem focuses on the possible mean and variance of a random variable X with support set $\{-1, 0, 1\}$. Let $p_X(-1) = a$ and $p_X(1) = b$ and $p_X(0) = 1 - a - b$. This is a valid pmf if and only if $a \geq 0$, $b \geq 0$, and $a + b \leq 1$. Let $\mu = E[X]$, $\sigma^2 = \text{Var}(X)$, and $m_2 = E[X^2]$.

   (a) Find (a, b) so that $\mu = \frac{1}{2}$ and $\sigma^2 = \frac{3}{4}$.

   (b) Express a and b in terms of μ and m_2. Determine and sketch the region of (μ, m_2) pairs for which there is a valid choice of (a, b).

   (c) Determine and sketch the set of (μ, σ^2) pairs for which there is a valid choice of (a, b). (Hint: For μ fixed, the set of possible values of σ^2 is the set of possible values of $m_2 - \mu^2$.)

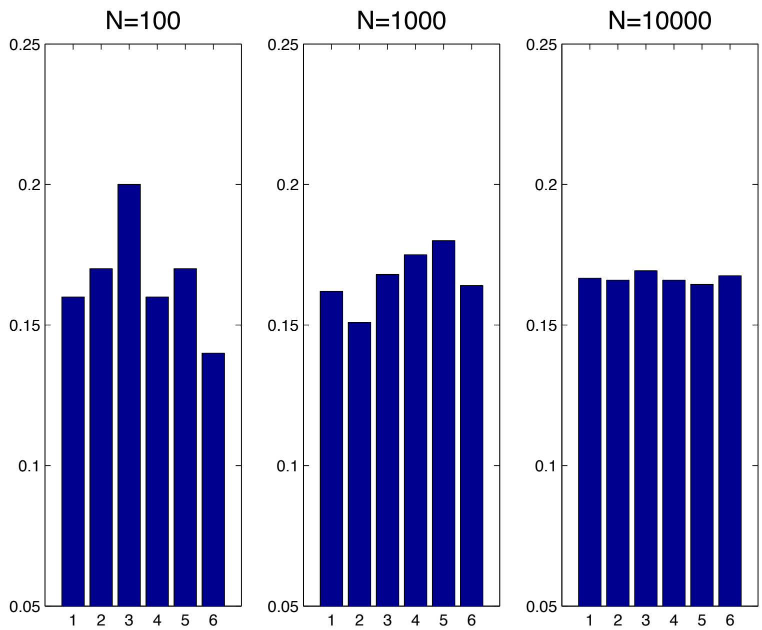
5. **[The Zipf distribution]**
The Zipf distribution has been found to model well the distribution of popularity of items such as books or videos in a library. The Zipf distribution with parameters M and α, is the distribution supported on $\{1, \ldots, M\}$ with pmf $p(k) = \frac{k^{-\alpha}}{Z}$ for $1 \leq k \leq M$, where Z is the constant chosen to make the pmf sum to one, i.e. $Z = \sum_{k=1}^{M} k^{-\alpha}$. The interpretation is that if the videos in a library are numbered 1 through M, in order of decreasing popularity, then a random request to view a video would be for video k with probability $p(k)$. (Light use of a computer or programable calculator is recommended for this problem.)

   (a) If X has the Zipf distribution with parameters $M = 2000$ and $\alpha = 0.8$ (suitable for a typical video library), what is $P\{X \leq 500\}$?

   (b) If Y has the Zipf distribution with parameters $M = 2000$ and $\alpha > 0$, for what numerical value of α is it true that $P\{Y \leq 100\} = 30\%$?

6. **[Producing some simple empirical distributions by computer simulation]**
This problem is intended to be solved using MATLAB; other methods such as the use of another computer language or spreadsheet software is permitted but not supported. We begin by describing a computer simulation of N rolls of a fair die for some particular value of N. The *empirical distribution* for such an experiment is the fraction of rolls that are one, the fraction of rolls that are two, and so on, up to the fraction of rolls that are six. Start up MATLAB and place a copy of the file dice_histograms.m (available in the ECE 313 homework/MATLAB directory) into the MATLAB current directory (or one of the other directories in a MATLAB search path). Run the command dice_histograms by typing it into the command window and pressing the return key. It should produce on your screen (i.e. standard output) the three

empirical distributions shown below, and also produce a corresponding .eps figure file in the MATLAB working directory.



Examine the file dice_histograms.m to see how it produces the output shown.

(a) The script file dice_histograms.m uses the three MATLAB functions: output=randi(i,j,n) for integers i,j, and n; output=hist(Y,x) for vectors Y and x; and bar(Z) for a vector Z. Explain what each of these functions produces.

(b) Run the command dice_histograms at least twenty times; the displayed triplet of empirical distributions should change each time. See how much the distributions change from one simulation to the next, for each of the three values of N. The empirical distribution for the case $N = 100$ can often be far from uniform. Here is what to turn in for this part: Find, print out, and turn in the printout, for an example triplet of plots produced by the program such that the first empirical distribution (i.e. the one for $N = 100$) is far from the uniform one.

(c) Produce a modification of the program so that for each N, two dice are rolled N times, and for each time the sum of the two dice is recorded. The empirical distributions will thus have support over the range from 2 to 12. Print out and turn in a figure showing empirical distributions for $N = 100$, $N = 1000$, and $N = 10000$, and also turn in a copy of the computer code you used. (Hint: Even though the sum of two dice is always at least 2, it might be easier to make the plot if one of the bins for the histogram is centered at one; the count for such bin will always be zero.)